Méthodes cinétiques appliquées à l'étude de certains comportements collectifs

Valentin Ayot

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR

DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE MATHÉMATIQUES ET INFORMATIQUE

SPÉCIALITÉ: MATHÉMATIQUES APPLIQUÉES

Par **AYOT VALENTIN**

Méthodes cinétiques appliquées à l'étude de certains comportements collectifs

Sous la direction de: **Philippe Thieullen**
Co-directeur: **Stéphane Brull**

Soutenue le 13 décembre 2022 à l'institut de mathématiques de Bordeaux
devant le jury composé de

Marc ARNAUDON	Professeur des Universités	Université de Bordeaux	Président
Stéphane BRULL	Maitre de Conférences	Université de Bordeaux	Co-Directeur
Seung-Yeal HA	Professeur	Seoul National University	Rapporteur
Christian LÉONARD	Professeur	Université Paris Nanterre	Examinateur
Gael RAOUL	Chargé de Recherche	École Polytechnique	Examinateur
Philippe THIEULLEN	Professeur des Universités	Université de Bordeaux	Directeur

Résumé

Dans divers domaines, certains phénomènes qui dépendent du temps sont représentés par des modèles mathématiques. En particulier les comportements collectif d'un groupe composé d'un certain nombre d'agents. Dans les modèles de la théorie cinétique, on étudie la position et la vitesse de chaque agent au cours du temps. L'équation de Boltzmann est l'une des équations les plus connues dans ce domaine. Dans la littérature moderne, on trouve beaucoup de modèles cinétiques *de type Boltzmann* qui décrivent au cours du temps le comportement collectif d'un groupe (gaz de particules, groupe d'individus, nuée d'oiseaux, banc de poissons,...). On retrouve dans la plupart de ces modèles les mêmes résultats que ceux établis pour l'équation de Boltzmann en utilisant les mêmes arguments. Le but de ce manuscrit est d'introduire puis d'étudier un modèle cinétique de type Boltzmann dans lequel les outils classiques utilisés pour l'équation de Boltzmann sont inopérants.

Le modèle présenté dans ce manuscrit est un modèle de type Boltzmann qui décrit au cours du temps le comportement collectif d'un grand groupe d'individus. Celui-ci considère un mécanisme où lorsque deux individus entrent en collision, ils vont adopter après la collision la même vitesse selon une distribution centrée sur la vitesse moyenne avant la collision. Le premier chapitre de ce manuscrit concerne la version la plus simplifiée de ce modèle: la version homogène en espace avec un taux de collision constant. Nous montrerons dans ce chapitre que dans ce cas particulier, les solutions du modèle sur $\mathbb{R}^d$ convergent exponentiellement vers l'état d'équilibre pour la distance de Wasserstein. Cette convergence sera obtenue grâce à un phénomène de contraction qui prend place dans le processus de collision. La convergence des solutions pour la norme forte L^1 sera également démontrée pour des conditions initiales vérifiant une propriété de régularité plus forte. Des illustrations numériques seront présentes afin de visualiser cette convergence.

Dans un second chapitre, on considèrera le même modèle mais avec un taux de collision quelconque. La non-constance de celui-ci complique la preuve puisqu'il fait disparaitre le phénomène de contraction dans le processus de collision. Cependant, on peut écrire le modèle comme une équation différentielle ordinaire en dimension infinie, ce qui fait apparaître un opérateur qui satisfait la condition de Hölder pour $\alpha = 1/2$. Nous montrerons dans ce second chapitre l'existence d'états d'équilibre et des simulations numériques seront présentées. Celles-ci montrerons l'existence d'une solution en tout temps ainsi que la convergence vers un unique état d'équilibre. La convergence théorique n'est pas démontré rigoureusement.

Mots-clés: Modèles cinétiques, Dynamique collective, Distance de Wasserstein, Transport optimal.

Laboratoire d'accueil: Institut mathématiques de Bordeaux UMR 5251, Université de Bordeaux, Bâtiment A33, 351 cours de la Libération, F-33405 Talence cedex France.

Abstract

In various domains, certain phenomena that depend on time are represented by mathematical models. In particular the collective behavior of a group composed of a certain number of agents. In kinetic theory models, we study the position and velocity of each agent over time. The Boltzmann equation is one of the best-known equations in this domain. In modern literature, there are many *Boltzmann type* kinetic models which describe the collective behavior of a group over time (gas of particles, group of individuals, flock of birds, school of fish,...). We find in most of these models the same results as those established for the Boltzmann equation using the same arguments. The goal of this manuscript is to introduce and then to study a kinetic Boltzmann-type model in which the classical tools used for the Boltzmann equation are inoperative.

The model presented in this manuscript is a Boltzmann-type model that describes the collective behavior of a large group of individuals over time. This one considers a mechanism where as two individuals collide, they will adopt after the collision the same post-collisionnal velocity according to a distribution centered at the mid pre-collisional velocity. The first chapter of this manuscript concerns the most simplified version of this model: the spatially homogeneous version with a constant collision rate. We will show in this chapter that in this particular case, the solutions of the model on $\mathbb{R}^d$ converge exponentially towards the equilibrium state for the Wasserstein metric. This convergence will be obtained thanks to the phenomenon of contraction which takes place in the collision process. The convergence of solutions for the strong norm L^1 will also be proved for initial conditions satisfying a stronger regularity property. Numerical illustrations will be present to visualize this convergence.

In a second chapter, we will consider the same model but with any collision rate. The non-constancy of this one complicates the proof since it makes disappear the phenomenon of contraction in the collision process. However, we can write the model as an ordinary differential equation in infinite dimension, which gives rise to an operator which satisfies the Hölder condition with $\alpha = 1/2$. We will show in this second chapter the existence of equilibrium state and numerical simulations will be presented. These will show the existence of a solution at all times as well as the convergence towards a unique equilibrium state. Unfortunately we do not know how to show rigorously the convergence.

Keywords: Kinetic models, Collective dynamics, Wasserstein metric, Optimal transport.

Host laboratory: Institut mathématiques de Bordeaux UMR 5251, Université de Bordeaux, Bâtiment A33, 351 cours de la Libération, F-33405 Talence cedex France.

Remerciements

Quand j'étais au lycée, un de mes enseignants en mathématiques nous donnait régulièrement du travail évalué à la maison. Ces exercices étaient relativement longs, difficiles et très calculatoires. La majorité de ma classe était après cette année là, dégoutée des mathématiques, sauf moi. En effet, l'aspect très calculatoire de ces devoirs maison me stimulait énormément. A partir de l'année qui a suivi, je ne souhaitais faire que des mathématiques. Mes premiers remerciements vont donc à cet enseignant.

Mes trois années de thèse auront été fortement impactées par la pandémie de COVID 19. Travailler dans ce contexte aura souvent été mentalement très difficile. C'est pourquoi je ne saurais suffisamment remercier mon directeur de thèse, Philippe Thieullen, pour son accompagnement et son encadrement hors du commun, et ce depuis ma deuxième année de Master. Ce manuscrit qui a été pour moi un plaisir à écrire n'aurait jamais vu le jour sans ses connaissances, sa rigueur, sa persévérance, sa curiosité, son enthousiasme et sa critique. Je remercie aussi mon co-directeur, Stéphane Brull, de m'avoir co-encadré avec Philippe depuis ma deuxième année de Master.

Je remercie José A. Carrillo de nous avoir suggéré l'étude de ce modèle lors de sa visite à l'IMB en Septembre 2019. Je le remercie également, ainsi que Seung-Yeal Ha d'avoir bien voulu rapporter ma thèse. Je tiens également à remercie Seung-Yeal Ha une deuxième fois, ainsi qu'à Marc Arnaudon, à Christian Léonard et à Gaël Raoul d'avoir accepté de prendre part à mon jury de thèse.

Certaines personnes auront joué un rôle très important. J'aimerais d'abord remercier infiniment Nicole Bergerot la secrétaire de mon ancien Master dont je ne compte plus le nombre d'heures passées dans son bureau à l'époque. Nicole est la personne qui m'a fortement recommandé mon directeur de thèse donc par conséquent tout ce travail n'aurait jamais vu le jour sans son expérience, ses conseils, sa bienveillance et son humour. Ensuite, j'aimerais remercier mon médecin et également ami David, dont la gentillesse, la culture, la curiosité et l'esprit critique atteignent des records. Et qui pense que je devrais me faire examiner le cerveau. J'aimerais aussi remercier Nicolas pour nos nombreuses discussions et pour m'avoir épauler durant les périodes difficiles.

Je souhaiterais également remercier tout le monde présent au CEMRACS 2022. Les 6 semaines passées là bas auront été fabuleuses. Merci à Frederik pour les sessions psicobloc, à Nono pour les apéros et les soirées pétanques, puis à tous mes collègues doctorants: Mickaël, Ignacio, Damien, Giuseppe, Kévin, Brieuc, Léo, Léopold, Averil et tous les autres. Je remercie également les organisateurs ainsi que le personnel du CIRM.

Et pour terminer je remercie ma famille, en particulier ma grand mère et mon frère. Ainsi que mes amis: amis du lycée, de soirées, de grimpe, de mon frère et j'en passe.

Cette thèse est dédiée à la mémoire de mon arrière mamie Gisèle, qui est malheureusement partie quelques mois avant le début de cette aventure.

Contents

Conventions and recalls

Sets and structures

• $\mathbb{N}$ is the set of non-negative integers: $\mathbb{N} := \{0, 1, 2, 3, \cdots\}$, a sequence of elements of a set E is denoted $(x_n)_{n\in\mathbb{N}}$ or more simply $(x_n)_n$. And $\mathbb{Z}$ is the set of integers: $\mathbb{Z} := \{\cdots, -1, 0, 1, 2, \cdots\}$.
• $\mathbb{Q}$ is the set of rational number: $\mathbb{Q} := \{a/b,\, a \in \mathbb{Z},\, b \in \mathbb{N} \text{ and } b > 0.\}$.
• $\mathbb{R}$ is the set of real numbers. The notation $\mathbb{R}^d$ obviously implies that d is a positive integer. For two vectors $x = (x_1, \cdots, x_d)$ and $y = (y_1, \cdots, y_d)$ in $\mathbb{R}^d$, we denote $\langle x, y\rangle := \sum x_i y_i$ the Euclidian scalar product between x and y. All vectors considered in this text are column vector. The norm in $\mathbb{R}^d$ is denoted by $|\cdot|$. Since all the norms are equivalent in $\mathbb{R}^d$, $|\cdot|$ will denote the Euclidian norm, whatever the dimension d: $|x| := (\sum x_i^2)^{1/2}$. And $\mathbb{S}^d$ is the unit sphere in $\mathbb{R}^{d+1}$: $\mathbb{S}^d := \{x \in \mathbb{R}^{d+1},\, |x| = 1\}$.
• $\mathbb{C}$ is the set of complex numbers: $\mathbb{C} := \{z = a + ib,\, (a, b) \in \mathbb{R}^2\}$. The real number a is the real part of z denoted $\Re(z)$, the real number b is the imaginary part of z denoted $\Im(z)$ and i is the imaginary unit: $i^2 = -1$. The modulus of a complex number $z = a + ib$ is denoted by $|z|$. It is the Euclidian norm of the vector (a, b) in $\mathbb{R}^2$. And the conjugate of a complex number $z = a + ib$ is the complex number $\bar{z} := a - ib$.
• $M_n(\mathbb{R})$ is the set of real $n \times n$ matrices. The matrix I_n is the $n \times n$ identity matrix and we denote by $\mathrm{diag}(d_1, \cdots, d_n)$ the diagonal matrix formed by the d_i on the diagonal. A matrix $M \in M_n(\mathbb{R})$ is said to be invertible if there exists a matrix P such that $MP = PM = I_n$. In this case, P is unique and it is denoted by M^{-1}. For a matrix $M \in M_n(\mathbb{R})$, we denote ${}^tM$, $\det(M)$ and $\mathrm{Tr}(M)$ the transposate, the determinant and the trace of M. For two vectors x, $y \in \mathbb{R}^d$ and for two matrices A, $B \in M_n(\mathbb{R})$, some authors write $x \otimes y$ to designate the matrix $x\,{}^ty$ and $A : B$ to designate $\mathrm{Tr}({}^tAB)$.
• We say that a symmetric matrix M is positive semi-definite (positive definite), denoted by $M \succeq 0$ ($M \succ 0$) if for every real column vector x, ${}^txMx \geq 0$ (for every non-zero real column vector x, ${}^txMx > 0$). For M, $N \in M_d(\mathbb{R})$, we write $M \succeq N$ ($M \succ N$) if $M - N \succeq 0$ ($M - N \succ 0$). For a symmetric matrix $M \succeq 0$ ($M \succ 0$), $M^{1/2}$ denote the square root of M. It is the unique symmetric positive semi-definite (positive definite) matrix such that $(M^{1/2})^2 = M$.
• If $A \subset E$ is a set, then A^c is the complement of A in E: $A^c := \{x \in E,\, x \notin A\}$.
• A subset $A \subset \mathbb{R}$ is said to be bounded above if there exists $M \in \mathbb{R}$ such that for every $x \in A$, $x \geq M$. The smallest possible constant M is called the supremum of A and is denoted $\sup(A)$. A is said to be bounded below if there exists $m \in \mathbb{R}$ such that for every $x \in A$, $x \leq m$. The greatest possible constant m is called the infimum of A and is denoted $\inf(A)$.
• If E is a metric space, then we denote by (E, d) the space E equipped with the metric d. And if E is a normed vector space, then we denote by $(E, \|\cdot\|)$ the vector space E equipped with the norm $\|\cdot\|$.
• $E \times F$ denotes the Cartesian product between two sets E and F: $E \times F := \{(x, y),\, x \in E \text{ and } y \in F\}$. If (E, d_E) and (F, d_F) are two metrics spaces, then we can equip $E \times F$ with the distance $d_E + d_F$. If $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ are two normed vector spaces, then we can equip $E \times F$ with the norm $\|\cdot\|_E + \|\cdot\|_F$.
• The open ball of radius $r > 0$ and center x_0 in a metric space (E, d) is denoted by $B(x_0, r)$: $B(x_0, r) := \{x \in E,\, d(x, x_0) < r\}$. The unit ball $B(0, 1)$ in $\mathbb{R}^d$ is denoted by B_d.
• For a set A in a metric space (E, d), the interior of A is denoted by $\mathrm{Int}(A)$: $\mathrm{Int}(A) := \{x \in A,\, \exists \varepsilon > 0,$

$B(x,\varepsilon) \subset A\}$. And the closure of A is denoted by $\overline{A}$, this is the set of all limits of sequences taking values in A.
• A set C in a real vector space is said to be convex if $tx + (1-t)y \in C$ for all $x, y \in C$ and $t \in [0,1]$.

Functions and sequences

• For a space E, id is the identity mapping on E. For $x \in E$, $\mathrm{id}(x) := x$.
• For $u \in \mathbb{R}^d$, τ_u is the translation of vector u: $\tau_u(x) := x + u$.
• If A is a set, then the function $\mathbb{1}_A$ is the indicator function of A: $\mathbb{1}_A(x) := 1$ is $x \in A$ and 0 otherwise. If $\mathscr{F}$ is a formula, then $\mathbb{1}_{\mathscr{F}}$ is the indicator function of the set defined by the formula $\mathscr{F}$.
• In a Cartesian product $X_1 \times \cdots \times X_j \times \cdots \times X_n$, proj_j is the projection map into the j-th component: $\mathrm{proj}_j(x_1, \cdots, x_j, \cdots, x_n) := x_j$.
• The maximum and the minimum between a and $b \in \mathbb{R}$ is denoted by $\max(a,b)$ and $\min(a,b)$. The positive and negative parts of $x \in \mathbb{R}$ are defined respectively by $x_+ := \max(x,0)$ and $x_- := \max(-x,0)$. Both are non-negative and we naturally have $|x| = x_+ + x_-$.
• The floor of a real number x, denoted by $\lfloor x \rfloor$, is the largest integer that is less than or equal to x. And the ceiling of a real number x, denoted by $\lceil x \rceil$, is the smallest integer that is greater than or equal to x.
• We write $x_n \longrightarrow x$ if the real-sequence $(x_n)_n$ converge to x. An element $l \in \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ is a subsequential limit of $(x_n)_n$ if there exists a subsequence $(x_{\sigma(n)})_n$ which converges to l. The greatest possible limit l is called the limit superior of $(x_n)_n$ and the smallest possible limit l is called the limit inferior of $(x_n)_n$. They are denoted $\limsup x_n$ and $\liminf x_n$.
• A sequence $(x_n)_n$ in a metric space (E,d) converge to $x \in \overline{E}$ if the real sequence $(d(x_n,x))_n$ converge to 0. A sequence $(x_n)_n$ in a normed vector space $(E, \|\cdot\|)$ converge to $x \in \overline{E}$ if the real sequence $(\|x_n - x\|)_n$ converge to 0.
• If f and g are two functions, then $f+g$, fg and f/g denotes the sum, product and quotient of f by g: $(f+g)(x) := f(x) + g(x)$, $(fg)(x) := f(x)g(x)$ and $(f/g)(x) := f(x)/g(x)$. The composition of g by f is denoted by $f \circ g$: $f \circ g(x) := f(g(x))$ and (f,g) denote the function $(f,g)(x) := (f(x), g(x))$.
• If f is a function $X \longrightarrow Y$, then the set $\mathrm{Im}(f)$ is the image of f: $\mathrm{Im}(f) := \{y \in Y, \exists x \in X, y = f(x)\}$ and for a subset $B \subset Y$, $f^{-1}(B)$ denote the inverse image of B under f: $f^{-1}(B) := \{x \in X, f(x) \in B\}$. If f is a function $E \longrightarrow \overline{\mathbb{R}}$, then the set $\mathrm{Dom}(f)$ is the domain of f: $\mathrm{Dom}(f) := \{x \in E, f(x) \neq \pm\infty\}$. And if f is a continuous function $E \longrightarrow \mathbb{R}$, then the set $\mathrm{supp}(f)$ is the support of f: $\mathrm{supp}(f) := \overline{\{x \in E, f(x) \neq 0\}}$.
• A function $f : E \longrightarrow \mathbb{R}$ is said to be bounded above on E if there exists $M \in \mathbb{R}$ such that for all $x \in E$, $f(x) \leq M$. The smallest possible constant M is called the supremum of f on E and is denoted $\sup f$ or $\sup_x f(x)$. f is said to be bounded below on E if there exists $m \in \mathbb{R}$ such that for all $x \in E$, $f(x) \geq m$. The greatest possible constant m is called the infimum of f on E and is denoted $\inf f$ or $\inf_x f(x)$.
• If f is a real function, then we write $f(x) \longrightarrow L$ when $x \longrightarrow l$ if $f(x)$ converges to L when x goes to l. If f is a function $\mathbb{R} \longrightarrow (E,d)$, then $f(x)$ converges to L when x goes to l if the real function $d_E(f(x), L)$ converges to 0 when x goes to l.
• The notation $f = o(g)$ means that f becomes insignifiant relative to g in the neighborhood of 0. In other words, $f/g \longrightarrow 0$ when $x \longrightarrow 0$.
• The derivative of a function $f : I \longrightarrow \mathbb{R}$ where I is an open interval of $\mathbb{R}$ is denoted by f' or df/dx: $f'(x) := \lim_{h \to 0}(f(x+h) - f(x))/h$. If f is a vector valued function $I \longrightarrow \mathbb{R}^d$, then the derivative of $f = (f_1, \cdots, f_d)$ is the vector valued function $f' = (f_1', \cdots, f_d')$. The derivative of order n of f is denoted by $f^{(n)}$ or df^n/dx^n.
• If f is a function of severable variables, the partial derivative of f with respect to the variable x will be denoted by $\partial f/\partial x$ or $\partial_x f$. The differential of f at the point x is denoted by $D_x f$ and the differential of order n of f at the point x is denoted by $D_x^n f$.

• The gradient of a function $f : U \longrightarrow \mathbb{R}$ where U is an open set of $\mathbb{R}^d$ is denoted by ∇f. It is the vector composed of all the partial derivatives of f: $\nabla f := (\partial f/\partial x_i)_{1\leq i\leq d}$. The value of the gradient of f at the point x will be denoted by $\nabla f(x)$. If f is a vector valued function, then ∇f is the vector composed of the gradient of the components of f.
• The divergence of a function $f : U \longrightarrow \mathbb{R}^d$ where U is an open set of $\mathbb{R}^d$ is denoted by $\mathrm{div}(f)$ or $\nabla \cdot f$: $\mathrm{div}(f) := \sum(\partial f_i/\partial x_i)$. If f is a matrix valued function, then $\mathrm{div}(f)$ is the vector composed of the divergence of the rows of f.
• The Laplacian of a function $f : U \longrightarrow \mathbb{R}$ where U is an open set of $\mathbb{R}^d$ is denoted by Δf. It is the divergence of the gradient of f: $\Delta f := \mathrm{div}(\nabla f)$.
• If f is a function $\mathbb{R}^d \longrightarrow \mathbb{R}^d$, then J_f denotes the Jacobian matrix of f. It is the matrix composed of all the partial derivatives of f: $J_f := (\partial f_i/\partial x_j)_{1\leq i,j\leq d}$. And if f is a function $\mathbb{R}^d \longrightarrow \mathbb{R}$, then H_f denotes the Hessian matrix of f. It is the matrix composed of all the partial derivatives of order 2 of f: $H_f := (\partial f^2/\partial x_i \partial x_j)_{1\leq i,j\leq d}$. The value of the Jacobian matrix of f at the point x will be denoted by $J_f(x)$ and the value of the Hessian matrix of f at the point x will be denoted by $H_f(x)$.
• A function $f : E \longrightarrow \mathbb{R}$ is said to be convex if $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$ for all $x, y \in E$ and $t \in [0, 1]$.
• A map f between two metric spaces (E, d_E) and (F, d_F) is said to be M-Lipschitzian if for all $x, y \in E$, $d_F(f(x), f(y)) \leq M d_E(x, y)$. The smallest possible constant M will be denoted by $\|f\|_{\mathrm{Lip}}$.

Integration and measures

• A σ-algebra on a set $\mathcal{X}$ is a family of subsets of $\mathcal{X}$ which included $\mathcal{X}$, is closed under complement, and is closed under countable unions. If $\mathcal{X}$ is a topological space, then the Borel σ-algebra is the smallest σ-algebra on $\mathcal{X}$ which contains all the open sets of $\mathcal{X}$. An element of the σ-algebra is said to be measurable, and an element of a Borel σ-algebra is called Borelian.
• A Borel measure is a measure μ defined on a Borel σ-algebra which for any Borelian $\mathcal{A}$ associates a positive number (or possibly $+\infty$) $\mu(\mathcal{A}) \geq 0$, and satisfies the axiom of countable additivity: $\mu(\cup \mathcal{A}_i) = \sum \mu(\mathcal{A}_i)$ for all countable family $(\mathcal{A}_i)_{i\in\mathbb{N}}$ of pairwise disjoints sets. We speak of a signed measure if μ takes negative values. All measures considered in this text are Borel measures.
• For a measure μ, we define the integral of a real valued measurable function f with respect to μ the real number (possibly $\pm\infty$) denoted $\int f(x)d\mu(x)$, $\int f(x)\mu(dx)$ or $\int f d\mu$. If $f = (f_1, \cdots, f_d)$ is a real vector valued measurable function, then $\int f d\mu$ is the integral of all components of f: $\int f d\mu = (\int f_1 d\mu, \cdots, \int f_d d\mu)$. We say that f is integrable with respect to the measure μ if the quantity $\int |f| d\mu$ is finite.
• For a measure μ on $\mathcal{X}$, $\mathrm{supp}(\mu)$ is the support of μ. It is the intersection of all closed sets of full measure. Some authors define $\mathrm{supp}(\mu)$ by its complement: $\mathrm{supp}(\mu)^c := \bigcup\{O \in \mathcal{O},\, \mu(O) = 0\}$ where $\mathcal{O}$ is the set of open sets of $\mathcal{X}$.
• A probability measure on $\mathcal{X}$ is a measure μ such that $\mu(\mathcal{X}) = 1$. In this case, the measured space $(\mathcal{X}, \mu)$ is a probability space.
• For a probability measure μ on $\mathbb{R}^d$, the mean of μ is the real vector $\int x d\mu(x)$ and the covariance matrix is the positive semi-definite matrix $\int (x - \int x d\mu(x))^{\,t}(x - \int x d\mu(x)) d\mu(x)$. If $d = 1$, then the covariance matrix is a non-negative number called the variance of μ and is denoted σ_μ^2: $\sigma_\mu^2 := \int (x - \int x d\mu(x))^2 d\mu(x)$. And the moment of order $\alpha > 0$ of μ, denoted by $M_\alpha(\mu)$ is the real non-negative number $M_\alpha(\mu) := \int |x|^\alpha d\mu(x)$.
• $\mathcal{P}(\mathcal{X})$ is the space of probability measures on $\mathcal{X}$, $\mathcal{P}_\alpha(\mathcal{X})$ the space of probability measures on $\mathcal{X}$ which admit a finite moment of order $\alpha > 0$ and $\mathcal{P}_\alpha^m(\mathcal{X})$ the space of probability measures on $\mathcal{X}$ which admit a finite moment of order $\alpha \geq 1$ and of mean m.
• The cumulative distribution function of a probability measure μ on $\mathbb{R}$ is the real function denoted by F_μ: $F_\mu(x) := \mu(]-\infty, x])$.
• δ_{x_0} is the Dirac mass at the point $x_0 \in \mathcal{X}$. For any Borelian $\mathcal{A} \subset \mathcal{X}$, $\delta_{x_0}(\mathcal{A}) := \mathbb{1}_{x_0\in\mathcal{A}}$. In terms of

integrals, for any measurable function f, $\int f\delta_{x_0} = f(x_0)$.
• λ is the Lebesgue measure. If $\mathcal{A}$ is a Borelian on $\mathbb{R}^d$, then $\text{Vol}(\mathcal{A}) := \lambda(\mathcal{A})$ is the Euclidian volume of $\mathcal{A}$.
• A measure μ is said to be absolutely continuous with respect to a measure ν if there exists a measurable function f with respect to ν such that for any Borelian $\mathcal{A}$, $\mu(\mathcal{A}) = \int_{\mathcal{A}} f(x)d\nu(x)$. We denote $\mu = f\nu$. We say that μ is a density measure if it is absolutely continuous with respect to the Lebesgue measure and that f is a probability density if μ is a probability measure.
• If μ is a measure on $\mathcal{X}$ and ν a measure on $\mathcal{Y}$, then $\mu \otimes \nu$ is the product measure between μ and ν on $\mathcal{X} \times \mathcal{Y}$. For any Borelian $\mathcal{A} \times \mathcal{B} \subset \mathcal{X} \times \mathcal{Y}$, $\mu \otimes \nu(\mathcal{A} \times \mathcal{B}) := \mu(\mathcal{A})\nu(\mathcal{B})$.
• If μ and ν are two measures on $\mathcal{X}$, then $\mu * \nu$ is the convolution product between μ and ν on $\mathcal{X}$. For any Borelian $\mathcal{A} \subset \mathcal{X}$, $\mu * \nu(\mathcal{A}) := \iint \mathbb{1}_{\mathcal{A}}(x+y)d\mu(x)d\nu(y)$.
• If μ is a measure on $\mathcal{X}$ and T a Borel map $\mathcal{X} \longrightarrow \mathcal{Y}$, then $T\#\mu$ designates the image measure (or push-forward) of μ by T (see Definition A.1.1). The law of a random variable X defined on a probability space $(\Omega, \mathbb{P})$ is the measure $X\#\mathbb{P}$.
• If μ is a measure on $\mathbb{R}^d$ of finite mass: $\mu(\mathbb{R}^d) < +\infty$, then the function $\hat{\mu} : \mathbb{R}^d \longrightarrow \mathbb{C}$ is the Fourier transform of μ: $\hat{\mu}(\xi) := \int e^{-i\langle x,\xi\rangle} d\mu(x)$. If X is a random variable which taking values in $\mathbb{R}^d$ of law μ, then the function $\varphi_X : \mathbb{R}^d \longrightarrow \mathbb{C}$ is the characteristic function of X: $\varphi_X(t) := \hat{\mu}(-t)$.

Function spaces

• If (E, d_E) and (F, d_F) are two metric spaces, then $\mathcal{C}(E, F)$ is the space of continuous functions $E \longrightarrow F$. We can equip $\mathcal{C}(E, F)$ with the uniform metric d_∞: $d_\infty(f, g) := \sup d_F(f(x), g(x))$. If $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ are two normed vector spaces, then we can equip $\mathcal{C}(E, F)$ with the uniform norm $\|\cdot\|_\infty$: $\|f\|_\infty := \sup \|f(x)\|_F$. If $F = \mathbb{R}$, we will denote by $\mathcal{C}(E, \mathbb{R}) = \mathcal{C}(E)$.
• $\mathcal{C}_b(\mathbb{R}^d)$ is the space of bounded continuous functions $\mathbb{R}^d \longrightarrow \mathbb{R}$, $\mathcal{C}_0(\mathbb{R}^d)$ the space of continuous functions $\mathbb{R}^d \longrightarrow \mathbb{R}$ which tend to 0 at infinity and $\mathcal{C}_c(\mathbb{R}^d)$ the space of continuous functions $\mathbb{R}^d \longrightarrow \mathbb{R}$ with compact support. We can equip these three spaces with the uniform norm $\|\cdot\|_\infty$.
• If n is a positive integer, then $\mathcal{C}^n(\mathbb{R}^d)$ is the space of n-times continuously differentiable functions $\mathbb{R}^d \longrightarrow \mathbb{R}$, $\mathcal{C}^n_b(\mathbb{R}^d)$ the space of n-times continuously differentiable functions $\mathbb{R}^d \longrightarrow \mathbb{R}$ such that all the partial derivatives of f up to the order n are bounded and $\mathcal{C}^n_c(\mathbb{R}^d)$ the space of n-times continuously differentiable functions $\mathbb{R}^d \longrightarrow \mathbb{R}$ with compact support. We can equip these spaces with the norm $\|\cdot\|_{\mathcal{C}^n(\mathbb{R}^d)}$: $\|f\|_{\mathcal{C}^n(\mathbb{R}^d)} := \sum_{k=0}^n \|D^k f\|_\infty$.
• If $(\mathcal{X}, \mu)$ is a measured space, then $L^p(\mathcal{X}, \mu)$ with $p \geq 1$ is the Lebesgue space of order p. It is the quotient of the function space $\mathcal{X} \longrightarrow \mathbb{R}$ which are of power p integrable with respect to μ by the space of zero functions μ-almost everywhere on $\mathcal{X}$. Two functions f and g are in the same equivalence class in the quotient $L^p(\mathcal{X}, \mu)$ if and only if $f = g$ μ-almost everywhere. We can equip $L^p(\mathcal{X}, \mu)$ with the norm $\|\cdot\|_{L^p(\mathcal{X},\mu)}$: $\|f\|_{L^p(\mathcal{X},\mu)} := \int |f|^p d\mu$. If $(\mathcal{X}, \mu) = (\mathbb{R}^d, \lambda)$, we will denote by $L^p(\mathbb{R}^d, \lambda) = L^p(\mathbb{R}^d)$.
• $H^s(\mathbb{R}^d)$ is the Sobolev space of fractionnal order $s \geq 0$. It is the space of functions $f \in L^2(\mathbb{R}^d, \mu)$ such that the function $\xi \longmapsto (1 + |\xi|^2)^{s/2}|\hat{f}(\xi)|$ is in $L^2(\mathbb{R}^d)$. We can equip $H^s(\mathbb{R}^d)$ with the norm $\|\cdot\|_{H^s(\mathbb{R}^d)}$: $\|f\|_{H^s(\mathbb{R}^d)} := (\int (1 + |\xi|^2)^s |\hat{f}(\xi)|^2 d\xi)^{1/2}$.
• If $(E, \|\cdot\|_E)$ is a normed vector space, then E^* is the dual space of E. It is the space of continuous linear forms $E \longrightarrow \mathbb{R}$. The image of $x \in E$ by $f \in E^*$ will be denoted by $f(x)$ or $\langle f, x\rangle$. We can equip E^* with the dual norm $\|\cdot\|_{E^*}$: $\|f\|_{E^*} := \sup\{|\langle f, x\rangle|,\ x \in E \text{ and } \|x\|_E \leq 1\}$.

Introduction

Chapter 1

Introduction

We consider a group of agents subject to various interactions, and we would like to study the collective behavior over time. There are three scales depending on the precision of the desired observation: microscopic, mesoscopic and macroscopic. In this introduction, we will begin by explaining what these three descriptions mean, what their advantages and inconvenients are, and which description is adopted depending on the physical context. We will give examples of models for each, and we will explain how to pass from one description to another by passing to the limit. Then we will present the model that will be considered throughout this manuscript, with the results obtained chapter by chapter.

1.1 Mathematical models in kinetic theory

1.1.1 Presentation

We consider a mathematical model which describes over time the evolution of a system composed of a large number of agents. This model is a partial differential equation (PDE) whose unknown is a distribution function in the space X of the phases of the agents. In the case of a dilute gas for example, the PDE describes its evolution over time and the agents are the particles of the gas. We admit that the system is observed in a time interval $[0, T]$, $(T \in]0, +\infty])$, then the solution of the PDE is a nonnegative function $f : (t, x) \longmapsto f(t, x)$ where $t \geq 0$ and $x \in X$.

The phase space X depends on the context and on how the state of an agent is determined. For example in the case of a gas, if we consider that it is contained in a domain $\Omega \subset \mathbb{R}^d$ ($d = 3$ in general), then each particle is characterized by its position $x \in \Omega$ and its velocity $v \in \mathbb{R}^d$. The phase space is therefore $\Omega \times \mathbb{R}^d$. In population dynamics, we can take the phenotypic traits $y \in \mathbb{R}$ of each agent within a population contained in a domain $\Omega \subset \mathbb{R}^d$ ($d = 3$ in general). So in this case, the phase space is $\Omega \times \mathbb{R}$. We have an homogeneous model if the phase space does not contain the space variable. Some authors consider that the unknown is a probability density $f(t, x)$, $t \geq 0$, $x \in X$. It is more natural to see f as a time-dependent probability measure. Of course, if f is a time-dependent density measure, the two coincide. In the general framework, a such model is called a mesoscopic model and we speak of a kinetic model when we take into account the velocities of each agent. We will consider in this text only kinetic models.

Kinetic theory consists of studying kinetic models. In 1738, Daniel Bernoulli laid the foundations of the kinetic theory of gases by publishing his book *Hydrodynamica*. He explained there that a gas is composed of a large number of molecules moving in all directions, that their impact on a surface causes the pressure of the gas and that their average kinetic energy determines the temperature of the gas. In 1857 Rudolf Clausius developed a more sophisticated version of the theory. In which the molecular movements included translations, rotations and vibrations [27]. It was only a few years later that James Clerk Maxwell formulated one of the main bases of the kinetic theory of gases according to which molecular collisions lead to an equalization of temperatures and therefore a tendency to

equilibrium [63].

Let us now move on to the mathematical description of the movement of agents. When no force acts on the system, it is considered that the agents to move at constant velocity along a straight line. If an agent is moving at velocity v and is at the position x at time t, then it was at time $t = 0$ at the position $x - tv$ while moving at the same velocity v. Since $f(t, x, v)dxdv$ denotes the number of agents in an infinitesimal volume $(dxdv)$ centered at (x, v), then we have $f(t, x, v) = f(0, x - tv, v)$ therefore f is a solution in the weak sense of the classical transport equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = 0 \tag{1.1.1}$$

where $\nabla_x f$ denotes the gradient of the function $x \longmapsto f(t, x, v)$. In the case where a macroscopic force F acts on the system, then the trajectories of the agents are deviated by it. So f is solution of the following equation called Vlasov's linear equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + \nabla_v \cdot (Ff) = 0 \tag{1.1.2}$$

where $\nabla_v \cdot F$ denotes the divergence of the vector function $v \longmapsto F(t, x, v)$. The two equations (1.1.1) and (1.1.2) have one thing in common. The right-hand side is null, which means that no interaction takes place between the agents.

1.1.2 The Boltzmann equation

We now want to take into account the interactions between the agents of the system. We call binary collision a process in which when two agents are close enough, then their respective trajectory will be strongly deviated during a very short time. We will assume that only binary collisions take place.

In 1872, Ludwig Boltzmann [13] developed a quadratic collision operator Q describing the interactions of (assumed to be identical) particles within a dilute gas that occupies the three-dimensional space. For a distribution function $f = f(t, x, v) \geq 0$, the operator Q is defined by

$$Q(f, f) := \iint_{\mathbb{R}^3 \times \mathbb{S}^2} (f' f'_\star - f f_\star) q(v - v_\star, \omega) dv_\star d\omega \tag{1.1.3}$$

with $f' = f(t, x, v')$, $f_\star = f(t, x, v_\star)$ and $f'_\star = f(t, x, v'_\star)$. The velocities v and $v_\star$ represent the velocities of the two particles before the collision while v' and $v'_\star$ designate the velocities of the two particles after the collision and they are written

$$\begin{cases} v' = v - \langle v - v_\star, \omega \rangle \omega, \\ v'_\star = v_\star + \langle v - v_\star, \omega \rangle \omega \end{cases}$$

where the parameter $\omega \in \mathbb{S}^2$ corresponds to the angle of deviation θ taken between v and v'. We see that the map $(v, v_\star) \longmapsto (v', v'_\star)$ is invertible. We then say that the collisions are microreversible. If we have two particles with respective velocitiy v' and $v'_\star$ after a collision, then we can find their velocitiy v and $v_\star$ before the collision. Boltzmann also considered that collisions are elastic, that is to say that the momentum and the kinetic energy are preserved during a collision and we have therefore

$$\begin{cases} v' + v'_\star = v + v_\star \\ |v'|^2 + |v'_\star|^2 = |v|^2 + |v_\star|^2. \end{cases}$$

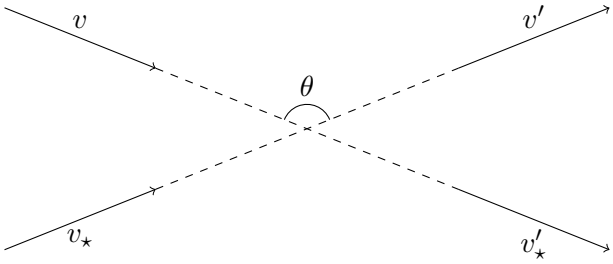


Figure 1.1: Example of an elastic binary collision. ω is the vector of norm 1 and angle θ.

The function q in (1.1.3) is called the Boltzmann collision kernel. It is a nonnegative function that only depends on $|v - v_\star|$ and $|\langle v - v_\star, \omega\rangle|$. Since $\langle v - v_\star, \omega\rangle = |v - v_\star| \cos(\theta)$, q therefore only depends on $|v - v_\star|$ and the cosine of the angle of deviation θ. It can be seen as a probability rate over all possible choices for the parameter $\omega \in \mathbb{S}^2$. In the particular case of hard spheres, in other words if we consider that the molecules of gas are assimilated to billiard balls of radius r, then the collision kernel is proportional to $|\langle v - v_\star, \omega\rangle|$. We have $q(v - v_\star, \omega) = 2r^2|\langle v - v_\star, \omega\rangle|$.

By definition of the collision operator Q (1.1.3), we notice that we can write it as a difference between a gain term Q_+ and a loss term Q_-

$$Q(f, f) = Q_+(f, f) - Q_-(f, f) \tag{1.1.4}$$

where

$$Q_+(f, f) := \iint_{\mathbb{R}^3\times\mathbb{S}^2} f' f'_\star q(v - v_\star, \omega) dv_\star d\omega \qquad \text{and} \qquad Q_-(f, f) := \iint_{\mathbb{R}^3\times\mathbb{S}^2} f f_\star q(v - v_\star, \omega) dv_\star d\omega.$$

Q_+ is a gain rate of particles at velocity v after collision of two particles at velocities v' et $v'_\star$. And Q_- is a loss rate of particles at velocity v due to collisions of particles of velocity $v_\star$.

In the end, the Boltzmann equation is written

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q(f, f) \tag{1.1.5}$$

or, when a macroscopic force F is present,

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + \nabla_v \cdot (Ff) = Q(f, f).$$

The Boltzmann equation (1.1.5) means that in a dilute gas, the particles are subjected to binary elastic collisions. And apart from any interaction, the particles move at constant velocity along a straight line or only under the influence of the force F in the case where a macroscopic force F is present.

This equation is very important in the kinetic theory of gases since it has made it possible to check the conservation over time of certain quantities: mass, momentum and total energy. But the most important result is the H Theorem, which says that the quantity $\int f \log(f) dv$ which is interpreted as the entropy of the system, decreases over time. Boltzmann also finds the result [14], [15] established by Maxwell a few years earlier [61], [62]. That is, at equilibrium, f is of the form

$$f(v) = \frac{\rho}{(2\pi\theta)^{3/2}} \exp\left(-\frac{|v - u|^2}{2\theta}\right)$$

where ρ, u and θ denote respectively the density, mean velocity and temperature of the gas (independent of (t, x)) at equilibrium.

The first global existence result of the Cauchy problem for the Boltzmann equation is due to Ukai [84] for the small datum inhomogeneous. The greatest breakthrough remains in 1988 where DiPerna and Lions showed the existence of a so-called renormalized solution to the Cauchy problem for the Boltzmann equation for any initial data f_0 satisfying natural estimates [26], [39].

1.1.3 Some kinetic models

Vlasov equation

The Vlasov equation, introduced by Anatoly Vlassov [91] in 1938, is the following kinetic equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + \nabla_v \cdot (F[f]f) = 0 \tag{1.1.6}$$

where the unknown $f = f(t, x, v)$, $t \geq 0$, $x \in \mathbb{R}^d$ and $v \in \mathbb{R}^d$ is a probability distribution and the map $F : \mu \longmapsto F[\mu]$ a given force. In some cases, the vector function $F[f]$ is given by a convolution product between a kernel $k = k(x, v)$ and the unknown f. In other words, for a measure π with two variables we have

$$F[\pi](x, v) := (k * \pi)(x, v) = \iint_{\mathbb{R}^d \times \mathbb{R}^d} k(x - y, v - w) d\pi(y, w). \tag{1.1.7}$$

The force F can only depend on x if for example the kernel k is independent of v. In this case, the force term simplifies and becomes $\nabla_v \cdot (F[f]f) = F[f] \cdot \nabla_v f$. For example if $k(x, v) := -\nabla U(x)$ where $U(x) = 1/(4\pi|x|)$, $x \in \mathbb{R}^3$, then (1.1.6) becomes what is called the Vlasov-Poisson equation and we have $\nabla_v \cdot (F[f]f) = -\nabla\Phi(t, x) \cdot \nabla_v f$ where $-\Delta\Phi(t, x) = \int f(t, x, v)dv$.

This equation is very important in plasma physics because it describes the evolution over time of the distribution of particles in a plasma, neglecting the effect of binary collisions. We can all the same consider in this model that the binary collisions are not negligible. To do this, simply add the Boltzmann collision operator (1.1.3) to the right-hand side of the equation (1.1.6). We say that the equation (1.1.2) is the linear form of the Vlasov equation (1.1.6) because the force F does not depend on the unknown f.

Fokker-Planck equation

This is the following kinetic equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = \nabla_v \cdot (\nabla_v f + fv). \tag{1.1.8}$$

where the unknown $f = f(t, x, v)$, $t \geq 0$, $x \in \mathbb{R}^d$ and $v \in \mathbb{R}^d$ is a probability distribution. It is a diffusion equation. The term on the right-hand side is non-zero, which means that the agents interact with each other. But contrary to the Boltzmann equation, the collective behavior is influenced by the random movement of each agent, in a way each agent undergoes multiple small collisions. The term on the right-hand side is therefore not a collision term, but a sum between a diffusion term $\Delta_v f$, the Laplacian of the function $v \longmapsto f(t, x, v)$ (to be taken in the weak sense) and a drift term $\nabla_v \cdot (fv)$. The equation (1.1.8) is named in honor of Adriaan Fokker and Max Planck. We can combine the two equations (1.1.6) and (1.1.8). One obtains the following equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + \nabla_v \cdot (F[f]f) = \nabla_v \cdot (\nabla_v f + fv). \tag{1.1.9}$$

which is the Vlasov-Fokker-Planck equation.

A kinetic equation for traffic flow

Dimarco, Tosin, Zanella [38] propose a kinetic description of traffic flow. In their model, it is assumed that the vehicles modify their velocity via interactions with other vehicles located at a given constant headway $H > 0$. Each vehicle will adapt its velocity according to the vehicle in front of it but will not

feel the influence of the one behind. Denoting v the velocity of a vehicle, and $v_\star$ the velocity of the vehicle in front of it, then the new velocities v' and $v'_\star$ after an interaction are given by

$$\begin{cases} v' = v + \gamma\lambda(\rho)(v_\star - v) + D(v)\eta, \\ v'_\star = v_\star. \end{cases} \tag{1.1.10}$$

In (1.1.10), $\gamma > 0$ is a time-scale factor, $\lambda(\rho) > 0$ is the reactiveness of the driver which is expressed as a function of the local vehicle density ρ : $\rho(t,x) := \int f(t,x,v)dv$, $D : [0,1] \longrightarrow \mathbb{R}^+$ is a function that measures the local relevance of the random fluctuations and η is a centered random variable with positive variance (in other words its law belongs to $\mathcal{P}_2^0(\mathbb{R})$ and its variance σ^2 is positive: $\sigma^2 > 0$), which models random fluctuations in the driver behavior. The (1.1.10) system is a system inspired by traffic models of type *follow-the-leader* [49].

Therefore, the distribution function of the vehicles located at the point $x \in \mathbb{R}^2$ and circulating at the (non-dimensional) velocity $v \in [0,1]$ at the instant $t \geq 0$ is solution of the following Enskog-type equation

$$\begin{aligned} \partial_t \int_0^1 \varphi(v) f(t,x,v)dv + \partial_x \int_0^1 v\varphi(v) f(t,x,v)dv \\ = \frac{1}{2}\int_0^1 \int_0^1 \langle \varphi(v') - \varphi(v) \rangle f(t,x,v) f(t,x+H,v_\star) dv dv_\star \end{aligned} \tag{1.1.11}$$

where $\varphi : [0,1] \longrightarrow \mathbb{R}$ is a test function.

1.1.4 Other models

We have presented in this section only kinetic models, in other words models involving the velocity of each agent. But there are many other "kinetic" models that do not take velocities into account. These models are called mesoscopic.

In biology, there exist many mesoscopic models that describe the growth of a population (individuals, cells, bacteria,...) over time. For example, the model introduced by Raoul in Ref. [72] describes the dynamics of a sexual population. This model considers that at each instant $t \geq 0$, the population is structured by a phenotypic trait $y \in \mathbb{R}$ and a space variable x. The velocity is therefore not taken into account. The population is therefore represented by a density $n = n(t,x,y)$ and is the solution of what is called the *Spatially structured Infinitesimal Model* (SIM) [64]. Moreover, there exists a connection between the SIM and the Kirkpatrick-Barton [57] model, a model that was introduced in 1997 and is widely used in evolutionary ecology.

In financial mathematics, one could, for example, cite a simple kinetic model introduced by Cordier, Pareschi, Toscani [29] involving both exchanges between agents and speculative exchanges. We could also cite another model, that of Cordier, Pareschi, Piatecki [28]. In Ref. [28], they describe the behavior of a simple financial market where agents can create their own portfolio according to two investment alternatives: a stock and a bond. Their model uses the methods of kinetic theory and consists of a Boltzmann-type equation of the wealth distribution of the agents coupled with an equation for the price of the stock. In these models, we take into account the wealth of the agents $w \in \mathbb{R}^+$ at time $t \geq 0$, and not their velocity.

1.2 Macroscopic models

1.2.1 Introduction to fluid dynamics

Let's go further. We would like to act on a system of agents or see what impact it has on its environment. In both cases, the quantities of interest are the same and are those that can be observed naturally or with measuring devices. They are obtained by carrying out means on the agents: mean density,

mean velocity, temperature. A macroscopic model is a PDE system whose unknowns are these mean quantities.

The classic macroscopic models are those of fluid dynamics (hydrodynamic models). The best known are the Euler and Navier-Stokes equations. They were established by applying Newton's second law to each infinitesimal volume of the considered fluid [58]. This method has an advantage because all hydrodynamic models can be obtained in this way. But it has an inconvenient. This is because the equations of state and the transport coefficients such as viscosity and thermal conductivity are given in the form of experimental data and not in the form of data linked to the dynamics of agents. However, in the case of gases and plasmas for example, the interactions between the particles are very elementary. So that one can hope to express thermodynamic functions and transport coefficients in terms of purely mechanical data concerning the collisions between gas molecules.

We will consider in this text that all macroscopic models are hydrodynamic models. Usually, a fluid is defined as a continuous medium. In other words, it is a set of points which, at a time t, fills a smooth domain of $\mathbb{R}^d$ ($d = 2$ or 3 in general). Fluid dynamics is based on three fundamental laws (or equations): The continuity equation, the momentum equation and the balance energy equation.

The continuity equation is the following

$$\partial_t \rho + \mathrm{div}_x(\rho u) = 0 \tag{1.2.1}$$

where $\rho = \rho(t, x)$ is the density of the fluid at point $x \in \mathbb{R}^d$ and at time $t \geq 0$, and $u = u(t, x)$ the velocity (in $\mathbb{R}^d$) of a fluid particle at point $x \in \mathbb{R}^d$ and at time $t \geq 0$.

The momentum equation derives from the fundamental principle of dynamics applied to the fluid under consideration. This equation is written as follows

$$\partial_t(\rho u) + \mathrm{div}_x(\rho u \otimes u - S) = \rho F \tag{1.2.2}$$

where S is the stress tensor and F an external force field (for example gravity or the Lorentz force in the case of a plasma).

The last equation involves the internal energy of the fluid per unit mass $e = e(t, x)$. The total energy of the system $E = E(t, x)$ is $|u|^2/2 + e$. It is the sum of kinetic energy and internal energy. The balance energy equation reads as

$$\partial_t \left(\rho e + \frac{\rho |u|^2}{2} \right) + \mathrm{div}_x \left(\rho e u + \rho u \frac{|u|^2}{2} \right) = -\mathrm{div}_x(Q) + \mathrm{div}_x(Su) + \rho F \cdot u \tag{1.2.3}$$

where Q is the heat flux.

In the momentum and balance energy equations (1.2.2), (1.2.3), F is a given vector field. While the density ρ, the velocity field u, the internal energy e, the stress tensor S and the heat flow Q are unknown. However, these quantities are generally not independent. They are linked by equations of state which depend on the fluid considered.

1.2.2 Classical models in fluid dynamics

The compressible Euler system

The compressible Euler equations are the system (1.2.1), (1.2.2) and (1.2.3) in the case of an ideal fluid. An ideal fluid is one in which the effects of viscosity and thermal conductivity can be neglected. In this case, $Q = 0$ and the stress tensor is written $S = -pI_d$ where the scalar p is the pressure. Thus, the system of equations (1.2.1), (1.2.2) and (1.2.3) becomes

$$\begin{cases} \partial_t \rho + \mathrm{div}_x(\rho u) = 0, \\ \partial_t(\rho u) + \mathrm{div}_x(\rho u \otimes u + pI_d) = \rho F, \\ \partial_t \left(\rho e + \frac{\rho |u|^2}{2} \right) + \mathrm{div}_x \left(\rho e u + \rho u \frac{|u|^2}{2} \right) + \mathrm{div}_x(pu) = \rho F \cdot u. \end{cases} \tag{1.2.4}$$

The unknowns in the (1.2.4) system are ρ, u, e and p. But ρ, e and p are not independent. These three quantities are linked by the equations of state. We can express the pressure p and the internal energy e as a function of the density ρ and the temperature θ: $p = p(\rho, \theta)$ and $e = e(\rho, \theta)$. Thus, the compressible Euler system (1.2.4) is a system of $d+2$ PDE on the unknowns ρ, u and θ (one equation for ρ, one for θ and one for each coordinate of u).

In the case of an ideal gas for example, the equations of state are $p = k\rho\theta$ where k is the Boltzmann constant and $e = k\theta/(\gamma - 1)$ where $\gamma > 1$ is a constant called the adiabatic exponent. For an ideal gas where the particles have n degrees of freedom, $\gamma = 1 + 2/n$. We have for example $\gamma = 5/3$ for a monatomic gas and $\gamma = 7/5$ for a diatomic gas. If we choose a temperature scale such that $k = 1$, then the system (1.2.4) for an ideal gas is

$$\begin{cases} \partial_t \rho + \mathrm{div}_x(\rho u) = 0, \\ \\ \partial_t(\rho u) + \mathrm{div}_x(\rho u \otimes u + \rho\theta I_d) = \rho F, \\ \\ \partial_t \left(\dfrac{\rho\theta}{\gamma - 1} + \dfrac{\rho|u|^2}{2} \right) + \mathrm{div}_x \left(\dfrac{\gamma\rho\theta u}{\gamma - 1} + \rho u \dfrac{|u|^2}{2} \right) = \rho F \cdot u. \end{cases} \tag{1.2.5}$$

If the temperature θ is constant, then the last equation in (1.2.5) is obtained directly with the other two. Thus, the system (1.2.5) becomes closed in the unknowns ρ and u and we have

$$\begin{cases} \partial_t \rho + \mathrm{div}_x(\rho u) = 0, \\ \\ \partial_t(\rho u) + \mathrm{div}_x(\rho u \otimes u + \rho\theta I_d) = \rho F \end{cases} \tag{1.2.6}$$

with θ a real constant which is therefore the constant temperature of the system. This system (1.2.6) is called isothermal compressible Euler system.

The compressible Navier-Stokes system

Here, the fluid is no longer ideal. The effects of viscosity and thermal conduction must be taken into account. Generally, thermal conduction is modeled by Fourier's law, in other words $Q = -\kappa\nabla_x\theta$ where κ is the thermal conductivity. It is a constant that depends on temperature and pressure. But since the pressure depends on the density and the temperature, κ therefore depends on the density and the temperature: $\kappa = \kappa(\rho, \theta) > 0$. While the viscous forces are modeled by adding a corrective term to the pressure in the stress tensor S. Still following the Fourier law, we have $S = -pI_d + \lambda \mathrm{div}_x(u) I_d + \mu D(u)$ where D is defined by

$$D(u) := \nabla_x u + {}^t\nabla_x u - \frac{2}{d}\mathrm{div}_x(u) I_d$$

and λ, μ are two scalars called viscosity coefficients. They also depend on pressure and temperature, and therefore as for κ, $\lambda = \lambda(\rho, \theta) > 0$ and $\mu = \mu(\rho, \theta) > 0$. The corrective term is therefore a linear combination between $\mathrm{div}_x(u) I_d$ and $D(u)$.

We thus obtain the Navier-Stokes equations for compressible fluids

$$\begin{cases} \partial_t \rho + \mathrm{div}_x(\rho u) = 0, \\ \\ \partial_t(\rho u) + \mathrm{div}_x(\rho u \otimes u) = -\nabla_x p(\rho, \theta) + \nabla_x(\lambda(\rho, \theta)\mathrm{div}_x(u)) + \mathrm{div}_x(\mu(\rho, \theta) D(u)) + \rho F, \\ \\ \partial_t \left(\rho e(\rho, \theta) + \dfrac{\rho|u|^2}{2} \right) + \mathrm{div}_x \left(\rho e(\rho, \theta) u + \rho u \dfrac{|u|^2}{2} \right) = -p(\rho, \theta)\mathrm{div}_x(u) + \mathrm{div}_x(\kappa \nabla_x \theta) \\ \qquad\qquad + \dfrac{1}{2}\mu(\rho, \theta) D(u) : D(u) \\ \qquad\qquad + \lambda(\rho, \theta)\mathrm{div}_x(u)^2. \end{cases} \tag{1.2.7}$$

Hydrodynamics models for incompressible fluids

We now turn to the case where the fluid is incompressible. A fluid is incompressible if its volume is constant whatever the pressure it undergoes. In fluid mechanics, it is often assumed that the studied fluid is incompressible to simplify the equations. Mathematically, this means that ρ is constant.

In the case of an ideal fluid, the first two equations of the compressible Euler system (1.2.4) under the constraint ρ is constant becomes

$$\begin{cases} \operatorname{div}_x(u) = 0, \\ \\ \partial_t u + \operatorname{div}_x(u \otimes u) = -\nabla_x \pi + F \end{cases} \tag{1.2.8}$$

where $\pi := p/\rho$. Since ρ is constant, the pressure p only depends on the temperature θ, so π depends on the temperature. However, there is no need for an equation of state to determine π. Indeed, by taking the divergence on both side of the second equation in (1.2.8), we have

$$-\Delta_x \pi = \operatorname{div}_x\left(\partial_t u + \operatorname{div}_x(u \otimes u)\right) - \operatorname{div}_x(F) = \operatorname{Tr}((\nabla_x u)^2) - \operatorname{div}_x(F).$$

So π is the solution of a Laplace equation with unknown u. The system (1.2.8) is called incompressible Euler system.

Next, if the considered fluid is no longer ideal, we start from the Navier-Stokes equations (1.2.7) and we assume that the density ρ is constant. The first equation of (1.2.7) gives $\operatorname{div}_x(u) = 0$. Moreover, if we suppose that the coefficient of viscosity μ is constant, the second equation of (1.2.7) becomes

$$\rho(\partial_t u + \operatorname{div}_x(u \otimes u)) + \nabla_x p(\rho, \theta) = \mu \Delta_x u + \rho F.$$

We define the kinematic viscosity $\nu := \mu/\rho$ and we set $\pi := p/\rho$. We arrive at the incompressible Navier-Stokes equations

$$\begin{cases} \operatorname{div}_x(u) = 0, \\ \\ \partial_t u + \operatorname{div}_x(u \otimes u) + \nabla_x \pi = F + \nu \Delta_x u. \end{cases} \tag{1.2.9}$$

1.2.3 Hydrodynamic limit for the Boltzmann equation

A kinetic description is much more complex than a hydrodynamic description, because the phase space is much larger. In practical applications, kinetic equations require extremely expensive calculations. This is the case for the Boltzmann equation because of the collision integral. It is therefore advantageous, when possible, to replace a kinetic model with a hydrodynamic model.

Let $f = f(t, x, v)$ be a solution of the Boltzmann equation (1.1.5). We define the macroscopic quantities ρ, u and θ by

$$\rho(t, x) := \int_{\mathbb{R}^d} f(t, x, v) dv, \tag{1.2.10}$$

$$u(t, x) := \frac{1}{\rho(t, x)} \int_{\mathbb{R}^d} v f(t, x, v) dv, \tag{1.2.11}$$

$$\theta(t, x) := \frac{1}{d\rho(t, x)} \int_{\mathbb{R}^d} |v - u(t, x)|^2 f(t, x, v) dv. \tag{1.2.12}$$

ρ represents the density of particles, u the mean velocity and θ the temperature. In the Boltzmann equation, we have conservation of mass, momentum and energy. That is, $\int \varphi(v) Q(f, f) dv = 0$ for

$\varphi(v) = 1$, v and $|v|^2$ where Q is the Boltzmann collision operator defined by (1.1.3).We multiply the Boltzmann equation by 1, v, $|v|^2/2$ then we integrate in v. We obtain

$$\begin{cases} \partial_t \left(\displaystyle\int_{\mathbb{R}^d} f dv \right) + \operatorname{div}_x \left(\displaystyle\int_{\mathbb{R}^d} v f dv \right) = 0, \\ \partial_t \left(\displaystyle\int_{\mathbb{R}^d} v f dv \right) + \operatorname{div}_x \left(\displaystyle\int_{\mathbb{R}^d} v \otimes v f dv \right) = 0, \\ \partial_t \left(\displaystyle\int_{\mathbb{R}^d} \frac{|v|^2}{2} f dv \right) + \operatorname{div}_x \left(\displaystyle\int_{\mathbb{R}^d} v \frac{|v|^2}{2} f dv \right) = 0. \end{cases} \tag{1.2.13}$$

Then by definition of ρ, u and θ, (1.2.13) takes the form

$$\begin{cases} \partial_t \rho + \operatorname{div}_x(\rho u) = 0, \\ \partial_t(\rho u) + \operatorname{div}_x(\rho u \otimes u) + \nabla_x(\rho\theta) = -\operatorname{div}_x \left(\displaystyle\int_{\mathbb{R}^d} A(v-u) f dv \right), \\ \partial_t \left(\rho \dfrac{|u|^2}{2} + \dfrac{d\rho\theta}{2} \right) + \operatorname{div}_x \left(\rho u \dfrac{|u|^2}{2} + \dfrac{(d+2)\rho\theta u}{2} \right) = -\operatorname{div}_x \left(\displaystyle\int_{\mathbb{R}^d} B(v-u) f dv \right) \\ \qquad\qquad - \operatorname{div}_x \left(\displaystyle\int_{\mathbb{R}^d} A(v-u) u f dv \right) \end{cases} \tag{1.2.14}$$

with $A(v) := v \otimes v - |v|^2 I_d/d$ and $B(v) := (|v|^2 - (d+2))v/2$. The left-hand side in (1.2.14) coincides with the compressible Euler system (1.2.5) for gases with $\gamma = 1 + 2/d$. The right-hand side, on the contrary, depends on the solution of the Boltzmann equation f and in general cannot be determined as a function of the macroscopic variables ρ, u and θ.

The ideal would be that f vanishes the Boltzmann collision operator, $Q(f, f) = 0$. In this case, f would be a local Maxwellian (or a local thermodynamic equilibrium) in other words

$$f = \frac{\rho}{(2\pi\theta)^{d/2}} \exp\left(-\frac{|v-u|^2}{2\theta} \right).$$

And the system (1.2.14) would be closed on the unknowns ρ, u and θ because the right hand side would be null. In other words, we would have

$$\begin{cases} \partial_t \rho + \operatorname{div}_x(\rho u) = 0, \\ \partial_t(\rho u) + \operatorname{div}_x(\rho u \otimes u) + \nabla_x(\rho\theta) = 0, \\ \partial_t \left(\rho \dfrac{|u|^2}{2} + \dfrac{d\rho\theta}{2} \right) + \operatorname{div}_x \left(\rho u \dfrac{|u|^2}{2} + \dfrac{(d+2)\rho\theta u}{2} \right) = 0. \end{cases} \tag{1.2.15}$$

We introduce into the Boltzmann equation the Knudsen number Kn defined as being the mean length l that a particle travels without undergoing a collision over the length L of the considered object $Kn := l/L$. After scaling the constants, the Boltzmann equation becomes

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = \frac{1}{Kn} Q(f, f). \tag{1.2.16}$$

The hydrodynamic limit consists in making Kn tend towards 0. Then to show that the solution f_ε, $\varepsilon = Kn$ of (1.2.16) tends towards a local thermodynamic equilibrium f when $\varepsilon \longrightarrow 0$ and to deduce from it the limits equations for the hydrodynamic fields associated with f.

The first hydrodynamic limit studies for the Boltzmann equation are due to Hilbert [54], then to Chapman-Enskog [46] who were inspired by it. Their methods make it possible to derive the Boltzmann equation in order to formally recover the Euler and Navier-Stokes equations (compressible or not).

1.2.4 Macroscopic models for traffic flows

To model traffic flow, we use macroscopic models when we consider the flow of vehicles as a continuous medium. We are therefore dealing with equations derived from fluid dynamics.

A typical example of a first-order model is the famous model independently developed by Lightill, Whitham [60] and Richards [75] (called the LWR model) in the 1950s. In this model, we are interested in the density of vehicles $\rho(t,x)$ along the road. The LWR model is the following equation of unknown ρ

$$\partial_t \rho + \partial_x(\rho V(\rho)) = 0 \tag{1.2.17}$$

where V is the corresponding preferential velocity. It is a non-increasing given function in ρ, non-negative for $\rho \in [0, \rho_m]$ where ρ_m is the density of vehicles during a total traffic jam. However, such an equilibrium model is unable to describe flows in which there are few slow drivers.

Let us now move on to second-order models with the example of the Payne-Whitham model [69], [92] (called the PW model). These models are always macroscopic but unlike first-order models such as the LWR (1.2.17) model, these include two equations which are strongly inspired by one-dimensional fluid equations. The PW model is the following system of two equations

$$\begin{cases} \partial_t \rho + \partial_x(\rho v) = 0, \\ \\ \partial_t v + v\partial_x v + \dfrac{1}{\rho} p'(\rho)\partial_x \rho = \dfrac{1}{\tau}(V(\rho) - v) + \nu \partial_x^2 v \end{cases} \tag{1.2.18}$$

of unknowns ρ and v which are respectively the density and the velocity. The first equation in (1.2.18) is conservation of mass while the second equation in (1.2.18) mimics the momentum equation (1.2.2). However, there are big differences between traffic flow and fluid flows. In reality, there is no conservation of momentum in the first. The pressure term p that we have in fluid dynamics is replaced by an anticipation factor. That is, a term that describes the reaction of a mean driver to a spatial variation in the concentration of vehicles in front of him. In the second equation of the system (1.2.18), we have $p = p(\rho)$ which is inspired by the dynamics of gases by an isothermal law: $p(\rho) = \rho$. And τ, ν are non-negative constants.

However, the PW (1.2.18) model has an inconvenient [33]. The characteristic velocities in (1.2.18) are $v = \pm\sqrt{p'(\rho)}$. Consequently part of the information travels faster than the velocity v of the vehicles. Since $p(\rho) = \rho$, the anticipation factor involves the derivative of the pressure with respect to x. This is incorrect because if we suppose for example that in front of a driver moving at velocity v, the density increases with respect to x, but decreases with respect to $x - tv$, then the model PW (1.2.18) then predicts that this conductor will slow down since the density increases relative to x. On the contrary, any reasonable driver would accelerate since this denser traffic is moving faster than him. This problem has been rectified by Aw, Rascle, Zhang [5], [93] with the following model (called ARZ model) of unknowns ρ and v

$$\begin{cases} \partial_t \rho + \partial_x(\rho v) = 0, \\ \\ \partial_t(v + p(\rho)) + v\partial_x(v + p(\rho)) = 0 \end{cases} \tag{1.2.19}$$

where this time the pressure term p is a smooth function which vanishes at 0: $p(0) = 0$, which has a strictly increasing derivative: $p'(\rho) > 0$ for all $\rho > 0$ and such that the function $\rho \longmapsto \rho p(\rho)$ is strictly convex. A typical example is $p(\rho) = \rho^\gamma$ with $\gamma > 0$.

Example of an hydrodynamic limit

In Ref. [38], an ARZ-type model is obtained from a hydrodynamic limit of the equation (1.1.11). If consistent with a hydrodynamic regime, we assume that the progress H is small, then we approach

$f(t,x+H,v_\star) \simeq f(t,x,v_\star) + H\partial_x f(t,x,v_\star)$ by a truncated Taylor expansion of order 1. Therefore, we can approximate the equation (1.1.11) by

$$\begin{aligned}\partial_t \int_0^1 \varphi(v) f(t,x,v)dv + \partial_x \int_0^1 v\varphi(v) f(t,x,v)dv \\ = \frac{1}{2}\int_0^1\int_0^1 \langle \varphi(v') - \varphi(v)\rangle f(t,x,v) f(t,x,v_\star)dvdv_\star \\ + \frac{H}{2}\int_0^1\int_0^1 \langle \varphi(v') - \varphi(v)\rangle f(t,x,v)\partial_x f(t,x,v_\star)dvdv_\star.\end{aligned} \tag{1.2.20}$$

After scaling space and time, the model (1.2.20) can be written for any $\varepsilon > 0$

$$\partial_t \int_0^1 \varphi(v) f(t,x,v)dv + \partial_x \int_0^1 v\varphi(v) f(t,x,v)dv = \frac{1}{\varepsilon}(Q(f,f),\varphi) + \frac{H}{2}(Q(f,\partial_x f),\varphi) \tag{1.2.21}$$

where Q is the collision operator defined for every test function $\varphi : [0,1] \longrightarrow \mathbb{R}$ by

$$\langle Q(f,g),\varphi\rangle := \int_0^1\int_0^1 \langle \varphi(v') - \varphi(v)\rangle f(t,x,v) g(t,x,v_\star)dvdv_\star. \tag{1.2.22}$$

with v' is defined in (1.1.10). On the right hand side in (1.2.21), due to the presence of the partial derivative at x of f, the time scale of the second term is different from that of the first term. Two time scales are detected but this can be resolved by means of the following splitting (for details, see Ref. [44])

$$\begin{cases} \partial_t \int_0^1 \varphi(v) f(t,x,v)dv = \frac{1}{\varepsilon}(Q(f,f),\varphi), \\ \\ \partial_t \int_0^1 \varphi(v) f(t,x,v)dv + \partial_x \int_0^1 v\varphi(v) f(t,x,v)dv = \frac{H}{2}(Q(f,\partial_x f),\varphi). \end{cases} \tag{1.2.23}$$

The first equation in (1.2.23) describes quick local interactions among the vehicles. By letting $\varepsilon \longrightarrow 0$, we formally obtain $Q(f,f) = 0$. So $f = M_{\rho,u}$ is a local Maxwellian which is spanned by the local density of vehicles ρ and the bulk velocity of vehicles u defined as

$$\rho(t,x) := \int_0^1 f(t,x,v)dv \qquad \text{and} \qquad u(t,x) := \frac{1}{\rho(t,x)}\int_0^1 v f(t,x,v)dv.$$

In particular, we have

$$\int_0^1 M_{\rho,u}(v)dv = \rho \qquad \text{and} \qquad \frac{1}{\rho}\int_0^1 v M_{\rho,u}(v)dv = u.$$

In the second equation in (1.2.23), we have by (1.1.10) that $(Q(f,f),1) = (Q(f,f),v) = 0$. This expresses the slower transport of the local Maxwellian on the hydrodynamic spatio-temporal scale. By injecting $M_{\rho,u}$ into the second equation of (1.2.23), one obtains $(Q(M_{\rho,u},\partial_x M_{\rho,u}),1) = 0$ and $(Q(M_{\rho,u},\partial_x M_{\rho,u}),v) = \gamma\rho^2\lambda(\rho)\partial_x u$. We obtain the closed system of equations on ρ and u

$$\begin{cases} \partial_t\rho + \partial_x(\rho u) = 0, \\ \\ \partial_t(\rho u) + \partial_x(\rho E) = \frac{\gamma H}{2}\rho^2\lambda(\rho)\partial_x u \end{cases} \tag{1.2.24}$$

where

$$E := \frac{1}{\rho}\int_0^1 v^2 M_{\rho,u}(v)dv$$

is the energy of the local equilibrium distribution. The system (1.2.24) is an ARZ-type system.

1.3 Microscopic models

1.3.1 Definition

Consider a system composed of a number n of agents (n is large but finite). A microscopic model is a mathematical model that describes the evolution of each agent one by one. We therefore have a system of n equations with for each equation the unknown which is the state of the agent considered. Since we are in the kinetic case, the state of an agent i is determined by its position $x_i \in \mathbb{R}^d$ and its velocity $v_i \in \mathbb{R}^d$ ($d = 2$ or 3 in general). The natural phase space considered is $(\mathbb{R}^d \times \mathbb{R}^d)^n$.

The fundamental principle of dynamics makes it possible to describe the movement of agents in a classical way. The sum of the forces exerted on an agent is proportional to its acceleration (we will assume for simplicity that all the agents have the same mass equal to 1). Then each agent exerts a force on the others and is subject to external forces independent of the other agents. By noting $x_i(t)$ the position of the agent i at time $t \geq 0$, and $v_i(t)$ the velocity of the agent i at time $t \geq 0$, we obtain the following system

$$\begin{cases} \dfrac{d}{dt}x_i(t) = v_i(t), \\ \\ \dfrac{d}{dt}v_i(t) = G(x_i(t), v_i(t)) + \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} F(x_i(t) - x_j(t), v_i(t) - v_j(t)) \end{cases} \tag{1.3.1}$$

which is a system of $2n$ first-order differential equations. We add to this system the initial positions $x_0 := (x_1(0), \cdots, x_n(0)) \in (\mathbb{R}^d)^n$ and the initial velocities $v_0 := (v_1(0), \cdots, v_n(0)) \in (\mathbb{R}^d)^n$.

In (1.3.1), the term $G(x_i(t), v_i(t))$ represents the external forces acting on the agent i at time t. These forces are independent of the other agents. For example, we can have $G(x, v) = -\nabla U(x)$. It is the force imposed at the point x in a potential $U(x)$. This is the case for $U(x) = \alpha/|x - x_0|$, which is the gravitational potential of a star placed at the point $x_0 \in \mathbb{R}^3$ with $\alpha > 0$. The term $F(x_i(t) - x_j(t), v_i(t) - v_j(t))$ describes the force exerted by the agent j on the agent i. If $F(0, 0) = 0$, then no agent exerts any force on itself, we say in this case that there is no self-interaction. The coefficient $1/n$ placed in front just allows to induce a total force of order 1 which will lead to the mean-field limit. In some models, the function F may only depend on x and derive from a potential U: $F(x, v) = -\nabla U(x)$. This is the case for the stars where $U(x) = \alpha/|x|$, $\alpha > 0$ and $x \in \mathbb{R}^3$. It is the gravitational potential for the stars.

Microscopic models can also be random, in other words for each agent i, its position at time t and its velocity at time t depend on a random parameter. They are therefore no longer elements of $\mathbb{R}^d$ but random variables which taking values in $\mathbb{R}^d$, denoted $X_i(t)$ and $V_i(t)$. The map $t \longmapsto X_i(t)$ are the trajectories of the agent i.

A fundamental example of trajectories is the *Brownian motion* $(B(t))_t$, $t \geq 0$, which we usually start at 0: $B(0) = 0$. Consider a system of n agents where each follows an independent Brownian trajectory, and each starting from a random initial position $X_i(0)$. So we have for each i, $X_i(t) = X_i(0) + B_i(t)$. Since the random drifts are infinitesimal, it is more natural to write

$$\frac{dX_i(t)}{dt} = \frac{dB_i(t)}{dt}.$$

Only that doesn't make sense because the trajectories of Brownian motions are almost surely nowhere differentiable. Stochastic calculus makes it possible to give a sense to equations of this form. By convention, the system is therefore written for all $1 \leq i \leq n$

$$dX_i(t) = dB_i(t).$$

It can very well be made more complex by adding drift terms. For all $1 \leq i \leq n$,

$$dX_i(t) = dB_i(t) + \varphi(X_i(t))dt.$$

These equations can be written in integral form,

$$X_i(t) = X_0(t) + B_i(t) + \int_0^s \varphi(X_i(s))ds$$

for all $1 \leq i \leq n$.

We will rather suppose that the position and the velocity of each agent i satisfy the following system of stochastic differential equations

$$\begin{cases} dX_i(t) = V_i(t)dt, \\ \\ dV_i(t) = G(X_i(t), V_i(t))dt + \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} F(X_i(t) - X_j(t), V_i(t) - V_j(t))dt + \sigma B_i(t) \end{cases} \tag{1.3.2}$$

where σ is a diffusion coefficient that we will assume to be constant and $(B_i(t))_t$ independent Brownian motions. We add the random initial conditions $(X_1(0), \cdots, X_n(0))$, $(V_1(0), \cdots, V_n(0))$. We say that the microscopic model (1.3.2) is random when $\sigma \neq 0$ and that it is deterministic when $\sigma = 0$. In the deterministic case, the system (1.3.2) is simply the system (1.3.1).

1.3.2 Mean-field limit

We give ourselves a microscopic system composed of n agents. One can ask the following question: Is there a limit when n tends to infinity ? To study a such passage to the limit, it is first necessary to consider the problem at the right scale so that the interactions between the agents are weak enough so that the forces and influences exerted on a given agent remain finite at the limit. We want to identify a limit, in other words a mathematical object that is defined as the solution of a kinetic (or mesoscopic) model.

Most of the time, microscopic models have a very high number of agents. There can be up to 10^9 insects in a swarm or 10^{12} stars in a galaxy. What interests us is rather the global state of the system and not the state of each agent. Because each agent taken individually has a very weak action on the system and cannot generally be observed. We therefore need to observe a microscopic system on a statistical scale where, rather than studying the state of each agent one by one, we will rather ask ourselves what is the probability of falling on an agent in a given state or rather in a part of the phase space. We therefore wish to establish a statistical description of a model from an individual description. This corresponds to a limit of the law of large numbers type in a probabilistic framework. The simplest mathematical object allowing to describe a microscopic system at the limit is a probability density f which represents the statistical distribution of the agents in the system. The mean-field limit is a limit on the number of agents n in which the larger n, the weaker the action of each agent becomes while preserving the mean interaction, in other words say that the average action of the agents is of order 1.

From the outset, we are faced with a difficulty. Indeed, if we consider for example a system where the state of an agent i is its position $x_i \in \mathbb{R}^d$, then the phase space in which the solution lives is $(\mathbb{R}^d)^n$ whose dimension tends to infinity with n. We are therefore led to consider the empirical measure

$$\mu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}.$$

The advantage of considering the empirical measure is that it lives in the same space whatever n, the space of probability measures on $\mathbb{R}^d$: $\mathcal{P}(\mathbb{R}^d)$. If we add the velocity v_i to characterize the state of the agent i, then the empirical measure is

$$\mu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{(x_i,v_i)}.$$

It is a probability measure on $\mathbb{R}^d \times \mathbb{R}^d$.

We give ourselves a non-random microscopic model composed of n agents. Each characterized at time t by its position $x_i(t)$ and its velocity $v_i(t)$, with initial positions $x_0 := (x_1(0), \cdots, x_n(0))$ and initial velocities $v_0 := (v_1(0), \cdots, v_n(0))$. We define at time $t \geq 0$ the empirical measure

$$\mu_n(t) := \frac{1}{n} \sum_{i=1}^{n} \delta_{(x_i(t), v_i(t))}.$$

We assume that $\mu_n(0)$ converges to a probability measure μ_0. Solving a mean-field limit problem means finding a time-dependent probability measure μ, which is the solution of a kinetic model with initial condition μ_0 and such that $\mu_n(t)$ converges to μ_t for all $t \geq 0$.

Mean-field limit problems also exist for random microscopic models. In this case, for any instant $t \geq 0$, the empirical measure associated to the system

$$\mu_n(t) := \frac{1}{n} \sum_{i=1}^{n} \delta_{(X_i(t), V_i(t))}$$

is no longer a probability measure in $\mathbb{R}^d \times \mathbb{R}^d$ but a random variable taking values in the space of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$.

The Boltzmann equation, which is a kinetic model, is not a mean-field limit. Because the particles only interact with those with which they collide. That within a mean-field limit, each agent feels the influence of all the others. Nevertheless, the Boltzmann equation can be derived from the fundamental law of dynamics. This is the Grad limit, which is not a mean-field limit. This problem has been solved for small times by Lanford [59].

Example in the deterministic case

Consider Newton's system of equations (1.3.1) (which has already been scaled with the term $1/n$) with the functions $G = 0$ and F smooth for simplicity. Then (1.3.1) becomes

$$\begin{cases} \dfrac{d}{dt} x_i(t) = v_i(t), \\ \\ \dfrac{d}{dt} v_i(t) = (F * \mu_n(t))(x_i(t), v_i(t)). \end{cases} \tag{1.3.3}$$

where $\mu_n(t)$ is the associated empirical measure. It is easy to check that it is a solution of the Vlasov equation (1.1.6) (in a weak sense) where the force term is the convolution product defined by (1.1.7) with the kernel k the function F in (1.3.3). In other words μ_n is solution in the weak sense of

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + \nabla_v \cdot ((F * f) f) = 0. \tag{1.3.4}$$

The mean-field limit problem can be stated as follows. Consider a sequence of initial conditions $(x_1(0), \cdots, x_n(0), v_1(0), \cdots, v_n(0))$ in $(\mathbb{R}^d \times \mathbb{R}^d)^n$ of associated empirical measure $\mu_n(0)$ which converges to a probability measure $\mu_0 \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. So does at each instant $t \geq 0$, the measure $\mu_n(t)$ converges towards the measure $\mu(t)$ solution of (1.3.4) of initial condition μ_0 ? Knowing that $\mu_n(t)$ is itself a solution of (1.3.4). This question comes down to a problem of stability of the solutions of (1.3.4), which was solved at the end of the 70s independently by Braun, Hepp [17], Dobrushin [41] and Neunzert [66].

1.3.3 Some microscopic models

Cucker-Smale model

This microscopic model introduced by Cucker, Smale [30], [31] describes the collective behavior over time of a group of individuals (flocks of birds, schools of fish, etc.) without any leader. In this model,

the velocity of each individual is influenced by the others depending on the distance. We say that there is *flocking* when all the individuals tend to have the same velocity and to keep the group. From a mathematical point of view, this means that

$$\forall 1 \leq i \leq n, \qquad \sup_{t\geq 0}\left|x_i(t) - \frac{1}{n}\sum_{j=1}^{n} x_j(t)\right|^2 < +\infty \qquad \text{and} \qquad \lim_{t\to+\infty}\left|v_i(t) - \frac{1}{n}\sum_{j=1}^{n} v_j(t)\right|^2 = 0.$$

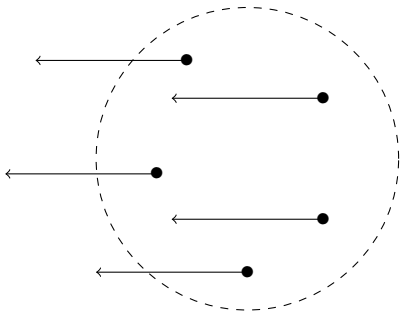


Figure 1.2: Illustration of flocking.

The idea of Cucker and Smale is that each individual will tend to average its velocity with that of the others so that there is flocking. In addition, the influence of other individuals decreases with the distance. So we have the system

$$\begin{cases} \dfrac{d}{dt}x_i(t) = v_i(t), \\ \\ \dfrac{d}{dt}v_i(t) = \dfrac{\lambda}{n}\displaystyle\sum_{j=1}^{n}\psi(|x_i(t) - x_j(t)|)(v_j(t) - v_i(t)) \end{cases} \tag{1.3.5}$$

where $\psi : \mathbb{R} \longmapsto \mathbb{R}$ is a positive decreasing function called communication rate and λ a positive real which measures the strength of interaction between individuals. The system (1.3.5) is deterministic and corresponds to (1.3.1) with $G = 0$ and $F(x, v) = -\lambda\psi(|x|)v$, so it is a problem of mean-field limit. In Ref. [30], [31], Cucker and Smale showed that there is flocking in the case where the communication rate ψ is defined by $\psi(r) := 1/(1 + r^2)^\gamma$ with $\gamma \in [0, 1/2[$. This result was found, then extended to $\gamma = 1/2$ by Ha, Tadmor [53].

In Ref. [53], we have also a hydrodynamic description of flocking. By defining ρ, u by (1.2.10), (1.2.11), and E by $E(t, x) := e(t, x) + |u(t, x)|^2/2$ with $\rho(t, x)e(t, x) := \int |v|^2 f(t, x, v)dv/2$, we have the closed system of equations on ρ, u and E

$$\begin{cases} \partial_t\rho + \mathrm{div}_x(\rho u) = 0, \\ \\ \partial_t(\rho u) + \mathrm{div}_x(\rho u \otimes u + P) = S_1(t, x), \\ \\ \partial_t(\rho E) + \mathrm{div}_x(\rho u E + Pu + Q) = S_2(t, x). \end{cases} \tag{1.3.6}$$

Here, $S_1(x, t)$ and $S_2(x, t)$ are the nonlocal source terms given by

$$S_1(t, x) := -\lambda\int_{\mathbb{R}^d}\psi(|x - y|)(u(t, x) - u(t, y))\rho(t, x)\rho(t, y)dxdy,$$
$$S_2(t, x) := -\lambda\int_{\mathbb{R}^d}\psi(|x - y|)(E(t, x) + E(t, y) - u(t, x)\cdot u(t, y))\rho(t, x)\rho(t, y)dxdy$$

and $P = (p_{i,j})_{i,j}$, $Q = (q_i)_i$ denote respectively the stress tensor and the heat flux vector

$$p_{i,j} := \int_{\mathbb{R}^d} (v_i - u_i)(v_j - u_j) f dv \qquad \text{and} \qquad q_i := \int_{\mathbb{R}^d} (v_i - u_i)|v - u|^2 f dv.$$

There are random versions of the Cucker-Smale model. Ha, Lee and Levy [52] proposed for example the following model. For $1 \leq i \leq n$

$$\begin{cases} dX_i(t) = V_i(t)dt, \\ \\ dV_i(t) = \dfrac{\lambda}{n} \displaystyle\sum_{j=1}^{n} \psi(|X_i(t) - X_j(t)|)(V_j(t) - V_i(t)) + \sqrt{D} dB_i(t) \end{cases}$$

with D a non-negative number representing the intensity of the noise and $B_1, \cdots, B_n$ independent d-dimensional Brownian motions. This is the system (1.3.2) with $G = 0$, $F(x, v) = -\lambda\psi(|x|)v$ and $\sigma = \sqrt{D}$. The random perturbation introduced is independent for each individual. It can be seen as a model of the degree of freedom or madness of each individual.

Self-propelled model with interactions

This is a microscopic model proposed by D'Orsogna, Chuang, Bertozzi, Chayes [45] reading as

$$\begin{cases} \dfrac{d}{dt} x_i(t) = v_i(t), \\ \\ \dfrac{d}{dt} v_i(t) = (\alpha - \beta|v_i|^2)v_i - \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} \nabla U(x_i - x_j). \end{cases} \tag{1.3.7}$$

This is again a deterministic Newton system (1.3.1) with $G(x, v) = (\alpha - \beta|v|^2)v$ and $F(x, v) = -\nabla U(x)$. The term G where α and β are two non-negative parameters is a force of self-propelled and friction to model the tendency of agents to reach a certain velocity limit of norm $\sqrt{\alpha/\beta}$. The force F for its part derives from a potential U. It models the tendency of agents to get closer if they are too far apart and vice versa. A typical potential used in this model is the Morse potential

$$U(x) = -C_A e^{-|x|/l_A} + C_R e^{-|x|/l_R}$$

where C_A, C_R, $l_R < l_A$ are non-negative constants which indicate the intensity and the characteristic length of attraction and repulsion.

This model (1.3.7) is also a mean-field limit. The associated empirical measure $\mu_n(t)$ is the solution of the Vlasov equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f - (\nabla U * \rho) \cdot \nabla_v f + \nabla_v \cdot ((\alpha - \beta|v|^2)vf) = 0 \tag{1.3.8}$$

with $\rho := \int f(t, x, v)dv$. We take a sequence of initial conditions $(x_1(0), \cdots, x_n(0), v_1(0), \cdots, v_n(0))$ in $(\mathbb{R}^d \times \mathbb{R}^d)^n$ with empirical measure $\mu_n(0)$ which converges to a probability measure $\mu_0 \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. Then for each instant $t \geq 0$, the measure $\mu_n(t)$ converges to the measure $\mu(t)$ solution of (1.3.8) with initial condition μ_0 [21], [25].

There is a hydrodynamic limit result for the equation (1.3.8) in Ref. [21], [25]. For a potential U, we consider the closed system of equations in ρ and u

$$\begin{cases} \partial_t \rho + \mathrm{div}_x(\rho u) = 0, \\ \\ \partial_t u + \mathrm{div}_x(u \otimes u) = (\alpha - \beta|u|^2)u - \nabla_x U * \rho. \end{cases} \tag{1.3.9}$$

with initial conditions (ρ_0, u_0). We assume that there exists (ρ, u) solutions of (1.3.9) with initial conditions (ρ_0, u_0) defined on the interval $[0,T]$ and we define $f : [0,T] \longrightarrow \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ by

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x,v) f(t,x,v) dx dv := \int_{\mathbb{R}^d} \varphi(x, u(t,x)) \rho(t,x) dx.$$

In other words, $f(t,x,v)dv := \rho(t,x)\delta_{u(t,x)}(dv)$ is a monokinetic distribution. Then f is the unique solution of (1.3.8) with initial condition $f_0(x,v)dv = \rho_0(x)\delta_{u_0(x)}(dv)$.

Vicsek model

This model proposed by Vicsek in 1995 [86] is a random microscopic model. The random positions $X_i(t)$ and $V_i(t)$ are solutions of the system of Stratonovich stochastic differential equations

$$\begin{cases} dX_i(t) = V_i(t)dt, \\ \\ dV_i(t) = \sqrt{2} P(V_i(t)) \circ dB_i(t) - P(V_i(t)) \left(\dfrac{1}{n} \displaystyle\sum_{j=1}^{n} K(X_i(t) - X_j(t))(V_i(t) - V_j(t)) \right) dt \end{cases} \tag{1.3.10}$$

where $P(v)$ is the projection on the tangent space to $v/|v|$ on the unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$. In other words $P(v) := I_d - (v \otimes v)/|v|^2$. It is kind of a random version of the Cucker-Smale model where the individuals evolve at constant velocity in norm.

The problem of the mean-field limit for this model has been traited in [12]. Consider the system of equations (1.3.10) where the initial data $(X_i(0), V_i(0))$ are independent and have the same law μ_0. For $1 \leq i \leq n$, we define the fictitious process $(\bar{X}_i(t), \bar{V}_i(t))$ which is the solution of

$$\begin{cases} d\bar{X}_i(t) = \bar{V}_i(t)dt, \\ \\ d\bar{V}_i(t) = \sqrt{2} P(\bar{V}_i(t)) \circ dB_i(t) - P(\bar{V}_i(t))(H * \mu(t))(\bar{X}_i(t), \bar{V}_i(t))dt \end{cases} \tag{1.3.11}$$

where $H(x,v) := K(x)v$, $(B_i(t))_t$ the Brownian motion which directs the evolution of the process $(X_i(t), V_i(t))$ and $\mu(t)$ the solution of the Fokker-Planck equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = \Delta_v f + \nabla_v \cdot (f P(v)(H * f)) \tag{1.3.12}$$

with $x \in \mathbb{R}^d$, $v \in \mathbb{S}^{d-1}$ and initial data $\mu_0 \in \mathcal{P}(\mathbb{R}^d \times \mathbb{S}^{d-1})$. The fictitious individual $(\bar{X}_i(t), \bar{V}_i(t))$ evolves in the field $H * \mu(t)$ generated by $\mu(t)$ when the physical individual $(X_i(t), V_i(t))$ evolves in the field $H * \mu_n(t)$ generated by $\mu_n(t)$, which is close to $\mu(t)$. This method is presented in [79] and makes it possible to obtain convergence rates.

The question of the hydrodynamic limit has been treating for this model. Let $X_i(t)$ be the random position in $\mathbb{R}^d$ and $V_i(t)$ the random velocity in $\mathbb{S}^{d-1}$ of the agent i. All agents have a constant norm velocity equal to c. Then the system (1.3.10) is written as

$$\begin{cases} dX_i(t) = cV_i(t)dt, \\ \\ dV_i(t) = P(V_i(t)) \circ \left(\nu \dfrac{J_i(t)}{|J_i(t)|} dt + \sqrt{2D} dB_i(t) dt \right) \end{cases} \tag{1.3.13}$$

where the parameter ν is the interaction frequency, D the intensity of the noise and

$$J_i(t) := \sum_{j,\ |X_i - X_j| \leq R} V_j(t)$$

with R the interaction range. When n tends to infinity, we obtain the mean-field model (1.3.12) of unknown $f = f(t, x, v)$, $t \geq 0$, $x \in R^d$ and $v \in \mathbb{S}^{d-1}$ which in the case of (1.3.13) is written

$$\frac{\partial f}{\partial t} + cv \cdot \nabla_x f = -\nabla_v \cdot (F[f]f) + D\Delta_v f \tag{1.3.14}$$

with $F[f](t, x) := \nu P(v) I(t, x)/|I(t, x)|$ where

$$I(t, x) := \int_{|x-y|<R} \int_{\mathbb{S}^{d-1}} v f(t, y, v) dv dy$$

and where the constants c, ν, D and R are the same as in (1.3.13). The derivation of the mean-field model (1.3.14) from the discrete model (1.3.13) has been justified in [12]. After scaling the constants (for detail, see Ref. [34]) there exists $k = O(1)$ such that the model (1.3.14) can be written for all $\varepsilon > 0$

$$\frac{\partial f_\varepsilon}{\partial t} + v \cdot \nabla_x f_\varepsilon = \frac{1}{\varepsilon} Q(f_\varepsilon) \tag{1.3.15}$$

where the collision operator Q is defined by $Q(f) := -\nabla_v((kP(v)J_f/|J_f|)f) + \Delta_v f$ with $k = O(1)$ and

$$J_f := \int_{\mathbb{S}^{d-1}} v f(t, x, v) dv.$$

When ε tends to 0, f_ε formally tends to f which satisfies $Q(f) = 0$. By defining ρ and u by (1.2.10) and (1.2.11), then there exist two positive constants C_1 and C_2 such that we have the closed system of equations on ρ and u

$$\begin{cases} \partial_t \rho + C_1 \mathrm{div}_x(\rho u) = 0, \\ \partial_t(\rho u) + C_2 \mathrm{div}_x(\rho u \otimes u) + P(u)\nabla_x \rho = 0, \\ |u| = 1. \end{cases} \tag{1.3.16}$$

The limit $f_\varepsilon \longrightarrow f$ when $\varepsilon \longrightarrow 0$ has been rigorously proved in Ref. [55].

1.3.4 Microscopic models for traffic flow

Microscopic models are used to describe a traffic flow when considering individual vehicles. Typically, these microscopic models are based on the method called *follow-the-leader* and are stated as a system of ordinary differential equations. They are usually based on velocity or acceleration functions that depend on spacing distance, velocity, predecessor velocity, relative velocity and so on. These models have a different form than (1.3.1).

One of the simplest approaches is a model based only on the distance between vehicles proposed by Pipes [70]

$$\frac{d}{dt} x_i(t) = W(\Delta x_i(t)) \tag{1.3.17}$$

where $\Delta x_i(t) := x_{i+1}(t) - x_i(t)$ denotes the spacing between vehicle i and its predecessor $i+1$ and W represents the optimum velocity function as a function of the spacing. The microscopic behavior can be improved by introducing a reaction and relaxation time. The simplest model is the one proposed by Newell [67]

$$\frac{d}{dt} x_i(t + \tau) = W(\Delta x_i(t)) \tag{1.3.18}$$

with $\tau > 0$ the reaction time. By applying a Taylor expansion on the left-hand side in (1.3.18), we obtain the optimal velocity model (called the OVM model) introduced by Bando et al [6]

$$\begin{cases} \dfrac{d}{dt}x_i(t) = v_i(t), \\ \\ \dfrac{d}{dt}v_i(t) = \dfrac{1}{\tau}(W(\Delta x_i(t)) - v_i(t)). \end{cases} \tag{1.3.19}$$

However the inconvenient pointed out by Daganzo [33] for the PW model (1.2.18) can be observed in models of the type "follow-the-leader" such as the OVM model (1.3.19). Another microscopic model based on that of Newell (1.3.18) is the one introduced by Tordeux, Seyfried [80]

$$\frac{d}{dt}x_i(t) = W(\Delta x_i(t) - \tau[W(\Delta x_{i+1}(t)) - W(\Delta x_i(t))]). \tag{1.3.20}$$

The model (1.3.20) is a system of ordinary differential equations of first order with two interacting predecessors. It is calibrated by the delay time $\tau \in \mathbb{R}$, that is a reaction time if positive and an anticipation time if negative, and by the optimal velocity function W.

Another example of a second order microscopic model is the following

$$\begin{cases} \dfrac{d}{dt}x_i(t) = v_i(t), \\ \\ \dfrac{d}{dt}v_i(t) = C\dfrac{v_{i+1}(t) - v_i(t)}{(\Delta x_i(t))^{\gamma+1}} + \dfrac{A}{T_r}\left(V\left(\dfrac{\Delta X}{\Delta x_i(t)}\right) - v_i(t)\right) \end{cases} \tag{1.3.21}$$

with $\Delta x_i(t) = x_{i+1}(t) - x_i(t)$, ΔX the length of a car, T_r the relaxation time and $C > 0$, $A > 0$ and $\gamma \geq 0$ given constants. This model (1.3.21) can be derived in order to obtain the ARZ-type macroscopic model (1.2.19)

$$\begin{cases} \partial_t \rho + \partial_x(\rho v) = 0, \\ \\ \partial_t(\rho(v + p(\rho))) + \partial_x(\rho v(v + p(\rho))) = A\dfrac{\rho}{T_r}(V(\rho) - v). \end{cases} \tag{1.3.22}$$

The macroscopic model (1.3.22) can be viewed as the limit of the time discretization of the microscopic model (1.3.21) when the number of vehicles increases, with a scaling in space and time (a zoom) for which the density and the velocity remain fixed. This limit has been rigorously proved in Ref. [4].

1.4 Contributions

In this manuscrit, we are interested in a kinetic Boltzmann-type model introduced by Bertin, Droz, Gregoire [9], [10] which describes the movement of a group of individuals subject to a social interaction. In this model, each individual (bird, fish, rod,...) moves independently from the others outside the collisions and are indistinguishable. At the time of the collision if two individuals are close enough, then they will line up in velocity. For each $t \geq 0$, the evolution of the collective behavior is represented by a probability distribution $f_t = f(t, x, v)$ where x denotes the position, and v denotes the velocity of the individuals. The two velocities v, $v_\star$ adopted by the two individuals after the collision are equal $v = v_\star$, and randomly distributed according to a probability $K(\,\cdot\,, v', v'_\star)$ centered at the mid pre-collisional velocity $(v' + v'_\star)/2$. In general, the collision rate is represented by a symmetric function $\beta(v', v'_\star)$ taking its values close to 1 if the two individuals are almost aligned before the collision and taking its values close to 0 in the case of grazing collisions. The model may also take into account a velocity confinment as in Cañizo, Carrillo, Rosado [21]. The position space x is d-dimensional and the

velocity $v \in \mathbb{R}^d$ ($d = 2$ or 3) is of constant modulus equal to 1. In [73], Raoul studies a similar model. The population of the individuals is structured by a continuous one-dimensional trait. At the time two individuals meet, they interact sexually and the trait of the offspring is distributed according to a Gaussian measure centered at the mid trait of the parents.

The unknown probability distribution $f = f(t, x, v)$, $t \geq 0$, $x \in \mathbb{R}^d$ and $v \in \mathbb{S}^{d-1}$ satisfies the following Boltzmann like equation in the sense of distributions:

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q(f, f) = Q_+(f, f) - Q_-(f, f) \tag{1.4.1}$$

where $Q(f, f)$ is the collision operator which is decomposed as in the Boltzmann equation (1.1.5) into a gain term $Q_+(f, f)$ and a loss term $Q_-(f, f)$. For any test function $\varphi \in \mathcal{C}_c^\infty(\mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^{d-1})$,

$$\langle Q_+(f, f), \varphi \rangle := \int_0^{+\infty} \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} \varphi(t, x, v) K(dv, v', v'_\star) \beta(v', v'_\star) f(t, x, dv') f(t, x, dv'_\star) dt dx$$

and

$$\langle Q_-(f, f), \varphi \rangle := \int_0^{+\infty} \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} \varphi(t, x, v) \beta(v, v') f(t, x, dv) f(t, x, dv') dt dx.$$

In Ref. [9], [10] and Carlen, Carvalho, Degond, Wenneberg [22], the dimension of the velocity is $d = 1$, the direction taken after the collision is chosen according to a density probability distribution centered at the mean $(v' + v'_\star)/2$. This model is called the continuous midpoint model.

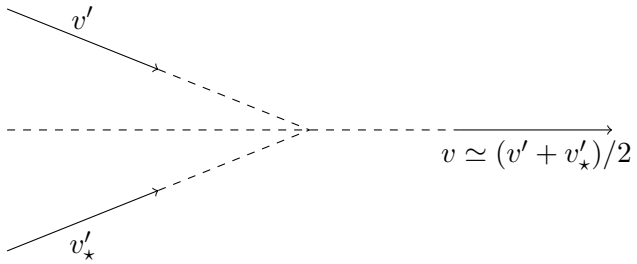


Figure 1.3: Scheme of a collision between two individuals of respective velocity v' and $v'_\star$. The velocity adopted by the two individuals after the collision will be the same, chosen according to a probability distribution centered at the mean $(v' + v'_\star)/2$.

In a probabilistic framework, in both cases, the velocity after the collision is written under the form $v = (v' + v'_\star)/2 + X$ where X is a random variable of law g, considered discrete or continuous. We choose from now on any probability $g(dv)$ on $\mathbb{R}^d$ with zero mean and covariance matrix Σ_g. For example $g(v)dv$, a density with respect to the Lebesgue measure as in ref. [22]. The model we consider in this manuscript is given by the following kernel K,

$$K(dv, v', v'_\star) := \tau_{(v'+v'_\star)/2} \# g(dv). \tag{1.4.2}$$

To establish the existence of solutions of equation (1.4.1), we will equip the space $\mathcal{P}_p(\mathbb{R}^d)$, $p \geq 1$ with the Wasserstein metric of order $p \geq 1$ W_p.

This model is new and interesting because it is located at the interface between collective dynamics and kinetic theory. The transport equation has no forcing or diffusion term in velocity, the change of velocity is computed as in Boltzmann framework. Since the collisions are not micro-reversible, it is not obvious to find an entropy functional. In the Boltzmann equations, micro-reversibility is a crucial element for obtaining the H Theorem. Consequently, the classical tools for dealing with the problems of returns to equilibrium, such as for example the Csiszàr-Kullback-Pinsker inequality [37], are inoperative. In our case, we have instead a phenomenon of contraction in the collision process

which does not take place for the Boltzmann operator but drives the density towards an equilibrium state.

We will start in chapter 2, by treating the most basic version of the model, and then we will assume in chapter 3 that the collision rate is no longer constant, which will add considerable difficulty. At the end of this manuscript, there are two appendices. The first is devoted to the Wasserstein metric W_p, a metric which will be of great use to us. And the second will state two elementary results on ODEs.

Chapter 2

In this chapter, we will study the BDG model (1.4.1) in the most simplified version possible. We will focus on the spatially homogeneous case with a constant collision rate β equal to 1, the so-called Maxwellian case. And the collision kernel K defined by (1.4.2) with $g \in \mathcal{P}_2^0(\mathbb{R}^d)$. We will also assume that the velocity v belongs to $\mathbb{R}^d$ and not $\mathbb{S}^{d-1}$. In Ref. [9], [10], [22], the dimension d of the space of velocities is equal to 1. And in Degond, Frouvelle, Raoul [35], the space of velocities may be a manifold of any dimension $d \geq 1$, but the probability $K(dv, v', v'_\star) = \delta_{(v'+v'_\star)/2}(dv)$ must be a Dirac mass at the mid velocity. The model in Ref. [35] is the BDG model in the particular case when $\beta = 1$ and when the direction taken by the two individuals is exactly the mean $(v' + v'_\star)/2$ of the two pre-collisionnal directions. In other words with $g = \delta_0$. This model is called the discrete midpoint model. Our results are more general in the sense that they are valid in any dimension and for any distribution $K(dv, v', v'_\star)$. However the technique of proofs we use assume that the space of velocities is Euclidean. A model where the velocity is constrained to be of norm 1 as in [35] is out of reached by our methods.

We will show in this chapter the existence of a unique solution called *Mild solution* (Definition 2.2.2 and Theorem 2.2.1) and that it converges exponentially towards the unique equilibrium state (Definition 2.3.1) for the Wasserstein metric (Theorem 2.3.1) and for the strong L^1 norm (Theorem 2.4.1). Since the collision rate β is constant, we will can use the Fourier transform in equation (1.4.1). We will show the existence and the uniqueness of the solutions and of the equilibrium state by a fixed point type argument. Moreover we will have an explicit formula for the Fourier transform of the equilibrium state (Proposition (2.3.1) of this manuscript and Theorem 1 in Ref. [22]). To prove the exponential convergence for the strong norm L^1, we will control the strong norm L^1 by the Fourier-Toscani-based distance d_2 introduced in Carrillo, Toscani [24] and Toscani, Villani [81]. Hence the interest of passing in Fourier (1.4.1). Then we will show that as for the Wasserstein metric, the solution converges exponentially towards the equilibrium state for the d_2 metric. To bound the strong L^1 norm by the d_2 metric, an estimate on the Sobolev norm $\| \cdot \|_{H^s(\mathbb{R}^d)}$ is needed for $s \geq 0$. We will also find the result concerning the discrete midpoint model obtained by Degond, Frouvelle, Raoul (Proposition 2.3 in Ref. [35]). Then we will end with numerical simulations.

Chapter 3

In this chapter, we consider the same model as in the previous chapter (1.4.1) but with a non-constant collision rate β. The collision rate β added is a symmetric continuous function of $\mathbb{R}^d \times \mathbb{R}^d$ in $[0, 1]$. This adds a considerable difficulty because in the previous model, the fact that β was constant made it possible to write the gain term Q_+ in the form of a double convolution, and therefore one could pass to Fourier. Unfortunately, the non-constancy of β gives Q_+ a more complicated expression than a double convolution, and no longer allows passing to Fourier. Therefore, all the tools used in the previous chapter will be inoperative in this chapter.

However, we can still write the model as a linear ordinary differential equation (B.1) in infinite dimension which gives rise to an operator $T_{g,\beta}$ which depends on g and β. We will show various properties concerning this operator. It is continuous, in particular it satisfies the Hölder condition with $\alpha = 1/2$ for some class of collision rate β, it preserves the space $\mathcal{P}_2^m(\mathbb{R}^d)$ and it is the double convolution from the previous chapter if β is constant equal to 1. With this operator, we can therefore

define a *Mild solution* (Definition (3.2.1)) as in the previous chapter. Only proving its existence is very difficult because this operator is a contraction only in the case where β is constant equal to 1.

We will show in this chapter that the operator $T_{g,\beta}$ admits a fixed point and therefore that the model admits an equilibrium state. Moreover we will have an explicit formula of the equilibrium state in the particular case where g and β are Gaussian with a condition on their variance. We will use for that a Schauder fixed point Theorem and therefore we will not have the uniqueness of the equilibrium state contrary to the previous chapter. We will end as for the previous chapter with numerical simulations. These will show the existence of the solution at all times as well as a convergence towards a unique equilibrium state. Unfortunately the convergence towards the equilibrium state is not yet proved.

Appendix

This manuscript has two appendices chapters. The first will be devoted to the Wasserstein metric. This metric will be very useful to us because it allows us to define a metric on the space $\mathcal{P}_p(\mathbb{R}^d)$. We will explain the link between this metric and the theory of optimal transport, then we will list various properties of this metric, including the famous Kantorovich Duality formula (Theorem A.3.1) for which we will give a rigorous proof for the particular compact case and when the cost function is continuous. We will give a characterization of what means *Converging in the Wasserstein sense* (Theorem A.5.1) with its proof then we will study the topology induced by this metric. And we will finish by explaining how to implement this metric numerically on the real line. In the second appendix chapter, we will recall two formulas on the ordinary differential equations (ODE): the Duhamel's formula and the Gronwall's Lemma. We will give the complete proof of these two results.

The alignment velocity model of Boltzmann type

Chapter 2

The homogeneous continuous midpoint model

2.1 Introduction

In this chapter, we consider a simplified version of the BDG model (1.4.1): the density f is independent of the position of the individuals, the velocity v belongs to $\mathbb{R}^d$ and not $\mathbb{S}^{d-1}$, the collision rate β is constant equal to 1 and the collision kernel K is defined by (1.4.2) with $g \in \mathcal{P}_2^0(\mathbb{R}^d)$ of covariance matrix Σ_g. Thus, the unknown probability distribution f satisfies the following Boltzmann like equation in the sense of distributions

$$\frac{\partial f}{\partial t} = Q_+(f,f) - Q_-(f,f)$$

where for any test function $\varphi \in \mathcal{C}_c^\infty(\mathbb{R} \times \mathbb{R}^d)$

$$\langle Q_+(f,f), \varphi \rangle := \int_0^{+\infty} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(t,v)(\tau_{(v'+v'_\star)/2}\#g)(dv) f(t,dv') f(t,dv'_\star) dt \tag{2.1.1}$$

and

$$\langle Q_-(f,f), \varphi \rangle := \int_0^{+\infty} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(t,v) f(t,dv) f(t,dv') dt. \tag{2.1.2}$$

For $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$, the evolution equation (1.4.1) becomes

$$\begin{cases} \dfrac{\partial f}{\partial t} = \displaystyle\iint_{\mathbb{R}^d \times \mathbb{R}^d} (\tau_{(v'+v'_\star)/2}\#g) f(t,dv') f(t,dv'_\star) - f(t,\cdot) \int_{\mathbb{R}^d} f(t,dv') \\ \\ f(0,\cdot) = f_0. \end{cases} \tag{2.1.3}$$

The plan of this chapter is the following. We start by establishing the existence of a mild solution of the equation (2.1.3) in section 2.2. We show the existence of the equilibrium state of the collision operator Q corresponding to probability measures f satisfying $Q(f,f) = 0$. And the proof of the exponential convergence of the solution towards the equilibrium state for the Wasserstein metric W_2 in section 2.3. We also make the link with the midpoint model in this same section. Then, we show the exponential convergence of the solution towards the equilibrium state for the d_2 metric, which will imply the convergence in L^1 in section 2.4. The last section is devoted to numerical simulations in dimension 1.

2.2 Existence of a Mild solution

To establish the existence of a solution to the equation (2.1.3), we equip the space $\mathcal{P}_2(\mathbb{R}^d)$ with the Wasserstein metric W_2. This metric metrizes the weak topology on $\mathcal{P}_2(\mathbb{R}^d)$ and makes that space complete (Theorem A.5.1 and A.6.1). Let $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$. For any test function $\varphi \in \mathcal{C}^\infty(\mathbb{R} \times \mathbb{R}^d)$ with compact support, we define

$$\langle f, \varphi \rangle := \int_0^{+\infty} \int_{\mathbb{R}^d} \varphi(t, v) f(t, dv) dt$$

which allows to define a solution in the sense of distributions as follows.

Definition 2.2.1

Let $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$. A **solution in the sense of distributions** of the equation (2.1.3) is a measured-valued function $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ satisfying for every test function $\varphi \in \mathcal{C}^\infty(\mathbb{R} \times \mathbb{R}^d)$ with compact support

$$-\left\langle f, \frac{\partial \varphi}{\partial t} \right\rangle = \int_{\mathbb{R}^d} \varphi(0, v) f_0(dv) + \langle Q_+(f, f), \varphi \rangle - \langle Q_-(f, f), \varphi \rangle. \tag{2.2.1}$$

We define a second notion of mild solution as follows, which is stronger than the first one.

Definition 2.2.2

A **mild solution** of the equation (2.1.3) is a function $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ taking values in the space of probability measures equipped with the Wasserstein metric W_2 satisfying for all $t \geq 0$

$$f(t, \cdot) = e^{-t} f_0 + \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{-(t-s)} (\tau_{(v'+v'_\star)/2} \# g) f(s, dv') f(s, dv'_\star) ds \tag{2.2.2}$$

For all $t \geq 0$, we denote by $\rho(t)$, $u(t)$ and $\Sigma_f(t)$ the mass, bulk velocity and covariance matrix at the instant t of the solution f:

$$\rho(t) := \int_{\mathbb{R}^d} f(t, dv), \tag{2.2.3}$$

$$u(t) := \frac{1}{\rho(t)} \int_{\mathbb{R}^d} v f(t, dv), \tag{2.2.4}$$

$$\Sigma_f(t) := \int_{\mathbb{R}^d} (\rho(t)v - \rho(t)u(t))^{\,t} (\rho(t)v - \rho(t)u(t)) f(t, dv). \tag{2.2.5}$$

For $g \in \mathcal{P}_2^0(\mathbb{R}^d)$, we define the following operator T_g for any measure μ by

$$T_g : \mu \longmapsto T_g[\mu] := g * (U\#\mu) * (U\#\mu) \tag{2.2.6}$$

where U is an uniform scaling with scaling factor $1/2$: $U : v \in \mathbb{R}^d \longmapsto v/2 \in \mathbb{R}^d$. Then for any test

function $\varphi \in \mathcal{C}_c^\infty(\mathbb{R} \times \mathbb{R}^d)$, we have by definition of the convolution product

$$\begin{aligned}
\langle T_g[f], \varphi \rangle &= \int_0^{+\infty} \int_{\mathbb{R}^d} \varphi(t,v) g * (U\#f(t,\cdot)) * (U\#f(t,\cdot))(dv)dt \\
&= \int_0^{+\infty} \iiint_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \varphi(t, v+v'+v'_\star) g(dv)(U\#f(t,\cdot))(dv')(U\#f(t,\cdot))(dv'_\star)dt \\
&= \int_0^{+\infty} \iiint_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \varphi\left(t, v + \frac{v'+v'_\star}{2}\right) g(dv) f(t,dv') f(t,dv'_\star)dt \\
&= \int_0^{+\infty} \iiint_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \varphi(t,v) (\tau_{(v'+v'_\star)/2}\#g)(dv) f(t,dv') f(t,dv'_\star)dt \\
&= \langle Q_+(f,f), \varphi \rangle.
\end{aligned}$$

Thus, the equation (2.1.3) can be written equivently as

$$\frac{\partial f}{\partial t} = T_g[f(t,\cdot)] - \rho(t) f(t,\cdot). \tag{2.2.7}$$

This equation (2.2.7) is interesting because this one is written as a linear ODE (B.1) with the initial condition $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$. By comparison with (B.1), we have only b depends on time and the unknown f. The Mild solution of the equation (2.1.3) is in a way the solution written by the Duhamel's formula (B.1.1). By direct computation, we have (with the notations $v\,{}^t v = (v_i v_j)_{1\le i,j\le d}$)

$$\int_{\mathbb{R}^d} T_g[f(t,\cdot)](dv) = \rho(t)^2, \tag{2.2.8}$$

$$\int_{\mathbb{R}^d} v T_g[f(t,\cdot)](dv) = \rho(t)^2 u(t), \tag{2.2.9}$$

$$\int_{\mathbb{R}^d} v\,{}^t v T_g[f(t,\cdot)](dv) = \rho(t)^2(\Sigma_g + u(t)\,{}^t u(t)) + \frac{\Sigma_f(t)}{2\rho(t)}. \tag{2.2.10}$$

The previous computation shows that T_g maps $\mathcal{P}_2^m(\mathbb{R}^d)$ into itself, hence show the mass and the mean velocity are preserved, but not the energy. Note that the gain term $Q_+(f,f)$ is a density if g is a density (even if f is a probability measure). Thus, $T_g[f(t,\cdot)]$ is a density for all $t \ge 0$ if g is a density and if moreover f_0 is a probability density, then by (2.2.2), $f(t,\cdot)$ is a probability density for all $t \ge 0$ where f is the mild solution of the equation (2.1.3). We will show in this section that the equation (2.1.3) admits a unique mild solution.

Theorem 2.2.1 (Existence of a mild solution)

Let $m \in \mathbb{R}^d$, $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$ and $g \in \mathcal{P}_2^0(\mathbb{R}^d)$. Then there exists a unique mild solution to the equation (2.1.3) $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ with $f(0,\cdot) = f_0$. Moreover, we have for all $t \ge 0$

$$\int_{\mathbb{R}^d} v f(t,dv) = \int_{\mathbb{R}^d} v f_0(dv) := m.$$

To prove this existence Theorem, we will use a fixed point type argument. So we'll need some properties on W_2. The following Lemma is the key step in the fixed point Theorem 2.2.1.

Lemma 2.2.1

For $\mu, \nu \in \mathcal{P}_2^m(\mathbb{R}^d)$, we have

$$W_2(T_g[\mu], T_g[\nu]) \le \frac{1}{\sqrt{2}} W_2(\mu,\nu). \tag{2.2.11}$$

Proof. Let μ, $\nu \in \mathcal{P}_2^m(\mathbb{R}^d)$. Then by definition of T_g operator (2.2.6), we have by sub-additivity of W_2 with respect to the convolution (A.4.13) and (A.4.14) that

$$W_2(T_g[\mu], T_g[\nu]) \leq W_2((U\#\mu) * (U\#\mu), (U\#\nu) * (U\#\nu)) \leq \sqrt{2} W_2(U\#\mu, U\#\nu).$$

Then, since U is a Lipschitzian map of Lipschitz norm $\|U\|_{\text{Lip}} = 1/2$, then by (A.4.5) we have $W_2(U\#\mu, U\#\nu) \leq W_2(\mu,\nu)/2$ and (2.2.11) follows. ■

This result is already present in Theorem 4.1 in ref. [72]. The proof presented here is different. We recall the following elementary fact.

Lemma 2.2.2

The space $\mathcal{P}_2^m(\mathbb{R}^d)$ is a complete metric space for W_2.

Proof. For $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, we define the map T by $T : \mu \longmapsto \int v d\mu(v)$. Let π be a coupling of (μ,ν) with μ, $\nu \in \mathcal{P}_2(\mathbb{R}^d)$. We have

$$|T(\mu) - T(\nu)| = \left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} (v-u) d\pi(u,v) \right| \leq \iint_{\mathbb{R}^d \times \mathbb{R}^d} |v-u| d\pi(u,v).$$

By taking the infimum over π, we obtain $|T(\mu) - T(\nu)| \leq W_1(\mu,\nu)$ and by Comparaison of the Wasserstein metrics (A.4.1), we have $W_1(\mu,\nu) \leq W_2(\mu,\nu)$. Therefore T is continuous. Since the space $\mathcal{P}_2(\mathbb{R}^d)$ is complete for the Wasserstein metric W_2 (Proposition A.6.1), it suffices to show that $\mathcal{P}_2^m(\mathbb{R}^d)$ is closed in $\mathcal{P}_2(\mathbb{R}^d)$. Let $(\mu_n)_n$ be a sequence in $\mathcal{P}_2^m(\mathbb{R}^d)$ which converges to μ for W_2. We have $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ by completeness and the continuity of T gives that $T(\mu_n)$ converges to $T(\mu)$. So $T(\mu) = m$ in other words $\mu \in \mathcal{P}_2^m(\mathbb{R}^d)$. ■

Let $\varphi \in \mathcal{C}([0,T])$ be a function such that $\int_0^T \varphi(t)dt = 1$. Then by convexity of W_2 with respect to the transition kernel (Proposition A.6.3) with μ the measure having for density $\varphi \mathbb{1}_{[0,T]}$, f_1 and $f_2 \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$, we obtain

$$W_2\left(\int_0^T \varphi(t) f_1(t,\cdot)dt, \int_0^T \varphi(t) f_2(t,\cdot)dt\right)^2 \leq \int_0^T \varphi(t) W_2(f_1(t,\cdot), f_2(t,\cdot))^2 dt. \tag{2.2.12}$$

The following Lemma shows that a mild solution can be seen as a fixed point of some contracting non-linear operator.

Lemma 2.2.3

Let $m \in \mathbb{R}^d$, $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$ and $g \in \mathcal{P}_2^0(\mathbb{R}^d)$. Define $E_{f_0} := \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ with $f(0,\cdot) = f_0$ equipped with the uniform norm. For $f \in E_{f_0}$, we define the map $\Phi : E_{f_0} \longrightarrow E_{f_0}$ by

$$\Phi[f](t,\cdot) := e^{-t} f(0,\cdot) + \int_0^t e^{-(t-s)} T_g[f(s,\cdot)] ds. \tag{2.2.13}$$

Then for all f_0^1, $f_0^2 \in \mathcal{P}_2^m(\mathbb{R}^d)$ and for $f^1 \in E_{f_0^1}$, $f^2 \in E_{f_0^2}$, we have

$$\begin{aligned} W_2(\Phi[f^1](t,\cdot), \Phi[f^2](t,\cdot))^2 \leq \frac{1}{2} \int_0^t e^{-(t-s)} W_2(f^1(s,\cdot), f^2(s,\cdot))^2 ds \\ + e^{-t} W_2(f_0^1, f_0^2)^2. \end{aligned} \tag{2.2.14}$$

Proof. Let $f \in E_{f_0}$. It is clear that $\Phi[f](0,\cdot) = f_0$ and since for all $t \geq 0$, $f(t,\cdot) \in \mathcal{P}_2^m(\mathbb{R}^d)$, we have $\rho(t) = 1$ and $u(t) = m$. And so by (2.2.8), (2.2.9) and (2.2.10), $\Phi[f](t,\cdot) \in \mathcal{P}_2^m(\mathbb{R}^d)$ for all $t \geq 0$. By writing

$$\Phi[f](t,\cdot) = e^{-t} f_0 + (1 - e^{-t}) \int_0^t \frac{e^{-(t-s)}}{1 - e^{-t}} T_g[f(s,\cdot)] ds,$$

we have by convexity of W_2 (A.4.12) that for all $f^1 \in E_{f_0^1}$, $f^2 \in E_{f_0^2}$ with f_0^1, $f_0^2 \in \mathcal{P}_2^m(\mathbb{R}^d)$,

$$W_2(\Phi[f^1](t,\cdot), \Phi[f^2](t,\cdot))^2 \leq (1 - e^{-t}) W_2 \left(\int_0^t \frac{e^{-(t-s)}}{1 - e^{-t}} T_g[f^1(s,\cdot)] ds, \int_0^t \frac{e^{-(t-s)}}{1 - e^{-t}} T_g[f^2(s,\cdot)] ds \right)^2 + e^{-t} W_2(f_0^1, f_0^2)^2$$

And using (2.2.12), it holds that

$$W_2(\Phi[f^1](t,\cdot), \Phi[f^2](t,\cdot))^2 \leq e^{-t} W_2(f_0^1, f_0^2)^2 + \int_0^t e^{-(t-s)} W_2(T_g[f^1(s,\cdot)], T_g[f^2(s,\cdot)])^2 ds.$$

And by (2.2.11), we obtain (2.2.14). ■

To prove Lemma 2.2.3, we used arguments that are used several times in ref. [73].

Proof of Theorem 2.2.1. Let $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$ and $g \in \mathcal{P}_2^0(\mathbb{R}^d)$. We take the map Φ defined by (2.2.13). Then for f_1, $f_2 \in E_{f_0}$, we have by (2.2.14)

$$W_2(\Phi[f_1](t,\cdot), \Phi[f_2](t,\cdot))^2 \leq \frac{1}{2} \int_0^t e^{-(t-s)} W_2(f_1(s,\cdot), f_2(s,\cdot))^2 ds.$$

Passing to the supremum in time,

$$\begin{aligned} W_2(\Phi[f_1](t,\cdot), \Phi[f_2](t,\cdot))^2 &\leq \frac{1}{2} \left(\sup_{t \in \mathbb{R}^+} W_2(f_1(t,\cdot), f_2(t,\cdot))^2 \right) \int_0^t e^{-(t-s)} ds \\ &= \frac{1 - e^{-t}}{2} \sup_{t \in \mathbb{R}^+} W_2(f_1(t,\cdot), f_2(t,\cdot))^2. \end{aligned}$$

One obtains

$$\sup_{t \in \mathbb{R}^+} W_2(\Phi[f_1](t,\cdot), \Phi[f_2](t,\cdot)) \leq \frac{1}{\sqrt{2}} \sup_{t \in \mathbb{R}^+} W_2(f_1(t,\cdot), f_2(t,\cdot)). \tag{2.2.15}$$

Hence Φ preserves the space E_{f_0} and is a contraction. Since $\mathcal{P}_2^m(\mathbb{R}^d)$ is a complete metric space for W_2, E_{f_0} is complete. Hence there exists a unique mild solution of the equation (2.1.3) belonging to E_{f_0}. ■

Proposition 2.2.1 (Properties of mild solutions)

Let $m \in \mathbb{R}^d$, $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ be the mild solution of equation (2.1.3) with $f(0,\cdot) = f_0$. Then we have the following properties.

(1) f is a solution of equation (2.1.3) in the sense of distributions (see Definition 2.2.1).

(2) For all $t \geq 0$,

$$\Sigma_f(t) = e^{-t/2} \Sigma_f(0) + 2(1 - e^{-t/2}) \Sigma_g. \tag{2.2.16}$$

(3) For every mild solution f_1, $f_2 \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$, we have for all $t \geq 0$

$$W_2(f_1(t,\cdot), f_2(t,\cdot)) \leq e^{-t/4} W_2(f_1(0,\cdot), f_2(0,\cdot)). \tag{2.2.17}$$

Proof. (1) By direct computation

$$
\begin{aligned}
-\left\langle f, \frac{\partial\varphi}{\partial t}\right\rangle &= -\int_0^{+\infty}\int_{\mathbb{R}^d}\int_0^t e^{-(t-s)}\frac{\partial\varphi}{\partial t}(t,v)T_g[f(s,\cdot)](dv)dsdt - \int_0^{+\infty}\int_{\mathbb{R}^d} e^{-t}\frac{\partial\varphi}{\partial t}(t,v)f_0(dv)dt\\
&= -\int_0^{+\infty}\int_{\mathbb{R}^d} e^s\left(\int_s^{+\infty} e^{-t}\frac{\partial\varphi}{\partial t}(t,v)dt\right)T_g[f(s,\cdot)](dv)ds + \int_{\mathbb{R}^d}\varphi(0,v)f_0(dv)\\
&\quad -\int_0^{+\infty}\int_{\mathbb{R}^d} e^{-t}\varphi(t,v)f_0(dv)dt\\
&= -\int_0^{+\infty}\int_{\mathbb{R}^d}\int_s^{+\infty}\varphi(t,v)e^{-(t-s)}T_g[f(s,\cdot)](dv)dtds + \int_0^{+\infty}\int_{\mathbb{R}^d}\varphi(s,v)T_g[f(s,\cdot)](dv)ds\\
&\quad +\int_{\mathbb{R}^d}\varphi(0,v)f_0(dv) - \int_0^{+\infty}\int_{\mathbb{R}^d} e^{-t}\varphi(t,v)f_0(dv)dt\\
&= \int_{\mathbb{R}^d}\varphi(0,v)f_0(dv) + \langle Q_+(f,f),\varphi\rangle - \int_0^{+\infty}\int_{\mathbb{R}^d}\varphi(t,v)f(t,dv)dt.
\end{aligned}
$$

Since for any $t\geq 0$, $f(t,\cdot)$ is a probability measure, we have

$$
\int_0^{+\infty}\int_{\mathbb{R}^d}\varphi(t,v)f(t,dv)dt = \int_0^{+\infty}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\varphi(t,v)f(t,dv)f(t,dv_\star)dt = \langle Q_-(f,f),\varphi\rangle.
$$

(2) We have by (2.2.2) and (2.2.10)

$$
\begin{aligned}
\int_{\mathbb{R}^d} v\,{}^t v f(t,dv) &= e^{-t}\int_{\mathbb{R}^d} v\,{}^t v f_0(dv) + \int_0^t e^{-(t-s)}\left(\int_{\mathbb{R}^d} v\,{}^t v T_g[f(s,\cdot)](dv)\right)ds\\
&= e^{-t}\int_{\mathbb{R}^d} v\,{}^t v f_0(dv) + \int_0^t e^{-(t-s)}\left(\Sigma_g + m\,{}^t m + \frac{\Sigma_f(s)}{2}\right)ds.
\end{aligned}
$$

So we have

$$
e^t\Sigma_f(t) = \Sigma_f(0) + (e^t-1)\Sigma_g + \int_0^t\frac{e^s}{2}\Sigma_f(s)ds.
$$

Applying the case of equality in Gronwall's Lemma (B.2.5), we have

$$
\begin{aligned}
e^t\Sigma_f(t) &= \Sigma_f(0) + (e^t-1)\Sigma_g + \int_0^t\frac{\Sigma_f(0) + (e^s-1)\Sigma_g}{2}e^{(t-s)/2}ds\\
&= \Sigma_f(0) + (e^t-1)\Sigma_g + (e^{t/2}-1)\Sigma_f(0) + (e^t+1-2e^{t/2})\Sigma_g.
\end{aligned}
$$

Which implies formula (2.2.16).

(3) Let $f_1\in\mathcal{C}(\mathbb{R}^+,\mathcal{P}_2^m(\mathbb{R}^d))$ be the mild solution with initial condition $f_0^1\in\mathcal{P}_2^m(\mathbb{R}^d)$ and let $f_2\in\mathcal{C}(\mathbb{R}^+,\mathcal{P}_2^m(\mathbb{R}^d))$ be the mild solution with initial condition $f_0^2\in\mathcal{P}_2^m(\mathbb{R}^d)$. By (2.2.14), we have

$$
W_2(f_1(t,\cdot),f_2(t,\cdot))^2 \leq e^{-t}W_2(f_0^1,f_0^2)^2 + \frac{1}{2}\int_0^t e^{-(t-s)}W_2(f_1(s,\cdot),f_2(s,\cdot))^2ds.
$$

By Gronwall's Lemma (Lemma B.2.1), we have

$$
\begin{aligned}
e^tW_2(f_1(t,\cdot),f_2(t,\cdot))^2 &\leq W_2(f_0^1,f_0^2)^2 + \int_0^t\frac{W_2(f_0^1,f_0^2)^2}{2}e^{(t-s)/2}ds\\
&= W_2(f_0^1,f_0^2)^2 + (e^{t/2}-1)W_2(f_0^1,f_0^2)^2\\
&= e^{t/2}W_2(f_0^1,f_0^2)^2.
\end{aligned}
$$

Hence the estimate (2.2.17). ■

2.3 Equilibrium state

This section is devoted to the determination of the equilibrium state of the collision operator of equation (2.1.3). Equilibrium states of the collision operator Q corresponding to the probability distribution functions satisfying $Q(f,f)=0$. We will mainly focus on the convergence to the unique equilibrium state of the equation (2.1.3) that is defined as follows.

Definition 2.3.1

An **equilibrium state** of the equation (2.1.3) is a probability distribution $f \in \mathcal{P}_2(\mathbb{R}^d)$ satisfying the fixed point equation

$$f = \iint_{\mathbb{R}^d\times\mathbb{R}^d} (\tau_{(v'+v'_\star)/2}\#g) f(dv') f(dv'_\star). \tag{2.3.1}$$

For $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ and for $t \geq 0$, the function $\hat{f}(t,\,\cdot\,) : \xi \longmapsto \hat{f}(t,\xi)$ is the Fourier transform of the probability measure $f(t,\,\cdot\,)$. So if $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ is the mild solution of equation (2.1.3) (see Definition 2.2.2) with initial condition $f(0,\,\cdot\,) = f_0$, then $\hat{f}$ is solution of the fixed point equation

$$\hat{f}(t,\xi) = e^{-t}\hat{f}(0,\xi) + \int_0^t e^{-(t-s)}\hat{g}(\xi)\hat{f}\left(s, \frac{\xi}{2}\right)^2 ds. \tag{2.3.2}$$

Note that the changeover in Fourier variable can be performed because β is constant. The equation (2.3.2) allows to differentiate $\hat{f}$ which satisfies therefore

$$\begin{cases} \dfrac{\partial \hat{f}}{\partial t} = \hat{g}(\xi)\hat{f}\left(t, \dfrac{\xi}{2}\right)^2 - \hat{f}(t,\xi), \\ \\ \hat{f}(0,\xi) = \hat{f}_0(\xi). \end{cases} \tag{2.3.3}$$

We notice that equation (2.3.1) can be written equivalently as

$$f = T_g[f].$$

In other words, f is a fixed point of T_g. Passing in Fourier, equation (2.3.1) is equivalent to

$$\hat{f}(\xi) = \hat{g}(\xi)\hat{f}\left(\frac{\xi}{2}\right)^2. \tag{2.3.4}$$

Theorem 2.3.1 (Exponential convergence for W_2)

Let $m \in \mathbb{R}^d$, $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$ and $g \in \mathcal{P}_2^0(\mathbb{R}^d)$. Then there exists a unique equilibrium state $f_m^\infty \in \mathcal{P}_2^m(\mathbb{R}^d)$, that is a probability measure f_m^∞ satisfying

$$Q(f_m^\infty, f_m^\infty) = 0.$$

And if $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ is the mild solution of equation (2.1.3) with $f(0,\,\cdot\,) = f_0$, then we have for all $t \geq 0$

$$W_2(f(t,\,\cdot\,), f_m^\infty) \leq e^{-t/4} W_2(f_0, f_m^\infty). \tag{2.3.5}$$

The existence of a unique equilibrium state is a consequence of the following proposition. Theorem 1 of Ref. [22] is improved by choosing a probability measure $g(dv)$ instead of a density $g(v)dv$ and by proving the uniqueness of the equilibrium state in this general setting. Moreover we do not use Levi's Theorem to recognize a Fourier transform of a probability measure. Results close to the existence of a unique equilibrium state are present in Ref. [22] and in Lemma 2.2 in Ref. [73] (the model is different and g is Gaussian). (2.3.5) is similar to step two of the proof of Lemma 2.1 in Ref. [73]. The author shows a perturbation result about the a Gaussian solution; we show an exponential convergence to an equilibrium state which may be non Gaussian. Our result is also valid in any dimension.

Proposition 2.3.1 (Theorem 1 in Ref. [22])

Let $g \in \mathcal{P}_2^0(\mathbb{R}^d)$. For all $m \in \mathbb{R}^d$, there exists a unique $f \in \mathcal{P}_2^m(\mathbb{R}^d)$ solution of (2.3.1). In addition we have

$$\hat{f}(\xi) = e^{-i\langle m,\xi\rangle} \prod_{n=0}^{+\infty} \hat{g}\left(\frac{\xi}{2^n}\right)^{2^n}. \tag{2.3.6}$$

Proof. Let $f \in \mathcal{P}_2^m(\mathbb{R}^d)$. Then $T_g[f] \in \mathcal{P}_2^m(\mathbb{R}^d)$ since T_g maps $\mathcal{P}_2^m(\mathbb{R}^d)$ into itself by (2.2.8), (2.2.9) and (2.2.10). And by (2.2.11), T_g is a contraction on the metric space $(\mathcal{P}_2^m(\mathbb{R}^d), W_2)$ which is complete by Lemma 2.2.2. Hence, there exists a unique $f \in \mathcal{P}_2^m(\mathbb{R}^d)$ such that $T_g[f] = f$. Let us now show that $\hat{f}$ satisfies (2.3.6). By iterating the equation (2.3.4), we have by a Taylor expansion

$$\begin{aligned}\hat{f}(\xi) &= \left(\prod_{k=0}^{n-1} \hat{g}\left(\frac{\xi}{2^k}\right)^{2^k}\right) \hat{f}\left(\frac{\xi}{2^n}\right)^{2^n} \\ &= \left(\prod_{k=0}^{n-1} \hat{g}\left(\frac{\xi}{2^k}\right)^{2^k}\right) \left(1 + \frac{1}{2^n}\left(-i\langle m,\xi\rangle - \frac{{}^t\xi(\Sigma_f + m\,{}^t m)\xi}{2^{n+1}} + o\left(\frac{|\xi|^2}{2^n}\right)\right)\right)^{2^n}.\end{aligned}$$

The second factor on the right-hand side converges to $e^{-i\langle m,\xi\rangle}$ when n tends to infinity since if the real sequence $(x_n)_n$ converges to x, then $(1 + x_n/n)^n$ converges to e^x. We thus obtain (2.3.6) by letting n tend to infinity. ■

We denote by f_m^∞ (which is a density if g is a density) the unique solution of (2.3.1) in $\mathcal{P}_2^m(\mathbb{R}^d)$. In particular, $\hat{f}_m^\infty$ satisfies (2.3.6). By calculating $H_{\hat{f}_m^\infty}(0)$, we notice that if $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ is the mild solution of equation (2.1.3), then by (2.2.16), $\Sigma_f(t)$ converges to $2\Sigma_g$ when t goes to infinity. Which corresponds well to the covariance matrix of f_m^∞. Indeed, by differentiating the function $\xi \longmapsto \log(\hat{f}_m^\infty(\xi))$, we have since $\hat{f}_m^\infty$ satisfies (2.3.6) that

$$\nabla \hat{f}_m^\infty(\xi) = \left(-im + \sum_{n=0}^{+\infty} \frac{\nabla \hat{g}(\xi/2^n)}{\hat{g}(\xi/2^n)}\right) \hat{f}_m^\infty(\xi).$$

By differentiating this formula, we have

$$H_{\hat{f}_m^\infty}(\xi) = \hat{f}_m^\infty(\xi) \sum_{n=0}^{+\infty} \frac{\hat{g}(\xi/2^n) H_{\hat{g}}(\xi/2^n) - \nabla \hat{g}(\xi/2^n)\,{}^t\nabla \hat{g}(\xi/2^n)}{2^n(\hat{g}(\xi/2^n))^2} + \nabla \hat{f}_m^\infty(\xi)^{\,t} \left(-im + \sum_{n=0}^{+\infty} \frac{\nabla \hat{g}(\xi/2^n)}{\hat{g}(\xi/2^n)}\right).$$

We obtain the covariance matrix of f_m^∞ by calculating this expression above at $\xi = 0$ since for a probability measure μ of mean vector m, $\hat{\mu}(0) = 1$, $i\nabla\hat{\mu}(0) = m$ and $-H_{\hat{\mu}}(0) = \int v\,{}^t v d\mu(v)$. And

in the particular case where $g \in \mathcal{P}_2^0(\mathbb{R}^d)$ is a Gaussian (centered of covariance matrix Σ_g), then $\hat{g}(\xi) = \exp(-({}^t\xi\Sigma_g\xi)/2)$ and by (2.3.6),

$$\hat{f}_m^\infty(\xi) = e^{-i\langle m,\xi\rangle}\prod_{n=0}^{+\infty}\exp\left(-\frac{{}^t\xi\Sigma_g v}{2^{n+1}}\right) = \exp\left(-i\langle m,\xi\rangle - \frac{{}^t\xi\Sigma_g\xi}{2}\sum_{n=0}^{+\infty}\frac{1}{2^n}\right) = \exp\left(-i\langle m,\xi\rangle - {}^t\xi\Sigma_g\xi\right).$$

So by the Fourier inverse transform, f_m^∞ is also a Gaussian with mean vector m and covariance matrix $2\Sigma_g$.

Proof of Theorem 2.3.1. The existence of a unique equilibrium state $f_m^\infty \in \mathcal{P}_2^m(\mathbb{R}^d)$ readily follows from Proposition 2.3.1. Let $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ be the mild solution of equation 2.1.3 with initial condition $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$. Note that

$$\forall t \geq 0, \quad f_m^\infty = e^{-t}f_m^\infty + \int_0^t e^{-(t-s)}T_g[f_m^\infty]ds.$$

By taking the map Φ defined in (2.2.13), we have by (2.2.14)

$$W_2(f(t,\cdot),f_m^\infty)^2 = W_2(\Phi[f](t,\cdot),\Phi[f_m^\infty])^2 \leq e^{-t}W_2(f_0,f_m^\infty)^2 + \frac{1}{2}\int_0^t e^{-(t-s)}W_2(f(s,\cdot),f_m^\infty)^2 ds.$$

By Gronwall's Lemma (Lemma B.2.1), we have

$$\begin{aligned} e^t W_2(f(t,\cdot),f_m^\infty)^2 &\leq W_2(f_0,f_m^\infty)^2 + \int_0^t \frac{W_2(f_0,f_m^\infty)^2}{2}e^{(t-s)/2}ds \\ &= W_2(f_0,f_m^\infty)^2 + (e^{t/2}-1)W_2(f_0,f_m^\infty)^2 \\ &= e^{t/2}W_2(f_0,f_m^\infty)^2. \end{aligned}$$

Which implies (2.3.5). ■

We now make the link with the result obtained in Ref. [35]. The discrete midpoint model corresponding to the equation (2.1.3) with $g = \delta_0$.

Proposition 2.3.2 (Proposition 2.3 in Ref. [35])

We consider equation (2.1.3) with $g = \delta_0$. For $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$, there exists a unique mild solution of equation (2.1.3) $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ with initial condition f_0. Moreover, we have the estimate

$$W_2(f(t,\cdot),\delta_m) \leq e^{-t/4}W_2(f_0,\delta_m). \tag{2.3.7}$$

Proof. Existence and uniqueness of the solution follow from Theorem 2.2.1. By Theorem 2.3.1, there exists a unique equilibrium state $f_m^\infty \in \mathcal{P}_2^m(\mathbb{R}^d)$ such that

$$W_2(f(t,\cdot),f_m^\infty) \leq e^{-t/4}W_2(f_0,f_m^\infty).$$

Since $g = \delta_0$, we have $\hat{g}(\xi) = 1$ for all ξ. And so by (2.3.6) we have $\hat{f}_m^\infty(\xi) = e^{-i\langle m,\xi\rangle}$. We recognize the Fourier transform of δ_m, so $f_m^\infty = \delta_m$ and (2.3.7) follows. ■

As mentioned in Ref. [35], the conservation of the center of mass m has played a fundamental role in the functional space $\mathcal{P}_2^m(\mathbb{R}^d)$. It would be much more difficult to prove these estimates on the sphere $\mathbb{S}^{d-1}$ because the center of mass is no longer conserved.

2.4 Convergence for the strong L^1 norm

This section is devoted to the proof of the exponential convergence of the solution of (2.1.3) towards the equilibrium state for the strong L^1 norm.

Theorem 2.4.1 (Exponential convergence for the strong L^1 norm)

Let $m \in \mathbb{R}^d$, $f_0 \in H^s(\mathbb{R}^d) \cap \mathcal{P}_2^m(\mathbb{R}^d)$ and $g \in H^s(\mathbb{R}^d) \cap \mathcal{P}_2^0(\mathbb{R}^d)$ densities with $s > 2 + d/2$. If $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ is the mild solution of equation (2.1.3) with $f(0, \cdot) = f_0$ and if $f_m^\infty \in \mathcal{P}_2^m(\mathbb{R}^d)$ is the equilibrium state of (2.1.3) with mean velocity m, then there exists a constant $C > 0$ explicitly computable such that for all $t \geq 0$,

$$\|f(t, \cdot) - f_m^\infty\|_{L^1(\mathbb{R}^d)} \leq Ce^{-t/(d+4)}. \tag{2.4.1}$$

To prove this theorem, we control the strong-norm L^1 by the Fourier-Toscani-based distance d_2 introduced in Carrillo, Toscani [24] and Toscani, Villani [81]. Then we show that the solution converges exponentially towards the equilibrium state of $Q(f, f)$ defined in (2.1.3, 2.1.1, 2.1.2) for the distance d_2.

Definition 2.4.1

For μ, $\nu \in \mathcal{P}_2^m(\mathbb{R}^d)$ having the same mean value, we define the **Fourier-Toscani-based distance** between μ and ν by

$$d_2(\mu, \nu) := \sup_{\xi \in \mathbb{R}^d} \frac{|\hat{\mu}(\xi) - \hat{\nu}(\xi)|}{|\xi|^2}. \tag{2.4.2}$$

A Taylor expansion shows that this metric is well-defined for μ, $\nu \in \mathcal{P}_2^m(\mathbb{R}^d)$ and metrizes the weak topology on $\mathcal{P}_2^m(\mathbb{R}^d)$ (see Ref. [81]). We proved in the previous section that $f_m^\infty \in \mathcal{P}_2^m(\mathbb{R}^d)$. So $d_2(f(t, \cdot), f_m^\infty)$ is well defined for all $t \geq 0$. The following result gives the exponential convergence of $f(t, \cdot)$ to f_m^∞ for the d_2 metric.

Proposition 2.4.1 (Exponential convergence for d_2)

Let $m \in \mathbb{R}^d$, $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$ and $g \in \mathcal{P}_2^0(\mathbb{R}^d)$. If $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ is the mild solution of equation (2.1.3) with $f(0, \cdot) = f_0$ and if $f_m^\infty \in \mathcal{P}_2^m(\mathbb{R}^d)$ is the equilibrium state of (2.1.3) with mean velocity m, then it holds that for all $t \geq 0$

$$d_2(f(t, \cdot), f_m^\infty) \leq \frac{M_2(f_0) + 2M_2(g) + |m|^2}{2} e^{-t/2}. \tag{2.4.3}$$

Proof. We set for $t \geq 0$ and for $\xi \in \mathbb{R}^d$,

$$H(t, \xi) := \frac{\hat{f}(t, \xi) - \hat{f}_m^\infty(\xi)}{|\xi|^2}$$

where $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ is the mild solution of equation (2.1.3). A Taylor expansion gives that

$$e^{-i\langle v, \xi \rangle} = 1 - i\langle v, \xi \rangle - \langle v, \xi \rangle^2 \int_0^1 (1 - s) e^{-is\langle v, \xi \rangle} ds.$$

So

$$|H(t,\xi)| = \left| \int_{\mathbb{R}^d} \int_0^1 \frac{\langle v,\xi\rangle^2}{|\xi|^2}(1-s)e^{-is\langle v,\xi\rangle} f(t,v)dsdv - \int_{\mathbb{R}^d} \int_0^1 \frac{\langle v,\xi\rangle^2}{|\xi|^2}(1-s)e^{-is\langle v,\xi\rangle} f_m^\infty(v)dsdv \right|$$
$$\leq \int_{\mathbb{R}^d} \int_0^1 \left| \frac{\langle v,\xi\rangle^2}{|\xi|^2}(1-s)e^{-is\langle v,\xi\rangle}(f(t,v) - f_m^\infty(v)) \right| dsdv.$$

Hence the previous inequality taken at $t = 0$ and Cauchy-Schwarz inequality leads to

$$|H(0,\xi)| \leq \int_{\mathbb{R}^d} \int_0^1 |v|^2(1-s)|f_0(v) - f_m^\infty(v)|dsdv = \frac{M_2(f_0) + M_2(f_m^\infty)}{2}.$$

And $M_2(f_m^\infty) = 2M_2(g) + |m|^2$ leads imediately

$$|H(0,\xi)| \leq \frac{M_2(f_0) + 2M_2(g) + |m|^2}{2}. \tag{2.4.4}$$

We define now G by

$$G(t,\xi) := \frac{\hat{g}(\xi)}{4}\left(\hat{f}\left(t,\frac{\xi}{2}\right) - \hat{f}_m^\infty\left(\frac{\xi}{2}\right)\right).$$

Using (2.3.3) and (2.3.4), it holds that

$$G(t,\xi)H\left(t,\frac{\xi}{2}\right) - H(t,\xi) = \frac{\hat{g}(\xi)}{4}\left(\hat{f}\left(t,\frac{\xi}{2}\right) + \hat{f}_m^\infty\left(\frac{\xi}{2}\right)\right)\left(\frac{\hat{f}(t,\xi/2) - \hat{f}_m^\infty(\xi/2)}{|\xi|^2/4}\right) - \frac{\hat{f}(t,\xi) - \hat{f}_m^\infty(\xi)}{|\xi|^2}$$
$$= \frac{\hat{g}(\xi)}{|\xi|^2}\hat{f}\left(t,\frac{\xi}{2}\right)^2 - \frac{\hat{f}(t,\xi)}{|\xi|^2} - \left(\frac{\hat{g}(\xi)}{|\xi|^2}\hat{f}_m^\infty\left(\frac{\xi}{2}\right)^2 - \frac{f_m^\infty(\xi)}{|\xi|^2}\right)$$
$$= \frac{1}{|\xi|^2}\frac{\partial \hat{f}}{\partial t}.$$

So H satisfies

$$\frac{\partial H}{\partial t}(t,\xi) = G(t,\xi)H\left(t,\frac{\xi}{2}\right) - H(t,\xi).$$

And by Duhamel's formula B.1.1, we get

$$H(t,\xi) = e^{-t}H(0,\xi) + \int_0^t e^{-(t-s)}G(s,\xi)H\left(s,\frac{\xi}{2}\right)ds.$$

For $R > 0$, we set

$$y(t) := e^t \sup_{|\xi|\leq R} |H(t,\xi)|$$

since the map $\xi \longmapsto \xi/2$ maps $\overline{B(0,R)}$ into $\overline{B(0,R/2)} \subset \overline{B(0,R)}$. Since $G(t,\xi) \leq 1/2$, Gronwall's Lemma (Lemma B.2.1) applied to inequality

$$y(t) \leq y(0) + \int_0^t (y(s)/2)ds$$

gives that for all $R \geq 0$,

$$\sup_{|\xi|\leq R} |H(t,\xi)| \leq \sup_{|\xi|\leq R} |H(0,\xi)|e^{-t/2}.$$

So by using the estimate (2.4.4) we get formula (2.4.3). ■

To bound the L^1 norm by the d_2 metric, an estimate on the Sobolev norm $\|\cdot\|_{H^s(\mathbb{R}^d)}$ is needed for $s \geq 0$. The initial condition $f(0, \cdot)$ is assumed to be a regular function and we prove the exponential convergence in L^1 of $f(t, \cdot)$ to the equilibrium state determined in section 2.3. The regularity of the initial condition is measured in term of the Sobolev norm.

Proposition 2.4.2

Let $s \geq 0$ and $f_0,\, g \in H^s(\mathbb{R}^d) \cap \mathcal{P}_2(\mathbb{R}^d)$. Let $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ be the mild solution of equation (2.1.3) with initial condition f_0. Then for every $t \geq 0$, $f(t, \cdot) \in H^s(\mathbb{R}^d)$ and

$$\|f(t, \cdot)\|_{H^s(\mathbb{R}^d)} \leq e^{-t}\|f_0\|_{H^s(\mathbb{R}^d)} + (1 - e^{-t})\|g\|_{H^s(\mathbb{R}^d)}. \tag{2.4.5}$$

Proof. Let $R > 0$. Define

$$Z_R(t) := \int_{|\xi|<R} (1 + |\xi|^2)^s |\hat{f}(t, \xi)|^2 d\xi.$$

Using (2.3.3) and the inequality $|\hat{\mu}(\xi)| \leq 1$ for μ a probability measure, it comes that $\partial_t \hat{f}(t, \xi)$ is uniformly bounded by 2. By differentiation, it holds that

$$\begin{aligned}
\frac{d}{dt} Z_R(t) &= \int_{|\xi|<R} (1 + |\xi|^2)^s \frac{\partial}{\partial t} |\hat{f}(t, \xi)|^2 d\xi \\
&= 2 \int_{|\xi|<R} (1 + |\xi|^2)^s \Re \left(\overline{\hat{f}(t, \xi)} \frac{\partial \hat{f}}{\partial t}(t, \xi) \right) d\xi \\
&= 2 \int_{|\xi|<R} (1 + |\xi|^2)^s \Re \left(\hat{g}(\xi) \hat{f}\left(t, \frac{\xi}{2}\right)^2 \overline{\hat{f}(t, \xi)} \right) d\xi - 2Z_R(t).
\end{aligned}$$

Using the classical inequality $\Re(z) \leq |z|$ for all $z \in \mathbb{C}$, we have

$$\begin{aligned}
\frac{d}{dt} Z_R(t) &\leq 2 \int_{|\xi|<R} (1 + |\xi|^2)^s \left| \hat{g}(\xi) \hat{f}\left(t, \frac{\xi}{2}\right)^2 \overline{\hat{f}(t, \xi)} \right| d\xi - 2Z_R(t) \\
&\leq 2 \int_{|\xi|<R} (1 + |\xi|^2)^s |\hat{g}(\xi)| \cdot |\hat{f}(t, \xi)| d\xi - 2Z_R(t).
\end{aligned}$$

Cauchy-Schwarz inequality applied to the right-hand side gives that

$$\begin{aligned}
\frac{d}{dt} Z_R(t) &\leq 2 \left(\int_{|\xi|<R} (1 + |\xi|^2)^s |\hat{g}(\xi)|^2 d\xi \right)^{1/2} \left(\int_{|\xi|<R} (1 + |\xi|^2)^s |\hat{f}(t, \xi)|^2 d\xi \right)^{1/2} - 2Z_R(t) \\
&\leq 2\|g\|_{H^s(\mathbb{R}^d)} \sqrt{Z_R(t)} - 2Z_R(t).
\end{aligned}$$

Since $Z_R(t)$ never vanishes,

$$\frac{d}{dt} \sqrt{Z_R(t)} \leq \frac{2\|g\|_{H^s(\mathbb{R})} \sqrt{Z_R(t)} - 2Z_R(t)}{2\sqrt{Z_R(t)}} = \|g\|_{H^s(\mathbb{R}^d)} - \sqrt{Z_R(t)}.$$

And by Gronwall's Lemma (Lemma B.2.1),

$$\sqrt{Z_R(t)} \leq e^{-t} \sqrt{Z_R(0)} + \int_0^t e^{-(t-s)} \|g\|_{H^s(\mathbb{R}^d)} ds = e^{-t} \sqrt{Z_R(0)} + (1 - e^{-t}) \|g\|_{H^s(\mathbb{R}^d)}.$$

We conclude the proof by letting R to infinity. ■

Since g is a probability density, $\|\hat{g}\|_\infty \leq 1$, using the explicit definition $\hat{f}_m^\infty$ in (2.3.6), we obtain $|\hat{f}_m^\infty| \leq |\hat{g}|$ and the following result:

$$\forall s \geq 0, \quad \|f_m^\infty\|_{H^s(\mathbb{R}^d)} \leq \|g\|_{H^s(\mathbb{R}^d)}. \tag{2.4.6}$$

We also need the following two interpolation inequalities (Theorem 4.1 and 4.2 in Ref. [23]) that we prove for the reader convenience.

Lemma 2.4.1

Let $\alpha > 0$ and $d \geq 1$ be an integer. Then there exists a constant $C(\alpha, d) > 0$ such that for every function $f \in L^2(\mathbb{R}^d) \cap \mathcal{P}_\alpha(\mathbb{R}^d)$

$$\|f\|_{L^1(\mathbb{R}^d)} \leq C(\alpha, d)\|f\|_{L^2(\mathbb{R}^d)}^{\alpha'} M_\alpha(f)^{1-\alpha'} \tag{2.4.7}$$

with $\alpha' := 2\alpha/(2\alpha + d)$.

Proof. We have for every $R > 0$

$$\begin{aligned}\int_{\mathbb{R}^d} |f(v)|dv &= \int_{|v|\leq R} |f(v)|dv + \int_{|v|>R} |f(v)|dv \\ &\leq \sqrt{\mathrm{Vol}(B(0,R))}\|f\|_{L^2(\mathbb{R}^d)} + \frac{1}{R^\alpha}\int_{|v|>R} |v|^\alpha |f(v)|dv.\end{aligned}$$

Recall that $\mathrm{Vol}(B(0,R)) = R^d \mathrm{Vol}(B_d)$. We have

$$\int_{\mathbb{R}^d} |f(v)|dv \leq R^{d/2}\sqrt{\mathrm{Vol}(B_d)}\|f\|_{L^2(\mathbb{R}^d)} + \frac{1}{R^\alpha} M_\alpha(f).$$

We choose R such that

$$R^{d/2}\sqrt{\mathrm{Vol}(B_d)}\|f\|_{L^2(\mathbb{R}^d)} = \frac{1}{R^\alpha} M_\alpha(f) \qquad \Longleftrightarrow \qquad R = \left(\frac{M_\alpha(f)}{\sqrt{\mathrm{Vol}(B_d)}\|f\|_{L^2(\mathbb{R}^d)}}\right)^{2/(2\alpha+d)}.$$

We substitute in the previous inequality, we obtain

$$\begin{aligned}\|f\|_{L^1(\mathbb{R}^d)} \leq &\left(\frac{M_\alpha(f)}{\sqrt{\mathrm{Vol}(B_d)}\|f\|_{L^2(\mathbb{R}^d)}}\right)^{d/(2\alpha+d)} \sqrt{\mathrm{Vol}(B_d)}\|f\|_{L^2(\mathbb{R}^d)} \\ &+ \left(\frac{M_\alpha(f)}{\sqrt{\mathrm{Vol}(B_d)}\|f\|_{L^2(\mathbb{R}^d)}}\right)^{-2\alpha/(2\alpha+d)} M_\alpha(f).\end{aligned}$$

By setting $\alpha' := 2\alpha/(2\alpha + d)$, we have

$$\begin{aligned}\|f\|_{L^1(\mathbb{R}^d)} &\leq \mathrm{Vol}(B_d)^{-(1-\alpha')/2+1/2}\|f\|_{L^2(\mathbb{R}^d)}^{1+(\alpha'-1)} M_\alpha(f)^{1-\alpha'} + \mathrm{Vol}(B_d)^{\alpha'/2}\|f\|_{L^2(\mathbb{R}^d)}^{\alpha'} M_\alpha(f)^{1-\alpha'} \\ &= 2\mathrm{Vol}(B_d)^{\alpha'/2}\|f\|_{L^2(\mathbb{R}^d)}^{\alpha'} M_\alpha(f)^{1-\alpha'}.\end{aligned}$$

So we obtain (2.4.7) with $C(\alpha, d) = 2\mathrm{Vol}(B_d)^{\alpha'/2}$. ■

Lemma 2.4.2

Let $s \geq 0$ and $d \geq 1$ be an integer. For every $s' > d/2 + 2s + 2$ there exists a constant $C(s, s', d) > 0$ such that for every f_1, $f_2 \in H^{s'}(\mathbb{R}^d) \cap \mathcal{P}_2(\mathbb{R}^d)$ having the same mean,

$$\|f_1 - f_2\|_{H^s(\mathbb{R}^d)} \leq C(s, s', d)\sqrt{d_2(f_1, f_2)}\sqrt{\|f_1 - f_2\|_{H^{s'}(\mathbb{R}^d)}}. \tag{2.4.8}$$

Proof. For $s' > d/2 + 2s + 2$, we have

$$\begin{aligned}\|f_1 - f_2\|^2_{H^s(\mathbb{R}^d)} &= \int_{\mathbb{R}^d} (1+|\xi|^2)^s |\hat{f}_1(\xi) - \hat{f}_2(\xi)|^2 d\xi \\ &\leq \int_{\mathbb{R}^d} \frac{|\hat{f}_1(\xi) - \hat{f}_2(\xi)|}{|\xi|^2} \frac{(1+|\xi|^2)^{s+1}}{(1+|\xi|^2)^{s'/2}} (1+|\xi|^2)^{s'/2} |\hat{f}_1(\xi) - \hat{f}_2(\xi)| d\xi.\end{aligned}$$

By Cauchy-Schwarz inequality, it holds that

$$\int_{\mathbb{R}^d} (1+|\xi|^2)^s |\hat{f}_1(\xi) - \hat{f}_2(\xi)|^2 d\xi \leq d_2(f_1, f_2) \left(\int_{\mathbb{R}^d} \frac{d\xi}{(1+|\xi|^2)^{s'-2s-2}} \right)^{1/2} \|f_1 - f_2\|_{H^{s'}(\mathbb{R}^d)}.$$

The assumption on s' implies that the integral in the right-hand side is finite. Because $\int d\xi/((1+|\xi|^2)^\alpha)$ is finite if $\alpha > d$. ■

Proof of Theorem 2.4.1. Applying (2.4.7) with the function $v \longmapsto |f(t,v) - f_m^\infty(v)|$ and with $\alpha = 2$ leads to

$$\|f(t,\cdot) - f_m^\infty\|_{L^1(\mathbb{R}^d)} \leq C(2,d) \|f(t,\cdot) - f_m^\infty\|^{4/(d+4)}_{L^2(\mathbb{R}^d)} M_2(|f(t,\cdot) - f_m^\infty|)^{d/(d+4)}.$$

Since $M_2(|f(t,\cdot) - f_m^\infty|) \leq M_2(f(t,\cdot)) + M_2(f_m^\infty)$, we have by (2.2.16) that $M_2(|f(t,\cdot) - f_m^\infty|) \leq M_2(f_0) + 4M_2(g) + |m|^2$ for all $t \geq 0$ and so

$$\|f(t,\cdot) - f_m^\infty\|_{L^1(\mathbb{R}^d)} \leq C_1 \|f(t,\cdot) - f_m^\infty\|^{4/(d+4)}_{L^2(\mathbb{R}^d)} (M_2(f_0) + M_2(g) + |m|^2)^{d/(d+4)} \tag{2.4.9}$$

with $C_1 := 4^{d/(d+4)} C(2,d)$. Then by (2.4.8) with $s = 0$ and $s' > 2 + d/2$, it comes that

$$\|f(t,\cdot) - f_m^\infty\|^2_{L^2(\mathbb{R}^d)} \leq C(0,s',d)^2 d_2(f(t,\cdot), f_m^\infty) \|f(t,\cdot) - f_m^\infty\|_{H^{s'}(\mathbb{R}^d)}.$$

Since $\|f(t,\cdot) - f_m^\infty\|_{H^{s'}(\mathbb{R}^d)} \leq \|f(t,\cdot)\|_{H^{s'}(\mathbb{R}^d)} + \|f_m^\infty\|_{H^{s'}(\mathbb{R}^d)}$, we have by (2.4.5) and (2.4.6) that $\|f(t,\cdot) - f_m^\infty\|_{H^{s'}(\mathbb{R}^d)} \leq \|f_0\|_{H^{s'}(\mathbb{R}^d)} + 2\|g\|_{H^{s'}(\mathbb{R}^d)}$ for all $t \geq 0$ and so

$$\|f(t,\cdot) - f_m^\infty\|^{4/(d+4)}_{L^2(\mathbb{R}^d)} \leq C_2 d_2(f(t,\cdot), f_m^\infty)^{2/(d+4)} (\|f_0\|_{H^{s'}(\mathbb{R}^d)} + \|g\|_{H^{s'}(\mathbb{R}^d)})^{2/(d+4)} \tag{2.4.10}$$

with $C_2 := 2^{2/(d+4)} C(0,s',d)^{4/(d+4)}$. Substituting (2.4.10) in (2.4.9), leads to

$$\|f(t,\cdot) - f_m^\infty\|_{L^1(\mathbb{R}^d)} \leq C_3 d_2(f(t,\cdot), f_m^\infty)^{2/(d+4)}$$

with $C_3 := C_1 C_2 (\|f_0\|_{H^{s'}(\mathbb{R}^d)} + \|g\|_{H^{s'}(\mathbb{R}^d)})^{2/(d+4)} (M_2(f_0) + M_2(g) + |m|^2)^{d/(d+4)}$. And then by (2.4.3),

$$\begin{aligned} d_2(f(t,\cdot), f_m^\infty)^{2/(d+4)} &\leq \left(\frac{M_2(f_0) + 2M_2(g) + |m|^2}{2} \right)^{2/(d+4)} e^{-t/(d+4)} \\ &\leq (M_2(f_0) + M_2(g) + |m|^2)^{2/(d+4)} e^{-t/(d+4)}. \end{aligned}$$

Thus, we obtain (2.4.1) with a constant

$$C := C_3 (M_2(f_0) + 2M_2(g) + |m|^2)^{2/(d+4)}.$$

And consequently the exponential convergence of $f(t,\cdot)$ towards f_m^∞ is obtained since $f_0 \in H^{s'}(\mathbb{R}^d) \cap \mathcal{P}_2^m(\mathbb{R}^d)$ and $g \in H^{s'}(\mathbb{R}^d) \cap \mathcal{P}_2^0(\mathbb{R}^d)$ with $s' > 2 + d/2$. ■

2.5 Numerical results

This section is devoted to the numerical resolution of (2.1.3) in dimension $d = 1$. We will present six tests cases for two different initial conditions f_0 and with three different values of g, where g is a density. For each test case, the solution f is depicted for different values of t and compared with the equilibrium state f_m^∞ theoretically found in order to characterize the exponential rate of convergence for the Wasserstein metric W_2 and for the strong-norm L^1. To represent the solution of (2.1.3) numerically, we use an Euler scheme in time for $\Delta t = 0.015$ followed by a Simpson rule on the interval $[-10, 10]$ with a uniform step $\Delta x = 0.1$. We will use a quadrature method to numerically represent the strong L^1 norm. And to represent numerically the Wasserstein metric W_2 between two measures with density $\mu := f_1\lambda$ and $\nu := f_2\lambda$ with compact support contained in a segment $[a, b]$, we discretize them this way

$$\mu := \sum_{i=1}^{n} x_i f_1(x_i)\delta_{x_i} \qquad \text{and} \qquad \nu := \sum_{i=1}^{n} x_i f_2(x_i)\delta_{x_i}$$

where $x_i := a + i(b - a)/n$, then we will apply (A.7.9) with $p = 2$. We will therefore represent numerically the solutions of the equation

$$\begin{cases} \dfrac{\partial f}{\partial t} = \displaystyle\iint_{\mathbb{R}\times\mathbb{R}} g\left(v - \frac{v' + v'_\star}{2}\right) f(t, dv') f(t, dv'_\star) - f(t, \cdot) \int_{\mathbb{R}} f(t, dv') \\ \\ f(0, \cdot) = f_0. \end{cases} \tag{2.5.1}$$

In the first case, g is a centered Gaussian of variance $\sigma_g^2 = 1$,

$$g(v) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right). \tag{2.5.2}$$

In the second case where g is an indicator function

$$g(v) = \frac{1}{2}\mathbb{1}_{[-1,1]}(v). \tag{2.5.3}$$

In the third case, g writes

$$g(v) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(v - m_i)^2}{2\sigma_i^2}\right), \tag{2.5.4}$$

with $\sum m_i = 0$ since g is zero mean. We take in (2.5.4) $n = 3$, $m_1 = 3$, $m_2 = m_3 = -3/2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 2$ and $\sigma_3^2 = 4$. In other words, g is a normalized sum of three Gaussians. For each g, we consider two different initial conditions f_0 defined as follows

- The first the initial condition f_0 is a normalized Gaussian of mean 2 given by

$$f_0(v) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(v - 2)^2}{2}\right). \tag{2.5.5}$$

- The second initial condition f_0 is equal to g defined by (2.5.4), with $n = 3$, $m_1 = 3$, $m_2 = m_3 = -3/2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 2$ and $\sigma_3^2 = 4$.

Test case 1

For the first test case, g is a normalized centered Gaussian (2.5.2). Since g is a Gaussian, Proposition 2.3.1 gives an explicit formula for the equilibrium state. In that case, f_m^∞ is a Gaussian of variance $2\sigma_g^2 = 2$ with the same mean as the initial condition f_0. Hence for f_0 defined by (2.5.5), f_m^∞ writes

$$f_m^\infty(v) = \frac{1}{\sqrt{4\pi}} \exp\left(-\frac{(v-2)^2}{4}\right). \tag{2.5.6}$$

From f_0 defined by (2.5.4), f_m^∞ writes

$$f_m^\infty(v) = \frac{1}{\sqrt{4\pi}} \exp\left(-\frac{v^2}{4}\right) \tag{2.5.7}$$

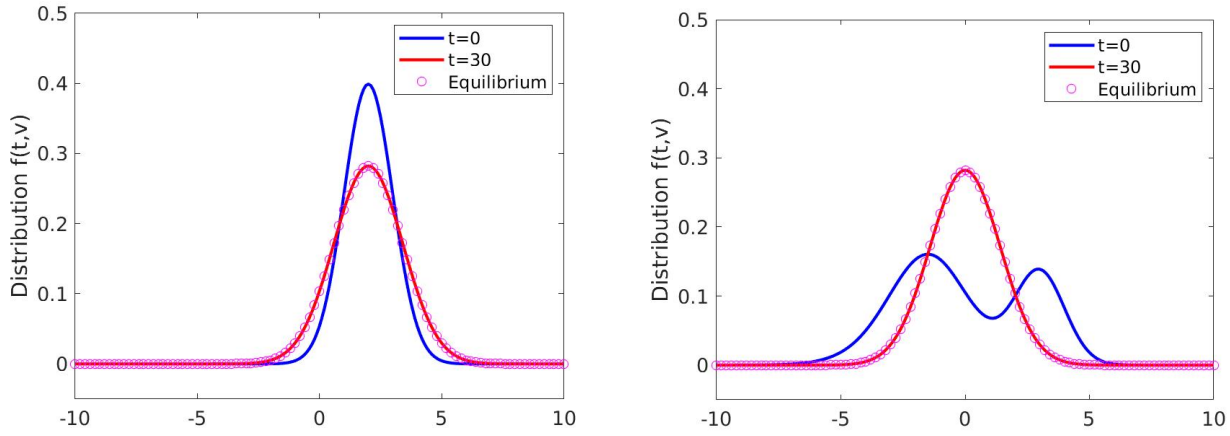


Figure 2.1: Distribution function of the solution of (2.5.1) with g defined by (2.5.2). On the left, solution for initial condition (2.5.5) at times $t = 0$, $t = 30$ and on the right, solution for initial condition (2.5.4) at times $t = 0$, $t = 30$. Equilibrium state given by (2.5.7) (in circle).

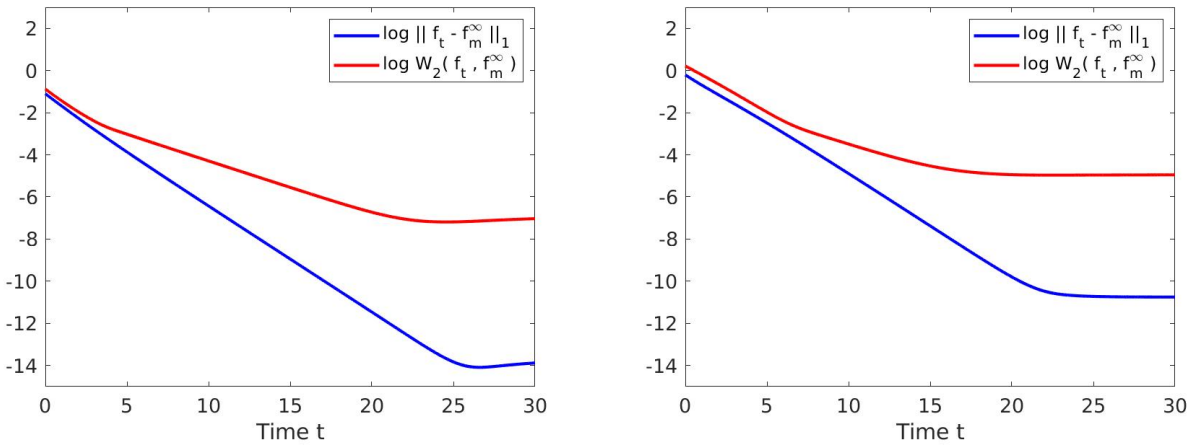


Figure 2.2: Functions $t \longmapsto \log \|f(t,\cdot) - f_m^\infty\|_{L^1}$ and $t \longmapsto \log W_2(f(t,\cdot), f_m^\infty)$ where f is the solution of (2.5.1) with g defined by (2.5.2). On the left, $f(t,\cdot)$ for initial condition (2.5.5) and f_m^∞ given by (2.5.6). On the right, $f(t,\cdot)$ for initial condition defined by (2.5.4) and f_m^∞ is given by (2.5.7).

Since we have shown that $f(t,.)$ converges exponentially to f_m^∞ for the Wasserstein metric W_2 and for the strong-norm L^1, then the two functions $t \longmapsto \log W_2(f(t,\cdot), f_m^\infty)$ and $t \longmapsto \log \|f(t,.) - f_m^\infty\|_{L^1}$ must be bounded by an affine function. For the strong L^1 norm, this is the case in Figure 2.2 but with a ratio 1/2 and not 1/5 theoretically found. Unfortunately for the Wasserstein metric W_2 it is a little less clear. But we still have at the beginning a linear decreasing as for the strong L^1 norm but with always a ratio 1/2 and not 1/4 theoretically found.

Test case 2

For the second test case, g is defined by (2.5.3). Since g is not a Gaussian, the expression of the equilibrium state f_m^∞ is not explicit. Then f_m^∞ is approached by $f(t,\cdot)$ at time $t = 35$ corresponding to a converged result.

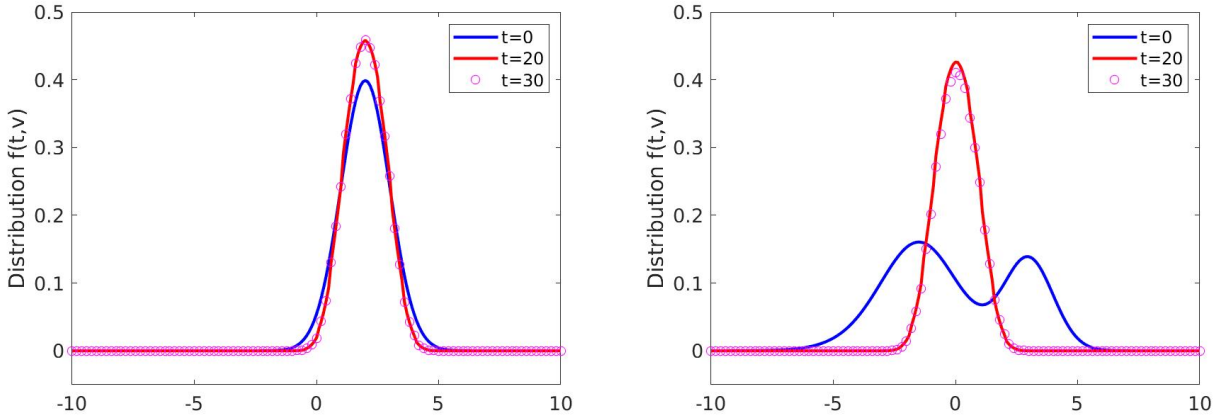


Figure 2.3: Distribution function of the solution of (2.5.1) with g defined by (2.5.3). On the left, solutions for initial condition (2.5.5) at times $t = 0$, $t = 20$, $t = 30$ and on the right, solutions for initial condition (2.5.4) at times $t = 0$, $t = 20$, $t = 30$.

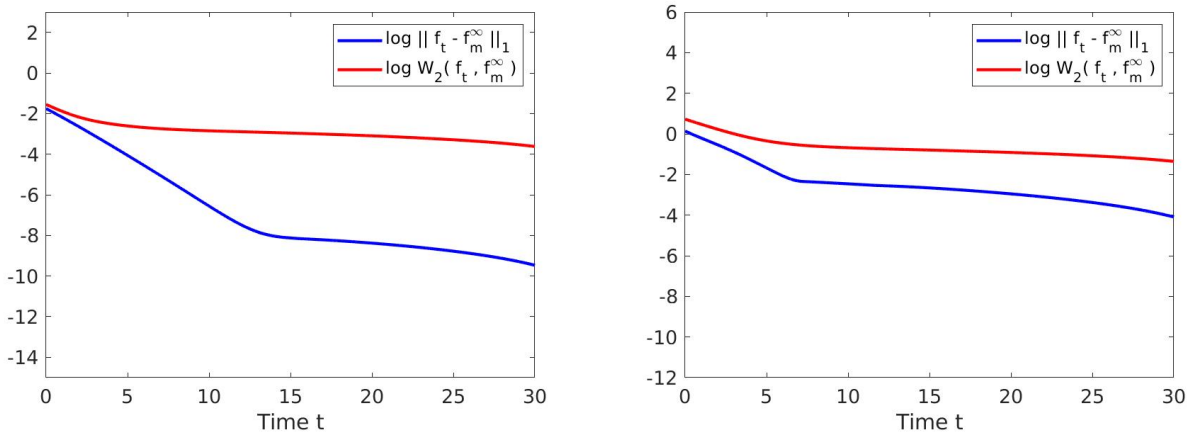


Figure 2.4: Functions $t \longmapsto \log \|f(t,\cdot) - f_m^\infty\|_{L^1}$ and $t \longmapsto \log W_2(f(t,\cdot), f_m^\infty)$ for f the solution of (2.5.1) with g defined by (2.5.3). On the left, $f(t,\cdot)$ for initial condition (2.5.5) and f_m^∞ is replaced by a converged solution $f(t,\cdot)$ at time $t = 35$. On the right, $f(t,\cdot)$ for initial condition (2.5.4) and f_m^∞ replaced by a converged solution $f(t,\cdot)$ at time $t = 35$.

Figure 2.3 shows that $f(t,\cdot)$ goes towards the same asymptotic limit for the two different initial conditions. This numerical result is consistent with Proposition 2.3.1 claiming that the equilibrium

state depends only on g. In Figure 2.4, the two curves are not rectilinear because there are two phenomenons. First of all, the distribution goes towards a Maxwellian distribution and next to the right one. However Figure 2.4 shows that the convergence remains with an exponential rate.

Test case 3

The third test case is devoted to g defined by (2.5.4). Since g is not a Gaussian, no explicit formula are again available for the equilibrium state f_m^∞. Hence f_m^∞ is again approached by converged solution obtained at time $t = 35$ as for test case 2.

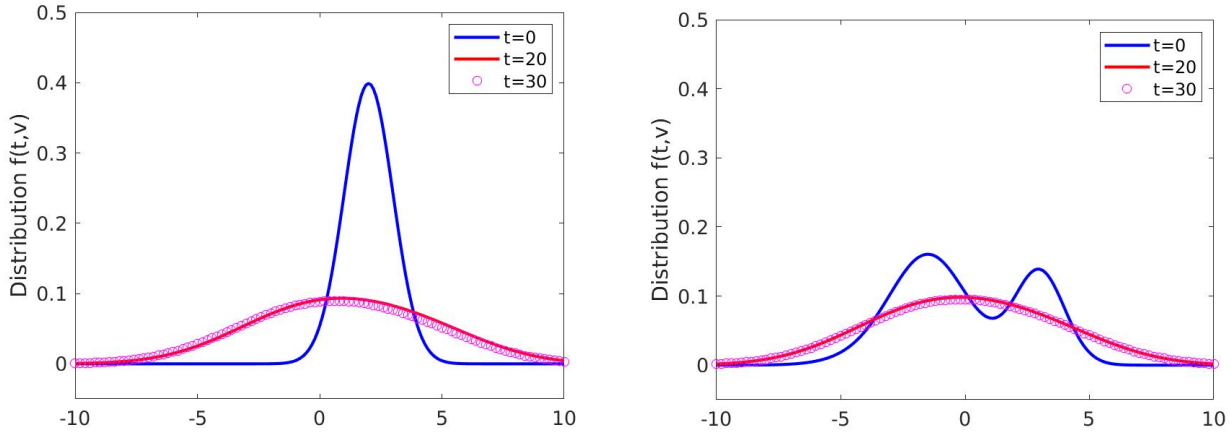


Figure 2.5: Distribution function of the solution of (2.5.1) with g defined by (2.5.4). On the left, solutions with initial condition (2.5.5) at times $t = 0$, $t = 20$, $t = 30$ and on the right, solutions with initial condition (2.5.4) at times $t = 0$, $t = 20$, $t = 30$.

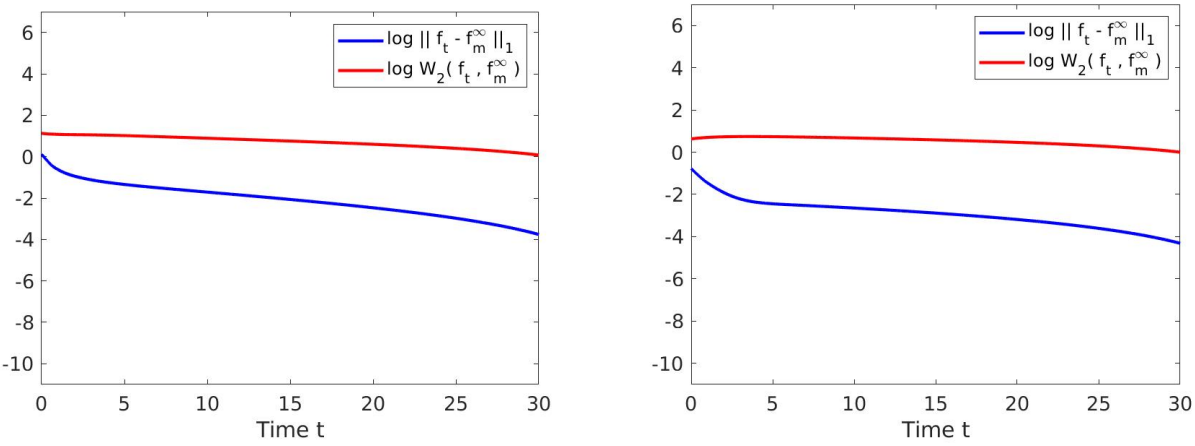


Figure 2.6: Functions $t \longmapsto \log \|f(t,\cdot) - f_m^\infty\|_{L^1}$ and $t \longmapsto \log W_2(f(t,\cdot), f_m^\infty)$ where f is the solution of (2.5.1) with g defined by (2.5.4) and f_m^∞ the equilibrium state. On the left, $f(t,\cdot)$ with initial condition (2.5.5) and f_m^∞ is replaced by a converged solution $f(t,\cdot)$ at time $t = 35$. On the right, $f(t,\cdot)$ for initial condition (2.5.4) and f_m^∞ replaced by a converged solution $f(t,\cdot)$ at time $t = 35$.

Figure 2.5 shows that the two different initial conditions lead to the same asymptotic state that theoretically only depends on the distribution g.

2.6 Conclusion

We have shown in this chapter the existence of a unique mild solution $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ of the equation (2.1.3) for any initial condition $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$. And the existence of a unique equilibrium state $f \in \mathcal{P}_2^m(\mathbb{R}^d)$ of the equation (2.1.3). We have also shown the exponential convergence of the mild solution towards the equilibrium state for the Wasserstein metric W_2 and for the strong L^1 norm, in the case $\beta = 1$. The result elaborated by Degond, Frouvelle, Raoul in Ref. [35] has been extended to the cases $g \in \mathcal{P}_2^0(\mathbb{R}^d)$.

2.A Appendix: Hydrodynamic description of the BDG Model

In this appendix section, we now consider the inhomogeneous version of the BDG model (1.4.1). We add the space variable. The position x and the velocity v belong to $\mathbb{R}^d$, the collision kernel K is always defined by (1.4.2) with $g \in \mathcal{P}_2^0(\mathbb{R}^d)$ of covariance matrix Σ_g and the collision rate β is always constant: $\beta = 1$. The density $f(t, x, v)$ satisfies the following equation in the sense of distributions

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q_+(f, f) - Q_-(f, f)$$

where for any test function $\varphi \in \mathcal{C}_c^\infty(\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d)$

$$\langle Q_+(f, f), \varphi \rangle := \int_0^{+\infty} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(t, x, v)(\tau_{(v'+v'_\star)/2} \# g)(dv) f(t, x, dv') f(t, x, dv'_\star) dt dx \quad (2.A.1)$$

and

$$\langle Q_-(f, f), \varphi \rangle := \int_0^{+\infty} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(t, x, v) f(t, x, dv) f(t, x, dv') dt dx. \quad (2.A.2)$$

In other words, f is solution of

$$\begin{cases} \dfrac{\partial f}{\partial t} + v \cdot \nabla_x f = \displaystyle\iint_{\mathbb{R}^d \times \mathbb{R}^d} (\tau_{(v'+v'_\star)/2} \# g) f(t, x, dv') f(t, x, dv'_\star) - f(t, x, \cdot) \int_{\mathbb{R}^d} f(t, x, dv') \\ \\ f(0, \cdot) = f_0 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d). \end{cases} \quad (2.A.3)$$

This equation (2.A.3) is to be taken in the weak sense. Let $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d))$. For any test function $\varphi \in \mathcal{C}^\infty(\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d)$ with compact support, we define

$$\langle f, \varphi \rangle := \int_0^{+\infty} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(t, x, v) f(t, x, dv) dt dx$$

which allows to define a solution in the sense of distributions as follows.

Definition 2.A.1

A **solution in the sense of distributions** of the equation (2.A.3) is a measured-valued function $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d))$ satisfying for every test function $\varphi \in \mathcal{C}^\infty(\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d)$ with compact support

$$-\left\langle f, \frac{\partial \varphi}{\partial t} + v \cdot \frac{\partial \varphi}{\partial x} \right\rangle = \iint_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(0, x, v) f_0(x, dv) dx + \langle Q_+(f, f), \varphi \rangle - \langle Q_-(f, f), \varphi \rangle. \quad (2.A.4)$$

Let $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d))$. Then for any $t \geq 0$, $x \in \mathbb{R}^d$, we define the hydrodynamic quantites ρ and u by

$$\rho(t,x) := \int_{\mathbb{R}^d} f(t,x,dv) \qquad \text{and} \qquad u(t,x) := \frac{1}{\rho(t,x)} \int_{\mathbb{R}^d} v f(t,x,dv). \tag{2.A.5}$$

It is understood that ρ is the function $\rho : \mathbb{R}^+ \longrightarrow \mathcal{P}_2(\mathbb{R}^d)$ obtained by taking the first marginal at each time t. Thus, the equation (2.A.3) can be written equivently as

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = T_g[f(t,x,\cdot)] - \rho(t,x) f(t,x,\cdot) \tag{2.A.6}$$

where T_g is the operator defined by (2.2.6).

We will now establish a hydrodynamic description of the BDG model (2.A.3). We formally take a function $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d))$ such that $Q(f,f) = 0$. By integrating in v the equation (2.A.6), we have the first equation

$$\partial_t \rho + \mathrm{div}_x(\rho u) = 0.$$

Then, we multiply by v the equation (2.A.6) and we integrate in v. One obtains the second equation

$$\partial_t(\rho u) + \mathrm{div}_x \left(\int_{\mathbb{R}^d} v \otimes v f(t,x,dv) \right) = 0.$$

Since $Q(f,f) = 0$, then $T_g[f] = \rho f$ and we have by direct computation

$$\begin{aligned}
\int_{\mathbb{R}^d} v \otimes v f(t,x,dv) &= \frac{1}{\rho(t,x)} \int_{\mathbb{R}^d} v \otimes v T_g[f(t,x,\cdot)](dv) \\
&= \frac{1}{\rho(t,x)} \left(\rho(t,x)^2 \Sigma_g + \frac{\rho(t,x)^2 u(t,x) \otimes u(t,x)}{2} + \frac{\rho(t,x)}{2} \int_{\mathbb{R}^d} v \otimes v f(t,x,dv) \right) \\
&= \rho(t,x)\Sigma_g + \frac{\rho(t,x) u(t,x) \otimes u(t,x)}{2} + \frac{1}{2} \int_{\mathbb{R}^d} v \otimes v f(t,x,dv).
\end{aligned}$$

We obtains the following closed system of equations on ρ and u

$$\begin{cases} \partial_t \rho + \mathrm{div}_x(\rho u) = 0, \\ \\ \partial_t(\rho u) + \mathrm{div}_x(\rho u \otimes u + 2\rho\Sigma_g) = 0. \end{cases} \tag{2.A.7}$$

This system (2.A.7) is very similar to the isothermal compressible Euler system (1.2.6). Except the fact that the stress tensor is not diagonal. This comes from the fact that the matrix Σ_g is not diagonal but only positive semi-definite. In the future, we intent to study the hydrodynamic limits of this system (2.A.7).

2.B Appendix: The biased mid-point model

In the model (2.1.3), the velocity chosen after the collision by the two individuals is a median velocity. It is a particular choice. But it is just as interesting to choose a biased velocity with respect to the median. That is to say that the model could consider that the velocities after the collision will no longer be chosen along the mean but along a convex combination of the pre-collisional velocities. In this case, the kernel K would take the following form

$$K(dv, v', v'_\star) := \tau_{\alpha v' + (1-\alpha) v'_\star} \# g(dv) \tag{2.B.1}$$

with $\alpha \in]0,1[$ and $g \in \mathcal{P}_2^0(\mathbb{R}^d)$ a probability measure, discrete or continuous.

Only such a formula (2.B.1) prohibits considering that the agents are indistinguishable. So the choice of the couple $(\alpha, 1-\alpha)$ must be chosen with a probability $(1/2, 1/2)$. We would obtain a kernel K in two terms where the first is when the agent v' has an influence on the agent $v'_\star$ and the second when the agent $v'_\star$ has an influence on the agent v'. Thus, we obtain a symmetry in $(v', v'_\star)$ only the kernel is complicated to write so we allow ourselves to keep the simple form (2.B.1) because we will see that the results obtained will be symmetric in $(\alpha, 1-\alpha)$.

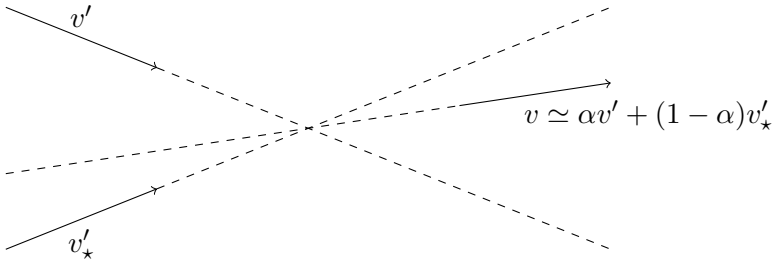


Figure 2.7: Scheme of a collision between two individuals of respective velocity v' and $v'_\star$. The velocity adopted by the two individuals after the collision will be the same, chosen according to a probability distribution centered at the convex combination $\alpha v' + (1-\alpha)v'_\star$.

Thus, the unknown probability distribution f satisfies the following Boltzmann like equation in the sense of distributions

$$\frac{\partial f}{\partial t} = Q_+(f,f) - Q_-(f,f)$$

where for any test function $\varphi \in \mathcal{C}_c^\infty(\mathbb{R} \times \mathbb{R}^d)$

$$\langle Q_+(f,f), \varphi \rangle := \int_0^{+\infty} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(t,v)(\tau_{\alpha v' + (1-\alpha)v'_\star} \# g)(dv) f(t,dv') f(t,dv'_\star) dt \tag{2.B.2}$$

and $Q_-(f,f)$ is defined by (2.1.2):

$$\langle Q_-(f,f), \varphi \rangle := \int_0^{+\infty} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(t,v) f(t,dv) f(t,dv') dt.$$

For $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$, the evolution equation (1.4.1) becomes

$$\begin{cases} \dfrac{\partial f}{\partial t} = \displaystyle\iint_{\mathbb{R}^d \times \mathbb{R}^d} (\tau_{\alpha v' + (1-\alpha)v'_\star} \# g) f(t,dv') f(t,dv'_\star) - f(t,\cdot) \int_{\mathbb{R}^d} f(t,dv') \\ \\ f(0,\cdot) = f_0. \end{cases} \tag{2.B.3}$$

This equation (2.B.3) is to be taken in the weak sense. We define a solution in the weak sense of the equation (2.B.3) in the same way as for the equation (2.1.3) but with Q_+ defined by (2.B.2) instead of (2.1.1).

Definition 2.B.1

A **solution in the sense of distributions** of the equation (2.B.3) is a measured-valued function $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ satisfying for every test function $\varphi \in \mathcal{C}^\infty(\mathbb{R} \times \mathbb{R}^d)$ with compact support

$$-\left\langle f, \frac{\partial \varphi}{\partial t} \right\rangle = \int_{\mathbb{R}^d} \varphi(0,v) f_0(dv) + \langle Q_+(f,f), \varphi \rangle - \langle Q_-(f,f), \varphi \rangle. \tag{2.B.4}$$

For $g \in \mathcal{P}_2^0(\mathbb{R}^d)$ and for $\alpha \in]0,1[$, we define the measured-valued operator $T_{g,\alpha}$ by

$$T_{g,\alpha} : \mu \longmapsto T_{g,\alpha}[\mu] := g * (U_\alpha \# \mu) * (U_{1-\alpha} \# \mu) \tag{2.B.5}$$

where for $a \in \mathbb{R}$, U_a is an uniform scaling with scaling factor a: $U_a : v \in \mathbb{R}^d \longmapsto av \in \mathbb{R}^d$. Thus, the equation (2.B.3) can be written equivently as

$$\frac{\partial f}{\partial t} = T_{g,\alpha}[f(t,\cdot)] - \rho(t) f(t,\cdot) \tag{2.B.6}$$

where for all $t \geq 0$, $\rho(t)$ is the mass at the instant t of the solution f defined by (2.2.3):

$$\rho(t) := \int_{\mathbb{R}^d} f(t, dv).$$

The gain term is always written as a double convolution. Only the function that appears in the second term of the double convolution (2.B.5) and that in the third term are different. This will complicate the computations.

Definition 2.B.2

A **mild solution** of the equation (2.B.3) is a function $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ taking values in the space of probability measures equipped with the Wasserstein metric W_2 satisfying for all $t \geq 0$

$$f(t,\cdot) = e^{-t} f_0 + \int_0^t e^{-(t-s)} T_{g,\alpha}[f(s,\cdot)] ds. \tag{2.B.7}$$

The notion of mild solution is stronger than the first one since a mild solution is a weak solution (See Proof of item (1) of Proposition 2.2.1). By direct computation, we have

$$\int_{\mathbb{R}^d} T_{g,\alpha}[f(t,\cdot)](dv) = \rho(t)^2, \tag{2.B.8}$$

$$\int_{\mathbb{R}^d} v T_{g,\alpha}[f(t,\cdot)](dv) = \rho(t)^2 u(t), \tag{2.B.9}$$

$$\int_{\mathbb{R}^d} v\, {}^t v T_{g,\alpha}[f(t,\cdot)](dv) = \rho(t)^2 (\Sigma_g + u(t)\, {}^t u(t)) + \frac{(\alpha^2 + (1-\alpha)^2)\Sigma_f(t)}{\rho(t)} \tag{2.B.10}$$

where for all $t \geq 0$, $u(t)$ and $\Sigma_f(t)$ are the bulk velocity and covariance matrix at the instant t of the solution f defined respectively by (2.2.4) and (2.2.5)

$$u(t) := \frac{1}{\rho(t)} \int_{\mathbb{R}^d} v f(t, dv) \qquad \text{and} \qquad \Sigma_f(t) := \int_{\mathbb{R}^d} (\rho(t)v - \rho(t)u(t))\, {}^t \rho(t)v - \rho(t)u(t) f(t, dv).$$

This calculation shows that for any $\alpha \in]0,1[$, $T_{g,\alpha}$ maps $\mathcal{P}_2^m(\mathbb{R}^d)$ into itself, hence shows the mass and the mean velocity are preserved, but not the energy. The gain term $Q_+(f,f)$ is a density if g is a density (even if f is a probability measure). Thus, $T_{g,\alpha}[f(t,\cdot)]$ is a density for all $t \geq 0$ if g is a density and if moreover f_0 is a probability density, then by (2.B.7), $f(t,\cdot)$ is a probability density for all $t \geq 0$ where f is the mild solution of the equation (2.B.3).

We now consider the equilibrium state of the equation (2.B.3). Equilibrium states of equation (2.B.3) corresponding to the probability distribution functions satisfying $Q(f,f) = 0$.

Definition 2.B.3

An **equilibrium state** of the equation (2.B.3) is a probability distribution $f \in \mathcal{P}_2(\mathbb{R}^d)$ satisfying the fixed point equation

$$f = \iint_{\mathbb{R}^d \times \mathbb{R}^d} (\tau_{\alpha v' + (1-\alpha)v'_\star} \# g) f(dv') f(dv'_\star). \tag{2.B.11}$$

Equivalently, the equilibrium states of the equation (2.B.3) are the fixed points of the operator $T_{g,\alpha}$. We always have the existence of a unique Mild solution and that this one converges exponentially towards a unique equilibrium state. This is the following Theorem.

Theorem 2.B.1

Let $m \in \mathbb{R}^d$, $\alpha \in]0,1[$, $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$ and $g \in \mathcal{P}_2^0(\mathbb{R}^d)$.

(1) There exists a unique mild solution of the equation (2.B.3) $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ with $f(0, \cdot) = f_0$. In addition, we have for all $t \geq 0$

$$\int_{\mathbb{R}^d} v f(t, dv) = \int_{\mathbb{R}^d} v f_0(dv) := m.$$

(2) There exists a unique equilibrium state $f_m^\infty \in \mathcal{P}_2^m(\mathbb{R}^d)$.

(3) If $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ is the mild solution of equation (2.B.3) with $f(0, \cdot) = f_0$, then for all $t \geq 0$

$$\begin{aligned} \Sigma_f(t) = 2(\alpha^2 + (1-\alpha)^2)((e^{-t/2} - e^{-t})\Sigma_f(0) + (1 + e^{-t} - 2e^{-t/2})\Sigma_g) \\ + e^{-t}\Sigma_f(0) + (1 - e^{-t})\Sigma_g. \end{aligned} \tag{2.B.12}$$

(4) If $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ is the mild solution of equation (2.B.3) with $f(0, \cdot) = f_0$, then we have for all $t \geq 0$

$$W_2(f(t, \cdot), f_m^\infty) \leq e^{-\alpha(1-\alpha)t} W_2(f_0, f_m^\infty). \tag{2.B.13}$$

The key point for showing item (1) and (2) is the following result.

Lemma 2.B.1

For $\mu, \nu \in \mathcal{P}_2^m(\mathbb{R}^d)$ and for $\alpha \in]0,1[$, we have

$$W_2(T_{g,\alpha}[\mu], T_{g,\alpha}[\nu]) \leq \sqrt{\alpha^2 + (1-\alpha)^2} W_2(\mu, \nu). \tag{2.B.14}$$

Proof. Let $\mu, \nu \in \mathcal{P}_2^m(\mathbb{R}^d)$. Then by definition of $T_{g,\alpha}$ operator (2.B.5), we have by sub-additivity of W_2 with respect to the convolution (A.4.13) and (A.4.14) that

$$\begin{aligned} W_2(T_{g,\alpha}[\mu], T_{g,\alpha}[\nu])^2 &\leq W_2((U_\alpha \# \mu) * (U_{1-\alpha} \# \mu), (U_\alpha \# \nu) * (U_{1-\alpha} \# \nu))^2 \\ &\leq W_2(U_\alpha \# \mu, U_\alpha \# \nu)^2 + W_2(U_{1-\alpha} \# \mu, U_{1-\alpha} \# \nu)^2. \end{aligned}$$

Then, since U_α and $U_{1-\alpha}$ are two Lipschitzians maps of respectives Lipschitz norm $\|U_\alpha\|_{\text{Lip}} = \alpha$ and $\|U_{1-\alpha}\|_{\text{Lip}} = 1 - \alpha$, then by (A.4.5) we have

$$W_2(U_\alpha \# \mu, U_\alpha \# \nu)^2 \leq \alpha^2 W_2(\mu, \nu)^2 \qquad \text{and} \qquad W_2(U_{1-\alpha} \# \mu, U_{1-\alpha} \# \nu)^2 \leq (1-\alpha)^2 W_2(\mu, \nu)^2$$

and (2.B.14) follows. ■

Proof of Theorem 2.B.1. (1) Let $m \in \mathbb{R}^d$, $\alpha \in]0,1[$, $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$ and $g \in \mathcal{P}_2^0(\mathbb{R}^d)$. Define $E_{f_0} := \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ with $f(0,\cdot) = f_0$ equipped with the uniform norm. For $f \in E_{f_0}$, we take the map $\Phi : E_{f_0} \longrightarrow E_{f_0}$ defined by (2.2.13):

$$\Phi[f](t,\cdot) := e^{-t} f(0,\cdot) + \int_0^t e^{-(t-s)} T_{g,\alpha}[f(s,\cdot)]ds.$$

Let $f \in E_{f_0}$. It is clear that $\Phi[f](0,\cdot) = f_0$ and since for all $t \geq 0$, $f(t,\cdot) \in \mathcal{P}_2^m(\mathbb{R}^d)$, we have $\rho(t) = 1$ and $u(t) = m$. And so by (2.B.8), (2.B.9) and (2.B.10), $\Phi[f](t,\cdot) \in \mathcal{P}_2^m(\mathbb{R}^d)$ for all $t \geq 0$. By writing

$$\Phi[f](t,\cdot) = e^{-t} f_0 + (1 - e^{-t}) \int_0^t \frac{e^{-(t-s)}}{1 - e^{-t}} T_{g,\alpha}(s,\cdot)]ds,$$

we have by convexity of W_2 (A.4.12) that for all $f^1 \in E_{f_0^1}$, $f^2 \in E_{f_0^2}$ with f_0^1, $f_0^2 \in \mathcal{P}_2^m(\mathbb{R}^d)$,

$$\begin{aligned} W_2(\Phi[f^1](t,\cdot), \Phi[f^2](t,\cdot))^2 \leq (1 - e^{-t}) W_2 \left(\int_0^t \frac{e^{-(t-s)}}{1 - e^{-t}} T^1_{g,\alpha}(s,\cdot)]ds, \int_0^t \frac{e^{-(t-s)}}{1 - e^{-t}} T^2_{g,\alpha}(s,\cdot)]ds \right)^2 \\ + e^{-t} W_2(f_0^1, f_0^2)^2 \end{aligned}$$

And using (2.2.12), it holds that

$$W_2(\Phi[f^1](t,\cdot), \Phi[f^2](t,\cdot))^2 \leq e^{-t} W_2(f_0^1, f_0^2)^2 + \int_0^t e^{-(t-s)} W_2(T_{g,\alpha}[f^1(s,\cdot)], T_{g,\alpha}[f^2(s,\cdot)])^2 ds.$$

And by (2.B.14), we obtain

$$\begin{aligned} W_2(\Phi[f^1](t,\cdot), \Phi[f^2](t,\cdot))^2 \leq (\alpha^2 + (1-\alpha)^2) \int_0^t e^{-(t-s)} W_2(f^1(s,\cdot), f^2(s,\cdot))^2 ds \\ + e^{-t} W_2(f_0^1, f_0^2)^2. \end{aligned} \tag{2.B.15}$$

Now, we use the Banach fixed point Theorem since a Mild solution of the equation (2.B.3) corresponds to a fixed point of Φ in the space E_{f_0}. Let f_1, $f_2 \in E_{f_0}$. We have by (2.B.15)

$$W_2(\Phi[f_1](t,\cdot), \Phi[f_2](t,\cdot))^2 \leq (\alpha^2 + (1-\alpha)^2) \int_0^t e^{-(t-s)} W_2(f_1(s,\cdot), f_2(s,\cdot))^2 ds.$$

Passing to the supremum in time,

$$\begin{aligned} W_2(\Phi[f_1](t,\cdot), \Phi[f_2](t,\cdot))^2 &\leq (\alpha^2 + (1-\alpha)^2) \left(\sup_{t \in \mathbb{R}^+} W_2(f_1(t,\cdot), f_2(t,\cdot))^2 \right) \int_0^t e^{-(t-s)} ds \\ &= (\alpha^2 + (1-\alpha)^2)(1 - e^{-t}) \sup_{t \in \mathbb{R}^+} W_2(f_1(t,\cdot), f_2(t,\cdot))^2. \end{aligned}$$

One obtains

$$\sup_{t \in \mathbb{R}^+} W_2(\Phi[f_1](t,\cdot), \Phi[f_2](t,\cdot)) \leq \sqrt{\alpha^2 + (1-\alpha)^2} \sup_{t \in \mathbb{R}^+} W_2(f_1(t,\cdot), f_2(t,\cdot)). \tag{2.B.16}$$

Hence Φ preserves the space E_{f_0} and is a contraction since $\alpha \in]0,1[$. By Lemma 2.2.2, $\mathcal{P}_2^m(\mathbb{R}^d)$ is a complete metric space for W_2 so E_{f_0} is complete. Hence there exists a unique mild solution of the equation (2.B.3) belonging to E_{f_0}.

(2) Let $f \in \mathcal{P}_2^m(\mathbb{R}^d)$. Then $T_{g,\alpha}[f] \in \mathcal{P}_2^m(\mathbb{R}^d)$ since $T_{g,\alpha}$ maps $\mathcal{P}_2^m(\mathbb{R}^d)$ into itself by (2.B.8), (2.B.9) and (2.B.10). And by (2.B.14), $T_{g,\alpha}$ is a contraction on the metric space $(\mathcal{P}_2^m(\mathbb{R}^d), W_2)$ which is complete by Lemma 2.2.2. Therefore by The Banach fixed point Theorem, there exists a unique $f \in \mathcal{P}_2^m(\mathbb{R}^d)$ such that $T_{g,\alpha}[f] = f$. And so there exists a unique equilibrium state in $\mathcal{P}_2^m(\mathbb{R}^d)$ since the equilibrium states correspond to the fixed points of $T_{g,\alpha}$.

(3) We have by (2.B.7) and (2.B.10)

$$\begin{aligned}\int_{\mathbb{R}^d} v\,{}^t v f(t,dv) &= e^{-t}\int_{\mathbb{R}^d} v\,{}^t v f_0(dv) + \int_0^t e^{-(t-s)}\left(\int_{\mathbb{R}^d} v\,{}^t v T_{g,\alpha}[f(s,\,\cdot\,)](dv)\right) ds \\ &= e^{-t}\int_{\mathbb{R}^d} v\,{}^t v f_0(dv) + \int_0^t e^{-(t-s)}(\Sigma_g + m\,{}^t m + (\alpha^2+(1-\alpha)^2)\Sigma_f(s))ds.\end{aligned}$$

So we have

$$e^t\Sigma_f(t) = \Sigma_f(0) + (e^t-1)\Sigma_g + \int_0^t (\alpha^2+(1-\alpha)^2)e^s\Sigma_f(s)ds.$$

Applying the case of equality in Gronwall's Lemma (B.2.5), we have

$$\begin{aligned}e^t\Sigma_f(t) &= \Sigma_f(0) + (e^t-1)\Sigma_g + \int_0^t (\alpha^2+(1-\alpha)^2)(\Sigma_f(0) + (e^s-1)\Sigma_g)e^{(t-s)/2}ds \\ &= \Sigma_f(0) + (e^t-1)\Sigma_g + 2(\alpha^2+(1-\alpha)^2)((e^{t/2}-1)\Sigma_f(0) + (e^t+1-2e^{t/2})\Sigma_g).\end{aligned}$$

Which implies formula (2.B.12).

(4) Let $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ be the mild solution of equation 2.B.3 with initial condition $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$. Note that

$$\forall t \geq 0, \quad f_m^\infty = e^{-t} f_m^\infty + \int_0^t e^{-(t-s)} T_{g,\alpha}[f_m^\infty]ds.$$

By taking the map Φ defined in (2.2.13), we have by (2.B.15)

$$\begin{aligned}W_2(f(t,\,\cdot\,), f_m^\infty)^2 &= W_2(\Phi[f](t,\,\cdot\,), \Phi[f_m^\infty])^2 \\ &\leq e^{-t}W_2(f_0, f_m^\infty)^2 + (\alpha^2+(1-\alpha)^2)\int_0^t e^{-(t-s)}W_2(f(s,\,\cdot\,), f_m^\infty)^2 ds.\end{aligned}$$

By Gronwall's Lemma (Lemma B.2.1), we have

$$\begin{aligned}e^t W_2(f(t,\,\cdot\,), f_m^\infty)^2 &\leq W_2(f_0, f_m^\infty)^2 + \int_0^t (\alpha^2+(1-\alpha)^2)W_2(f_0, f_m^\infty)^2 e^{(\alpha^2+(1-\alpha)^2)(t-s)}ds \\ &= W_2(f_0, f_m^\infty)^2 + (e^{t(\alpha^2+(1-\alpha)^2)} - 1)W_2(f_0, f_m^\infty)^2 \\ &= e^{t(\alpha^2+(1-\alpha)^2)}W_2(f_0, f_m^\infty)^2.\end{aligned}$$

Which implies (2.B.13). ■

For $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ and for $t \geq 0$, the function $\hat{f}(t,\,\cdot\,) : \xi \longmapsto \hat{f}(t,\xi)$ is the Fourier transform of the probability measure $f(t,\,\cdot\,)$. So if $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2^m(\mathbb{R}^d))$ is the mild solution of equation (2.B.3) with initial condition $f(0,\,\cdot\,) = f_0$, then $\hat{f}$ is solution of the fixed point equation

$$\hat{f}(t,\xi) = e^{-t}\hat{f}(0,\xi) + \int_0^t e^{-(t-s)}\hat{g}(\xi)\hat{f}(s,\alpha\xi)\hat{f}(s,(1-\alpha)\xi)ds. \tag{2.B.17}$$

This equation (2.B.17) allows to differentiate $\hat{f}$ which satisfies therefore

$$\begin{cases} \dfrac{\partial \hat{f}}{\partial t} = \hat{g}(\xi)\hat{f}(s,\alpha\xi)\hat{f}(s,(1-\alpha)\xi) - \hat{f}(t,\xi), \\ \\ \hat{f}(0,\xi) = \hat{f}_0(\xi). \end{cases} \tag{2.B.18}$$

Since an equilibrium state of the equation (2.B.3) is a fixed point of $T_{g,\alpha}$, then passing in Fourier the equation (2.B.11) we obtain that the Fourier transform of a fixed point of $T_{g,\alpha}$ satisfies

$$\hat{f}(\xi) = \hat{g}(\xi)\hat{f}(s, \alpha\xi)\hat{f}(s, (1-\alpha)\xi). \tag{2.B.19}$$

By iterating this equation (2.B.19), we obtain a recurrence relation for the Fourier transform of the equilibrium state of the equation (2.B.3):

$$\forall n \in \mathbb{N}^*, \qquad \hat{f}(\xi) = \left(\prod_{k=0}^{n-1}\prod_{l=0}^{k} \hat{g}(\alpha^l(1-\alpha)^{k-l}\xi)^{C_l^k}\right)\left(\prod_{k=0}^{n} \hat{f}(\alpha^k(1-\alpha)^{n-k}\xi)^{C_k^n}\right) \tag{2.B.20}$$

where $C_k^n := n!/(k!(n-k)!)$ are the binomial coefficients.

Chapter 3

The homogeneous continuous midpoint model with collision rate

3.1 Introduction

In this chapter, we still consider the simplified version of the BDG model (1.4.1) studied in the previous chapter, but this time with a non-constant collision rate: the density f is independent of the position of the individuals, the velocity v belongs to $\mathbb{R}^d$ and not $\mathbb{S}^{d-1}$, the collision kernel K is defined by (1.4.2) with $g \in \mathcal{P}_2^0(\mathbb{R}^d)$ of covariance matrix Σ_g and the collision rate $\beta : (v, v') \in \mathbb{R}^d \times \mathbb{R}^d \longmapsto \beta(v, v') \in [0, 1]$ is a continuous symmetric function. Thus, the unknown probability distribution f satisfies the following Boltzmann like equation in the sense of distributions

$$\frac{\partial f}{\partial t} = Q_+(f, f) - Q_-(f, f)$$

where for any test function $\varphi \in \mathcal{C}_c^\infty(\mathbb{R} \times \mathbb{R}^d)$

$$\langle Q_+(f, f), \varphi \rangle := \int_0^{+\infty} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(t, v)\beta(v', v'_\star)(\tau_{(v'+v'_\star)/2}\#g)(dv)f(t, dv')f(t, dv'_\star)dt \tag{3.1.1}$$

and

$$\langle Q_-(f, f), \varphi \rangle := \int_0^{+\infty} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(t, v)\beta(v, v')f(t, dv)f(t, dv')dt. \tag{3.1.2}$$

For $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$, the evolution equation (1.4.1) becomes

$$\begin{cases} \dfrac{\partial f}{\partial t} = \displaystyle\iint_{\mathbb{R}^d \times \mathbb{R}^d} \beta(v', v'_\star)(\tau_{(v'+v'_\star)/2}\#g)f(t, dv')f(t, dv'_\star) - f(t, \cdot) \int_{\mathbb{R}^d} \beta(v, v')f(t, dv') \\ \\ f(0, \cdot) = f_0 \in \mathcal{P}_2^m(\mathbb{R}^d). \end{cases} \tag{3.1.3}$$

This equation (3.1.3) is to be taken in the weak sense. We define a solution in the weak sense of the equation (3.1.3) in the same way as for the equation (2.1.3) but with Q_+ and Q_- defined by (3.1.1) and (3.1.2) instead of (2.1.1) and (2.1.2).

Definition 3.1.1

A **solution in the sense of distributions** of the equation (3.1.3) is a measured-valued function $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ satisfying for every test function $\varphi \in \mathcal{C}^\infty(\mathbb{R} \times \mathbb{R}^d)$ with compact support

$$-\left\langle f, \frac{\partial \varphi}{\partial t} \right\rangle = \int_{\mathbb{R}^d} \varphi(0, v)f_0(dv) + \langle Q_+(f, f), \varphi \rangle - \langle Q_-(f, f), \varphi \rangle. \tag{3.1.4}$$

The plan of this chapter is the following. We start in section 3.2 by introducing the operator $T_{g,\beta}$ depending on g and β which will allow us to write the model (3.1.3) as a linear ODE (B.1). We will show in this same section various properties of this operator. We will then show in the next section that the operator $T_{g,\beta}$ admits a fixed point, which amounts to saying that the model (3.1.3) admits an equilibrium state. The last section will first explain how to implement numerically the operator $T_{g,\beta}$ and will then be devoted to numerical simulations in dimension 1.

3.2 The $T_{g,\beta}$ operator

3.2.1 Definition of the operator

As in the previous chapter, we denote for all $t \geq 0$, $\rho(t)$, $u(t)$ and $\Sigma_f(t)$ the mass, bulk velocity and covariance matrix at the instant t of the solution f defined respectively by (2.2.3), (2.2.4) and (2.2.5),

$$\begin{aligned}
\rho(t) &:= \int_{\mathbb{R}^d} f(t,dv),\\
u(t) &:= \frac{1}{\rho(t)} \int_{\mathbb{R}^d} v f(t,dv),\\
\Sigma_f(t) &:= \int_{\mathbb{R}^d} (\rho(t)v - \rho(t)u(t))\,{}^t(\rho(t)v - \rho(t)u(t)) f(t,dv).
\end{aligned}$$

We define the following measured-valued operator $T_{g,\beta}$ for any measure μ by

$$\begin{aligned}
T_{g,\beta} : \mu \longmapsto T_{g,\beta}[\mu] := &\iint_{\mathbb{R}^d\times\mathbb{R}^d} \beta(v',v'_\star)(\tau_{(v'+v'_\star)/2}\#g)\mu(dv')\mu(dv'_\star)\\
&+ \iint_{\mathbb{R}^d\times\mathbb{R}^d} (1-\beta(v,v'))\frac{\delta_v+\delta_{v'}}{2}\mu(dv)\mu(dv').
\end{aligned} \tag{3.2.1}$$

In other words for any test function φ we have

$$\begin{aligned}
\int_{\mathbb{R}^d} \varphi(v) T_{g,\beta}[\mu](dv) = &\iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \varphi\left(v + \frac{v'+v'_\star}{2}\right)\beta(v',v'_\star)g(dv)\mu(dv')\mu(dv'_\star)\\
&+ \iint_{\mathbb{R}^d\times\mathbb{R}^d} (1-\beta(v,v'))\frac{\varphi(v)+\varphi(v')}{2}\mu(dv)\mu(dv').
\end{aligned} \tag{3.2.2}$$

Note that if $\beta = 1$, then we have $T_{g,\beta} = T_g$ where T_g is the operator in the previous chapter defined by (2.2.6). For any test function $\varphi \in \mathcal{C}^\infty_c(\mathbb{R}\times\mathbb{R}^d)$ we have

$$\begin{aligned}
\langle T_{g,\beta}[f],\varphi\rangle &= \int_0^{+\infty}\int_{\mathbb{R}^d} \varphi(t,v) T_{g,\beta}[f(t,\cdot)](dv)dt\\
&= \int_0^{+\infty} \iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \varphi\left(t, v + \frac{v'+v'_\star}{2}\right)\beta(v',v'_\star)g(dv)f(t,dv')f(t,dv'_\star)dt\\
&+ \int_0^{+\infty} \iint_{\mathbb{R}^d\times\mathbb{R}^d} (1-\beta(v,v'))\frac{\varphi(t,v)+\varphi(t,v')}{2} f(t,dv)f(t,dv')dt.
\end{aligned}$$

Since β is symmetric, then the second term is symmetric on (v,v'). One obtains

$$\begin{aligned}
\langle T_{g,\beta}[f],\varphi\rangle &= \int_0^{+\infty} \iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \varphi(t,v)\beta(v',v'_\star)(\tau_{(v'+v'_\star)/2}\#g)(dv)f(t,dv')f(t,dv'_\star)dt\\
&+ \int_0^{+\infty} \iint_{\mathbb{R}^d\times\mathbb{R}^d} (1-\beta(v,v'))\varphi(t,v)f(t,dv)f(t,dv')dt\\
&= \langle Q_+(f,f),\varphi\rangle - \langle Q_-(f,f),\varphi\rangle + \langle \rho f,\varphi\rangle.
\end{aligned}$$

And so the equation (3.1.3) can be written equivently as

$$\frac{\partial f}{\partial t} = T_{g,\beta}[f(t,\cdot)] - \rho(t) f(t,\cdot). \tag{3.2.3}$$

The symmetry of the β function allowed us to write the equation (3.1.3) in the form (3.2.3), which is a linear ODE (B.1) in infinite dimension with initial condition $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$. By comparison with (B.1), we have $a = -\rho$ and $b = T_{g,\beta}$ which depends on time and the unknown f. With the equation (3.2.3), we can, as in the previous chapter, define a mild solution of the equation (3.1.3) using the Duhamel's Formula (B.1.1).

Definition 3.2.1

A **mild solution** of the equation (3.1.3) is a function $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ taking values in the space of probability measures equipped with the Wasserstein metric W_2 satisfying for all $t \geq 0$

$$f(t,\cdot) = e^{-t} f_0 + \int_0^t e^{-(t-s)} T_{g,\beta}[f(s,\cdot)] ds. \tag{3.2.4}$$

By direct computation, we have

$$\int_{\mathbb{R}^d} T_{g,\beta}[f(t,\cdot)](dv) = \rho(t)^2, \tag{3.2.5}$$

$$\int_{\mathbb{R}^d} v T_{g,\beta}[f(t,\cdot)](dv) = \rho(t)^2 u(t), \tag{3.2.6}$$

$$\begin{aligned}\int_{\mathbb{R}^d} |v|^2 T_{g,\beta}[f(t,\cdot)](dv) \leq \rho(t)^2 \|\beta\|_\infty \int_{\mathbb{R}^d} |v|^2 g(dv) + \frac{\rho(t)(1+\|1-\beta\|_\infty)}{2} \int_{\mathbb{R}^d} |v|^2 f(t,dv) \\ + \rho(t)^2 |u(t)|^2 \frac{1-\|1-\beta\|_\infty}{2}.\end{aligned} \tag{3.2.7}$$

This calculation shows that the equation (3.1.3) preserves the mass and the mean velocity. We obtain (3.2.5) and (3.2.6) immediately by definition of the $T_{g,\beta}$ operator (3.2.2) and by the fact that $g \in \mathcal{P}_2^0(\mathbb{R}^d)$. The energy is however not preserved but we can control the moment of order 2 of $T_{g,\beta}[f(t,\cdot)]$ if we control the moment of order 2 of $f(t,\cdot)$. The estimate (3.2.7) is a little more complex to obtain. We perform the same calculation as in the proof of Proposition 3.2.1 to obtain the estimate (3.2.8) using the expressions that define ρ and u.

Note that if β is not symmetric, then we can still reduce the equation (3.1.3) to a linear ODE. It suffices to symmetrize the $T_{g,\beta}$ operator in the following way. By setting $\tilde{\beta}$ the function $\tilde{\beta}(v, v_\star) := \beta(v_\star, v)$, we define the operator $\overline{T}_{g,\beta} := T_{g,(\beta+\tilde{\beta})/2}$. We have $\tilde{\beta} = \beta$ and $\overline{T}_{g,\beta} = T_{g,\beta}$ if β is symmetric and in all cases the equation (3.1.3) is equivalent to

$$\frac{\partial f}{\partial t} = \overline{T}_{g,\beta}[f(t,\cdot)] - \rho(t) f(t,\cdot).$$

However, considering β non-symmetric is unjustifiable from a physical point of view. There is no reason to consider that the agent v' has more influence than the agent $v'_\star$. We will assume from now on that $\rho(t) = 1$ for all $t \geq 0$.

3.2.2 Properties of the operator

We state in this section some properties of $T_{g,\beta}$ operator. Here, the function β is not necessarily assumed to be symmetric. The first property of $T_{g,\beta}$ is that it maps $\mathcal{P}_2^m(\mathbb{R}^d)$ into itself. We recall that $M_2(\mu)$ designates the moment of order 2 of μ.

Proposition 3.2.1

If $\mu \in \mathcal{P}_2^m(\mathbb{R}^d)$, then $T_{g,\beta}[\mu] \in \mathcal{P}_2^m(\mathbb{R}^d)$ and we have

$$M_2(T_{g,\beta}[\mu]) \leq M_2(g) + \frac{(1 + \|1-\beta\|_\infty)(M_2(\mu) - |m|^2)}{2} + |m|^2. \tag{3.2.8}$$

Proof. By (3.2.5) and (3.2.6), we have immediately that $T_{g,\beta}[\mu]$ is a probability measure of mean m. Let's now to prove (3.2.8). By definition of $T_{g,\beta}$ we have

$$\begin{aligned}\int_{\mathbb{R}^d} |v|^2 T_{g,\beta}[\mu](dv) = &\iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \left|v + \frac{v' + v'_\star}{2}\right|^2 \beta(v', v'_\star) g(dv)\mu(dv')\mu(dv'_\star) \\ &+ \iint_{\mathbb{R}^d\times\mathbb{R}^d} (1-\beta(v,v'))\frac{|v|^2 + |v'|^2}{2}\mu(dv)\mu(dv').\end{aligned}$$

By expanding, we have since g is zero mean that

$$\begin{aligned}\int_{\mathbb{R}^d} |v|^2 T_{g,\beta}[\mu](dv) = M_2(g)\iint_{\mathbb{R}^d\times\mathbb{R}^d} \beta(v,v')\mu(dv)\mu(dv') + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \left|\frac{v+v'}{2}\right|^2 \beta(v,v')\mu(dv)\mu(dv') \\ + \iint_{\mathbb{R}^d\times\mathbb{R}^d} (1-\beta(v,v'))\frac{|v|^2+|v'|^2}{2}\mu(dv)\mu(dv').\end{aligned}$$

We will add and substract the quantity

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} (1-\beta(v,v'))\left|\frac{v+v'}{2}\right|^2 \mu(dv)\mu(dv').$$

One obtains

$$\begin{aligned}\int_{\mathbb{R}^d} |v|^2 T_\beta[\mu](dv) &= M_2(g)\iint_{\mathbb{R}^d\times\mathbb{R}^d} \beta(v,v')\mu(dv)\mu(dv') + \frac{M_2(\mu)-|m|^2}{2} + m^2 \\ &\quad + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \left|\frac{v-v'}{2}\right|^2 (1-\beta(v,v'))\mu(dv)\mu(dv') \\ &\leq M_2(g) + \frac{1+\|1-\beta\|_\infty}{2}(M_2(\mu)-|m|^2) + |m|^2\end{aligned}$$

and (3.2.8) follows. ■

Since the $T_{g,\beta}$ operator maps $\mathcal{P}_2^m(\mathbb{R}^d)$ into itself, then the convex combination $(T_{g,\beta} + T_{g,1-\beta})/2$ maps $\mathcal{P}_2^m(\mathbb{R}^d)$ into itself. By definition of the $T_{g,\beta}$ operator, we have immediately the following identity

$$T_{g,\beta} + T_{g,1-\beta} = T_g + \mathrm{id}. \tag{3.2.9}$$

And by Proposition 2.3.1, the unique fixed point in $\mathcal{P}_2^m(\mathbb{R}^d)$ of the T_g operator, f_m^∞, is the unique solution of the following fixed-point equation

$$\frac{1}{2}(T_{g,\beta}[\mu] + T_{g,1-\beta}[\mu]) = \mu$$

in $\mathcal{P}_2^m(\mathbb{R}^d)$. The following Proposition gives an explicit formula of $\hat{T}_{g,\beta}[\mu]$ in the Gaussian case.

Proposition 3.2.2

Consider the operator where g is a Gaussian with zero mean and covariance matrix Σ_g and β is a "non-normalized Gaussian" defined by

$$\beta(v, v_\star) := \exp\left(-\frac{{}^t(v - v_\star)\Sigma_\beta^{-1}(v - v_\star)}{2}\right) \tag{3.2.10}$$

where $\Sigma_\beta \succeq 0$. If μ is a probability measure with density f which is Gaussian, with mean m and covariance matrix Σ_f, then there exists a constant C which depends on m, Σ_f and Σ_β such that

$$\hat{T}_{g,\beta}[\mu](\xi) = C\left(\exp\left(-\frac{{}^t\xi\Sigma_g\xi}{2}\right) - \exp\left(-\frac{1}{4}\,{}^t\xi\left(2\Sigma_\beta^{-1} + \Sigma_f^{-1}\right)^{-1}\xi\right)\right) + \hat{\mu}(\xi). \tag{3.2.11}$$

Proof. By definition of the $T_{g,\beta}$ operator and the Fourier transform, we have

$$\begin{aligned}\hat{T}_{g,\beta}[\mu](\xi) = &\iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} e^{-i\langle v+(v'+v'_\star)/2,\xi\rangle}\beta(v', v'_\star)g(dv)f(v')f(v'_\star)dv'dv'_\star \\ &+ \iint_{\mathbb{R}^d\times\mathbb{R}^d}(1-\beta(v,v'))\frac{e^{-i\langle v,\xi\rangle} + e^{-i\langle v',\xi\rangle}}{2}f(v)f(v')dvdv'.\end{aligned}$$

We set $Z[f] := (2\pi)^{d/2}\sqrt{|\det(\Sigma_f)|}$ the normalization factor of f. One obtains

$$\begin{aligned}\hat{T}_{g,\beta}[\mu](\xi) = &\frac{\hat{g}(\xi)}{Z[f]^2}\iint_{\mathbb{R}^d\times\mathbb{R}^d} e^{-i\langle(v+v_\star)/2,\xi\rangle}e^{-({}^t(v-v_\star)\Sigma_\beta^{-1}(v-v_\star)+{}^t(v-m)\Sigma_f^{-1}(v-m)+{}^t(v_\star-m)\Sigma_f^{-1}(v_\star-m))/2}dvdv_\star \\ &- \frac{1}{Z[f]^2}\iint_{\mathbb{R}^d\times\mathbb{R}^d} e^{-i\langle v,\xi\rangle}e^{-({}^t(v-v_\star)\Sigma_\beta^{-1}(v-v_\star)+{}^t(v-m)\Sigma_f^{-1}(v-m)+{}^t(v_\star-m)\Sigma_f^{-1}(v_\star-m))/2}dvdv_\star \\ &+ \hat{f}(\xi).\end{aligned}$$

We make the following change of variables

$$\begin{cases} v - m := \dfrac{u + u_\star}{\sqrt{2}}, \\ \\ v_\star - m := \dfrac{-u + u_\star}{\sqrt{2}}. \end{cases}$$

To compute $dvdv_\star/(dudu_\star)$, we calculate the determinant of the following block matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, B, C and $D \in M_d(\mathbb{R})$ are 4 diagonal matrices: $A = B = D := \mathrm{diag}(1/\sqrt{2}, \cdots, 1/\sqrt{2})$ and $C := -A$. We have

$$\frac{dvdv_\star}{dudu_\star} = |\det(A - BD^{-1}C)\det(D)| = |\det(\mathrm{diag}(\sqrt{2}, \cdots, \sqrt{2}))\det(\mathrm{diag}(1/\sqrt{2}, \cdots, 1/\sqrt{2}))| = 1.$$

This gives

$$\begin{aligned}&\hat{T}_{g,\beta}[\mu](\xi) \\ &= \frac{\hat{g}(\xi)}{Z[f]^2}\iint_{\mathbb{R}^d\times\mathbb{R}^d} e^{-i\langle u_\star/\sqrt{2}+m,\xi\rangle}e^{-(2\,{}^tu\Sigma_\beta^{-1}u+({}^t(u+u_\star)\Sigma_f^{-1}(u+u_\star)+{}^t(-u+u_\star)\Sigma_f^{-1}(-u+u_\star))/2)/2}dudu_\star + \hat{f}(\xi) \\ &- \frac{1}{Z[f]^2}\iint_{\mathbb{R}^d\times\mathbb{R}^d} e^{-i\langle(u+u_\star)/\sqrt{2}+m,\xi\rangle}e^{-(2\,{}^tu\Sigma_\beta^{-1}u+({}^t(u+u_\star)\Sigma_f^{-1}(u+u_\star)+{}^t(-u+u_\star)\Sigma_f^{-1}(-u+u_\star))/2)/2}dudu_\star.\end{aligned}$$

By Fubini Theorem,

$$\hat{T}_{g,\beta}[\mu](\xi) = \frac{e^{-i\langle m,\xi\rangle}\hat{g}(\xi)}{Z[f]^2}\left(\int_{\mathbb{R}^d} e^{-{}^t u(2\Sigma_\beta^{-1}+\Sigma_f^{-1})u/2}du\right)\left(\int_{\mathbb{R}^d} e^{-i\langle u_\star,\xi/\sqrt{2}\rangle}e^{-{}^t u_\star\Sigma_f^{-1}u_\star/2}du_\star\right) + \hat{f}(\xi)$$
$$- \frac{e^{-i\langle m,\xi\rangle}}{Z[f]^2}\left(\int_{\mathbb{R}^d} e^{-i\langle u,\xi/\sqrt{2}\rangle}e^{-{}^t u(2\Sigma_\beta^{-1}+\Sigma_f^{-1})u/2}du\right)\left(\int_{\mathbb{R}^d} e^{-i\langle u_\star,\xi/\sqrt{2}\rangle}e^{-{}^t u_\star\Sigma_f^{-1}u_\star/2}du_\star\right).$$

We set $h(v) := \exp(-{}^t u(2\Sigma_\beta^{-1} + \Sigma_f^{-1})u/2)$. We recognize in the above right-hand side the Fourier transform of h and f. Therefore we have

$$\hat{T}_{g,\beta}[\mu](\xi) = \frac{e^{-i\langle m(\sqrt{2}-1),\xi/\sqrt{2}\rangle}\hat{g}(\xi)}{Z[f]}\hat{f}\left(\frac{\xi}{\sqrt{2}}\right)\int_{\mathbb{R}^d} h(u)du - \frac{e^{-i\langle m(\sqrt{2}-1),\xi/\sqrt{2}\rangle}}{Z[f]}\hat{f}\left(\frac{\xi}{\sqrt{2}}\right)\hat{h}\left(\frac{\xi}{\sqrt{2}}\right) + \hat{f}(\xi).$$

Computing $\hat{h}$, $\hat{T}_{g,\beta}[\mu]$ becomes

$$\hat{T}_{g,\beta}[\mu](\xi) = \frac{e^{-i\langle m(\sqrt{2}-1),\xi/\sqrt{2}\rangle}}{Z[f]}\hat{f}\left(\frac{\xi}{\sqrt{2}}\right)\exp\left(-\frac{{}^t\xi\Sigma_g\xi}{2}\right)\int_{\mathbb{R}^d} h(u)du + \hat{f}(\xi)$$
$$- \frac{e^{-i\langle m(\sqrt{2}-1),\xi/\sqrt{2}\rangle}}{Z[f]}\hat{f}\left(\frac{\xi}{\sqrt{2}}\right)\exp\left(-\frac{1}{4}{}^t\xi\left(2\Sigma_\beta^{-1}+\Sigma_f^{-1}\right)^{-1}\xi\right)\int_{\mathbb{R}^d} h(u)du$$

and (3.2.11) follows. ■

The following Proposition is the most important of this section. We showed in the previous chapter that $T_{g,\beta}$ is a contraction if β is constant equal to 1. In the general case, the $T_{g,\beta}$ operator is unfortunately no longer a contraction but the Hölder condition is satisfied with $\alpha = 1/2$ for some class of β.

Proposition 3.2.3

Let $\mu,\ \nu \in \mathcal{P}_2^m(\mathbb{R}^d)$. Then for every continuous functions $\beta_1,\ \beta_2 : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow [0,1]$ we have

$$W_2(T_{g,\beta_1}[\mu], T_{g,\beta_2}[\nu])^2 \leq \frac{1+\|\min(1-\beta_1, 1-\beta_2)\|_\infty}{2}W_2(\mu,\nu)^2 + \mathrm{Tr}(\Sigma_g)\|\beta_{1,2}\|_\infty$$
$$+ \frac{\mathrm{Tr}(\Sigma_\mu)}{2}\|(\beta_{1,2})_-\|_\infty + \frac{\mathrm{Tr}(\Sigma_\nu)}{2}\|(\beta_{1,2})_+\|_\infty. \tag{3.2.12}$$

where $\beta_{1,2}(v, v_\star, v', v'_\star) := \beta_1(v, v_\star) - \beta_2(v', v'_\star)$.

Proof. Let $\mu,\ \nu \in \mathcal{P}_2^m(\mathbb{R}^d)$ and let $\pi_* \in \Pi(\mu,\nu)$ be an optimal coupling for W_2. By setting $\beta_1 := \beta_1(v, v')$, $\beta_2 := \beta_2(v_\star, v'_\star)$ and $\gamma(dvdv_\star dv'dv'_\star) := \pi_*(dvdv_\star)\pi_*(dv'dv'_\star)$, we define the coupling π for every test function φ by

$$\iint \varphi(v,v_\star)d\pi(v,v_\star) := \iiiint \min(\beta_1,\beta_2)\varphi\left(\frac{v+v'}{2}+w, \frac{v_\star+v'_\star}{2}+w\right)g(dw)d\gamma(v,v_\star,v',v'_\star)$$
$$+ \frac{1}{2}\iiiint (\beta_1-\beta_2)_+\left(\varphi\left(\frac{v+v'}{2}+w, v_\star\right)+\varphi\left(\frac{v+v'}{2}+w, v'_\star\right)\right)g(dw)d\gamma(v,v_\star,v',v'_\star)$$
$$+ \frac{1}{2}\iiiint (\beta_1-\beta_2)_-\left(\varphi\left(v, \frac{v_\star+v'_\star}{2}+w\right)+\varphi\left(v', \frac{v_\star+v'_\star}{2}+w\right)\right)g(dw)d\gamma(v,v_\star,v',v'_\star)$$
$$+ \frac{1}{2}\iiint \min(1-\beta_1, 1-\beta_2)(\varphi(v,v_\star)+\varphi(v',v'_\star))d\gamma(v,v_\star,v',v'_\star).$$

Using the formulas $\min(a,b)+(a-b)_+ = a$ and $\min(1-a,1-b)+(a-b)_- = 1-a$, we have

$$\begin{aligned}
&\iint \varphi(v)d\pi(v,v_\star) = \iiiint (\min(\beta_1,\beta_2)+(\beta_1-\beta_2)_+)\varphi\left(\frac{v+v'}{2}+w\right) g(dw)d\gamma(v,v_\star,v',v'_\star)\\
&+\iiiint (\min(1-\beta_1,1-\beta_2)+(\beta_1-\beta_2)_-)\frac{\varphi(v)+\varphi(v')}{2}d\gamma(v,v_\star,v',v'_\star)\\
&=\iiint \varphi(w)\beta_1(v,v')(\tau_{(v+v')/2}\#g)(dw)\mu(dv)\mu(dv') + \iint (1-\beta_1(v,v'))\frac{\varphi(v)+\varphi(v')}{2}\mu(dv)\mu(dv')\\
&=\int \varphi(v)T_{g,\beta_1}[\mu](dv).
\end{aligned}$$

By doing the same calculation, we obtain by symmetry that $\iint \varphi(v_\star)d\pi(v,v_\star) = \int \varphi(v_\star)T_{g,\beta_2}[\nu](dv_\star)$ so π is a coupling of $(T_{g,\beta_1}[\mu], T_{g,\beta_2}[\nu])$. We take $\varphi(v,v_\star)=|v-v_\star|^2$ to obtains

$$\begin{aligned}
W_2(T_{g,\beta_1}[\mu],T_{g,\beta_2}[\nu])^2 &\le \iiiint \min(\beta_1,\beta_2)\left|\frac{(v-v_\star)+(v'-v'_\star)}{2}\right|^2 d\gamma(v,v_\star,v',v'_\star)\\
&+\frac{1}{2}\iiiint (\beta_1-\beta_2)_+\left(\left|\frac{v+v'}{2}-v_\star+w\right|^2+\left|\frac{v+v'}{2}-v'_\star+w\right|^2\right) g(dw)d\gamma(v,v_\star,v',v'_\star)\\
&+\frac{1}{2}\iiiint (\beta_1-\beta_2)_-\left(\left|\frac{v_\star+v'_\star}{2}-v+w\right|^2+\left|\frac{v_\star+v'_\star}{2}-v'+w\right|^2\right) g(dw)d\gamma(v,v_\star,v',v'_\star)\\
&+\frac{1}{2}\iiiint \min(1-\beta_1,1-\beta_2)(|v-v_\star|^2+|v'-v'_\star|^2)d\gamma(v,v_\star,v',v'_\star).
\end{aligned} \tag{3.2.13}$$

In the right-hand side of (3.2.13), we note I_1 the first term, I_2 the second, I_3 the third and I_4 the last. For the first term, we have by expanding

$$\begin{aligned}
I_1 &= \frac{1}{4}\iiiint \min(\beta_1,\beta_2)(|v-v_\star|^2+|v'-v'_\star|^2)d\gamma(v,v_\star,v',v'_\star)\\
&+\frac{1}{2}\iiiint \min(\beta_1,\beta_2)\langle v-v_\star,v'-v'_\star\rangle d\gamma(v,v_\star,v',v'_\star).
\end{aligned} \tag{3.2.14}$$

Then for the second integral, we have

$$\begin{aligned}
I_2 &= \frac{1}{8}\iiiint (\beta_1-\beta_2)_+(|v-v_\star|^2+|v'-v_\star|^2+|v-v'_\star|^2+|v'-v'_\star|^2)d\gamma(v,v_\star,v',v'_\star)\\
&+\frac{1}{4}\iiiint (\beta_1-\beta_2)_+(\langle v-v_\star,v'-v_\star\rangle+\langle v-v'_\star,v'-v'_\star\rangle)d\gamma(v,v_\star,v',v'_\star)\\
&+M_2(g)\iiiint (\beta_1-\beta_2)_+ d\gamma(v,v_\star,v',v'_\star).
\end{aligned}$$

We add and substract the quantity

$$\frac{1}{2}\iiiint (\beta_1-\beta_2)_+\langle v-v_\star,v'-v'_\star\rangle d\gamma(v,v_\star,v',v'_\star)$$

and the identity $|v-v_\star|^2+|v'-v_\star|^2+|v-v'_\star|^2+|v'-v'_\star|^2+2\langle v-v_\star,v'-v_\star\rangle+2\langle v-v'_\star,v'-v'_\star\rangle-4\langle v-v_\star,v'-v'_\star\rangle = 2|v-v_\star|^2+2|v'-v'_\star|^2+2|v_\star-v'_\star|^2$ gives that

$$\begin{aligned}
I_2 &= \iiiint (\beta_1-\beta_2)_+\left(\frac{\langle v-v_\star,v'-v'_\star\rangle}{2}+M_2(g)\right)d\gamma(v,v_\star,v',v'_\star)\\
&+\frac{1}{4}\iiiint (\beta_1-\beta_2)_+(|v-v_\star|^2+|v'-v'_\star|^2+|v_\star-v'_\star|^2)d\gamma(v,v_\star,v',v'_\star).
\end{aligned} \tag{3.2.15}$$

For the third integral, we have

$$\begin{aligned}I_3 &= \frac{1}{8}\iiiint (\beta_1-\beta_2)_-(|v-v_\star|^2+|v-v'_\star|^2+|v_\star-v'|^2+|v'-v'_\star|^2)d\gamma(v,v_\star,v',v'_\star)\\ &+\frac{1}{4}\iiiint (\beta_1-\beta_2)_-(\langle v-v_\star,v-v'_\star\rangle+\langle v_\star-v',v'_\star-v'\rangle)d\gamma(v,v_\star,v',v'_\star)\\ &+M_2(g)\iiiint (\beta_1-\beta_2)_-d\gamma(v,v_\star,v',v'_\star).\end{aligned}$$

We add and substract the quantity

$$\frac{1}{2}\iiiint (\beta_1-\beta_2)_-\langle v-v_\star,v'-v'_\star\rangle d\gamma(v,v_\star,v',v'_\star)$$

and the identity $|v-v_\star|^2+|v-v'_\star|^2+|v_\star-v'|^2+|v'-v'_\star|^2+2\langle v-v_\star,v-v'_\star\rangle+2\langle v_\star-v',v'_\star-v'\rangle-4\langle v-v_\star,v'-v'_\star\rangle=2|v-v_\star|^2+2|v'-v'_\star|^2+2|v-v'|^2$ gives that

$$\begin{aligned}I_3 &= \iiiint (\beta_1-\beta_2)_-\left(\frac{\langle v-v_\star,v'-v'_\star\rangle}{2}+M_2(g)\right)d\gamma(v,v_\star,v',v'_\star)\\ &+\frac{1}{4}\iiiint (\beta_1-\beta_2)_-(|v-v_\star|^2+|v'-v'_\star|^2+|v-v'|^2)d\gamma(v,v_\star,v',v'_\star).\end{aligned}\tag{3.2.16}$$

For the last integral, we add and substract the quantity

$$\frac{1}{2}\iiiint \min(1-\beta_1,1-\beta_2)\langle v-v_\star,v'-v'_\star\rangle d\gamma(v,v_\star,v',v'_\star)$$

and the identity $2|v-v_\star|^2+2|v'-v'_\star|^2-2\langle v-v_\star,v'-v'_\star\rangle=|v-v_\star|^2+|v'-v'_\star|^2+|(v-v_\star)-(v'-v'_\star)|^2$ gives that

$$\begin{aligned}I_4 &= \frac{1}{4}\iiiint \min(1-\beta_1,1-\beta_2)(|v-v_\star|^2+|v'-v'_\star|^2)d\gamma(v,v_\star,v',v'_\star)\\ &+\frac{1}{4}\iiiint \min(1-\beta_1,1-\beta_2)|(v-v_\star)-(v'-v'_\star)|^2d\gamma(v,v_\star,v',v'_\star)\\ &+\frac{1}{2}\iiiint \min(1-\beta_1,1-\beta_2)\langle v-v_\star,v'-v'_\star\rangle d\gamma(v,v_\star,v',v'_\star).\end{aligned}\tag{3.2.17}$$

We sum (3.2.14), (3.2.15), (3.2.16) and (3.2.17). Using the identities $a_++a_-=|a|$, $\min(a,b)+\min(1-a,1-b)+|a-b|=1$ and that μ and ν have the same mean, we obtain

$$\begin{aligned}W_2(T_{g,\beta_1}[\mu],T_{g,\beta_2}[\nu])^2 &\le \frac{1}{4}\iiiint (|v-v_\star|^2+|v'-v'_\star|^2)d\gamma(v,v_\star,v',v'_\star)\\ &+\frac{1}{4}\iiiint (\beta_1-\beta_2)_+|v_\star-v'_\star|^2d\gamma(v,v_\star,v',v'_\star)+\frac{1}{4}\iiiint (\beta_1-\beta_2)_-|v-v'|^2d\gamma(v,v_\star,v',v'_\star)\\ &+\frac{1}{4}\iiiint \min(1-\beta_1,1-\beta_2)(|v-v_\star|^2+|v'-v'_\star|^2)d\gamma(v,v_\star,v',v'_\star)\\ &+M_2(g)\iiiint |\beta_1-\beta_2|d\gamma(v,v_\star,v',v'_\star).\end{aligned}$$

And (3.2.12) follows since $\mathrm{Tr}(\Sigma_g)=M_2(g)$, $\mathrm{Tr}(\Sigma_\mu)=M_2(\mu)-|m|^2$ and $\mathrm{Tr}(\Sigma_\nu)=M_2(\nu)-|m|^2$. ■

There are several interesting particular cases of (3.2.12). The first is when $\beta_1=\beta_2=1$. In this case, we find (2.2.11). Another is the following Proposition.

Corollary 3.2.1

We assume that β is of the form $\beta(v, v_\star) = \tilde{\beta}(v - v_\star)$ where $\tilde{\beta} : \mathbb{R}^d \longrightarrow [0, 1]$ is a function which satisfies the following Lipschitzian condition. There exists a constant $C > 0$ such that for every v, $v_\star \in \mathbb{R}^d$,

$$|\tilde{\beta}(v) - \tilde{\beta}(v_\star)| \leq C \frac{|v - v_\star|}{\sqrt{1 + |v|^2 + |v_\star|^2}}. \tag{3.2.18}$$

Then for any μ, $\nu \in \mathcal{P}_2^m(\mathbb{R}^d)$ we have

$$\begin{aligned} W_2(T_{g,\beta}[\mu], T_{g,\beta}[\nu])^2 \leq C \left(\sqrt{2}\mathrm{Tr}(\Sigma_g) + \frac{1}{2}\sqrt{\mathrm{Tr}(\Sigma_\mu) + \mathrm{Tr}(\Sigma_\nu)} \right) W_2(\mu, \nu) \\ + \frac{1 + \|1 - \beta\|_\infty}{2} W_2(\mu, \nu)^2. \end{aligned} \tag{3.2.19}$$

Proof. Let μ, $\nu \in \mathcal{P}_2^m(\mathbb{R}^d)$ and let $\pi_* \in \Pi(\mu, \nu)$ be an optimal coupling for W_2. By setting $\beta_1 := \beta(v, v')$, $\beta_2 := \beta(v_\star, v'_\star)$ and $\gamma := \pi_* \otimes \pi_*$, we have by Cauchy-Schwarz

$$\begin{aligned} \iiiint |\beta_1 - \beta_2| d\gamma(v, v_\star, v', v'_\star) &\leq C \iiiint |(v - v') - (v_\star - v'_\star)| d\gamma(v, v_\star, v', v'_\star) \\ &\leq C \left(\iiiint |(v - v') - (v_\star - v'_\star)|^2 d\gamma(v, v_\star, v', v'_\star) \right)^{1/2} \\ &= \sqrt{2} C W_2(\mu, \nu). \end{aligned}$$

And we also have

$$\begin{aligned} &\iiiint (\beta_1 - \beta_2)_+ |v_\star - v'_\star|^2 d\gamma(v, v_\star, v', v'_\star) + \iiiint (\beta_1 - \beta_2)_- |v - v'|^2 d\gamma(v, v_\star, v', v'_\star) \\ &\leq \iiiint |\beta_1 - \beta_2| (|v_\star - v'_\star|^2 + |v - v'|^2) d\gamma(v, v_\star, v', v'_\star) \\ &\leq C \iiiint \frac{|(v - v') - (v_\star - v'_\star)|}{\sqrt{1 + |v - v'|^2 + |v_\star - v'_\star|^2}} (|v_\star - v'_\star|^2 + |v - v'|^2) d\gamma(v, v_\star, v', v'_\star). \end{aligned}$$

One obtains by Cauchy-Schwarz inequality

$$\begin{aligned} &\iiiint (\beta_1 - \beta_2)_+ |v_\star - v'_\star|^2 d\gamma(v, v_\star, v', v'_\star) + \iiiint (\beta_1 - \beta_2)_- |v - v'|^2 d\gamma(v, v_\star, v', v'_\star) \\ &\leq \sqrt{2} C W_2(\mu, \nu) \left(\iiiint (|v - v'|^2 + |v_\star - v'_\star|^2) d\gamma(v, v_\star, v', v'_\star) \right)^{1/2} \\ &= 2 C W_2(\mu, \nu) \sqrt{\mathrm{Tr}(\Sigma_\mu) + \mathrm{Tr}(\Sigma_\nu)}. \end{aligned}$$

So (3.2.12) becomes in this case (3.2.19). ■

3.3 Equilibrium state

3.3.1 Existence of an equilibrium state

We consider the equilibrium state of the equation (3.1.3). As in the previous chapter, the equilibrium states of equation (3.1.3) corresponding to the probability distribution functions satisfying $Q(f, f) = 0$.

Definition 3.3.1

An **equilibrium state** of the equation (3.1.3) is a probability distribution $f \in \mathcal{P}_2(\mathbb{R}^d)$ satisfying the fixed point equation

$$f \int_{\mathbb{R}^d} \beta(v, v') f(dv') = \iint_{\mathbb{R}^d \times \mathbb{R}^d} (\tau_{(v'+v'_\star)/2} \# g) \beta(v', v'_\star) f(dv') f(dv'_\star). \tag{3.3.1}$$

Note that the equilibrium states are densities if g is a density. The goal of this subsection is to show that the equation (3.1.3) admits an equilibrium state.

Theorem 3.3.1 (Existence of an equilibrium state)

There exists an equilibrium state of the equation (3.1.3) $f \in \mathcal{P}_2^m(\mathbb{R}^d)$. Moreover we have

$$M_2(f) \leq \frac{2M_2(g)}{1 - \|1 - \beta\|_\infty} + |m|^2.$$

Rather than considering the equation (3.3.1), we will rather study the existence of fixed points of the operator $T_{g,\beta}$. It is quite difficult to prove that $T_{g,\beta}$ is a contracting operator. So we will use a Schauder fixed-point Theorem to show the existence of an equilibrium state.

We recall that the Schauder fixed-point Theorem asserts that if C is a non-empty closed convex subset of a normed vector space X and if T is a continuous mapping of C into a compact subset of C, then T has a fixed point in C. This result was proved by Schauder (Ref. [78] and Theorem 2.2 in Ref. [16]) then he conjectured the general case in the Scottish book. In 1934, Tychonoff proved the Theorem in the case where C is compact convex and X is locally convex (Ref. [83] and Theorem 3.2 in Ref. [16]). This result is known as Tychonoff fixed-point Theorem. Then, B.V. Singbal extended the Tychonoff result by removing the compactness assumption on C (Appendix in Ref. [16]). It is this result that we will use.

Theorem 3.3.2 (Schauder fixed-point Theorem)

Let X be a locally convex Hausdorff linear topological space, C a non-empty closed convex subset of X, T a continuous mapping of C into a compact subset of C. Then T has a fixed point in C.

Proof of Theorem 3.3.1. Let E be the space $(\mathcal{C}_0(\mathbb{R}^d), \|\cdot\|_\infty)$. We will apply the Schauder fixed-point Theorem (Theorem 3.3.2) with $T = T_{g,\beta}$ and $X = E^*$. By the Riesz-Markov Theorem (Theorem A.3.4), the dual space of E, E^* is the set of bounded Radon measures on $\mathbb{R}^d$ (i.e the set of regular bounded countable additivity "measures" on $\mathbb{R}^d$, $rca(\mathbb{R}^d)$) equipped with the weak* topology, which is the topology inducted by the variational norm. We consider the following space

$$C := \{\mu \in \mathcal{P}_2^m(\mathbb{R}^d),\ M_2(\mu) \leq \Sigma\}$$

with $\Sigma \geq 2M_2(g)/(1 - \|1 - \beta\|_\infty) + |m|^2$. Since probability measures are Radon measures, we have $C \subset E^*$, so we will show that the operator $T_{g,\beta}$ has a fixed point in C. For beginning, let $\mu \in C$, we have by Proposition 3.2.1 that $\int T_{g,\beta}[\mu](dv) = 1$, $\int v T_{g,\beta}[\mu](dv) = m$ and

$$M_2(T_{g,\beta}[\mu]) \leq M_2(g) + \frac{(1 + \|1 - \beta\|_\infty)(M_2(\mu) - |m|^2)}{2} + |m|^2 \leq \frac{2M_2(g)}{1 - \|1 - \beta\|_\infty} + |m|^2 \leq \Sigma.$$

Thus, $T_{g,\beta}$ preserves the space C.

Let us show that C is convex and closed. By linearity of the integral, it is easy to show that $\mathcal{P}_2^m(\mathbb{R}^d)$ is convex. Let μ, $\nu \in C$ and $t \in [0,1]$. Then, since $t\mu + (1-t)\nu \in \mathcal{P}_2^m(\mathbb{R}^d)$, we have $M_2(t\mu + (1-t)\nu) = tM_2(\mu) + (1-t)M_2(\nu)$ and therefore $M_2(t\mu + (1-t)\nu) \leq \Sigma$. Thus, C is convex. Next, let $(\mu_n)_n$ be a sequence in C which converges to μ for the weak* topology. For $R > 0$, we define $\varphi_R : \mathbb{R}^d \longrightarrow [0,1]$ a continuous function such that $\varphi_R(v) = 1$ for $|v| \leq R$ and $\varphi_R(v) = 0$ for $|v| \geq R+1$. We have

$$\left|1 - \int_{\mathbb{R}^d} \varphi_R(v) d\mu_n(v)\right| = \int_{|v|\geq R} (1 - \varphi_R(v)) d\mu_n(v) \leq \frac{1}{R^2} \int_{|v|\geq R} (1 - \varphi_R(v))|v|^2 d\mu_n(v) \leq \frac{\Sigma}{R^2}$$

which is valid for any $R > 0$ therefore μ is a probability measure. Then,

$$\left|m - \int_{\mathbb{R}^d} \varphi_R(v) d\mu_n(v)\right| \leq \int_{|v|\geq R} (1 - \varphi_R(v))|v| d\mu_n(v) \leq \frac{1}{R} \int_{|v|\geq R} (1 - \varphi_R(v))|v|^2 d\mu_n(v) \leq \frac{\Sigma}{R}$$

which is valid for any $R > 0$ therefore μ has for mean m. And finally we have since $\varphi_R \in \mathcal{C}_c(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} \varphi_R(v)|v|^2 d\mu(v) = \lim_{n\to+\infty} \int_{\mathbb{R}^d} \varphi_R(v)|v|^2 d\mu_n(v) \leq \Sigma.$$

By letting $R \longrightarrow +\infty$, we have $M_2(\mu) \leq \Sigma$ and hence C is closed for the weak* topology.

Let us show that the map $T_{g,\beta}$ is continuous for the weak* topology. Let $(\mu_n)_n$ be a sequence in C which converge to $\mu \in C$ for the weak* topology. To show that $T_{g,\beta}$ is continuous for the weak* topology, it suffices to show that the sequence $(T_{g,\beta}[\mu_n])_n$ converges to $T_{g,\beta}[\mu]$ for the weak* topology. For any test function $\varphi \in E$ we have

$$\begin{aligned}
\int_{\mathbb{R}^d} \varphi(v) T_{g,\beta}[\mu_n](dv) &= \iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \varphi\left(v + \frac{v' + v'_\star}{2}\right) \beta(v, v'_\star) g(dv) \mu_n(dv') \mu_n(dv'_\star) \\
&\quad + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \frac{\varphi(v) + \varphi(v')}{2} (1 - \beta(v, v')) \mu_n(dv) \mu_n(dv') \\
&= \iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \Phi(v, v', v'_\star) g(dv) \mu_n(dv') \mu_n(dv'_\star) + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \Psi(v, v') \mu_n(dv) \mu_n(dv')
\end{aligned}$$

where $\Phi(v, v', v'_\star) := \varphi(v + (v' + v'_\star)/2)\beta(v', v'_\star)$ and $\Psi(v, v') := (\varphi(v) + \varphi(v'))(1 - \beta(v, v'))/2$. Since β is continuous, then $\Phi \in \mathcal{C}_c(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$, $\Psi \in \mathcal{C}_c(\mathbb{R}^d \times \mathbb{R}^d)$. So by Lemma A.5.1 we have

$$\begin{aligned}
&\lim_{n\to+\infty} \int_{\mathbb{R}^d} \varphi(v) T_{g,\beta}[\mu_n](dv) \\
&= \lim_{n\to+\infty} \iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \Phi(v, v', v'_\star) g(dv) \mu_n(dv') \mu_n(dv'_\star) + \lim_{n\to+\infty} \iint_{\mathbb{R}^d\times\mathbb{R}^d} \Psi(v, v') \mu_n(dv) \mu_n(dv') \\
&= \iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \Phi(v, v', v'_\star) g(dv) \mu(dv') \mu(dv'_\star) + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \Psi(v, v') \mu(dv) \mu(dv') \\
&= \int_{\mathbb{R}^d} \varphi(v) T_{g,\beta}[\mu](dv).
\end{aligned}$$

We have $T_{g,\beta}[\mu_n] \longrightarrow T_{g,\beta}[\mu]$ for the weak* topology when $n \longrightarrow +\infty$ hence $T_{g,\beta}$ is continuous for the weak* topology.

It remains to show that there exists a compact $K \subset C$ such that $T_{g,\beta} : C \longrightarrow K$. We take $K := \overline{T_{g,\beta}(C)}$. K is closed and included in C since the operator $T_{g,\beta}$ preserves the space C which is closed. Then, we define $B_{E^*} := \{\mu \in E^*,\ \|\mu\| \leq 1\}$ the unit closed ball in E^*. By the Banach-Alaoglu Theorem, B_{E^*} is compact for the weak* topology. Let $\mu \in C$, by setting $\Phi := \{\varphi \in E,\ \|\varphi\|_\infty \leq 1\}$, we have

$$\|\mu\| := \sup_{\varphi\in\Phi} \left|\int_{\mathbb{R}^d} \varphi(v) d\mu(v)\right| = |\mu|(\mathbb{R}^d) = 1$$

since μ is a probability measure therefore $\mu \in B_{E^*}$. So C is compact for the weak* topology since it is closed and contained in a compact set for the weak* topology. This immediately implies that K is compact for the weak* topology.

We have shown that C is a closed convex set and that $T_{g,\beta} : C \longrightarrow K$ is continuous with $K \subset C$ which is compact. So by Schauder fixed-point Theorem (Theorem 3.3.2), the operator $T_{g,\beta}$ has a fixed point in C. ■

In the particular Gaussian case, in other words if g is a Gaussian with zero mean and covariance matrix Σ_g, and if β is a "non-normalized Gaussian" defined by (3.2.10) where $\Sigma_\beta \succeq 0$, then we can explicitly determine an equilibrium state under certain conditions.

Proposition 3.3.1

Consider g a Gaussian with zero mean and covariance matrix Σ_g and β a "non-normalized Gaussian" defined by (3.2.10) where $\Sigma_\beta \succ 4\Sigma_g$. Then there exists an equilibrium state f of the equation (3.1.3) which is Gaussian of mean m and such that this covariance matrix Σ_f satisfies

$$2\Sigma_g = (2\Sigma_\beta^{-1} + \Sigma_f^{-1})^{-1}. \tag{3.3.2}$$

Proof. Let g be a Gaussian with zero mean and covariance matrix Σ_g, and β a "non-normalized Gaussian" defined by (3.2.10) where $\Sigma_\beta \succeq 0$. If we assume that the equilibrium state is Gaussian, then by Proposition 3.2.2, we obtain a fixed point of $T_{g,\beta}$ if we manage to vanishe the first term of the right-hand side in (3.2.11). But this can only be vanished if $\Sigma_\beta \succ 4\Sigma_g$ and in this case, the covariance matrix of f, Σ_f must satisfy (3.3.2). ■

If otherwise we have $\Sigma_\beta \preceq 4\Sigma_g$, then the first term of the right-hand side in (3.2.11) cannot vanish and thus Σ_f cannot satisfy the condition (3.3.2). It just means that there are no equilibrium states which are Gaussian of the equation (3.3.1). If $d = 1$, then g is a Gaussian with zero mean and variance σ_g^2 and β is defined by $\beta(v, v_\star) := \exp(-(v - v_\star)/2b^2)$. And if we have $b^2 > 4\sigma_g^2$, then by Proposition 3.2.2, there exist equilibrium states of the equation (3.1.3) which are Gaussians such that their variance σ_f^2 satisfies

$$\frac{1}{2\sigma_g^2} = \frac{2}{b^2} + \frac{1}{\sigma_f^2}. \tag{3.3.3}$$

Another particular case is when β is of the form $1 - \tilde{\beta}$ where $\tilde{\beta} : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow [0, 1]$ is a symmetric continuous function. In this case, we have $T_{g,\beta} = T_{g,1-\tilde{\beta}}$ and the identity (3.2.9) implies that a fixed point of $T_{g,1-\tilde{\beta}}$ is solution of the equation $T_{g,\tilde{\beta}}[\mu] = T_g[\mu]$.

3.3.2 Discussion on the existence of an equilibrium state in the Wasserstein space

The existence of a fixed point for the $T_{g,\beta}$ operator leads to the existence of an equilibrium state of the equation (3.1.3). One can wonder if one could have directly applied the Schauder fixed-point Theorem (Theorem 3.3.2) on the Wasserstein space $X := (\mathcal{P}_2(\mathbb{R}^d), W_2)$. Taking as in the proof of Theorem 3.3.1 the space

$$C := \{\mu \in \mathcal{P}_2^m(\mathbb{R}^d),\ M_2(\mu) \leq \Sigma\}$$

with $\Sigma \geq 2M_2(g)/(1 - \|1 - \beta\|_\infty) + |m|^2$, we have C which is a subspace of $\mathcal{P}_2^m(\mathbb{R}^d)$ so $C \subset \mathcal{P}_2(\mathbb{R}^d)$. And C is a convex closed for W_2 since $\mathcal{P}_2^m(\mathbb{R}^d)$ is a complete metric space for W_2 (Lemma 2.2.2) and that the convergence for the Wasserstein metric W_2 implies the convergence of the moments of order 2 (Theorem A.5.1).

Proposition 3.3.2

The map $T_{g,\beta} : C \longrightarrow C$ is continuous for the Wasserstein metric W_2.

Proof. By Theorem A.5.1, $T_{g,\beta}$ is continuous for W_2 if for every sequence $(\mu_n)_n$ in C which converge weakly to $\mu \in C$, we have $T_{g,\beta}[\mu_n] \longrightarrow T_{g,\beta}[\mu]$ weakly and $M_2(T_{g,\beta}[\mu_n]) \longrightarrow M_2(T_{g,\beta}[\mu])$ when $n \longrightarrow +\infty$. The weak convergence of $T_{g,\beta}[\mu_n]$ to $T_{g,\beta}[\mu]$ has been established in the proof of Theorem 3.3.1. To prove the convergence of moments of order 2, we use Lemma A.5.4. Since we have shown weak convergence, it suffices to show that the family $\{T_{g,\beta}[\mu_n],\, n \in \mathbb{N}\}$ is tight for moments of order 2, in other words that

$$\lim_{R\to+\infty} \sup_n \int_{|v|>R} |v|^2 T_{g,\beta}[\mu_n](dv) = 0.$$

Since $g \in \mathcal{P}_2^0(\mathbb{R}^d)$, we have

$$\lim_{R\to+\infty} \int_{|v|>R} \varphi_i(v) g(dv) = 0, \qquad i \in \{1,\, 2,\, 3\} \tag{3.3.4}$$

where $\varphi_1(v) = 1$, $\varphi_2(v) = v$ and $\varphi_3(v) = |v|^2$. Then, since $W_2(\mu_n, \mu) \longrightarrow 0$ when n tends to infinity, we have by Lemma A.5.4

$$\lim_{R\to+\infty} \sup_n \int_{|v|>R} \varphi_i(v) \mu_n(dv) = 0, \qquad i \in \{1,\, 2,\, 3\} \tag{3.3.5}$$

where $\varphi_1(v) = 1$, $\varphi_2(v) = |v|$ and $\varphi_3(v) = |v|^2$. And finally, since $\mu_n \in C$, we have

$$\sup_n \int_{\mathbb{R}^d} \mu_n(dv) = 1, \qquad \sup_n \int_{\mathbb{R}^d} v\mu_n(dv) = m \qquad \text{and} \qquad \sup_n \int_{\mathbb{R}^d} |v|^2 \mu_n(dv) \le \Sigma + |m|^2.$$

By definition of $T_{g,\beta}$ we have

$$\begin{aligned}
\int_{|v|>R} |v|^2 T_{g,\beta}[\mu_n](dv) = &\iiint_{|v+(v'+v'_\star)/2|^2>R^2} \left| v + \frac{v'+v'_\star}{2} \right|^2 \beta(v', v'_\star) g(dv) \mu_n(dv') \mu_n(dv'_\star) \\
&+ \frac{1}{2} \iint_{|v|>R} |v|^2 (1 - \beta(v, v')) \mu_n(dv) \mu_n(dv') \\
&+ \frac{1}{2} \iint_{|v'|>R} |v'|^2 (1 - \beta(v, v')) \mu_n(dv) \mu_n(dv').
\end{aligned}$$

In the right-hand side, we respectively define I_1 the first term, I_2 the second and I_3 the third. We will show that each of these three terms tends to 0 when R tends to infinity. For I_1, we use the following inclusion

$$\{(x,y,z) \in \mathbb{R}^3,\, |x+y+z|^2 > R^2\} \subset \left\{ |x|^2 > \left(\frac{R}{3}\right)^2 \right\} \cap \left\{ |y|^2 > \left(\frac{R}{3}\right)^2 \right\} \cap \left\{ |z|^2 > \left(\frac{R}{3}\right)^2 \right\}.$$

One obtains

$$\begin{aligned}
I_1 \le &\iiint_{|v|^2>(R/3)^2} \left| v + \frac{v'+v'_\star}{2} \right|^2 \beta(v', v'_\star) g(dv) \mu_n(dv') \mu_n(dv'_\star) \\
&+ \iiint_{|v'|^2>(2R/3)^2} \left| v + \frac{v'+v'_\star}{2} \right|^2 \beta(v', v'_\star) g(dv) \mu_n(dv') \mu_n(dv'_\star) \\
&+ \iiint_{|v'_\star|^2>(2R/3)^2} \left| v + \frac{v'+v'_\star}{2} \right|^2 \beta(v', v'_\star) g(dv) \mu_n(dv') \mu_n(dv'_\star).
\end{aligned}$$

The right-hand side is a sum of three terms. We denote J_1 the first, J_2 the second and J_3 the third. We use the classical convexity inequality $|a+b+c|^2 \leq 3|a|^2+3|b|^2+3|c|^2$ in the three integrals. For J_1 we have

$$\begin{aligned} J_1 &\leq 3\|\beta\|_\infty \iiint_{|v|^2>(R/3)^2} \left(|v|^2 + \frac{|v'|^2}{4} + \frac{|v'_\star|^2}{4}\right) g(dv)\mu_n(dv')\mu_n(dv'_\star) \\ &\leq 3\|\beta\|_\infty \int_{|v|^2>(R/3)^2} |v|^2 g(dv) + \frac{3\|\beta\|_\infty}{2}\left(\sup_n \int_{\mathbb{R}} |v|^2\mu_n(dv)\right) \int_{|v|^2>(R/3)^2} g(dv). \end{aligned}$$

So by (3.3.4), we have that $\sup_n J_1$ tends to 0 when R tends to infinity. For J_2 we have

$$\begin{aligned} J_2 &\leq 3\|\beta\|_\infty \iiint_{|v'|^2>(2R/3)^2} \left(|v|^2 + \frac{|v'|^2}{4} + \frac{|v'_\star|^2}{4}\right) g(dv)\mu_n(dv')\mu_n(dv'_\star) \\ &\leq 3\|\beta\|_\infty \left(M_2(g) + \frac{1}{4}\sup_n \int_{\mathbb{R}} |v'_\star|^2\mu_n(dv'_\star)\right) \sup_n \int_{|v'|^2>(2R/3)^2} \mu_n(dv') \\ &+ \frac{3\|\beta\|_\infty}{4} \sup_n \int_{|v'|^2>(2R/3)^2} |v'|^2\mu_n(dv'). \end{aligned}$$

So by (3.3.5), we have that $\sup_n J_2$ tends to 0 when R tends to infinity. For J_3, we do the same calculation as J_2 and we obtains the same thing by symmetry on $(v', v'_\star)$. Next for I_2 we have

$$I_2 \leq \frac{\|1-\beta\|_\infty}{2} \iint_{|v|>R} |v|^2\mu_n(dv)\mu_n(dv') \leq \frac{\|1-\beta\|_\infty}{2} \sup_n \int_{|v|>R} |v|^2\mu_n(dv).$$

So we have $\sup_n I_2$ which tends to 0 when R tends to infinity by (3.3.5) and we have the same thing for I_3 by symmetry on (v, v'). We have just shown the convergence of the moments of order 2 by Lemma A.5.4 and therefore the operator $T_{g,\beta}$ is continuous for W_2. ■

Is there a compact $K \subset C$ such that $T_{g,\beta} : C \longrightarrow K$? We come up against a considerable difficulty. For the weak* topology, it was the Banach-Alaoglu Theorem which made it possible to answer positively to this question. Unfortunately for the topology induced by W_2, the closed balls of $\mathcal{P}_2(\mathbb{R}^d)$ are not compact, in other words the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is not locally compact (Proposition 2.2.9 in Ref. [68]).

3.4 Numerical results

3.4.1 Implementation of $T_{g,\beta}$ operator

We consider the non-linear operator $T_{g,\beta}$ on $\mathbb{R}$ introduced in (3.2.1) with g a density $g(v)dv$ and $\beta(v,v') := \tilde{\beta}(v-v')$ where $\tilde{\beta} : \mathbb{R} \longrightarrow [0,1]$ is a continuous even function. We want to solve numerically the fixed point equation $\mu = T_{g,\beta}[\mu]$ where μ is a probability measure supported on the interval $[-R,R]$, $R>0$. For this, we need to write $T_{g,\beta}$ as a linear combination of Dirac masses. Assume that μ is a linear combination of Dirac masses, in other words

$$\mu := \sum_{i=1}^{2n} a_i \delta_{v_i}$$

with $\sum a_i = 1$ and $(v_i)_{1\leq i\leq 2n}$ a family of points of $\mathbb{R}$ such that $-R \leq v_1 < v_2 < \cdots < v_{2n} \leq R$, there are $2n$ points. We choose an uniform subdivision on the interval $[-R,R]$, for all i, $v_i := -R + iR/n$.

The kernel g is supposed to be supported on $[-R,R]$. The non-linear operator $T_{g,\beta}$ is a sum of two operators $T^{(1)}_{g,\beta}$ and $T^{(2)}_{g,\beta}$ defined for any test function φ by

$$\int_{\mathbb{R}} \varphi(v) T^{(1)}_{g,\beta}[\mu](dv) := \int_{\mathbb{R}} \varphi(v) \left(\sum_{i=1}^{2n} \sum_{j=1}^{2n} a_i a_j \tilde{\beta}(v_i - v_j) g\left(v - \frac{v_i + v_j}{2} \right) \right) dv,$$

$$\int_{\mathbb{R}} \varphi(v) T^{(2)}_{g,\beta}[\mu](dv) := \sum_{i=1}^{2n} \sum_{j=1}^{2n} a_i a_j (1 - \tilde{\beta}(v_i - v_j)) \varphi(v_i).$$

$T^{(1)}_{g,\beta}[f]$ is a measure with density h supported in $[-2R,2R]$ where

$$h(v) := \sum_{i=1}^{2n} \sum_{j=1}^{2n} a_i a_j \tilde{\beta}(v_i - v_j) g\left(v - \frac{v_i + v_j}{2} \right)$$

and $T^{(2)}_{g,\beta}[f]$ is a discrete measure:

$$T^{(2)}_{g,\beta}[f] = \sum_{i=1}^{2n} a_i \left(\sum_{j=1}^{2n} a_j (1 - \tilde{\beta}(v_i - v_j) \right) \delta_{v_i}.$$

We assume that g and $\tilde{\beta}$ are known functions that can computed at any value. We discretize any integral of the form $\int_{-2R}^{2R} h(x)dx$ using a Riemann sum on the subdivision $(x_i)_{1 \le i \le 4n}$

$$\int_{-2R}^{2R} h(x)dx = \sum_{i=1}^{4n} \alpha_i h(x_i)$$

with $\sum \alpha_i = 4R$. For the Riemann algorithm, we use an uniform step, we have $\alpha_i = R/n$ and $x_i := -2R + iR/n$ for all $i \in [\![1, 4n]\!]$. In order to eliminate the mass in (x_i) for $i \in [\![1,n]\!]$ and $i \in [\![3n+1, 4n]\!]$, we use the truncation method. The external masses on $[-2R,-R]$ and $[R,2R]$ are discarded and transported uniformly inside $[-R,R]$. The measure $T^{(1)}_{g,\beta}[\mu]$ becomes then a discrete measure with support on $[-R,R]$. Combining with $T^{(2)}_{g,\beta}[\mu]$, we obtain a discrete probability

$$T_{g,\beta}[\mu] = \frac{1}{Z_{g,\beta}[\mu]} \sum_{i=1}^{2n} \left(\sum_{j=1}^{2n} \sum_{k=1}^{2n} \alpha_i a_j a_k \tilde{\beta}(v_j - v_k) g\left(v_i - \frac{v_j + v_k}{2} \right) + a_i a_j a_k (1 - \tilde{\beta}(v_i - v_j)) \right) \delta_{v_i}$$

where $Z_{g,\beta}[\mu]$ is a normalizing factor. We decrease the computational cost by memorazing a g-kernel and a $\tilde{\beta}$-kernel. By setting $g_i := g(-2R + iR/2n)$ for $i \in [\![2, 8n-2]\!]$ and $\tilde{\beta}_i := \tilde{\beta}(-2R + iR/n)$ for $i \in [\![1, 4n-1]\!]$, the non-linear operator $T_{g,\beta}$ becomes

$$T_{g,\beta}[\mu] = \frac{1}{Z_{g,\beta}[\mu]} \sum_{i=1}^{2n} \left(\sum_{j=1}^{2n} \sum_{k=1}^{2n} \alpha_i a_j a_k \tilde{\beta}_{2n+j-k} g_{4n+2i-j-k} + a_i a_j a_k (1 - \tilde{\beta}_{2n+i-j}) \right) \delta_{v_i}. \tag{3.4.1}$$

However, the numerical scheme (3.4.1) may decenter $T_{g,\beta}[\mu]$. For example, if μ has zero mean and

$$\frac{1}{Z_{g,\beta}[\mu]} \sum_{i=1}^{2n} v_i \left(\sum_{j=1}^{2n} \sum_{k=1}^{2n} \alpha_i a_j a_k \tilde{\beta}_{2n+j-k} g_{4n+2i-j-k} + a_i a_j a_k (1 - \tilde{\beta}_{2n+i-j}) \right) = m \neq 0,$$

then it is necessary to recenter the numerical scheme.

Lemma 3.4.1

Let μ be a discrete probability measure on $\mathbb{R}$, $\mu := \sum a_i \delta_{x_i}$. We assume that there exists $R > 0$ such that for all $x \in \mathbb{Z}$, $x_i := -R + iR/n$. If there exists i_0 such that μ has mean x_{i_0}, then the measure μ_0,

$$\mu_0 := \sum_{i=-\infty}^{+\infty} a_{i+i_0-n} \delta_{x_i}$$

has zero mean.

Proof. Since μ has mean x_{i_0}, we have $\sum a_i x_i = x_{i_0}$. And by direct computation we have

$$\begin{aligned}
\sum_{i=-\infty}^{+\infty} a_{i+i_0-n} x_i &= \sum_{i=-\infty}^{+\infty} a_i \left(-R + (i - i_0 + n)\frac{R}{n} \right) \\
&= -R \sum_{i=-\infty}^{+\infty} a_i + \sum_{i=-\infty}^{+\infty} a_i \frac{iR}{n} - \sum_{i=-\infty}^{+\infty} a_i \frac{i_0 R}{n} + R \sum_{i=-\infty}^{+\infty} a_i \\
&= \sum_{i=-\infty}^{+\infty} a_i \left(-R + \frac{iR}{n} \right) - \sum_{i=-\infty}^{+\infty} a_i \left(-R + \frac{i_0 R}{n} \right) \\
&= 0
\end{aligned}$$

since $\sum a_i = 1$. ∎

We are looking for the first index i_0 such that $v_{i_0} \geq m$. We naturally have $i_0 := \lceil n(m+R)/R \rceil$. If $m > 0$, then $i_0 > n$ and the mass must be pushed to the left so that the index i_0 takes the place of the index n. We have therefore pushed the mass to the left by $i_0 - n$ points, which means that the values for i going from 1 to $3n - i_0$ take the values for i going from $i_0 - n + 1$ to $2n$ and the values for i going from $3n - i_0 + 1$ to $2n$ take the value 0 since μ is supported on the interval $[-R, R]$. Hence formula (3.4.1) becomes

$$\frac{1}{Z_{g,\beta}[\mu]} \sum_{i=1}^{3n-i_0} \left(\sum_{j=1}^{2n} \sum_{k=1}^{2n} \alpha_{i_0-n+i} a_j a_k \tilde{\beta}_{2n+j-k} g_{2(n+i_0+i)-j-k} + a_{i_0-n+i} a_j a_k (1 - \tilde{\beta}_{n+i_0+i-j}) \right) \delta_{v_i}.$$

If $m < 0$, then $i_0 \leq n$ and the mass must be pushed to the right so that the index i_0 takes the place of the index n. We have therefore pushed the mass to the right by $n - i_0$ points, which means that the values for i going from 1 to $n - i_0$ take the value 0 since μ is supported on the interval $[-R, R]$ and the values for i going from $n - i_0 + 1$ to $2n$ take the values for i going from 1 to $n + i_0$. Hence formula (3.4.1) becomes

$$\frac{1}{Z_{g,\beta}[\mu]} \sum_{i=n-i_0+1}^{2n} \left(\sum_{j=1}^{2n} \sum_{k=1}^{2n} \alpha_{i_0-n+i} a_j a_k \tilde{\beta}_{2n+j-k} g_{2(n+i_0+i)-j-k} + a_{i_0-n+i} a_j a_k (1 - \tilde{\beta}_{n+i_0+i-j}) \right) \delta_{v_i}.$$

3.4.2 Numerical simulations

This subsection is devoted to the numerical resolution of (3.1.3) in dimension $d = 1$ with g a density and $\beta(v, v') := \tilde{\beta}(v - v')$ where $\tilde{\beta} : \mathbb{R} \longrightarrow [0, 1]$ is a continuous even function. We will present three test cases for different functions $\tilde{\beta}$. For each test case, the solution f is depicted for different values of t and compared with the solution of the fixed point equation $T_{g,\beta}[f] = f$ obtained by iterating $T^n_{g,\beta}$. And we will show numerically the convergence of the solution towards the state of equilibrium for the strong

norm L^1. To represent the solution of (3.1.3) numerically, we use as in the previous chapter an Euler scheme in time for $\Delta t = 0.015$ followed by a Simpson rule on the interval $[-10, 10]$ with a uniform step $\Delta x = 0.1$. To represent numerically the solution of the fixed point of the $T_{g,\beta}$ operator, we will iterate the formula (3.4.1). We will use a quadrature method to numerically represent the strong L^1 norm. We will therefore represent numerically the solutions of the equation

$$\begin{cases} \dfrac{\partial f}{\partial t} = \displaystyle\iint_{\mathbb{R}\times\mathbb{R}} g\left(v - \frac{v' + v'_\star}{2}\right) \tilde{\beta}(v' - v'_\star) f(t, dv') f(t, dv'_\star) - f(t, \cdot) \int_{\mathbb{R}} \tilde{\beta}(v - v') f(t, dv') \\ \\ f(0, \cdot) = f_0. \end{cases} \tag{3.4.2}$$

For each of the five cases, the density g is a centered Gaussian of variance $\sigma_g^2 = 1$,

$$g(v) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) \tag{3.4.3}$$

and the initial condition f_0 is the normalized sum of three Gaussians defined in the previous chapter by (2.5.4),

$$f_0(v) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(v - m_i)^2}{2\sigma_i^2}\right), \tag{3.4.4}$$

with $n = 3$, $m_1 = 3$, $m_2 = m_3 = -3/2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 2$ and $\sigma_3^2 = 4$. We consider for each of the five cases the function $\tilde{\beta}$ defined as follows

- For the first case, $\tilde{\beta}$ is a non-normalized Gaussian given by

$$\tilde{\beta}(v) = \exp\left(-\frac{v^2}{2b^2}\right) \tag{3.4.5}$$

  where $b^2 = 8$, $b^2 = 4$ and $b^2 = 1$.

- For the second case, we take for $\tilde{\beta}$ the function

$$\tilde{\beta}(v) = 1 - \exp\left(-\frac{v^2}{2b^2}\right) \tag{3.4.6}$$

  where $b^2 = 8$.

- And for the last case, the function $\tilde{\beta}$ is equal to

$$\tilde{\beta}(v) = \frac{C}{\sqrt{1 + |v|^2}} + (1 - C). \tag{3.4.7}$$

  where $C = 1$, $C = 1/2$ and $C = 0$.

Test case 1

We will start with the Gaussian case. The function $\tilde{\beta}$ is a non-normalized Gaussian defined by (3.4.5). Since g is a centered normalized Gaussian with variance $\sigma_g^2 = 1$, then we will distinguish three cases: The case $b^2 > 4$, the case $b^2 = 4$ and the case $b^2 < 4$. For the case $b^2 > 4$, we take $b^2 = 8$. In this case, Proposition 3.2.2 gives an explicit formula for the equilibrium state. In this case, the equilibrium state f_m^∞ is a centered Gaussian such that its variance σ_f^2 satisfies (3.3.3). In other words $\sigma_f^2 = 4$ and hence f_m^∞ writes

$$f_m^\infty(v) = \frac{1}{\sqrt{8\pi}} \exp\left(-\frac{v^2}{8}\right). \tag{3.4.8}$$

For the case $b^2 < 4$, we take $b^2 = 1$. When $b^2 = 4$ and $b^2 = 1$, the condition $b^2 > 4$ is not satisfied, then according to the proposition 3.2.2, there is no equilibrium state which is Gaussian for these two cases and we do not have an explicit formula for this. Then f_m^∞ is approached by $T_{g,\beta}^n[f_0]$ with $n = 400$ and with f_0 the initial condition (3.4.4) corresponding to a converged result.

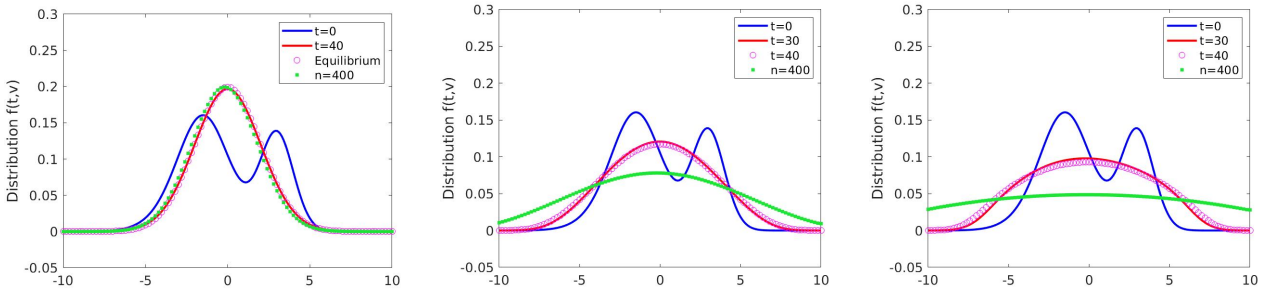


Figure 3.1: Distribution function solution of (3.4.2) with initial condition f_0 given by (3.4.4) with g defined by (3.4.3) and $\tilde{\beta}$ by (3.4.5). On the left, solutions for $b^2 = 8$ at times $t = 0$, $t = 40$ with the equilibrium state given by (3.4.8) compared to the fixed point of $T_{g,\beta}$ obtained by calculating $T_{g,\beta}^n[f_0]$ with $n = 400$ and f_0 the initial condition (3.4.4). On the middle, solutions for $b^2 = 4$ at times $t = 0$, $t = 30$ and $t = 40$ compared to the fixed point of $T_{g,\beta}$ obtained by computing $T_{g,\beta}^n[f_0]$ with $n = 400$ and f_0 the initial condition (3.4.4). On the right, solutions for $b^2 = 1$ at times $t = 0$, $t = 30$ and $t = 40$ compared to the fixed point of $T_{g,\beta}$ obtained by computing $T_{g,\beta}^n[f_0]$ with $n = 400$ and f_0 the initial condition (3.4.4).

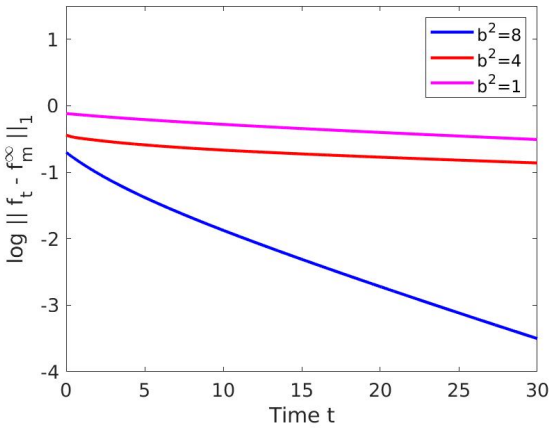


Figure 3.2: Function $t \longmapsto \log \|f(t,\cdot) - f_m^\infty\|_{L^1}$ where f is the solution of (3.4.2) with initial condition f_0 given by (3.4.4) with g defined by (3.4.3) and $\tilde{\beta}$ by (3.4.5). For $b^2 = 8$, f_m^∞ is given by (3.4.8) while for $b^2 = 4$ and $b^2 = 1$, f_m^∞ is replaced by $T_{g,\beta}^n[f_0]$ with $n = 400$ and with f_0 the initial condition (3.4.4).

We see in the three cases that the solution of the time equation (3.4.2) seems to converge towards the fixed point of $T_{g,\beta}$. The convergence seems to be much faster when we are in the case $b^2 > 4$.

Test case 2

This test is devoted to the case $1 - \beta$. Here, the function $\tilde{\beta}$ is defined by (3.4.6). In other words $\tilde{\beta}$ is of the form $1 - \tilde{\beta}_1$ where $\tilde{\beta}_1$ is the function of the first test which is defined by (3.4.5) with $b^2 = 8$. We have no explicit formula for the equilibrium state, we only know from the identity (3.2.9) that it is the solution of the equation $T_{g,\tilde{\beta}_1}[\mu] = T_g[\mu]$. Then f_m^∞ is approached as in the previous case by $T_{g,\beta}^n[f_0]$ with $n = 400$ and with f_0 the initial condition (3.4.4).

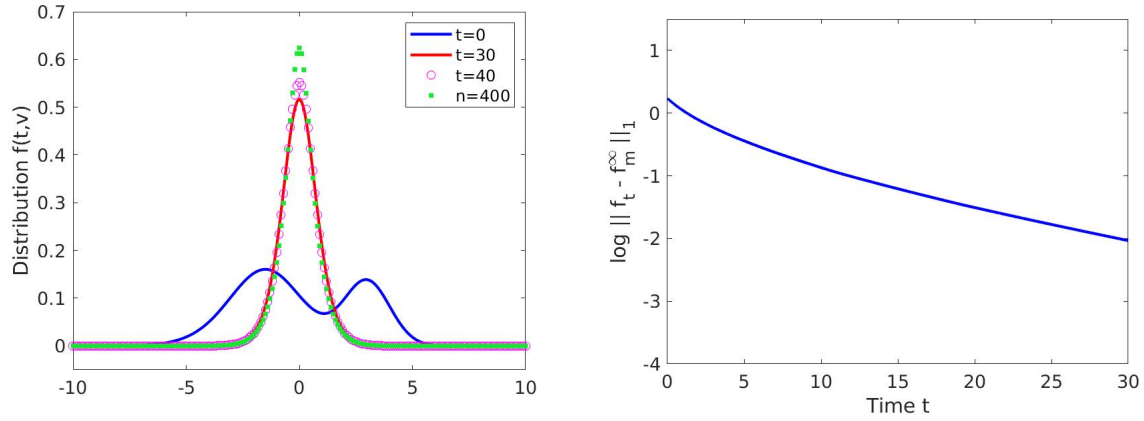


Figure 3.3: Distribution function solution of (3.4.2) with initial condition f_0 given by (3.4.4) with g defined by (3.4.3) and $\tilde{\beta}$ by (3.4.6). On the left, solutions at times $t = 0$, $t = 30$ and $t = 40$ compared to the fixed point of $T_{g,\beta}$ obtained by computing $T^n_{g,\beta}[f_0]$ with $n = 400$ and f_0 the initial condition (3.4.4). On the right, function $t \longmapsto \log \|f(t,\cdot) - f^\infty_m\|_{L^1}$ where f is the solution of (3.4.2) with initial condition f_0 given by (3.4.4) with g defined by (3.4.3) and $\tilde{\beta}$ by (3.4.6), and where f^∞_m is replaced by $T^n_{g,\beta}[f_0]$ with $n = 400$ and with f_0 the initial condition (3.4.4).

We see also in this case that the solution of the time equation (3.4.2) seems to converge towards the fixed point of $T_{g,\beta}$. We took as function $\tilde{\beta}$ a perturbation of 1. The larger b^2 is, the weaker the perturbation is, the faster the convergence is.

Test case 3

For this last test case, we take a function $\tilde{\beta}$ which satisfies the Lipschitz condition (3.2.18). Thus, we take $\tilde{\beta}$ is defined by (3.4.7). The case $C = 0$ is the case β constant equal to 1. We have an explicit formula for the equilibrium state in this case (see Figure 2.1). For the other two cases $C = 1$ and $C = 1/2$, no explicit formula are again available for the equilibrium state. Hence f^∞_m is again substituted by some $T^n_{g,\beta}[f_0]$ for a large n ($n = 400$ in the simulations) and with an initial condition f_0 (3.4.4) as in test case 2.

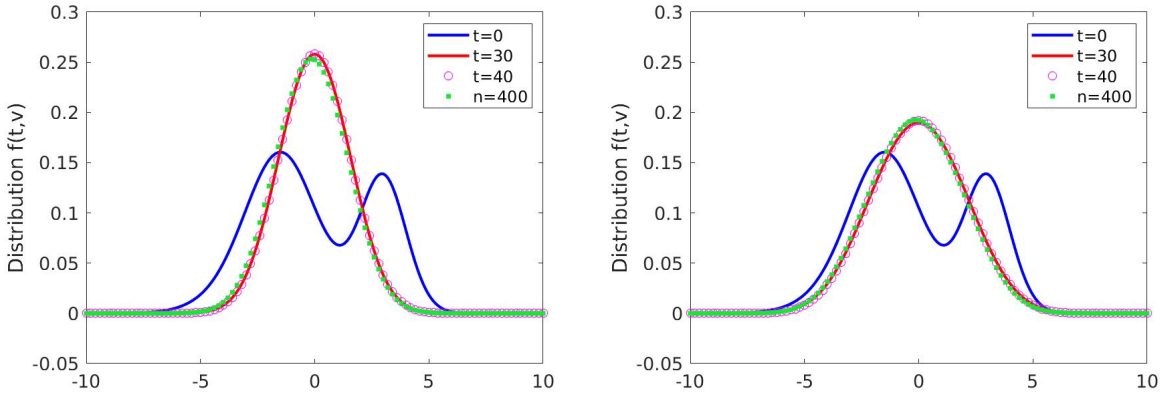


Figure 3.4: Distribution function solution of (3.4.2) with initial condition f_0 given by (3.4.4) with g defined by (3.4.3) and $\tilde{\beta}$ by (3.4.7). On the left, solutions for $C = 1/2$ at times $t = 0$, $t = 30$ and $t = 40$ compared to the fixed point of $T_{g,\beta}$ obtained by computing $T^n_{g,\beta}[f_0]$ with $n = 400$ and f_0 the initial condition (3.4.4). On the right, solutions for $C = 1$ at times $t = 0$, $t = 30$ and $t = 40$ compared to the fixed point of $T_{g,\beta}$ obtained by computing $T^n_{g,\beta}[f_0]$ with $n = 400$ and f_0 the initial condition (3.4.4).

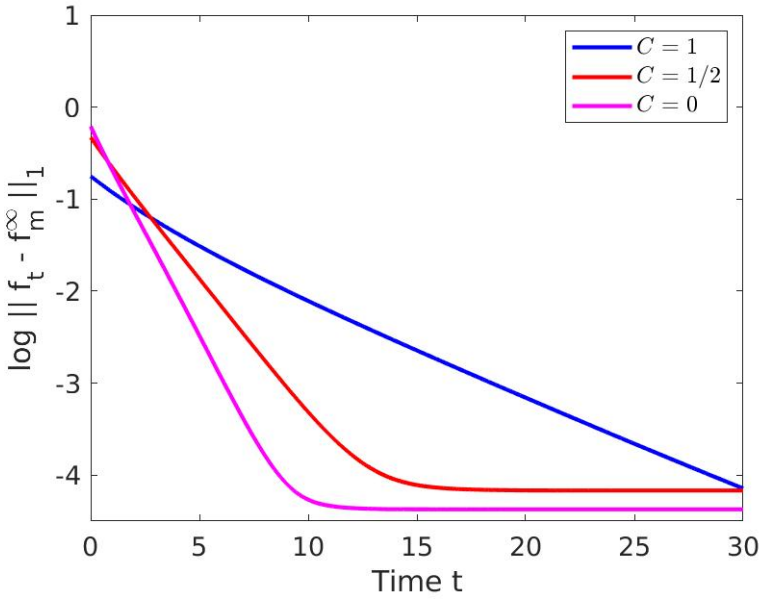


Figure 3.5: Function $t \longmapsto \log \|f(t,\cdot) - f_m^\infty\|_{L^1}$ where f is the solution of (3.4.2) with initial condition f_0 given by (3.4.4) and g defined by (3.4.3). The function $\tilde{\beta}$ is defined by (3.4.7) with $C = 1$ (in blue), $C = 1/2$ (in red) and $C = 0$ (in purple). For $C = 1$ and $C = 1/2$, f_m^∞ is replaced by $T_{g,\beta}^n[f_0]$ with $n = 400$ and with f_0 the initial condition (3.4.4) while for $C = 0$, f_m^∞ is given by (2.5.7).

Figure 3.5 shows that the solution of the time equation (3.4.2) converges exponentially towards the fixed point of $T_{g,\beta}$. The convergence seems to be faster when C is close to 0, in other words when $\tilde{\beta}$ is close to 1.

3.5 Conclusion

We have shown in this chapter the existence of an equilibrium state $f \in \mathcal{P}_2^m(\mathbb{R}^d)$. Furthermore the numerical simulations show the existence of a solution at all times that converges towards a unique equilibrium state. We hope in the future to show rigorously the uniqueness of the equilibrium state, then the existence of a solution at all times and the convergence of this one towards the unique equilibrium state.

Conclusion and perspectives

In this manuscript, we have studied a Boltzmann-type kinetic model of velocity alignment (1.4.1) introduced by Bertin, Droz, Gregoire [9], [10]. This model is new and interesting because it is located at the interface between collective dynamics and kinetic theory. The transport equation has no forcing or diffusion term in velocity, the change of velocity is computed as in Boltzmann framework. In this model, the collisions are not micro-reversible and it is not obvious to find an entropy functional. In the Boltzmann equations, micro-reversibility is a crucial element for obtaining the H Theorem. Consequently, the classical tools for dealing with the problems of returns to equilibrium, such as for example the Csiszàr-Kullback-Pinsker inequality [37], are inoperative . In this model, we have instead a phenomenon of contraction in the collision process which does not take place for the Boltzmann operator but drives the density towards an equilibrium state.

We have considered the spatially homogeneous version of the model and we have assumed that the space of velocities is Euclidean. We started in the chapter 2 by treating the Maxwellian case where the collision rate β is constant equal to 1, then we treated the case for any β in chapter 3.

In chapter 2, the collision rate β is therefore constant equal to 1. This version of the model has already been discussed in Ref. [35] where it is considered that the direction taken by the two individuals is exactly the mean $(v' + v'_\star)/2$ of the pre-collisionnal directions, and where the space of velocities may be a manifold of any dimension. Our results are more general, however they assume that the space of velocities is Euclidean. We have shown in this chapter that our model (2.1.3) has a unique mild solution $f \in \mathcal{C}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^d))$ for any initial condition $f_0 \in \mathcal{P}_2^m(\mathbb{R}^d)$. The existence has been proved by using a fixed point type argument for the Wasserstein metric W_2. An argument of the same style had already been used in Ref. [72] for a model similar to ours. Furthermore, there is a unique equilibrium state $f \in \mathcal{P}_2^m(\mathbb{R}^d)$. This mild solution f converges exponentially to the equilibrium state for the Wasserstein metric W_2 and for the strong L^1 norm. The exponential convergence for the strong L^1 norm has been proved by bounding the strong L^1 norm by the Fourier-Toscani based distance d_2 thanks to estimates on the Sobolev norm.

However, a model where for example the velocity is constrained to be of norm 1 as in Ref. [35] is out of reached by our methods. A first perspective would be to extend our results to the case where the space of velocities is a manifold of any dimension. A second perspective would be to add the space variable to this model. In this case we could give a hydrodynamic description of the model. We have formally shown that this hydrodynamic description is similar to the isothermal compressible Euler system but with a non-diagonal stress tensor. A third perspective would be to study the hydrodynamic limit.

In chapter 3, the collision rate β is arbitrary. This adds a considerable difficulty because in the Maxwellian case, the fact that β is constant equal to 1 made it possible to write the gain term Q_+ in the form of a double convolution. If β is not constant, then Q_+ has a more complex expression and therefore all the tools used in chapter 2 are inoperative. We wrote our model as a linear ODE in infinite dimension, which gives rise to an operator $T_{g,\beta}$ which is the double convolution of the chapter 2 when $\beta = 1$. We have shown that this operator admits a fixed point, which is equivalent to say that we have shown the existence of an equilibrium state.

The numerical simulations show the existence of a solution at all times that converges towards

a unique equilibrium state. A first perspective would be to show rigorously the uniqueness of the equilibrium state, then the existence of a solution at all times and the convergence of this one towards the unique equilibrium state. To show the existence of the solution at all times, we tried to apply a result present in Ref. [1]. This result is an extension of Peano's Theorem in infinite dimension with the assumption that the vector field defining the differential equation is α-Lipschitz where α is Kuratowski index of non-compactness.

We would also like to show the existence of an equilibrium state using the Schauder fixed point Theorem in the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$. The problem is the non-compactness of the closed balls of $\mathcal{P}_2(\mathbb{R}^d)$ for the topology induced by W_2. We tried to compute the Kuratowski index of non-compactness of the set $T^n_{g,\beta}\{\mu \in \mathcal{P}^m_2(\mathbb{R}^d),\, M_2(\mu) \leq \Sigma\}$: $\alpha(T^n_{g,\beta}\{\mu \in \mathcal{P}^m_2(\mathbb{R}^d),\, M_2(\mu) \leq \Sigma\})$. A preliminary computation seems to show that $\alpha(T^n_{g,\beta}\{\mu \in \mathcal{P}^m_2(\mathbb{R}^d),\, M_2(\mu) \leq \Sigma\})$ tends to 0 when n tends to infinity which would mean that the set $T^n_{g,\beta}\{\mu \in \mathcal{P}^m_2(\mathbb{R}^d),\, M_2(\mu) \leq \Sigma\}$ will become more and more compact as n becomes large.

A second perspective would be, as for the Maxwellian case, to consider the case where the space of velocities is no longer Euclidean but would be a manifold of any dimension. In order to introduce for example a velocity confinement. Adding the space variable to the model is also considered, as well as the hydrodynamic description of the model with the study of the hydrodynamic limit.

Appendix

Appendix A

On the Wasserstein metric

A.1 Introduction to Optimal Transport

In 1781, the French mathematician Gaspard Monge published his memoir entitled *Mémoire sur la théorie des déblais et des remblais*. Memoir in which he introduces the following reflection: How to fill in the most economical way possible a hole with a pile of sand ? In this problem, it is very useful to consider the pile of sand as being a Borelian $\mathcal{X}$ of $\mathbb{R}^3$ and the hole as being a Borelian $\mathcal{Y}$ of $\mathbb{R}^3$ of the same volume: $\text{Vol}(\mathcal{X}) = \text{Vol}(\mathcal{Y}) := V$. We associate with them the probability measures μ and ν defined by $\mu := \mathbb{1}_{\mathcal{X}}\lambda/V$ and $\nu := \mathbb{1}_{\mathcal{Y}}\lambda/V$. In physical terms, we can see the measure μ as being the density of distribution of the grains of sand at the beginning (in infinitely large number) and the measure ν as being the density of distribution of the grains of sand in the hole once it is filled. Monge's problem consists in finding a Borel map $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ which transports μ on ν with a minimal cost.

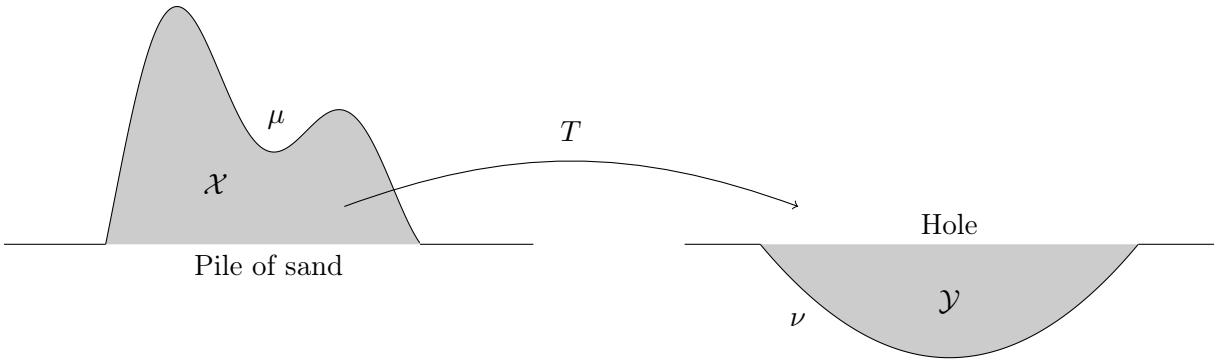


Figure A.1: Illustration of Monge's problem

Definition A.1.1

Let $T : \mathcal{X} \longrightarrow \mathcal{Y}$ be a Borel map and let μ be a measure on $\mathcal{X}$. We define the **image measure** (or push-forward) of μ by T as being the measure ν on $\mathcal{Y}$ such that for any Borelian $\mathcal{A} \subset \mathcal{Y}$, we have $\nu(\mathcal{A}) = \mu(T^{-1}(\mathcal{A}))$. We denote by $\nu := T\#\mu$ and in terms of integrals, we have for any test function φ

$$\int_{\mathcal{Y}} \varphi(y)d\nu(y) = \int_{\mathcal{X}} \varphi(T(x))d\mu(x).$$

To explain this term from a physical point of view, writing $\nu = T\#\mu$ means that if the grains of sand are initially distributed according to the configuration μ and that we move each grain of sand of the location x to the location $T(x)$, then these latters will be distributed, after transport, according to the configuration ν.

We give ourselves a cost function $c : (x, y) \in \mathbb{R}^3 \times \mathbb{R}^3 \longmapsto c(x, y) \in [0, +\infty[$ which gives the transport cost of a unit of volume from the point x to the point y. The cost of the infinitesimal volume transported from x to $y = T(x)$ is $c(x, T(x))dx$. So the total cost is

$$\int_{\mathcal{X}} c(x, T(x))dx = V \int_{\mathbb{R}^3} c(x, T(x))d\mu(x).$$

Since V is fixed, Monge's problem consists in minimizing the quantity $\int c(x, T(x))d\mu(x)$ on the set $\{T : \mathbb{R}^3 \to \mathbb{R}^3, T\#\mu = \nu\}$.

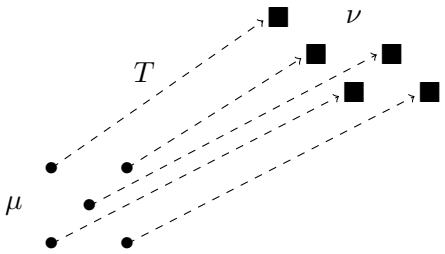


Figure A.2: Example in the discrete case in $\mathbb{R}^2$ of the Monge problem. μ is defined for any Borelian $\mathcal{A}$ by $\mu(\mathcal{A}) = \{\text{Number of points in } \mathcal{A}\}/5$ and T represents the map which associates each point of μ to a square of ν. The total transport cost is the sum of the transport costs for each of the five points of μ.

However, the condition $T\#\mu = \nu$ is actually very restrictive. Indeed, in the particular case where μ and ν are measures with for respective density f and g, then we have by the change of variable formula that $T\#\mu = \nu$ if and only if ν has for density the function $x \longmapsto f(T^{-1}(x))|\det(J_{T^{-1}}(x))|$. And by the classical equality $J_{T^{-1}}(x) = (J_T(T^{-1}(x)))^{-1}$, then $T\#\mu = \nu$ if and only if T is solution of the following Jacobian equation

$$f(x) = g(T(x))|\det(J_T(x))|. \tag{A.1.1}$$

So minimize $\int c(x, T(x))d\mu(x)$ over the set $\{T : \mathbb{R}^3 \to \mathbb{R}^3, T\#\mu = \nu\}$ consists in minimizing over the set of solutions of the equation (A.1.1). Worse, we can even end up in the case where the set $\{T : \mathbb{R}^3 \to \mathbb{R}^3, T\#\mu = \nu\}$ is empty. For example if we take $\mu = \delta_{x_1}$ and $\nu = (\delta_{y_1} + \delta_{y_2})/2$ with $y_1 \neq y_2$, then we have on one side $\nu(\{y_1\}) = 1/2$ and on the other hand $\mu(T^{-1}(y_1)) = 1$ if $T^{-1}(y_1) = x_1$ and 0 otherwise. In both cases, $\nu(\{y_1\}) \neq \mu(T^{-1}(y_1))$ so there is no map T such that $T\#\mu = \nu$. This problem has therefore been left abandoned during a very long time.

It was not until the 1940s that this problem resurfaced, thanks to the Russian mathematician Leonid Kantorovich. This one have seen the Monge's problem under another point of view, which made it possible to circumvent this constraining problem of the existence of map T such as $T\#\mu = \nu$. It was this new approach that won him the Nobel Prize in Economics in 1975. We can see the Monge's problem as follows. Rather than pouring the totality of the infinitesimal mass in x at the point y, it can be partially pour it in y but also at other locations. In the case $\mu = \delta_{x_1}$ and $\nu = (\delta_{y_1} + \delta_{y_2})/2$, one can for example pour half of the mass in x_1 into the point y_1 and the other half at point y_2. It is nevertheless necessary that the totality of the infinitesimal mass of x be poured (but not necessarily at the same point y), and that one pours in y the totality of its infinitesimal mass (but not necessarily taken on a single point x). From a mathematical point of view, if $d\pi(x, y)$ represents the quantity of infinitesimal mass at point x poured at point y (divided by V so that we have a probability measure), then π must have as marginals μ and ν. Thus, the cost of the infinitesimal volume transported from x to y is $Vc(x, y)d\pi(x, y)$ and therefore the total cost is

$$V \iint_{\mathbb{R}^3 \times \mathbb{R}^3} c(x, y)d\pi(x, y).$$

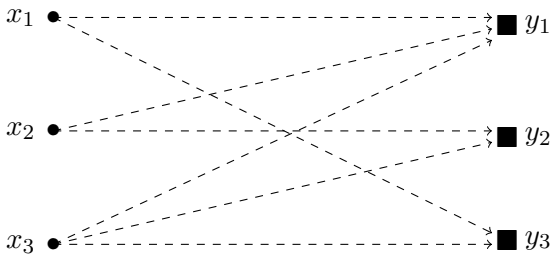
Definition A.1.2

Let μ, $\nu \in \mathcal{P}(\mathbb{R}^d)$. We say that $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is a **coupling** of (μ, ν) if π admits μ and ν as marginals. In other words if for any test functions φ and ψ, we have

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi(x) + \psi(y)) d\pi(x, y) = \int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y).$$

We denote $\Pi(\mu, \nu)$ the set of couplings of (μ, ν).

We can interpret the marginals in several ways. In terms of image measure, π is a coupling of (μ, ν) if $\text{proj}_1 \# \pi = \mu$ and $\text{proj}_2 \# \pi = \nu$. And in terms of Borelians, π is a coupling of (μ, ν) if for any Borelians $\mathcal{A}$, $\mathcal{B} \subset \mathbb{R}^d$, we have $\pi(\mathcal{A} \times \mathbb{R}^d) = \mu(\mathcal{A})$ and $\pi(\mathbb{R}^d \times \mathcal{B}) = \nu(\mathcal{B})$. Monge's problem from Kantorovich's point of view consists in minimizing the quantity $\iint c(x, y) d\pi(x, y)$ on the set $\Pi(\mu, \nu)$. And the constraint of existence of map T such that $T \# \mu = \nu$ is lifted since $\Pi(\mu, \nu)$ is always non-empty. Indeed, for all μ, $\nu \in \mathcal{P}(\mathbb{R}^d)$, $\mu \otimes \nu \in \Pi(\mu, \nu)$.



	y_1	y_2	y_3
x_1	1/10	0	9/10
x_2	5/10	5/10	0
x_3	4/10	5/10	1/10

Figure A.3: Illustration of Monge's problem from Kantorovich's point of view in the discrete case in $\mathbb{R}^2$. The array on the right gives the quantities of mass in x_i poured at point y_j (i.e $3d\pi(x_i, y_j)$). The coupling π is defined for any Borelian $\mathcal{A} \times \mathcal{B}$ by $\pi(\mathcal{A} \times \mathcal{B}) = \sum_{(x_i, y_j) \in \mathcal{A} \times \mathcal{B}} d\pi(x_i, y_j)$. The total transport cost is the sum of the transport costs for each pair (x_i, y_j) multiplied by the quantity of mass transported from x_i to y_j.

Another parameter that influences the cost of filling the hole is the mean distance between the pile of sand and this one. If we consider for a cost function c the optimal transport cost

$$\inf_{\pi \in \Pi(\mu, \nu)} \iint_{\mathbb{R}^3 \times \mathbb{R}^3} c(x, y) d\pi(x, y)$$

between the two measures μ and ν, we get a way to "measuring" the distance between μ and ν. However, the map to which two measures μ and ν associates the optimal transport cost between μ and ν does not always verify the axioms of a metric. It actually depends on the cost function and it is interesting to consider the case where this is the Euclidean norm.

Definition A.1.3

For μ, $\nu \in \mathcal{P}_p(\mathbb{R}^d)$, $p \geq 1$, we define the **Wasserstein metric** of order p between μ and ν by

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p d\pi(x, y) \right)^{1/p}. \tag{A.1.2}$$

It was the Russian mathematician Roland Dobrushin who named this metric *Wasserstein metric* after having discovered it in the works of the Russian mathematician Léonid Vasershtein [85]. However,

this distance had already been used by Kantorovich before [56]. The quantity $W_p(\mu,\nu)$ is always well-defined. Indeed, if π is a coupling between μ and ν, then by the classic convexity inequality $|x-y|^p \leq 2^{p-1}(|x|^p+|y|^p)$ we have

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\pi(x,y) \leq 2^{p-1}\left(\int_{\mathbb{R}^d}|x|^p d\mu(x)+\int_{\mathbb{R}^d}|y|^p d\nu(y)\right)$$

which is finite since μ and ν are in $\mathcal{P}_p(\mathbb{R}^d)$. If $p=1$, then $W_1(\mu,\nu)$ is the solution of Monge's problem when the transport cost between two points is the Euclidean distance between these two points. Note that if μ and ν are probability measures with bounded support, then we can define the Wasserstein metric W_p between μ and ν for $p=+\infty$ as the limit of $W_p(\mu,\nu)$ when $p\longrightarrow+\infty$. And we have

$$W_\infty(\mu,\nu) := \lim_{p\to+\infty} W_p(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \|c\|_{L^\infty(\mathbb{R}^d\times\mathbb{R}^d,\pi)}$$

where $c(x,y):=|x-y|$. This particular case $p=+\infty$ is given special attention in Ref. [77] but will not be discussed in this text. We will therefore assume that $p\in[1,+\infty[$. The question we can ask ourselves is: Is there a coupling π of (μ,ν) which attains the infimum in (A.1.2)?

Definition A.1.4

Let $\mu,\nu\in\mathcal{P}_p(\mathbb{R}^d)$ and let $\pi\in\Pi(\mu,\nu)$. We say that π is an **optimal coupling** for W_p if

$$W_p(\mu,\nu) = \left(\iint_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^p d\pi(x,y)\right)^{1/p}.$$

In other words if π attains the infimum in (A.1.2).

Let E be a metric space. We recall that a function $f:E\longrightarrow\mathbb{R}\cup\{+\infty\}$ is said to be lower semi-continuous at the point $x\in E$ if for any sequence $(x_n)_n$ converging to x, we have $f(x)\leq\liminf f(x_n)$. It is said to be lower semicontinuous on E if it is lower semicontinuous at any point x of E. In particular, a lower semicontinuous function on a compact set is bounded below and attains its lower bound. Thus, the existence of an optimal coupling for W_p is guaranteed by the fact that the map $\pi\in\Pi(\mu,\nu)\longmapsto\iint|x-y|^p d\pi(x,y)$ is lower semicontinuous on $\Pi(\mu,\nu)$ which is compact. More generally, for a cost function c, the map $\pi\in\Pi(\mu,\nu)\longmapsto\iint c(x,y)d\pi(x,y)$ is lower semicontinuous if the cost function c is lower semicontinuous (Theorem 4.1 in Ref. [88]). In the particular case $p=2$, if μ is absolutely continuous, then there exists a unique optimal coupling π_* for W_2. And it is given by $\pi_*=(\mathrm{id},T)\#\mu$ where $T:\mathbb{R}^d\longrightarrow\mathbb{R}^d$ is a map such that $T\#\mu=\nu$ [18].

A.2 Proof that W_p is a metric

Symmetry

Let $\mu,\nu\in\mathcal{P}_p(\mathbb{R}^d)$ and let $\pi\in\Pi(\mu,\nu)$ be an optimal coupling. We define the map f by $f:(x,y)\longmapsto(y,x)$. It is clear that the coupling $\pi_*:=f\#\pi$ is a coupling of (ν,μ) and we have

$$W_p(\nu,\mu)^p \leq \iint_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^p d\pi_*(x,y) = \iint_{\mathbb{R}^d\times\mathbb{R}^d}|y-x|^p d\pi(x,y) = W_p(\mu,\nu)^p.$$

Doing the same thing but taking $\pi\in\Pi(\nu,\mu)$ an optimal coupling and $\pi_*=f\#\pi$, we obtain $W_p(\mu,\nu)\leq W_p(\nu,\mu)$ and therefore $W_p(\mu,\nu)=W_p(\nu,\mu)$.

Identity of indiscernible

For $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, we take $\pi := (\mathrm{id}, \mathrm{id})\#\mu$. It is clear that $\pi \in \Pi(\mu, \mu)$ and therefore $\mu = \nu$ implies that

$$W_p(\mu,\nu)^p \leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\pi(x,y) = \int_{\mathbb{R}^d} |x-x|^p d\mu(x) = 0.$$

Conversely, if $W_p(\mu,\nu) = 0$, then there exists a coupling $\pi \in \Pi(\mu,\nu)$ such that $\iint |x-y|^p d\pi(x,y) = 0$. In other words, there exists a coupling π such that $x = y$ π-almost everywhere. If we denote $\Delta := \{(x,x),\, x \in \mathbb{R}^d\}$ the diagonal of $\mathbb{R}^d \times \mathbb{R}^d$, then we have $\pi(\Delta^c) = \pi(\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d,\, x \neq y\}) = 0$. Let $\varphi \in \mathcal{C}_b(\mathbb{R}^d)$, we have

$$\int_{\mathbb{R}^d} \varphi(x) d\mu(x) = \iint_{\Delta} \varphi(x) d\pi(x,y) + \iint_{\Delta^c} \varphi(x) d\pi(x,y).$$

The second term on the right-hand side is bounded by $\|\varphi\|_\infty \pi(\Delta^c) = 0$. Then

$$\int_{\mathbb{R}^d} \varphi(x) d\mu(x) = \iint_{\Delta} \varphi(x) d\pi(x,y) = \iint_{\Delta} \varphi(y) d\pi(x,y) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(y) d\pi(x,y) = \int_{\mathbb{R}^d} \varphi(y) d\nu(y).$$

The equality is valid for any $\varphi \in \mathcal{C}_b(\mathbb{R}^d)$ therefore $\mu = \nu$.

Triangle inequality

To prove the triangle inequality, we need to the following Lemma. We recall that a Polish space is a complete, separable metric space.

Lemma A.2.1 (Gluing Lemma)

Let X, Y, Z be three Polish spaces and let $\pi_1 \in \mathcal{P}(X \times Y)$, $\pi_2 \in \mathcal{P}(Y \times Z)$ and $\mu \in \mathcal{P}(Y)$ such that $\mathrm{proj}_2\#\pi_1 = \mathrm{proj}_1\#\pi_2 = \mu$. Then there exists $\pi \in \mathcal{P}(X \times Y \times Z)$ such that for any test functions φ and ψ,

$$\begin{aligned}\iiint_{X\times Y\times Z} (\varphi(x,y) + \psi(y,z)) d\pi(x,y,z) = &\iint_{X\times Y} \varphi(x,y) d\pi_1(x,y) \\ &+ \iint_{Y\times Z} \psi(y,z) d\pi_2(y,z).\end{aligned} \tag{A.2.1}$$

In other words, if π_1 is a coupling of (μ_1, μ) and if π_2 is a coupling of (μ, μ_2), then we can construct a coupling π with three variables which satisfies (A.2.1) by "gluing" π_1 and π_2 along their common marginal μ. Hence the name "Gluing Lemma". The proof use the following result [36].

Definition A.2.1

Let $\mathcal{X}$, $\mathcal{Y}$ be two Polish spaces. We say that the map $P : x \in \mathcal{X} \longmapsto P(x, \cdot) \in \mathcal{P}(\mathcal{Y})$ is a **transition kernel** if for any Borelian $\mathcal{B} \subset \mathcal{Y}$, the map $x \in \mathcal{X} \longmapsto P(x, \mathcal{B}) \in [0,1]$ is measurable.

Theorem A.2.1 (Disintegration Theorem)

Let $\mathcal{X}$, $\mathcal{Y}$ be two Polish spaces, $\mu \in \mathcal{P}(\mathcal{X})$, let $f : \mathcal{X} \longrightarrow \mathcal{Y}$ be a Borel map and let $\nu := f\#\mu \in \mathcal{P}(\mathcal{Y})$. Then there exists a transition kernel P such that for any test function φ, we have

$$\int_{\mathcal{X}} \varphi(x) d\mu(x) = \int_{\mathcal{Y}} \left(\int_{f^{-1}(y)} \varphi(x) P(y, dx) \right) d\nu(y)$$

and P is unique ν-almost everywhere.

In the particular case where $\mathcal{X} = X \times Y$ and $\mathcal{Y} = X$, then $\mu := \pi \in \mathcal{P}(X \times Y)$, $f = \mathrm{proj}_1$ and $\nu := \mu = \mathrm{proj}_1\#\pi \in \mathcal{P}(X)$. We can canonically identify each slice $\mathrm{proj}_1^{-1}(x)$ by Y and we obtain for any test function φ

$$\iint_{X\times Y} \varphi(x, y) d\pi(x, y) = \int_X \left(\int_Y \varphi(x, y) P(x, dy) \right) d\mu(x).$$

We denote $\pi(dxdy) = \mu(dx)P(x, dy)$.

Proof of Lemma A.2.1. We use the disintegration Theorem (Theorem A.2.1) on π_1 and π_2. Then there exists two transition kernels P_1 and P_2 such that $\pi_1(dxdy) = \mu(dy)P_1(y, dx)$ and $\pi_2(dydz) = \mu(dy)P_2(y, dz)$. We set $\pi := \mu \otimes P_1(y, \cdot) \otimes P_2(y, \cdot)$. In other words for any test function φ we have

$$\iiint_{X\times Y\times Z} \varphi(x, y, z) d\pi(x, y, z) = \int_Y \left(\int_Z \left(\int_X \varphi(x, y, z) P_1(y, dx) \right) P_2(y, dz) \right) d\mu(y).$$

It is clear that $\pi \in \mathcal{P}(X \times Y \times Z)$ since $P_1(y, \cdot) \in \mathcal{P}(X)$ and $P_2(y, \cdot) \in \mathcal{P}(Z)$ for all $y \in Y$. And it is easy to check that π satisfies (A.2.1). Therefore, π is suitable. ■

Let us now show the triangle inequality. Let μ_1, μ_2, $\mu_3 \in \mathcal{P}_p(\mathbb{R}^d)$ and let $\pi_{1,2} \in \Pi(\mu_1, \mu_2)$ and $\pi_{2,3} \in \Pi(\mu_2, \mu_3)$ two optimals couplings. By the Gluing Lemma (Lemma A.2.1), we can construct a coupling $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ with three variables such that $(\mathrm{proj}_1, \mathrm{proj}_2)\#\pi = \pi_{1,2}$ and $(\mathrm{proj}_2, \mathrm{proj}_3)\#\pi = \pi_{2,3}$. We set $\pi_{1,3} := (\mathrm{proj}_1, \mathrm{proj}_3)\#\pi$. It is clear that $\pi_{1,3} \in \Pi(\mu_1, \mu_3)$. So we have

$$\begin{aligned} W_p(\mu_1, \mu_3) &\leq \left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x - z|^p d\pi_{1,3}(x, z) \right)^{1/p} = \left(\iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} |x - z|^p d\pi(x, y, z) \right)^{1/p} \\ &\leq \left(\iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} (|x - y| + |y - z|)^p d\pi(x, y, z) \right)^{1/p}. \end{aligned}$$

By the Minkowski inequality, we have

$$\begin{aligned} W_p(\mu_1, \mu_3) &\leq \left(\iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} |x - y|^p d\pi(x, y, z) \right)^{1/p} + \left(\iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} |y - z|^p d\pi(x, y, z) \right)^{1/p} \\ &= \left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x - y|^p d\pi_{1,2}(x, y) \right)^{1/p} + \left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} |y - z|^p d\pi_{2,3}(y, z) \right)^{1/p} \\ &= W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3). \end{aligned}$$

We can thus equip the space $\mathcal{P}_p(\mathbb{R}^d)$ with the Wasserstein metric W_p. The metric space $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ is sometimes called the Wasserstein space of order p (see Ref. [88]).

A.3 Kantorovich formula

We consider factories being supplied with coal. We would like to minimize the transport cost of coal from mines to factories. We give ourselves a cost function $c : (x, y) \in \mathbb{R}^3 \times \mathbb{R}^3 \longmapsto c(x, y) \in [0, +\infty[$ which gives the transport cost of a unit of volume from mine x to factory y. If $d\pi(x, y)$ represents the quantity of infinitesimal mass of coal transported from mine x to factory y, then the total transport cost is $\iint c(x, y)d\pi(x, y)$. And the marginals of π are μ and ν where $d\mu(x)$ represents the infinitesimal quantity of coal produced in the mine x and $d\nu(y)$ the infinitesimal quantity of coal consumed in the factory y. It is assumed that mines produce as much as factories consume, in other words μ and ν are probability measures (even if it means normalizing). Since the marginals of π are fixed, the problem consists in minimizing the quantity $\iint c(x, y)d\pi(x, y)$ on the set $\Pi(\mu, \nu)$ so we are in the presence of a Monge-Kantorovich problem.

Suppose instead that a shipper offers to charge a price $\varphi(x)$ for loading a unit of volume into the mine x and a price $\psi(y)$ for unloading a unit of volume into the factory y. Knowing that the sum of these two prices will always be lower than the transport cost from mine x to factory y. The problem is therefore transferred to the shipper because the latter must fix its prices such that its profit is maximized, while respecting the constraint $\varphi(x) + \psi(y) \leq c(x, y)$ for that its proposition remains attractive. In mathematical terms, the shipper must maximize the quantity

$$\int_{\mathbb{R}^3} \varphi(x)d\mu(x) + \int_{\mathbb{R}^3} \psi(y)d\nu(y)$$

on the set of all pairs of price functions (φ, ψ) satisfying the inequality $\varphi(x) + \psi(y) \leq c(x, y)$ for all x, $y \in \mathbb{R}^3$.

Theorem A.3.1 (Kantorovich duality)

Let μ, $\nu \in \mathcal{P}(\mathbb{R}^d)$ and let $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow [0, +\infty[$ be a lower semicontinuous cost function. Then we have

$$\inf_{\pi \in \Pi(\mu,\nu)} \iint_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)d\pi(x, y) = \sup_{(\varphi,\psi)\in\Phi_c} \left(\int_{\mathbb{R}^d} \varphi(x)d\mu(x) + \int_{\mathbb{R}^d} \psi(y)d\nu(y) \right) \quad \text{(A.3.1)}$$

where $\Phi_c := \{\varphi,\ \psi \in \mathcal{C}_b(\mathbb{R}^d),\ \forall x, y \in \mathbb{R}^d,\ \varphi(x) + \psi(y) \leq c(x, y)\}$.

The formula (A.3.1) remains true if we extend the class Φ_c to the following class $\Phi_c(\mu, \nu) := \{\varphi \in L^1(\mathbb{R}^d, \mu),\ \psi \in L^1(\mathbb{R}^d, \nu),\ \forall x, y \in \mathbb{R}^d,\ \varphi(x) + \psi(y) \leq c(x, y)\}$. It would be tempting to prove the Theorem A.3.1 by a double inequality because the following inequality

$$\sup_{(\varphi,\psi)\in\Phi_c} \left(\int_{\mathbb{R}^d} \varphi(x)d\mu(x) + \int_{\mathbb{R}^d} \psi(y)d\nu(y) \right) \leq \inf_{\pi \in \Pi(\mu,\nu)} \iint_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)d\pi(x, y). \quad \text{(A.3.2)}$$

is very easy to prove. Indeed, let μ, $\nu \in \mathcal{P}(\mathbb{R}^d)$ and let $\pi_* \in \Pi(\mu, \nu)$ be a coupling such that

$$\inf_{\pi \in \Pi(\mu,\nu)} \iint_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)d\pi(x, y) = \iint_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)d\pi_*(x, y).$$

Such a coupling π_* exists since the map $\pi \in \Pi(\mu, \nu) \longmapsto \iint c(x, y)d\pi(x, y)$ is lower semicontinuous

(because c is lower semicontinuous) on $\Pi(\mu,\nu)$ which is compact. Let $(\varphi,\psi)\in\Phi_c$, we have

$$\begin{aligned}\int_{\mathbb{R}^d}\varphi(x)d\mu(x)+\int_{\mathbb{R}^d}\psi(y)d\nu(y) &= \iint_{\mathbb{R}^d\times\mathbb{R}^d}(\varphi(x)+\psi(y))d\pi_*(x,y)\\ &\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d}c(x,y)d\pi_*(x,y)\\ &= \inf_{\pi\in\Pi(\mu,\nu)}\iint_{\mathbb{R}^d\times\mathbb{R}^d}c(x,y)d\pi(x,y).\end{aligned}$$

And we take the supremum over (φ,ψ), we obtain (A.3.2). However, the reverse inequality is very complicated to prove. We will therefore directly show the equality (A.3.1) using convex analysis tools. The proof of this Theorem A.3.1 is given in Villani [89], where convex analysis plays a fundamental role. It is separated into three steps: first we treat the compact case and when the cost function c is continuous, then we prove the non-compact case but with c bounded and uniformly continuous and we end with the general case. We give a rigorous proof of this Theorem A.3.1 but only of the first step of the proof taken from Villani [89], in other words in the particular compact case and when the cost function c is continuous. We recall a preliminary result on convex sets.

Lemma A.3.1

Let E be a normed vector space and let $C\subset E$ be a convex set. Then $\mathrm{Int}(C)$ is convex, and if in addition $\mathrm{Int}(C)\neq\emptyset$, then we have $\overline{C}=\overline{\mathrm{Int}(C)}$.

Proof. Let $x,\ y\in\mathrm{Int}(C)$. Then there exists $\varepsilon>0$ and $u\in E$ with $|u|<1$ such that $x+\varepsilon u$ and $y+\varepsilon u\in C$. For $t\in[0,1]$, we set $z:=tx+(1-t)y$. By convexity of C, $t(x+\varepsilon u)+(1-t)(y+\varepsilon u)\in C$. But at the same time, we have $t(x+\varepsilon u)+(1-t)(y+\varepsilon u)=z+\varepsilon u$. So $B(z,\varepsilon)\subset C$, in other words, $z\in\mathrm{Int}(C)$.

Let $x\in\overline{\mathrm{Int}(C)}$, then there exists a sequence $(x_n)_n$ in $\mathrm{Int}(C)$ which converges to x. And $\mathrm{Int}(C)\subset C$ immediately results in $x\in\overline{C}$ thus $\overline{\mathrm{Int}(C)}\subset\overline{C}$. Conversely, if $x_0\in\mathrm{Int}(C)$, then there exists $r_0>0$ such that $B(x_0,r_0)\subset C$. For $x\in C$ and $0<\varepsilon<1$, we set $z:=(1-\varepsilon)(x-x_0)+x_0$. Then for $u\in E$ with $|u|<1$, we have by convexity of C that $z+\varepsilon r_0u\in C$, because it is a convex combination of x and x_0+r_0u. So $B(z,\varepsilon r_0)\subset C$, in other words $z\in\mathrm{Int}(C)$. Now let $x\in\overline{C}$, then there exists a sequence $(x_n)_n$ in C which converges to x. For $x_0\in\mathrm{Int}(C)$ (which exists since $\mathrm{Int}(C)$ is non-empty), we define the sequence $(z_n)_n$ by

$$z_n:=\left(1-\frac{1}{n}\right)(x_n-x_0)+x_0.$$

We have $z_n\in\mathrm{Int}(C)$ since $x_n\in C$, and $(z_n)_n$ converges to x. So $x\in\overline{\mathrm{Int}(C)}$, in other words $\overline{C}\subset\overline{\mathrm{Int}(C)}$. ■

To prove Theorem A.3.1, we use the following result which allows to interchange the infimum and the supremum in the specific case of convex functions (Theorem 1.9 in Ref. [89], Theorem 1.11 in Ref. [19]).

Definition A.3.1

Let E be a normed vector space and let $T:E\longrightarrow\mathbb{R}\cup\{+\infty\}$ a function which is not identically equal to $+\infty$ (i.e there exists $x_0\in\mathrm{Dom}(T)$ such that $T(x_0)<+\infty$). We define the **conjugate** of T the function $T^*:E^*\longrightarrow\mathbb{R}\cup\{+\infty\}$ by

$$T^*(f)=\sup_{x\in E}(f(x)-T(x)).$$

Theorem A.3.2 (Fenchel-Rockafellar)

Let E be a normed vector space and let T, $S : E \longrightarrow \mathbb{R} \cup \{+\infty\}$ be two convex functions. Assume that there exists $x_0 \in E$ such that $T(x_0) < +\infty$, $S(x_0) < +\infty$ and T is continuous at x_0. Then

$$\inf_{x\in E}(T(x) + S(x)) = \sup_{f\in E^*}(-T^*(-f) - S^*(f)) \tag{A.3.3}$$

and the supremum on the right-hand side is attained.

Proof. Let $f \in E^*$, we have

$$\begin{aligned}\inf_{x,y\in E}(T(x) + S(y) + f(x-y)) &= \inf_{x\in E}(f(x) + T(x)) + \inf_{y\in E}(-f(y) + S(y)) \\ &= -\sup_{x\in E}(-f(x) - T(x)) - \sup_{y\in E}(f(y) - S(y)) \\ &= -T^*(-f) - S^*(f).\end{aligned}$$

Thus,

$$-T^*(-f) - S^*(f) = \inf_{x,y\in E}(T(x) + S(y) + f(x-y)) \leq \inf_{x\in E}(T(x) + S(x)).$$

And by taking the supremum over f, we obtain

$$\sup_{f\in E^*}(-T^*(-f) - S^*(f)) \leq \inf_{x\in E}(T(x) + S(x)).$$

Conversely, let $m := \inf_x(T(x) + S(x))$. We set $A := \{(x,\lambda) \in E \times \mathbb{R},\ T(x) \leq \lambda\}$ the epigraph of T and $B := \{(x,\lambda) \in E \times \mathbb{R},\ \lambda \leq m - S(x)\}$ the hypograph of $m - S$. The convexity of T and S implies that A and B are convex. In addition, $T(x_0) < +\infty$, $S(x_0) < +\infty$ together with the continuity of T at x_0 gives that $\mathrm{Int}(A)$ and B are non-empty. By Lemma A.3.1, $\mathrm{Int}(A)$ is open and convex. And in addition, if $(x,\lambda) \in \mathrm{Int}(A)$, then $T(x) < \lambda$ and thus $m \leq T(x) + S(x) < \lambda + S(x)$, in other words $(x,\lambda) \notin B$. $\mathrm{Int}(A)$ and B are disjoint, so according to the Hahn-Banach Theorem, there exists a closed hyperplan that separates $\mathrm{Int}(A)$ and B. In other words there exists a linear form Φ on $E \times \mathbb{R}$ and $\alpha \in \mathbb{R}$ such that

$$\begin{aligned}&\Phi(x,\lambda) \geq \alpha \qquad \forall (x,\lambda) \in \mathrm{Int}(A) \\ &\Phi(x,\lambda) \leq \alpha \qquad \forall (x,\lambda) \in B.\end{aligned}$$

More precisely, there exists $k \in \mathbb{R}$ and $f \in E^*$ such that $\Phi(x,\lambda) = f(x) + k\lambda$. If $(x,\lambda) \in \overline{\mathrm{Int}(A)}$, then there exists a sequence $(x_n,\lambda_n)_n$ in $\mathrm{Int}(A)$ which converges to (x,λ). We have by continuity of f that $f(x) + k\lambda \geq \alpha$ for all $(x,\lambda) \in \overline{\mathrm{Int}(A)}$. We therefore have by Lemma A.3.1

$$f(x) + k\lambda \geq \alpha \qquad \forall (x,\lambda) \in A \tag{A.3.4}$$

$$f(x) + k\lambda \leq \alpha \qquad \forall (x,\lambda) \in B. \tag{A.3.5}$$

By taking $x = x_0$ in (A.3.4), we see by making λ tends to $+\infty$ that $k \geq 0$. Suppose that $k = 0$, then we have $f(x) \geq \alpha$ for all $(x,\lambda) \in A$ and $f(x) \leq \alpha$ for all $(x,\lambda) \in B$. In other words, we have $f(x) \geq \alpha$ for all $x \in \mathrm{Dom}(T)$ and $f(x) \leq \alpha$ for all $x \in \mathrm{Dom}(S)$. Since $x_0 \in \mathrm{Dom}(T) \cap \mathrm{Dom}(S)$, then $f(x_0) = \alpha$ and since T is continuous at x_0, then there exists a ball $B(x_0,\varepsilon) \subset \mathrm{Dom}(T)$. For all $x \in E$ with $|x| < \varepsilon$ and for all $\delta \in]-1,1[$, we have

$$f(x_0 + \delta x) \geq \alpha \qquad \Rightarrow \qquad f(x_0) + \delta f(x) \geq \alpha \qquad \Rightarrow \qquad \delta f(x) \geq 0.$$

So $f(x) = 0$ for all $x \in B(0, \varepsilon)$ and therefore $f = 0$ on E, which is a contradiction. So $k > 0$ and we have by (A.3.4)

$$T^*\left(-\frac{f}{k}\right) = \sup_{x\in E}\left(-\frac{1}{k}f(x) - T(x)\right) = -\inf_{x\in E}\left(\frac{1}{k}f(x) + T(x)\right) = -\frac{1}{k}\inf_{x\in E}(f(x) + kT(x)) \leq -\frac{\alpha}{k}$$

since $(x, T(x)) \in A$. Similary, we have by (A.3.5)

$$S^*\left(\frac{f}{k}\right) = \sup_{x\in E}\left(\frac{1}{k}f(x) - S(x)\right) = \frac{1}{k}\sup_{x\in E}(f(x) + k(m - S(x))) - m \leq \frac{\alpha}{k} - m$$

since $(x, m - S(x)) \in B$. So we have

$$m = \frac{\alpha}{k} + \left(m - \frac{\alpha}{k}\right) \leq -T^*\left(-\frac{f}{k}\right) - S^*\left(\frac{f}{k}\right) \leq \sup_{f\in E^*}(-T^*(-f) - S^*(f)) \leq m.$$

This results in (A.3.3) where the supremum is attained by f/k. ■

This Theorem A.3.2 involves the dual space. We will recall before moving on to the proof of Theorem A.3.1 what are the duals spaces of continuous functions. We first start by defining the total variation of a signed measure (i.e which can take negative values).

Definition A.3.2

For a signed measure μ on a measured space $\mathcal{X}$, the **total variation** of μ is defined for any Borelian $\mathcal{A}$ of $\mathcal{X}$ by

$$|\mu|(\mathcal{A}) := \sup \sum_{i=1}^{+\infty} |\mu(\mathcal{A}_i)| \tag{A.3.6}$$

where the supremum is taken over all the partitions of $\mathcal{A}$.

The total variation of a signed measure μ is in a way the absolute value of μ. Note that $|\mu|(\mathcal{A}) \geq |\mu(\mathcal{A})|$, but that in general $|\mu|(\mathcal{A})$ is not equal to $|\mu(\mathcal{A})|$. The total variation of a a signed measure μ, $|\mu|$ is a measure and if μ is a measure, we have $|\mu| = \mu$.

Definition A.3.3

Let μ be a signed measure (i.e it can take negative values) on a measured space $\mathcal{X}$. We define the **positive part** and the **negative part** of μ as being the two measures μ_+ and μ_- defined for any Borelian $\mathcal{A}$ of $\mathcal{X}$ by

$$\mu_+(\mathcal{A}) := \sup_{\mathcal{B}\subset\mathcal{A}} \mu(\mathcal{B}) \qquad \text{and} \qquad \mu_-(\mathcal{A}) := \sup_{\mathcal{B}\subset\mathcal{A}}(-\mu(\mathcal{B})). \tag{A.3.7}$$

The Jordan's decomposition Theorem guarantees the existence of a unique pair of measures (μ_+, μ_-) such that $|\mu| = \mu_+ + \mu_-$ and such that there exists a Borelian $\mathcal{A}$ satisfying $\mu_-(\mathcal{A}) = \mu_+(\mathcal{X}\backslash\mathcal{A}) = 0$. And μ_+ and μ_- are defined by (A.3.7). Definition A.3.3 extends the notion of positive and negative part to signed measures. It is easy to check that we have as for real numbers $\mu = \mu_+ - \mu_-$. We can define in a similar way to real numbers, the maximum and the minimum between two signed measures μ and ν by

$$\max(\mu, \nu) := \nu + (\mu - \nu)_+ \qquad \text{and} \qquad \min(\mu, \nu) := \mu - (\mu - \nu)_+. \tag{A.3.8}$$

Definition A.3.4

Let E be a topological space and let $\mathscr{B}(E)$ be the Borelian set of E. Consider a map $\mu : \mathscr{B}(E) \longrightarrow \mathbb{R}$. We say that μ is

(1) **Bounded** on E if for any Borelian $\mathcal{A} \in \mathscr{B}(E)$, $|\mu|(\mathcal{A}) < +\infty$.

(2) **Finitely (Countably) additive** if for any finite (countable) number of disjoint set $(\mathcal{A}_i)_i$ in $\mathscr{B}(E)$, it holds that $\mu(\bigcup \mathcal{A}_i) = \sum \mu(\mathcal{A}_i)$.

(3) **Regular** if for any Borelian $\mathcal{A} \in \mathscr{B}(E)$ and for any $\varepsilon > 0$, there exists an open set O and a closed set F such that $F \subset \mathcal{A} \subset O$ and such that for any Borelian $\mathcal{B} \subset O \backslash F$ it holds that $|\mu|(\mathcal{B}) < \varepsilon$.

We denote by $ba(E)$ the set of maps $\mu : \mathscr{B}(E) \longrightarrow \mathbb{R}$ bounded and finitely additive, $rba(E)$ that of regular bounded and finitely additive maps, and $rca(E)$ that of regular bounded and countably additive maps. Equipped with the variational norm $\|\cdot\|$: $\|\mu\| := |\mu|(E)$, $ba(E)$ is a Banach space if E is a normal topological space. The space $ba(E)$ is a fairly large space of "measures". Note that we did neither assume that the member are countably additive nor that the values $\mu(\mathcal{A})$ shall be non-negative. However, we assumed the all values $|\mu|(\mathcal{A})$ are finite which excludes for example the Lebesgue measure on $E = \mathbb{R}$. The space $ba(E)$ is slightly too large to be the dual space of $\mathcal{C}_b(E)$ for a normal topological space E and we need another restriction. We need that a "measure" in $ba(E)$ to be regular in other words that it be in $rba(E)$, which is a closed subspace of $ba(E)$.

Theorem A.3.3

For a normal topological space E, it holds that $\mathcal{C}_b(E)^* = rba(E)$.

This Theorem is Theorem IV.6.2 in Ref. [43]. The proof is lenghty and technical. Now we assume that E is a Hausdorff space (in other words separate) and locally compact. The absence of compactness will be compensated by the fact that the function "vanishes at the boundary". We obtain the following result called the Riesz-Markov Theorem.

Theorem A.3.4 (Riesz-Markov Theorem)

If E is a locally compact Hausdorff space, then every bounded linear functionnal on $\mathcal{C}_0(E)$ is represented by a unique "measure" $\mu \in rca(E)$. Moreover, the norm of μ is the total variation of μ: $\|\mu\| := |\mu|(E)$.

This result is best known when it states that the dual of $\mathcal{C}(E)$ when E is compact is $rca(E)$. This is Theorem IV.6.3 in Ref. [43]. But the locally compact case is more general. It is now time to move on to the proof of Kantorovich duality Theorem A.3.1.

Proof of Theorem A.3.1. We take the proof given by Villani in Ref. [89]. This is separated into three steps. The first step consists in proving the equality (A.3.1) in the compact case and when the cost function c is continuous. It is here that the convex analysis will play a fundamental role. In the second step, we extend (A.3.1) to the non-compact case but with the assumption that c is bounded and uniformly continuous. We will reduce to the compact case by a truncation procedure. And we end with the general case in a third step, by setting $c = \sup c_n$ where c_n is a non-decreasing sequence of non-negative and uniformly continuous cost functions.

We only give here the proof of the first step. Let X, Y be two compacts on $\mathbb{R}^d$ and $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. We set $E := \mathcal{C}_b(X \times Y)$ and we define the two functions T and $S : E \longrightarrow \mathbb{R} \cup \{+\infty\}$ by

$$T(f) := \begin{cases} 0 & \text{if } f(x,y) \geq -c(x,y), \\ +\infty & \text{else.} \end{cases}$$

$$S(f) := \begin{cases} \displaystyle\int_X \varphi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) & \text{if } f(x,y) = \varphi(x) + \psi(y),\ \varphi \in \mathcal{C}_b(X),\ \psi \in \mathcal{C}_b(Y), \\ +\infty & \text{else.} \end{cases}$$

We have

$$\begin{aligned} \mathrm{Dom}(T) &= \{f \in E,\ \forall (x,y) \in X \times Y,\ f(x,y) \geq -c(x,y)\} \\ \mathrm{Dom}(S) &= \{f \in E,\ \forall (x,y) \in X \times Y,\ f(x,y) = \varphi(x) + \psi(y),\ \varphi \in \mathcal{C}_b(X),\ \psi \in \mathcal{C}_b(Y)\}. \end{aligned}$$

Note that the decomposition $f(x,y) = \varphi(x) + \psi(y)$ in $\mathrm{Dom}(S)$ is not unique. If we have $\varphi_1(x) + \psi_1(y) = \varphi_2(x) + \psi_2(y)$ for every $x \in X$, $y \in Y$, then there exists a constant $\lambda \in \mathbb{R}$ such that $\varphi_1 = \varphi_2 + \lambda$ and $\psi_1 = \psi_2 - \lambda$ and therefore $\int \varphi_1 d\mu + \int \psi_1 d\nu = \int \varphi_2 d\mu + \int \psi_2 d\nu$. So S is well-defined. If we take $f_0 = 1$, then $f_0 \in \mathrm{Dom}(T) \cap \mathrm{Dom}(S)$ and T is continuous at f_0. Let us show that T and S are convex. Let f and g be two functions in E such that $T(f) = +\infty$ or $T(g) = +\infty$ and let $t \in [0,1]$. We have clearly

$$T(tf + (1-t)g) \leq tT(f) + (1-t)T(g) = +\infty.$$

If $T(f) < +\infty$ and $T(g) < +\infty$, then we have $f(x,y) \geq -c(x,y)$ and $g(x,y) \geq -c(x,y)$. Therefore $tf(x,y) + (1-t)g(x,y) \geq -c(x,y)$ for all $t \in [0,1]$ and this results in that

$$T(tf + (1-t)g) = 0 = tT(f) + (1-t)T(g).$$

So T is convex. For S, if $S(f) = +\infty$ or $S(g) = +\infty$, we have clearly

$$S(tf + (1-t)g) \leq tS(f) + (1-t)S(g) = +\infty.$$

If $S(f) < +\infty$ and $S(g) < +\infty$, then $f(x,y) = \varphi_1(x) + \psi_1(y)$ and $g(x,y) = \varphi_2(x) + \psi_2(y)$. Therefore $(tf + (1-t)g)(x,y) = (t\varphi_1 + (1-t)\varphi_2)(x) + (t\psi_1 + (1-t)\psi_2)(y)$ and this results in that

$$\begin{aligned} S(tf + (1-t)g) &= \int_X (t\varphi_1(x) + (1-t)\varphi_2(x))d\mu(x) + \int_Y (t\psi_1(y) + (1-t)\psi_2(y))d\mu(y) \\ &= tS(f) + (1-t)S(g). \end{aligned}$$

So S is convex, we can apply the Fenchel-Rockafellar Theorem (Theorem A.3.2). Let's now compute both sides of (A.3.3). Since

$$\begin{aligned} \mathrm{Dom}(T) \cap \mathrm{Dom}(S) = \{&f \in E,\ \exists (\varphi,\psi) \in \mathcal{C}_b(X) \times \mathcal{C}_b(Y),\ \forall (x,y) \in X \times Y,\ f(x,y) = \varphi(x) + \psi(y) \\ &\geq -c(x,y)\}, \end{aligned}$$

on one side we have

$$\begin{aligned} \inf_{f \in E}(T(f) + S(f)) &= \inf_{f \in \mathrm{Dom}(T) \cap \mathrm{Dom}(S)}(T(f) + S(f)) \\ &= \inf_{f \in \mathrm{Dom}(T) \cap \mathrm{Dom}(S)} \left(\int_X \varphi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) \right) \\ &= - \sup_{(\varphi,\psi) \in \Phi_c(X \times Y)} \left(\int_X \varphi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) \right) \end{aligned}$$

where $\Phi_c(X\times Y) := \{(\varphi,\psi)\in \mathcal{C}_b(X)\times\mathcal{C}_b(Y),\ \forall (x,y)\in X\times Y,\ \varphi(x)+\psi(y)\le c(x,y)\}$. And on the other hand, since $E=\mathcal{C}_b(X\times Y)$ with X and Y compact, then by the Riesz-Markov Theorem (Theorem A.3.4), the dual space of E, E^* is the set of Radon bounded measures on $X\times Y$ (i.e the set of regular bounded countable additivity "measures" on $X\times Y$, $rca(X\times Y)$). Let $\pi\in E^*$ with positive part π_+ and negative part π_-: $\pi=\pi_+-\pi_-$. We assume that $\pi_+\neq 0$. There are Z_+ and Z_- such that $X\times Y = Z_+\cup Z_-$ with $Z_+\cap Z_- = \emptyset$ and $\pi_+(Z_-)=\pi_-(Z_+)=0$. Since π_+ and π_- are regular measures, then for all $\lambda>0$, there exists a compact $K\subset Z_+$ such that $\pi_+(K)>0$ and an open set $O_\lambda\supset Z_+$ such that $\pi_-(O_\lambda)\le 1/\lambda$. Let $f_0\in E$ such that $\mathbb{1}_K\le f_0\le \mathbb{1}_{O_\lambda}$. In particular $f_0\ge -c$ since f_0 is non-negative and therefore it results that

$$\begin{aligned}
T^*(\pi) &= \sup_{f\in E}\left(\iint_{X\times Y} f(x,y)d\pi(x,y) - T(f)\right)\\
&\ge \lambda\iint_{X\times Y} f_0(x,y)d\pi_+(x,y) - \lambda\iint_{X\times Y} f_0(x,y)d\pi_-(x,y)\\
&\ge \lambda\iint_{X\times Y} f_0(x,y)d\pi_+(x,y) - \lambda\pi_-(O_\lambda) \longrightarrow +\infty
\end{aligned}$$

when $\lambda\longrightarrow+\infty$. So $\pi\notin \mathrm{Dom}(T^*)$ if $\pi_+\neq 0$ and $\pi_+=0$ immediately implies that

$$T^*(\pi) = \sup_{f\ge -c}\iint_{X\times Y} f(x,y)d\pi(x,y) = \sup_{f\le c}\iint_{X\times Y} f(x,y)d\pi_-(x,y) = \iint_{X\times Y} c(x,y)d\pi_-(x,y)$$

since π_- is a measure. We therefore have that $\mathrm{Dom}(T^*)$ is the set of Radon measures $\pi\in E^*$ with zero positive part: $\pi_+=0$. Thus we have

$$T^*(\pi) = \begin{cases} -\displaystyle\iint_{X\times Y} c(x,y)d\pi(x,y) & \text{if } \pi\in\mathrm{Dom}(T^*),\\ +\infty & \text{else.}\end{cases}$$

We will now show that $\Pi(\mu,\nu) = \{\pi\in E^*,\ \pi\in\mathrm{Dom}(S^*),\ -\pi\in\mathrm{Dom}(T^*)\}$. If $\pi\in\Pi(\mu,\nu)$, then π has zero negative part which means that $-\pi\in\mathrm{Dom}(T^*)$. Furthermore,

$$\begin{aligned}
S^*(\pi) &= \sup_{f\in E}\left(\iint_{X\times Y} f(x,y)d\pi(x,y) - S(f)\right)\\
&= \sup_{f=\varphi+\psi}\left(\iint_{X\times Y} f(x,y)d\pi(x,y) - \int_X\varphi(x)d\mu(x) - \int_Y\psi(y)d\nu(y)\right)\\
&= 0
\end{aligned}$$

since π has μ and ν as marginals, hence $\pi\in\mathrm{Dom}(S^*)$. Conversely, let $\pi\in E^*$ such that $\pi\in\mathrm{Dom}(S^*)$ and $-\pi\in\mathrm{Dom}(T^*)$. Then π has zero negative part. If we suppose that $\pi\notin\Pi(\mu,\nu)$, then there exists $\varphi_0\in\mathcal{C}_b(X)$ such that $\int\varphi_0 d\pi > \int\varphi_0 d\mu$ and this results in that

$$\begin{aligned}
S^*(\pi) &= \sup_{(\varphi,\psi)}\left(\iint_{X\times Y}\varphi(x)d\pi(x,y) - \int_X\varphi(x)d\mu(x) + \iint_{X\times Y}\psi(y)d\pi(x,y) - \int_Y\psi(y)d\nu(y)\right)\\
&\ge \lambda\left(\iint_{X\times Y}\varphi_0(x)d\pi(x,y) - \int_X\varphi_0(x)d\mu(x)\right)\longrightarrow+\infty
\end{aligned}$$

when $\lambda\longrightarrow+\infty$ then $\pi\notin\mathrm{Dom}(S^*)$ which is a contradiction. Therefore, $\pi\in\Pi(\mu,\nu)$ which results in

$$\begin{aligned}
\sup_{\pi\in E^*}\left(-T^*(-\pi)-S^*(\pi)\right) &= \sup_{(\pi,-\pi)\in\mathrm{Dom}(S^*)\times\mathrm{Dom}(T^*)}\left(-T^*(-\pi)-S^*(\pi)\right)\\
&= \sup_{\pi\in\Pi(\mu,\nu)}\left(-\iint_{X\times Y}c(x,y)d\pi(x,y)\right)\\
&= -\inf_{\pi\in\Pi(\mu,\nu)}\iint_{X\times Y}c(x,y)d\pi(x,y).
\end{aligned}$$

Putting everything together and changing signs, we obtain

$$\inf_{\pi\in\Pi(\mu,\nu)}\iint_{X\times Y}c(x,y)d\pi(x,y)=\sup_{(\varphi,\psi)\in\Phi_c(X\times Y)}\left(\int_X\varphi(x)d\mu(x)+\int_Y\psi(y)d\nu(y)\right)$$

which is (A.3.1) in the compact case. ■

Theorem A.3.1 gives an alternative expression for the Wasserstein metric called Kantorovich formula. For $\mu,\,\nu\in\mathcal{P}_p(\mathbb{R}^d)$, we have

$$W_p(\mu,\nu)=\left(\sup_{(\varphi,\psi)\in\Phi_p}\left(\int_{\mathbb{R}^d}\varphi(x)d\mu(x)+\int_{\mathbb{R}^d}\psi(y)d\nu(y)\right)\right)^{1/p}\tag{A.3.9}$$

where $\Phi_p:=\{\varphi,\,\psi\in\mathcal{C}_b(\mathbb{R}^d),\,\forall x,y\in\mathbb{R}^d,\,\varphi(x)+\psi(y)\leq|x-y|^p\}$. This formula (A.3.9) gives in the particular case $p=1$ a very nice expression.

Corollary A.3.1 (Dual representation of W_1)

Let $\mu,\,\nu\in\mathcal{P}_1(\mathbb{R}^d)$. By setting $\mathscr{F}:=\{f,\,\|f\|_{\text{Lip}}\leq 1\}$, we have

$$W_1(\mu,\nu)=\sup_{f\in\mathscr{F}}\left|\int_{\mathbb{R}^d}f(x)d\mu(x)-\int_{\mathbb{R}^d}f(y)d\nu(y)\right|.\tag{A.3.10}$$

Proof. Let $\mu,\,\nu\in\mathcal{P}_1(\mathbb{R}^d)$. We set

$$S:=\sup_{f\in\mathscr{F}}\left|\int_{\mathbb{R}^d}f(x)d\mu(x)-\int_{\mathbb{R}^d}f(y)d\nu(y)\right|.$$

Let $\pi\in\Pi(\mu,\nu)$ be an optimal coupling and let f be a 1-Lipschitzian function. Then we have

$$\begin{aligned}\left|\int_{\mathbb{R}^d}f(x)d\mu(x)-\int_{\mathbb{R}^d}f(y)d\nu(y)\right|&=\left|\iint_{\mathbb{R}^d\times\mathbb{R}^d}(f(x)-f(y))d\pi(x,y)\right|\\&\leq\iint_{\mathbb{R}^d\times\mathbb{R}^d}|f(x)-f(y)|d\pi(x,y)\\&\leq\iint_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|d\pi(x,y)\\&=W_1(\mu,\nu).\end{aligned}$$

And we take the supremum over f, we obtain $S\leq W_1(\mu,\nu)$.

Conversely, let $\varepsilon>0$. Then by (A.3.9) with $p=1$, there exists $\varphi\in L^1(\mathbb{R}^d,\mu)$ and $\psi\in L^1(\mathbb{R}^d,\nu)$ such that $\varphi(x)+\psi(y)\leq|x-y|$ and

$$W_1(\mu,\nu)-\varepsilon\leq\int_{\mathbb{R}^d}\varphi(x)d\mu(x)+\int_{\mathbb{R}^d}\psi(y)d\nu(y).$$

We set $f(x):=\sup_y(\psi(y)-|x-y|)$. Then by the triangle inequality we have

$$f(x)=\sup_{y\in\mathbb{R}^d}(\psi(y)-|x-y|)\leq\sup_{y\in\mathbb{R}^d}(\psi(y)-|x'-y|+|x-x'|)=f(x')+|x-x'|.$$

So f is 1-Lipschitz and $\mu,\,\nu\in\mathcal{P}_1(\mathbb{R}^d)$ implies that $f\in L^1(\mathbb{R}^d,\mu)\times L^1(\mathbb{R}^d,\nu)$ (because any Lipschitzian function is integrable with respect to μ if $\mu\in\mathcal{P}_1(\mathbb{R}^d)$). By taking $x=y$ in the supremum defining f, we have $f(x)\geq\psi(x)$. And since $\varphi(x)+\psi(y)\leq|x-y|$, we also have $f(x)\leq-\varphi(x)$. So we have

$$W_1(\mu,\nu)-\varepsilon\leq\int_{\mathbb{R}^d}\varphi(x)d\mu(x)+\int_{\mathbb{R}^d}\psi(y)d\nu(y)\leq-\int_{\mathbb{R}^d}f(x)d\mu(x)+\int_{\mathbb{R}^d}f(y)d\nu(y)\leq S.$$

Hence $W_1(\mu,\nu)\leq S$ since ε is arbitrary. ■

A.4 Properties of the Wasserstein metric

This paragraph lists various properties on the Wasserstein metric.

Proposition A.4.1 (Comparaison)

Let $1 \leq p \leq q$ and let μ, $\nu \in \mathcal{P}_q(\mathbb{R}^d)$. Then we have

$$W_p(\mu,\nu) \leq W_q(\mu,\nu). \tag{A.4.1}$$

Proof. Let $\pi \in \Pi(\mu,\nu)$ an optimal coupling for W_q. Then by the Hölder inequality, we have since $p \leq q$

$$W_p(\mu,\nu)^p \leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\pi(x,y) \leq \left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^q d\pi(x,y)\right)^{p/q} = W_q(\mu,\nu)^p.$$

■

The most useful exponents for the Wasserstein metric are $p = 1$ and $p = 2$. W_1 therefore is the weakest of all, but it is also the easiest to manipulate with the formula (A.3.10). W_2 has more interesting properties and is better adapted. But the results obtained for W_1 are generally difficult to show for W_2.

Proposition A.4.2 (Wasserstein metric from a Dirac measure)

If μ belongs to $\mathcal{P}_p(\mathbb{R}^d)$, then for all $x_0 \in \mathbb{R}^d$

$$W_p(\mu,\delta_{x_0}) = \left(\int_{\mathbb{R}^d} |x-x_0|^p d\mu(x)\right)^{1/p}. \tag{A.4.2}$$

Proof. Let $\pi \in \Pi(\mu,\delta_{x_0})$ and let $\pi_* := \mu\otimes\delta_{x_0}$. For $\varphi \in \mathcal{C}_c(\mathbb{R}^d\times\mathbb{R}^d)$, we have

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x,y) d\pi_*(x,y) = \int_{\mathbb{R}^d} \varphi(x,x_0) d\mu(x) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x,x_0) d\pi(x,y).$$

Let $\varepsilon > 0$, since $\varphi \in \mathcal{C}_c(\mathbb{R}^d\times\mathbb{R}^d)$, then φ is uniformly continuous. So there exists $\delta > 0$ such that for all (x,y), (x',y'), $\sqrt{|x-x'|^2+|y-y'|^2} < \delta$ implies that $|\varphi(x,y)-\varphi(x',y')| < \varepsilon$. So that results in

$$\begin{aligned}\left|\iint_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x,y)d\pi(x,y) - \iint_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x,y)d\pi_*(x,y)\right| &\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} |\varphi(x,y)-\varphi(x,x_0)|d\pi(x,y)\\ &= \iint_{|y-x_0|<\delta} |\varphi(x,y)-\varphi(x,x_0)|d\pi(x,y)\\ &+ \iint_{|y-x_0|\geq\delta} |\varphi(x,y)-\varphi(x,x_0)|d\pi(x,y).\end{aligned}$$

The second term on the right-hand side is bounded by $2\|\varphi\|_\infty \pi(\{(x,y),\, |y-x_0|\geq\delta\}) = 0$ since $x_0 \notin \{y,\, |y-x_0|\geq\delta\}$. For the first term on the right-hand side, we use the uniform continuity of φ with the pair (x,y), (x,x_0). Since $\sqrt{|x-x|^2+|y-x_0|^2} < \delta$, then $|\varphi(x,y)-\varphi(x,x_0)| < \varepsilon$, which results in

$$\left|\iint_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x,y)d\pi(x,y) - \iint_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x,y)d\pi_*(x,y)\right| < \varepsilon.$$

Valid for any $\varepsilon > 0$, for any $\varphi \in \mathcal{C}_c(\mathbb{R}^d\times\mathbb{R}^d)$ and for any coupling $\pi \in \Pi(\mu,\delta_{x_0})$ therefore $\pi = \pi_*$. ■

Proposition A.4.2 shows that the only coupling in $\Pi(\mu, \delta_{x_0})$ is $\mu \otimes \delta_{x_0}$ (all the mass should be transported to x_0). We can show this with the monotone class Theorem. Let $\nu \in \mathcal{P}(\mathbb{R}^d)$ be the measure defined for any Borelian $\mathcal{A} \subset \mathbb{R}^d$ by $\nu(\mathcal{A}) := \pi(\mathcal{A} \times \{x_0\})$. If $\pi \in \Pi(\mu, \delta_{x_0})$, then for any Borelian $\mathcal{A} \times \mathcal{B} \subset \mathbb{R}^d \times \mathbb{R}^d$ we have

$$\pi(\mathcal{A} \times \mathcal{B}) = \pi((\mathcal{A} \times (\mathcal{B} \cap \{x_0\})) \cup (\mathcal{A} \times (\mathcal{B} \backslash \{x_0\}))) = \pi(\mathcal{A} \times (\mathcal{B} \cap \{x_0\})) + \pi(\mathcal{A} \times (\mathcal{B} \backslash \{x_0\})).$$

And $\pi(\mathcal{A} \times (\mathcal{B} \backslash \{x_0\})) \leq \pi(\mathbb{R}^d \times (\mathcal{B} \backslash \{x_0\})) = 0$ so that implies that

$$\pi(\mathcal{A} \times \mathcal{B}) = \pi(\mathcal{A} \times (\mathcal{B} \cap \{x_0\}))\mathbb{1}_{x_0 \in \mathcal{B}} + \pi(\mathcal{A} \times (\mathcal{B} \cap \{x_0\}))\mathbb{1}_{x_0 \notin \mathcal{B}} = \pi(\mathcal{A} \times \{x_0\})\mathbb{1}_{x_0 \in \mathcal{B}} = \nu(\mathcal{A})\mathbb{1}_{x_0 \in \mathcal{B}}.$$

So by the monotone class Theorem, we have $\pi = \nu \otimes \delta_{x_0}$ and therefore we have for any Borelian $\mathcal{A} \subset \mathbb{R}^d$ that $\mu(\mathcal{A}) = \pi(\mathcal{A} \times \mathbb{R}^d) = \nu(\mathcal{A})\delta_{x_0}(\mathbb{R}^d) = \nu(\mathcal{A})$ so $\mu = \nu$.

Proposition A.4.3 (Non-decreasing coupling)

Let $\mu \in \mathcal{P}_2(\mathbb{R})$ with for density f and let $x_0 \in \mathbb{R}$ such that $\mu(\{x \leq x_0\}) = \alpha$. Then for all u, $v \in \mathbb{R}$ such that $u < v$ we have

$$W_2(\mu, \alpha\delta_u + (1-\alpha)\delta_v)^2 = \int_{-\infty}^{x_0} |x-u|^2 f(x)dx + \int_{x_0}^{+\infty} |x-v|^2 f(x)dx. \tag{A.4.3}$$

Proof. Let $\pi \in \Pi(\mu, \alpha\delta_u + (1-\alpha)\delta_v)$. We define the two measures μ_1 and μ_2 for any Borelian $\mathcal{A} \subset \mathbb{R}$ by $\mu_1(\mathcal{A}) := \pi(\mathcal{A} \times \{u\})/\alpha$ and $\mu_2(\mathcal{A}) := \pi(\mathcal{A} \times \{v\})/(1-\alpha)$. Let $\mathcal{A} \times \mathcal{B} \subset \mathbb{R}^2$ be a Borelian. We have

$$\begin{aligned}\pi(\mathcal{A} \times \mathcal{B}) &= \pi((\mathcal{A} \times (\mathcal{B} \cap \{u\})) \cup (\mathcal{A} \times (\mathcal{B} \cap \{v\})) \cap (\mathcal{A} \times (\mathcal{B} \backslash \{u, v\}))) \\ &= \pi(\mathcal{A} \times (\mathcal{B} \cap \{u\})) + \pi(\mathcal{A} \times (\mathcal{B} \cap \{v\})) + \pi(\mathcal{A} \times (\mathcal{B} \backslash \{u, v\})).\end{aligned}$$

And $\pi(\mathcal{A} \times (\mathcal{B} \backslash \{u, v\})) \leq \pi(\mathbb{R} \times (\mathcal{B} \backslash \{u, v\})) = 0$ so that implies that

$$\begin{aligned}\pi(\mathcal{A} \times \mathcal{B}) &= \pi(\mathcal{A} \times (\mathcal{B} \cap \{u\}))\mathbb{1}_{u \in \mathcal{B}} + \pi(\mathcal{A} \times (\mathcal{B} \cap \{u\}))\mathbb{1}_{u \notin \mathcal{B}} \\ &\quad + \pi(\mathcal{A} \times (\mathcal{B} \cap \{v\}))\mathbb{1}_{v \in \mathcal{B}} + \pi(\mathcal{A} \times (\mathcal{B} \cap \{v\}))\mathbb{1}_{v \notin \mathcal{B}} \\ &= \pi(\mathcal{A} \times \{u\})\mathbb{1}_{u \in \mathcal{B}} + \pi(\mathcal{A} \times \{v\})\mathbb{1}_{v \in \mathcal{B}}.\end{aligned}$$

So by the monotone class Theorem, we have $\pi = \alpha\mu_1 \otimes \delta_u + (1-\alpha)\mu_2 \otimes \delta_v$. Let $\mathcal{A} \subset \mathbb{R}$ be a Borelian, we have $\mu(\mathcal{A}) = \pi(\mathcal{A} \times \mathbb{R}) = \alpha\mu_1(\mathcal{A}) + (1-\alpha)\mu_2(\mathcal{A})$ so $\mu(\mathcal{A}) = 0$ implies that $\mu_1(\mathcal{A}) = \mu_2(\mathcal{A}) = 0$ since $\alpha \in [0, 1]$. So the two measures μ_1 and μ_2 are absolutely continuous with respect to μ and consequently there exists by the Radon-Nikodym Theorem two densities f_1 and f_2 such that for any test function φ

$$\int_{\mathbb{R}} \varphi(x)f(x)dx = \alpha \int_{\mathbb{R}} \varphi(x)f(x)f_1(x)dx + (1-\alpha) \int_{\mathbb{R}} \varphi(x)f(x)f_2(x)dx.$$

In other words, we have the equality $f = \alpha f f_1 + (1-\alpha) f f_2$. By setting $h := \alpha f_1$, we have in the end that for any test function φ

$$\iint_{\mathbb{R}\times\mathbb{R}} \varphi(x, y)d\pi(x, y) = \int_{\mathbb{R}} \varphi(x, u)h(x)f(x)dx + \int_{\mathbb{R}} \varphi(x, v)(1-h(x))f(x)dx. \tag{A.4.4}$$

Let π_* be the coupling defined by the formula (A.4.4) with $h : x \longmapsto h(x) := \mathbb{1}_{]-\infty, x_0]}(x)$. Then for all

$\pi \in \Pi(\mu, \alpha\delta_u + (1-\alpha)\delta_v)$ we have

$$\begin{aligned}\iint_{\mathbb{R}\times\mathbb{R}} |x-y|^2 d\pi(x,y) - \iint_{\mathbb{R}\times\mathbb{R}} |x-y|^2 d\pi_*(x,y) &= -2u\int_{\mathbb{R}} xh(x)f(x)dx - 2v\int_{\mathbb{R}} x(1-h(x))f(x)dx \\ &\quad + 2u\int_{-\infty}^{x_0} xf(x)dx + 2v\int_{x_0}^{+\infty} xf(x)dx \\ &= 2(v-u)\int_{x_0}^{+\infty} xh(x)f(x)dx \\ &\quad - 2(v-u)\int_{-\infty}^{x_0} x(1-h(x))f(x)dx.\end{aligned}$$

To know the sign of the right-hand side, we study the sign of the quantity

$$\int_{x_0}^{+\infty} \frac{xh(x)f(x)}{\int_{x_0}^{+\infty} h(x)f(x)dx}dx - \int_{-\infty}^{x_0} \frac{x(1-h(x))f(x)}{\int_{-\infty}^{x_0}(1-h(x))f(x)dx}dx$$

since $\int_{x_0}^{+\infty} h(x)f(x)dx = \int_{-\infty}^{x_0}(1-h(x))f(x)dx$. Since for all $x_0 \in \mathbb{R}$ and for all $\mu \in \mathcal{P}(]-\infty, x_0])$, $\nu \in \mathcal{P}([x_0, +\infty[)$, we have $\int_{-\infty}^{x_0} x d\mu(x) \leq \int_{x_0}^{+\infty} x d\nu(x)$, we conclude since $u < v$ that

$$\iint_{\mathbb{R}\times\mathbb{R}} |x-y|^2 d\pi(x,y) - \iint_{\mathbb{R}\times\mathbb{R}} |x-y|^2 d\pi_*(x,y) \geq 0$$

in other words π_* is optimal. ■

Proposition A.4.4 (Transfert)

Given μ, $\nu \in \mathcal{P}_p(\mathbb{R}^d)$ and f, $g : \mathbb{R}^d \longrightarrow \mathbb{R}^d$.

(1) If f is a Lipschitzian Borel map, then

$$W_p(f\#\mu, f\#\nu) \leq \|f\|_{\text{Lip}} W_p(\mu,\nu). \tag{A.4.5}$$

(2) If f and g are two μ-measurable maps in $L^p(\mathbb{R}^d)$, then

$$W_p(f\#\mu, g\#\mu) \leq \|f-g\|_{L^p(\mathbb{R}^d,\mu)}. \tag{A.4.6}$$

Proof. For (A.4.5), let $\pi \in \Pi(\mu,\nu)$ be an optimal coupling. We set $\pi_* := (f,f)\#\pi$, then $\pi_* \in \Pi(f\#\mu, f\#\nu)$ and therefore

$$\begin{aligned}W_p(f\#\mu, f\#\nu) &\leq \left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\pi_*(x,y)\right)^{1/p} = \left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} |f(x)-f(y)|^p d\pi(x,y)\right)^{1/p} \\ &\leq \|f\|_{\text{Lip}} \left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\pi(x,y)\right)^{1/p} \\ &= \|f\|_{\text{Lip}} W_p(\mu,\nu).\end{aligned}$$

To prove (A.4.6), we set $\pi := (f,g)\#\mu$. It is easy to check that $\pi \in \Pi(f\#\mu, g\#\mu)$ and therefore

$$W_p(f\#\mu, g\#\mu) \leq \left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\pi(x,y)\right)^{1/p} = \left(\int_{\mathbb{R}^d} |f(x)-g(x)|^p d\mu(x)\right)^{1/p} = \|f-g\|_{L^p(\mathbb{R}^d,\mu)}.$$

■

Corollary A.4.1 (W_p-Lipschitz continuity of p-moments)

Let μ, $\nu \in \mathcal{P}_p(\mathbb{R}^d)$ and let $f : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ be a Lipschitzian map. So we have

$$\left| \left(\int_{\mathbb{R}^d} |f(x)|^p d\mu(x) \right)^{1/p} - \left(\int_{\mathbb{R}^d} |f(x)|^p d\nu(x) \right)^{1/p} \right| \leq \|f\|_{\text{Lip}} W_p(\mu, \nu). \tag{A.4.7}$$

In other words, the map $\mu \in (\mathcal{P}_p(\mathbb{R}^d), W_p) \longmapsto \|f\|_{L^p(\mathbb{R}^d,\mu)}$ is Lipschitzian.

Proof. Using the classical inequality $|d(x,z) - d(y,z)| \leq d(x,y)$ for a distance d and (A.4.2), we have

$$\left| \left(\int_{\mathbb{R}^d} |f(x)|^p d\mu(x) \right)^{1/p} - \left(\int_{\mathbb{R}^d} |f(x)|^p d\nu(x) \right)^{1/p} \right| = |W_p(f\#\mu, \delta_0) - W_p(f\#\nu, \delta_0)| \leq W_p(f\#\mu, f\#\nu).$$

And by (A.4.5), we obtain (A.4.7). ■

An interesting property of the Wasserstein metric for $p = 2$ is that we can link $W_2(\mu_1, \nu_1)$ with $W_2(\mu_2, \nu_2)$ if μ_2 is a translation of μ_1 and ν_2 a translation of ν_1.

Proposition A.4.5 (Translation)

Let $\mu \in \mathcal{P}_2^{m_\mu}(\mathbb{R}^d)$ and $\nu \in \mathcal{P}_2^{m_\nu}(\mathbb{R}^d)$. Then for any u, $v \in \mathbb{R}^d$ we have

$$W_2(\tau_u\#\mu, \tau_v\#\nu)^2 = W_2(\mu, \nu)^2 + 2\langle m_\mu - m_\nu, u - v\rangle + |u - v|^2. \tag{A.4.8}$$

Proof. Let $\pi_* := (\tau_u, \tau_v)\#\pi$ with $\pi \in \Pi(\mu, \nu)$. By direct computation we have

$$\begin{aligned} \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi_*(x,y) &= \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x + u - y - v|^2 d\pi(x,y) \\ &= \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi(x,y) + 2 \iint_{\mathbb{R}^d \times \mathbb{R}^d} \langle x - y, u - v\rangle d\pi(x,y) + |u - v|^2. \end{aligned}$$

Expanding the scalar product and using $\iint x_i d\pi(x,y) = m_\mu(i)$, $\iint y_i d\pi(x,y) = m_\nu(i)$, one obtains

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi_*(x,y) = \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi(x,y) + 2\langle m_\mu - m_\nu, u - v\rangle + |u - v|^2.$$

Since the map $\pi \longmapsto (\tau_u, \tau_v)\#\pi$ maps $\Pi(\mu, \nu)$ into $\Pi(\tau_u\#\mu, \tau_v\#\nu)$ and is bijective, then taking the infimum over π we have

$$\begin{aligned} \inf_{\pi \in \Pi(\mu,\nu)} \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi_*(x,y) &= \inf_{\pi \in \Pi(\tau_u\#\mu, \tau_v\#\nu)} \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi(x,y) \\ &= W_2(\mu, \nu)^2 + 2\langle m_\mu - m_\nu, u - v\rangle + |u - v|^2 \end{aligned}$$

which is (A.4.8). ■

Proposition A.4.5 involves two interesting special cases. The first is when $u = 0$ and $\nu = \mu$, this is Monge's problem in the particular case of translation for quadratic cost and we have

$$W_2(\mu, \tau_v\#\mu) = |v|. \tag{A.4.9}$$

One can easily show by the formula (A.3.10) that (A.4.9) remains true for W_1.

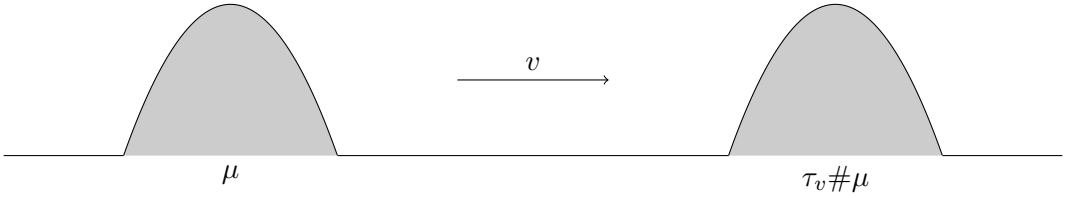


Figure A.4: Illustration of Monge's problem in the particular case of translation.

The second particular case is when $u = m_\mu$ and $v = m_\nu$, then Proposition A.4.5 implies that

$$W_2(\mu,\nu)^2 = W_2(\mu_0,\nu_0)^2 + |m_\mu - m_\nu|^2 \tag{A.4.10}$$

where μ_0 and ν_0 are the centered measures of zero mean of μ and ν. In other words $\mu_0 := \tau_{-m_\mu}\#\mu$ and $\nu_0 := \tau_{-m_\nu}\#\nu$. The proof taken from [42] of the following Proposition uses (A.4.10).

Proposition A.4.6 (Wasserstein metric between two Gaussians)

Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ with for density a Gaussian of mean vector m_μ and covariance matrix Σ_μ. And let $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ with for density a Gaussian of mean vector m_ν and covariance matrix Σ_ν. Then

$$W_2(\mu,\nu)^2 = |m_\mu - m_\nu|^2 + \mathrm{Tr}(\Sigma_\mu) + \mathrm{Tr}(\Sigma_\nu) - 2\mathrm{Tr}\left(\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}\right). \tag{A.4.11}$$

Proof. If m_μ, $m_\nu \neq 0$, then applying (A.4.10) we have

$$W_2(\mu,\nu)^2 = W_2(\tau_{-m_\mu}\#\mu, \tau_{-m_\nu}\#\nu)^2 + |m_\mu - m_\nu|^2.$$

Since $\tau_{-m_\mu}\#\mu$ is a Gaussian with zero mean and covariance matrix Σ_μ, and $\tau_{-m_\nu}\#\nu$ a Gaussian with zero mean and covariance matrix Σ_ν, it suffices to show (A.4.11) in the particular case where μ and ν are centered Gaussians. We therefore assume that $m_\mu = m_\nu = 0$. Let $\pi \in \Pi(\mu,\nu)$, we have

$$\begin{aligned}
\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\pi(x,y) &= \iint_{\mathbb{R}^d} |x|^2 d\mu(x) + \iint_{\mathbb{R}^d} |y|^2 d\nu(y) - 2\iint_{\mathbb{R}^d\times\mathbb{R}^d} \langle x,y\rangle d\pi(x,y) \\
&= \sum_{i=1}^d \left(\int_{\mathbb{R}^d} x_i^2 d\mu(x) + \int_{\mathbb{R}^d} y_i^2 d\nu(y) - 2\iint_{\mathbb{R}^d\times\mathbb{R}^d} x_i y_i d\pi(x,y)\right) \\
&= \mathrm{Tr}(\Sigma_1) + \mathrm{Tr}(\Sigma_2) - 2\mathrm{Tr}(C)
\end{aligned}$$

where $C = (c_{i,j})_{1\le i,j\le d}$ is a matrix defined by $c_{i,j} := \iint x_i y_j d\pi(x,y)$. The coupling π has for covariance matrix the block matrix Σ_π defined by

$$\Sigma_\pi := \begin{bmatrix} \Sigma_\mu & C \\ {}^tC & \Sigma_\nu \end{bmatrix}.$$

The main observation is that the matrix Σ_π is a symmetric positive semi-definite matrix. Indeed for any column vector u, $v \in \mathbb{R}^d$ we have

$$\begin{aligned}
Q(u,v) := \begin{bmatrix} {}^tu & {}^tv \end{bmatrix} \times \Sigma_\pi \times \begin{bmatrix} u \\ v \end{bmatrix} &= \sum_{i=1}^d\sum_{j=1}^d u_i u_j \Sigma_\mu(i,j) + 2u_i v_j c_{i,j} + v_i v_j \Sigma_\nu(i,j) \\
&= \iint_{\mathbb{R}^d\times\mathbb{R}^d} \left(\sum_{i=1}^d\sum_{j=1}^d u_i u_j x_i x_j + 2u_i v_j x_i y_j + v_i v_j y_i y_j\right) d\pi(x,y).
\end{aligned}$$

Using $(\sum a_i)^2 = \sum\sum a_i a_j$, one obtains

$$Q(u,v) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} \left|\sum_{i=1}^d x_i u_i + \sum_{i=1}^d y_i v_i\right|^2 d\pi(x,y).$$

Thus, $\Sigma_\pi \succeq 0$ but the quadratic form Q can be degenerate. Let $m \geq d$ be the rank of Q, there exists an orthogonal basis which diagonalizes it. So there exists P, $D \in M_{2d}(\mathbb{R})$ such that $\Sigma_\pi = {}^tPDP$ where $P = (p_{i,j})_{1\leq i,j\leq 2d}$ and $D := \mathrm{diag}(d_1,\cdots,d_m,0,\cdots,0)$ where for all i, $d_i \geq 0$. So for any column vector u, $v \in \mathbb{R}^d$ we have

$$Q(u,v) = \begin{bmatrix} {}^tu & {}^tv\end{bmatrix} \times {}^tPDP \times \begin{bmatrix} u \\ v\end{bmatrix} = \sum_{i=1}^m \left(\sum_{j=1}^d u_j\sqrt{d_i}p_{i,j} + \sum_{j=1}^d v_j\sqrt{d_i}p_{i,d+j}\right)^2.$$

For $1 \leq i \leq m$, we define the column vectors $a_i = (a_i(j))_{1\leq j\leq d}$, $b_i = (b_i(j))_{1\leq j\leq d}$ where $a_i(j) := \sqrt{d_i}p_{i,j}$ and $b_i(j) := \sqrt{d_i}p_{i,d+j}$. So we obtain for any column vector u, $v \in \mathbb{R}^d$

$$Q(u,v) = \sum_{i=1}^m |\,{}^ta_iu + {}^tb_iv|^2.$$

By expanding, we have on one side $Q(u,v) = {}^tu\Sigma_\mu u + 2\,{}^tuCv + {}^tv\Sigma_\nu v$ and on the other hand, since ${}^ta_iu \in \mathbb{R}$, then ${}^ta_iu = {}^tua_i$ and therefore we have $|\,{}^ta_iu + {}^tb_iv|^2 = {}^tua_i\,{}^ta_iu + {}^tvb_i\,{}^tb_iv + 2\,{}^tua_i\,{}^tb_iv$. We therefore obtain that for any column vector u, $v \in \mathbb{R}^d$

$${}^tu\Sigma_\mu u + 2\,{}^tuCv + {}^tv\Sigma_\nu v = \sum_{i=1}^m ({}^tua_i\,{}^ta_iu + {}^tvb_i\,{}^tb_iv + 2\,{}^tua_i\,{}^tb_iv).$$

By identification, we find $C = \sum a_i\,{}^tb_i$, $\Sigma_\mu = \sum a_i\,{}^ta_i$ and $\Sigma_\nu = \sum b_i\,{}^tb_i$. Calculating $W_2(\mu,\nu)$ consists in finding the maximum of the function $F : (a,b) \longmapsto \mathrm{Tr}\left(\sum a_i\,{}^tb_i\right)$ under the constraints $\sum a_i\,{}^ta_i = \Sigma_\mu$ and $\sum b_i\,{}^tb_i = \Sigma_\mu$. The set of constraints being compact, the maximum is therefore attained. For (i,j), we define the functions $g^1_{i,j} : (a,b) \longmapsto \sum a_i\,{}^ta_i - \Sigma_\mu(i,j)$ and $g^2_{i,j} : (a,b) \longmapsto \sum b_i\,{}^tb_i - \Sigma_\nu(i,j)$. Differentiate F, $g^1_{i,j}$, $g^2_{i,j}$, we have

$$\begin{aligned}
D_{(a,b)}F(h,k) &= \sum_{i=1}^m\sum_{j=1}^d b_i(j)h_i(j) + \sum_{i=1}^m\sum_{j=1}^d a_i(j)k_i(j) = \mathrm{Tr}\left(\sum_{i=1}^m (b_i\,{}^th_i + a_i\,{}^tk_i)\right),\\
D_{(a,b)}g^1_{i,j}(h,k) &= \sum_{l=1}^m (a_l(j)h_l(i) + a_l(i)h_l(j)),\\
D_{(a,b)}g^2_{i,j}(h,k) &= \sum_{l=1}^m (b_l(j)k_l(i) + b_l(i)k_l(j)).
\end{aligned}$$

So by the Lagrange multipliers Theorem there exists real constants $\lambda^1_{i,j}$, $\lambda^2_{i,j}$, $1 \leq i,j \leq d$ such that for any (h,k)

$$D_{(a,b)}F(h,k) = \sum_{i=1}^d\sum_{j=1}^d D_{(a,b)}g^1_{i,j}(h,k)\lambda^1_{i,j} + \sum_{i=1}^d\sum_{j=1}^d D_{(a,b)}g^2_{i,j}(h,k)\lambda^2_{i,j}.$$

Let $\Lambda_1 \in M_d(\mathbb{R})$ be the matrix formed by $\lambda^1_{i,j}$ and $\Lambda_2 \in M_d(\mathbb{R})$ the matrix formed by $\lambda^2_{i,j}$. Using trace

properties, $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$, $\mathrm{Tr}({}^tA) = \mathrm{Tr}(A)$ and $\mathrm{Tr}(A\,{}^tB) = \sum\sum a_{i,j}b_{i,j}$, we obtains

$$\begin{aligned}\mathrm{Tr}\left(\sum_{i=1}^m (b_i\,{}^th_i + a_i\,{}^tk_i)\right) &= \sum_{i=1}^d\sum_{j=1}^d\sum_{l=1}^m (a_l(j)h_l(i) + a_l(i)h_l(j))\lambda_{i,j}^1 \\ &\quad + \sum_{i=1}^d\sum_{j=1}^d\sum_{l=1}^m (b_l(j)k_l(i) + b_l(i)k_l(j))\lambda_{i,j}^2 \\ &= \mathrm{Tr}\left(\sum_{i=1}^m (h_i\,{}^ta_i + a_i\,{}^th_i)\,{}^t\Lambda_1\right) + \mathrm{Tr}\left(\sum_{i=1}^m (k_i\,{}^tb_i + b_i\,{}^tk_i)\,{}^t\Lambda_2\right) \\ &= \mathrm{Tr}\left(\sum_{i=1}^m a_i\,{}^th_i(\Lambda_1 + {}^t\Lambda_1)\right) + \mathrm{Tr}\left(\sum_{i=1}^m b_i\,{}^tk_i(\Lambda_2 + {}^t\Lambda_2)\right).\end{aligned}$$

By setting $M_1 := \Lambda_1 + {}^t\Lambda_1$ and $M_2 := \Lambda_2 + {}^t\Lambda_2$, we find by identification $b_i = M_1a_i$ and $a_i = M_2b_i$. Since M_1 and M_2 are symmetric, it is easy to check that $M_1\Sigma_\mu M_1 = \Sigma_2$ and $M_2\Sigma_\nu M_2 = \Sigma_1$. So that implies that

$$\mathrm{Tr}(C) = \mathrm{Tr}\left(\sum_{i=1}^m a_i\,{}^tb_i\right) = \mathrm{Tr}\left(\sum_{i=1}^m a_i\,{}^ta_iM_1\right) = \mathrm{Tr}(\Sigma_\mu M_1) = \mathrm{Tr}(\Sigma_\mu^{1/2}M_1\Sigma_\mu^{1/2}).$$

We set $R := \Sigma_\mu^{1/2}M_1\Sigma_\mu^{1/2}$. It is easy to check that $R^2 = \Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}$. So R^2 is a symmetric positive definite matrix since Σ_μ, $\Sigma_\nu \succ 0$. But R is not necessarily the square root matrix of R^2. Let $\lambda_1, \cdots, \lambda_d$ be the positive eigenvalues of R^2. So we have

$$\mathrm{Tr}(C) = \mathrm{Tr}(R) = \sum_{i=1}^d \varepsilon_i\sqrt{\lambda_i}$$

where for all i, $\varepsilon_i = \pm 1$. Thus, we therefore have

$$\begin{aligned}\max\left\{\mathrm{Tr}\left(\sum_{i=1}^m a_i\,{}^tb_i\right)\colon \sum_{i=1}^m a_i\,{}^ta_i = \Sigma_\mu \text{ et } \sum_{i=1}^m b_i\,{}^tb_i = \Sigma_\nu\right\} &= \max\left\{\sum_{i=1}^d \varepsilon_i\sqrt{\lambda_i}\colon\ \varepsilon_i = \pm 1\right\} = \sum_{i=1}^d \sqrt{\lambda_i} \\ &= \mathrm{Tr}\left(\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}\right).\end{aligned}$$

We deduce that for any coupling $\pi \in \Pi(\mu,\nu)$

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\pi(x,y) \geq \mathrm{Tr}(\Sigma_\mu) + \mathrm{Tr}(\Sigma_\nu) - 2\mathrm{Tr}\left(\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}\right).$$

Let $A := \Sigma_\mu^{-1/2}(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\Sigma_\mu^{-1/2}$, we define the coupling $\pi_* := (\mathrm{id}, T)\#\mu$ where $T : x \in \mathbb{R}^d \longmapsto Ax \in \mathbb{R}^d$. We recall that if a probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ has for density a Gaussian with mean vector m and covariance matrix Σ, then the image measure of μ by the map $x \longmapsto Ax + b$ has for density a Gaussian with mean vector $Am + b$ and covariance matrix $A\Sigma\,{}^tA$. We also recall that if M is a symmetric positive definite matrix, then its inverse M^{-1} and its square root $M^{1/2}$ are also symmetric positive definite matrices. By symmetry of the matrix $\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}$ we have

$${}^tA = {}^t\Sigma_\mu{}^{-1/2}\,{}^t\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}{}^t\Sigma_\mu{}^{-1/2} = \Sigma_\mu^{-1/2}\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}\Sigma_\mu^{-1/2} = A.$$

So A is symmetric. By direct computation, we have

$$\begin{aligned} A\Sigma_\mu {}^tA &= \Sigma_\mu^{-1/2}\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}\Sigma_\mu^{-1/2}\Sigma_\mu\Sigma_\mu^{-1/2}\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}\Sigma_\mu^{-1/2} \\ &= \Sigma_\mu^{-1/2}\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}\Sigma_\mu^{-1/2} \\ &= \Sigma_\mu^{-1/2}\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\Sigma_\mu^{-1/2} \\ &= \Sigma_\nu. \end{aligned}$$

We have $T\#\mu = \nu$ and

$$\begin{aligned} \iint_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2d\pi_*(x,y) &= \int_{\mathbb{R}^d}|x-Ax|^2d\mu(x) \\ &= \int_{\mathbb{R}^d}|x|^2d\mu(x)+\int_{\mathbb{R}^d}|Ax|^2d\mu(x)-2\int_{\mathbb{R}^d}\langle x,Ax\rangle d\mu(x) \\ &= \mathrm{Tr}(\Sigma_\mu)+\mathrm{Tr}(\Sigma_\nu)-2\int_{\mathbb{R}^d}\langle x,Ax\rangle d\mu(x). \end{aligned}$$

Using the equality $\int\langle x,Ax\rangle d\mu(x) = \mathrm{Tr}(\Sigma_\mu A)$, one obtains

$$\begin{aligned} \iint_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2d\pi_*(x,y) &= \mathrm{Tr}(\Sigma_\mu)+\mathrm{Tr}(\Sigma_\nu)-2\mathrm{Tr}(\Sigma_\mu A) \\ &= \mathrm{Tr}(\Sigma_\mu)+\mathrm{Tr}(\Sigma_\nu)-2\mathrm{Tr}\left(\Sigma_\mu^{1/2}\Sigma_\mu^{1/2}\Sigma_\mu^{-1/2}\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}\Sigma_\mu^{-1/2}\right) \\ &= \mathrm{Tr}(\Sigma_\mu)+\mathrm{Tr}(\Sigma_\nu)-2\mathrm{Tr}\left(\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}\right). \end{aligned}$$

So π_* is optimal. ■

In the particular case where $d = 1$, the proof is simplified. If $\mu \in \mathcal{P}_2(\mathbb{R})$ has for density a Gaussian with zero mean and variance σ_μ^2 and if $\nu \in \mathcal{P}_2(\mathbb{R})$ has for density a Gaussian with zero mean and variance σ_ν^2, then for any coupling $\pi \in \Pi(\mu,\nu)$ we have

$$\iint_{\mathbb{R}\times\mathbb{R}}|x-y|^2d\pi(x,y) = \sigma_\mu^2+\sigma_\nu^2-2\iint_{\mathbb{R}\times\mathbb{R}}xyd\pi(x,y).$$

And by the Cauchy-Schwarz inequality, we obtain the lower bound

$$\iint_{\mathbb{R}\times\mathbb{R}}|x-y|^2d\pi(x,y) \geq \sigma_\mu^2+\sigma_\nu^2-2\left(\int_{\mathbb{R}}|x|^2d\pi(x,y)\right)^{1/2}\left(\int_{\mathbb{R}}|y|^2d\pi(x,y)\right)^{1/2} = (\sigma_\mu-\sigma_\nu)^2.$$

By defining the coupling $\pi_* := (\mathrm{id},T)\#\mu$ where $T : x \longmapsto \sigma_\nu x/\sigma_\mu$, it is easy to check with a change of variable that $T\#\mu = \nu$ and we obtain

$$\iint_{\mathbb{R}\times\mathbb{R}}|x-y|^2d\pi_*(x,y) = \int_{\mathbb{R}}|x-T(x)|^2d\mu(x) = \left(1-\frac{\sigma_\nu}{\sigma_\mu}\right)^2\int_{\mathbb{R}}|x|^2d\mu(x) = (\sigma_\mu-\sigma_\nu)^2.$$

So π_* is optimal.

Proposition A.4.7 (Convexity)

Given μ_1, μ_2, ν_1 and ν_2 in $\mathcal{P}_p(\mathbb{R}^d)$ and $t \in [0,1]$, then

$$W_p(t\mu_1+(1-t)\mu_2, t\nu_1+(1-t)\nu_2)^p \leq tW_p(\mu_1,\nu_1)^p+(1-t)W_p(\mu_2,\nu_2)^p \qquad \text{(A.4.12)}$$

Proof. Let $\pi_1 \in \Pi(\mu_1, \nu_1)$ and $\pi_2 \in \Pi(\mu_2, \nu_2)$ be two optimal couplings. For $t \in [0,1]$, we set $\pi := t\pi_1 + (1-t)\pi_2$. It is easy to check that $\pi \in \Pi(t\mu_1 + (1-t)\mu_2, t\nu_1 + (1-t)\nu_2)$ and therefore

$$\begin{aligned} W_p(t\mu_1 + (1-t)\mu_2, t\nu_1 + (1-t)\nu_2)^p &\leq \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p d\pi(x,y) \\ &= t \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p d\pi_1(x,y) + (1-t) \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p d\pi_2(x,y) \\ &= tW_p(\mu_1, \nu_1)^p + (1-t)W_p(\mu_2, \nu_2)^p. \end{aligned}$$

■

Proposition A.4.8 (Sub-additivity with respect to convolution)

Given μ_1, $\nu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ have the same mean and μ_2, $\nu_2 \in \mathcal{P}_2(\mathbb{R}^d)$. Then

$$W_2(\mu_1 * \mu_2, \nu_1 * \nu_2)^2 \leq W_2(\mu_1, \nu_1)^2 + W_2(\mu_2, \nu_2)^2. \tag{A.4.13}$$

Proof. Let $\pi_1 \in \Pi(\mu_1, \nu_1)$ and $\pi_2 \in \Pi(\mu_2, \nu_2)$ be two optimal couplings. We set $\pi := \pi_1 * \pi_2$. It is easy to check that $\pi \in \Pi(\mu_1 * \mu_2, \nu_1 * \nu_2)$ and therefore

$$\begin{aligned} W_2(\mu_1 * \mu_2, \nu_1 * \nu_2)^2 &\leq \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 d\pi(x,y) \\ &= \iiiint_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} |x+u-y-v|^2 d\pi_1(x,y) d\pi_2(u,v). \end{aligned}$$

Using the classical equality $|x+y|^2 = |x|^2 + |y|^2 + 2\langle x, y\rangle$, we have

$$\begin{aligned} W_2(\mu_1 * \mu_2, \nu_1 * \nu_2)^2 &\leq W_2(\mu_1, \nu_1)^2 + W_2(\mu_2, \nu_2)^2 \\ &\quad + 2 \iiiint_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \langle x-y, u-v\rangle d\pi_1(x,y) d\pi_2(u,v). \end{aligned}$$

And by using the equality $\iint \langle x, y\rangle d\mu(x) d\nu(y) = \langle \int x d\mu(x), \int y d\nu(y)\rangle$, we therefore have that the second term on the right-hand side is equal to 0 since μ_1 and ν_1 have the same mean. ■

Inequality (A.4.13) is also true for W_1. But in this case, the condition of equality of means between μ_1 and ν_1 is not necessary. If $\pi_1 \in \Pi(\mu_1, \nu_1)$ and $\pi_2 \in \Pi(\mu_2, \nu_2)$ are two optimal couplings, then still setting $\pi := \pi_1 * \pi_2$, we have

$$\begin{aligned} W_1(\mu_1 * \mu_2, \nu_1 * \nu_2) &\leq \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x-y| d\pi(x,y) \\ &= \iiiint_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} |x+u-y-v| d\pi_1(x,y) d\pi_2(u,v). \end{aligned}$$

And by the triangle inequality,

$$\begin{aligned} W_1(\mu_1 * \mu_2, \nu_1 * \nu_2) &\leq \iiiint_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} (|x-y| + |u-v|) d\pi_1(x,y) d\pi_2(u,v) \\ &= \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x-y| d\pi_1(x,y) + \iint_{\mathbb{R}^d \times \mathbb{R}^d} |u-v| d\pi_2(u,v) \\ &= W_1(\mu_1, \nu_1) + W_1(\mu_2, \nu_2). \end{aligned}$$

For $p \in \{1, 2\}$, if μ_1, μ_2, $\nu \in \mathcal{P}_p(\mathbb{R}^d)$, then we obtain by (A.4.13)

$$W_p(\mu_1 * \nu, \mu_2 * \nu) \leq W_p(\mu_1, \mu_2). \tag{A.4.14}$$

A.5 Convergence in Wasserstein sense

In this section, we give a characterization so that a sequence of measures $(\mu_n)_n$ in $\mathcal{P}_p(\mathbb{R}^d)$ converges to $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ for the Wasserstein metric W_p. We recall before the different types of convergence for the measures. Let $(\mu_n)_n$ be a sequence of measures on $\mathbb{R}^d$ and μ a measure on $\mathbb{R}^d$. We consider a test function φ satisfying

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} \varphi(x) d\mu_n(x) = \int_{\mathbb{R}^d} \varphi(x) d\mu(x). \tag{A.5.1}$$

Definition A.5.1

We say that $(\mu_n)_n$ converges

(1) **Vaguely** to μ if (A.5.1) is satisfied for all $\varphi \in \mathcal{C}_c(\mathbb{R}^d)$.

(2) **Weakly** to μ if (A.5.1) is satisfied for all $\varphi \in \mathcal{C}_0(\mathbb{R}^d)$.

(3) **Narrowly** to μ if (A.5.1) is satisfied for all $\varphi \in \mathcal{C}_b(\mathbb{R}^d)$.

Since we have the obvious inclusions $\mathcal{C}_c(\mathbb{R}^d) \subset \mathcal{C}_0(\mathbb{R}^d) \subset \mathcal{C}_b(\mathbb{R}^d)$, then narrow convergence implies weak convergence and weak convergence implies vague convergence.

Proposition A.5.1

If $(\mu_n)_n$ is a sequence of measures which converges vaguely to μ such that $\mu_n(\mathbb{R}^d) \longrightarrow \mu(\mathbb{R}^d)$ when n goes to infinity, then $(\mu_n)_n$ converges narrowly to μ.

Proof. Let $\varphi \in \mathcal{C}_b(\mathbb{R}^d)$ and let $\psi : \mathbb{R}^d \longrightarrow [0,1]$ be a continuous function. We have

$$\begin{aligned}
&\left|\int_{\mathbb{R}^d} \varphi(x) d\mu_n(x) - \int_{\mathbb{R}^d} \varphi(x) d\mu(x)\right| \leq \int_{\mathbb{R}^d} |\varphi(x) - \varphi(x)\psi(x)| d\mu_n(x) \\
&+ \left|\int_{\mathbb{R}^d} \varphi(x)\psi(x) d\mu_n(x) - \int_{\mathbb{R}^d} \varphi(x)\psi(x) d\mu(x)\right| + \int_{\mathbb{R}^d} |\varphi(x)\psi(x) - \varphi(x)| d\mu(x) \\
&\leq \|\varphi\|_\infty \int_{\mathbb{R}^d} (1 - \psi(x)) d\mu_n(x) + \|\varphi\psi\|_\infty |\mu_n(\mathbb{R}^d) - \mu(\mathbb{R}^d)| + \|\varphi\|_\infty \int_{\mathbb{R}^d} (1 - \psi(x)) d\mu(x).
\end{aligned}$$

The second term tends to 0 by hypothesis and the first term tends to the third by vague convergence. So we have

$$\limsup_{n\to+\infty} \left|\int_{\mathbb{R}^d} \varphi(x) d\mu_n(x) - \int_{\mathbb{R}^d} \varphi(x) d\mu(x)\right| \leq 2\|\varphi\|_\infty \int_{\mathbb{R}^d} (1 - \psi(x)) d\mu(x).$$

The right-hand side can be made as small as we want. Because there exists an increasing sequence of functions $(\psi_k)_k$ in $\mathcal{C}_c(\mathbb{R}^d)$ such that for all k, $\psi_k(x) \in [0,1]$ and which converges to the constant function 1. We have by Lebesgue's dominated Theorem

$$\lim_{k\to+\infty} \int_{\mathbb{R}^d} (1 - \psi_k(x)) d\mu(x) = \mu(\mathbb{R}^d) - \lim_{k\to+\infty} \int_{\mathbb{R}^d} \psi_k(x) d\mu(x) = 0.$$

■

We have immediately that if $(\mu_n)_n$ is a sequence of probability measures which converges to a probability measure μ, then the three convergence are equivalent. We will say in this particular case that $(\mu_n)_n$ converges weakly to μ. A first result concerns the weak convergence of a product of probability measures.

Lemma A.5.1

Let $(\mu_n)_n$ be a sequence of probability measure on $\mathbb{R}^d$ which converges weakly to $\mu \in \mathcal{P}(\mathbb{R}^d)$ and let $(\nu_n)_n$ be a sequence of probability measure on $\mathbb{R}^d$ which converges weakly to $\nu \in \mathcal{P}(\mathbb{R}^d)$. Then the sequence $(\mu \otimes \nu)_n$ converges weakly to $\mu \otimes \nu$.

This Lemma is shown trivially with the Levy's Continuity Theorem. We assume that μ_n is the law of a random variable X_n and that ν_n is the law of a random variable Y_n. If X_n converges in distribution to X, and if Y_n converges in distribution to Y, then the independent random couple (X_n, Y_n) converges in distribution to the independent random couple (X, Y). Indeed, by using the characteristic function φ_n of (X_n, Y_n) and the characteristic function φ of (X, Y), we have for all t, $s \in \mathbb{R}^d$

$$\begin{aligned}
\lim_{n\to+\infty} \varphi_n(t,s) &= \lim_{n\to+\infty} \iint_{\mathbb{R}^d\times\mathbb{R}^d} e^{i\langle t,x\rangle + i\langle s,y\rangle} d\mu_n(x) d\nu_n(y) \\
&= \left(\lim_{n\to+\infty} \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} d\mu_n(x)\right)\left(\lim_{n\to+\infty} \int_{\mathbb{R}^d} e^{i\langle s,y\rangle} d\nu_n(y)\right) \\
&= \iint_{\mathbb{R}^d\times\mathbb{R}^d} e^{i\langle t,x\rangle + i\langle s,y\rangle} d\mu(x) d\nu(y) \\
&= \varphi(t,s).
\end{aligned}$$

We propose a rigorous proof of the Lemma A.5.1 without using the Levy's Continuity Theorem.

Proof of Lemma A.5.1. Let X and Y be two compacts on $\mathbb{R}^d$. We will show that the space E of finite sums of functions $(x,y) \longmapsto \varphi(x)\psi(y)$ where $\varphi \in \mathcal{C}(X)$, $\psi \in \mathcal{C}(Y)$ is dense in $\mathcal{C}(X \times Y)$. Let f and g belong to E and let λ, $\mu \in \mathbb{R}$. We have

$$\lambda f(x,y) + \mu g(x,y) = \lambda \sum_{i=1}^{n} \varphi_i^1(x)\psi_i^1(y) + \mu \sum_{i=1}^{m} \varphi_i^2(x)\psi_i^2(y) = \sum_{i=1}^{n+m} \varphi_i(x)\psi_i(y)$$

with $\varphi_i = \lambda\varphi_i^1$, $\psi_i = \psi_i^1$ if $i \leq n$ and $\varphi_i = \mu\varphi_{i-n}^2$, $\psi_i = \psi_{i-n}^2$ if $i > n$. So E is a linear subspace of $\mathcal{C}^0(X \times Y)$. Then, for f and g in E, we have

$$f(x,y)g(x,y) = \left(\sum_{i=1}^{n} \varphi_i^1(x)\psi_i^1(y)\right)\left(\sum_{i=1}^{m} \varphi_i^2(x)\psi_i^2(y)\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \varphi_i^1(x)\varphi_j^2(x)\psi_i^1(y)\psi_j^2(y).$$

Let σ be a bijection of $\{1, \cdots, n\} \times \{1, \cdots, m\}$ in $\{1, \cdots, nm\}$. Then by setting $\varphi_l(x) := \varphi_i^1(x)\varphi_j^2(x)$ and $\psi_l(y) := \psi_i^1(y)\psi_j^2(y)$ where $l = \sigma(i,j)$, we have

$$f(x,y)g(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{m} \varphi_i^1(x)\varphi_j^2(x)\psi_i^1(y)\psi_j^2(y) = \sum_{l=1}^{nm} \varphi_l(x)\psi_l(y).$$

Therefore, E is closed by multiplication. And then, it is clear that the constant function equal to 1 belongs to E and if (x_1, y_1), (x_2, y_2) are two points in $X \times Y$ such that $(x_1, y_1) \neq (x_2, y_2)$, we have for example if $x_1 \neq x_2$ that the function $f(x,y) := |x_1 - x|$ entails that $f(x_1, y_1) \neq f(x_2, y_2)$. Thus, E separates points so by Stone-Weierstrass Theorem, there exists for any $\varphi \in \mathcal{C}(X \times Y)$ a sequence of functions in E which converges uniformly to φ.

Let's now to show the weak convergence. Let $\varphi \in \mathcal{C}_c(\mathbb{R}^d \times \mathbb{R}^d)$, then there exists $R > 0$ such that $\text{supp}(\varphi) \subset \overline{B(0,R)} \times \overline{B(0,R)}$ and $\mu(\{|x| = R\}) = \nu(\{|x| = R\}) = 0$. Let $\varepsilon > 0$, then there exists $f \in E$ such that for any $(x,y) \in \overline{B(0,R)} \times \overline{B(0,R)}$, $|\varphi(x,y) - f(x,y)| < \varepsilon$. For $\rho \in]0, R[$, we define the function $\chi_{\rho,R} : \mathbb{R}^d \longrightarrow [0,1]$ such that $\chi_{\rho,R}(x) = 1$ if $|x| \leq \rho$, $\chi_{\rho,R}(x) = 0$ if $|x| \geq R$ and $\chi_{\rho,R} \in \mathcal{C}^\infty(\mathbb{R}^d)$.

By the Lebesgue's dominated Theorem, we can fix $\rho \in]0, R[$ such that $\mu(\{\rho \leq |x| \leq R\}) < \varepsilon$ and $\nu(\{\rho \leq |y| \leq R\}) < \varepsilon$. Then, since $(\mu_n)_n$ converges weakly to μ and $(\nu_n)_n$ converges weakly to ν, then by the Portremanteau Theorem, we have $\limsup \mu_n(\{\rho \leq |x| \leq R\}) \leq \mu(\{\rho \leq |x| \leq R\})$ and $\limsup \nu_n(\{\rho \leq |y| \leq R\}) \leq \nu(\{\rho \leq |y| \leq R\})$. So there exists $n_0 \in \mathbb{N}$ such that for all $n > n_0$, $\mu_n(\{\rho \leq |x| \leq R\}) < \varepsilon$ and $\nu_n(\{\rho \leq |y| \leq R\}) < \varepsilon$. By setting

$$u_n := \left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x,y) d\mu_n(x) d\nu_n(y) - \iint_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x,y) d\mu(x) d\nu(y) \right|,$$

we have

$$\begin{aligned} u_n \leq \|\varphi - f\|_\infty & \left| \iint_{\overline{B(0,R)}^2} \chi_{\rho,R}(x)\chi_{\rho,R}(y) d\mu_n(x) d\nu_n(y) - \iint_{\overline{B(0,R)}^2} \chi_{\rho,R}(x)\chi_{\rho,R}(y) d\mu(x) d\nu(y) \right| \\ & + \left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} f(x,y)\chi_{\rho,R}(x)\chi_{\rho,R}(y) d\mu_n(x) d\nu_n(y) - \iint_{\mathbb{R}^d \times \mathbb{R}^d} f(x,y)\chi_{\rho,R}(x)\chi_{\rho,R}(y) d\mu(x) d\nu(y) \right| \\ & + \|\varphi\|_\infty \left| \iint_{\overline{B(0,R)}^2} (1 - \chi_{\rho,R}(x)\chi_{\rho,R}(y)) d\mu_n(x) d\nu_n(y) - \iint_{\overline{B(0,R)}^2} (1 - \chi_{\rho,R}(x)\chi_{\rho,R}(y)) d\mu(x) d\nu(y) \right|. \end{aligned}$$

Since f belongs to E, there exists functions φ_i, $\psi_i \in \mathcal{C}^0(\overline{B(0,R)})$, $1 \leq i \leq p$ such that $f(x,y) = \sum \varphi_i(x)\psi_i(y)$. By setting $\square := (\overline{B(0,R)} \times \overline{B(0,R)}) \backslash (B(0,\rho) \times B(0,\rho))$, we have

$$\begin{aligned} u_n \leq \varepsilon & \left| \iint_{\overline{B(0,R)}^2} \chi_{\rho,R}(x)\chi_{\rho,R}(y) d\mu_n(x) d\nu_n(y) - \iint_{\overline{B(0,R)}^2} \chi_{\rho,R}(x)\chi_{\rho,R}(y) d\mu(x) d\nu(y) \right| \\ & + \sum_{i=1}^{p} \left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi_i\chi_{\rho,R})(x)(\psi_i\chi_{\rho,R})(y) d\mu_n(x) d\nu_n(y) - \iint_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi_i\chi_{\rho,R})(x)(\psi_i\chi_{\rho,R})(y) d\mu(x) d\nu(y) \right| \\ & + \|\varphi\|_\infty \left| \iint_{\square} (1 - \chi_{\rho,R}(x)\chi_{\rho,R}(y)) d\mu_n(x) d\nu_n(y) - \iint_{\square} (1 - \chi_{\rho,R}(x)\chi_{\rho,R}(y)) d\mu(x) d\nu(y) \right|. \end{aligned}$$

For the second term on the right-hand side, the functions $\varphi_i\chi_{\rho,R}$, $\psi_i\chi_{\rho,R}$ are continuous with compact support on $\mathbb{R}^d$ so there exists $n_1 \in \mathbb{N}$ such that for all $n > n_1$

$$\left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi_i\chi_{\rho,R})(x)(\psi_i\chi_{\rho,R})(y) d\mu_n(x) d\nu_n(y) - \iint_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi_i\chi_{\rho,R})(x)(\psi_i\chi_{\rho,R})(y) d\mu(x) d\nu(y) \right| < \varepsilon.$$

Therefore if $n > \max(n_0, n_1)$, one obtains

$$\begin{aligned} u_n &\leq \varepsilon(\mu_n \otimes \nu_n(\mathbb{R}^d \times \mathbb{R}^d) + \mu \otimes \nu(\mathbb{R}^d \times \mathbb{R}^d) + p) + \|\varphi\|_\infty(\mu_n \otimes \nu_n(\square) + \mu \otimes \nu(\square)) \\ &\leq \|\varphi\|_\infty(\mu_n(\{\rho \leq |x| \leq R\}) + \nu_n(\{\rho \leq |y| \leq R\}) + \mu(\{\rho \leq |x| \leq R\}) + \nu(\{\rho \leq |y| \leq R\})) \\ &\quad + \varepsilon(p+2) \\ &\leq \varepsilon(2 + p + 4\|\varphi\|_\infty). \end{aligned}$$

■

The following Theorem is the main result of this part taken from [88] and [89].

Theorem A.5.1 (Convergence in Wasserstein sense)

Let $(\mu_n)_n$ be a sequence of probability measures in $\mathcal{P}_p(\mathbb{R}^d)$ and let $\mu \in \mathcal{P}_p(\mathbb{R}^d)$. The following statements are equivalent.

(1) $W_p(\mu_n, \mu) \longrightarrow 0$ when n goes to infinity.

(2) $(\mu_n)_n$ converges weakly to μ and $\lim\limits_{n \to +\infty} \int_{\mathbb{R}^d} |x|^p d\mu_n(x) = \int_{\mathbb{R}^d} |x|^p d\mu(x)$.

This Theorem gives us a characterization of convergence in the Wasserstein sense. Thus, a sequence of probability measures $(\mu_n)_n$ in $\mathcal{P}_p(\mathbb{R}^d)$ converges in the Wasserstein sense to $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ if and only if $(\mu_n)_n$ converges weakly to μ and the moments of order p of μ_n converge to the moment of order p of μ. To prove this Theorem (A.5.1), we will need some Lemmas.

Definition A.5.2

We say that a sequence of functions $(\rho_n)_n$ is an **approximate of identity** if it satisfies

(1) For all n, $\int \rho_n(x)dx = 1$.

(2) There exists a constant C such that $\sup_n \|\rho_n\|_{L^1} \leq C$.

(3) For all $R > 0$, $\lim_{n\to+\infty} \int_{|x|>R} |\rho_n(x)|dx = 0$.

Lemma A.5.2

Let $\varphi \in \mathcal{C}_c(\mathbb{R}^d)$. Then there exists a sequence of function $(\psi_n)_n$ in $\mathcal{C}_c^\infty(\mathbb{R}^d)$ which converges uniformly to φ.

Proof. Let $\varphi \in \mathcal{C}_c(\mathbb{R}^d)$ and let $(\rho_n)_n$ be an approximate of identity. We set $\psi_n := \varphi * \rho_n$. Then for $\delta > 0$, we have

$$\begin{aligned}\psi_n(x) - \varphi(x) &= \int_{\mathbb{R}^d} (\varphi(x-y) - \varphi(x))\rho_n(y)dy - \varphi(x)\left(1 - \int_{\mathbb{R}^d} \rho_n(y)dy\right) \\ &= \int_{|y|<\delta} (\varphi(x-y) - \varphi(x))\rho_n(y)dy + \int_{|y|\geq\delta} (\varphi(x-y) - \varphi(x))\rho_n(y)dy \\ &\quad - \varphi(x)\left(1 - \int_{\mathbb{R}^d} \rho_n(y)dy\right).\end{aligned}$$

Which implies

$$|\psi_n(x) - \varphi(x)| \leq \int_{|y|<\delta} |\varphi(x-y) - \varphi(x)| \cdot |\rho_n(y)|dy + 2\|\varphi\|_\infty \int_{|y|\geq\delta} |\rho_n(y)|dy + \|\varphi\|_\infty \left|1 - \int_{\mathbb{R}^d} \rho_n(y)dy\right|.$$

By the uniform continuity of φ, we obtain

$$\|\psi_n - \varphi\|_\infty \leq \varepsilon \int_{|y|<\delta} |\rho_n(y)|dy + 2\|\varphi\|_\infty \int_{|y|\geq\delta} |\rho_n(y)|dy + \|\varphi\|_\infty \left|1 - \int_{\mathbb{R}^d} \rho_n(y)dy\right|.$$

The three terms on the right-hand side can be made as small as we want by definition of the approximate of identity. It remains to show that $\psi_n \in \mathcal{C}_c^\infty(\mathbb{R}^d)$. If ρ is a function in $\mathcal{C}_c^\infty(\mathbb{R}^d)$ such that $\int \rho(x)dx = 1$, then $\rho_n : x \longmapsto \rho_n(x) = n^d\rho(nx)$ is an approximate of identity and for any n, $\rho_n \in \mathcal{C}_c^\infty(\mathbb{R}^d)$. We have by definition of the convolution that $\psi_n \in \mathcal{C}_c^\infty(\mathbb{R}^d)$. ■

Lemma A.5.3

Let $(f_n)_n$ be a sequence of continuous functions on $\mathbb{R}^d$ such that $f_n(0) = 0$ for all n. If there exists a constant M independant of n such that $|f_n(x) - f_n(y)| \leq M|x-y|$, then there exists a subsequence $(f_{\sigma(n)})_n$ which converges uniformly on compact sets to a M-Lipschitzian function f which satisfies $f(0) = 0$.

Proof. It suffices to show that $(f_n)_n$ converges uniformly for all closed balls with center 0 and radius $N \in \mathbb{N}^*$. Let $N \in \mathbb{N}^*$, then the family $(f_n \mathbb{1}_{\{|x|<N\}})_n$ is uniformly bounded and uniformly equicontinuous on the ball $B(0,N)$. Then by Ascoli's Theorem, we can extract a subsequence $(f_{\sigma(n)}\mathbb{1}_{\{|x|<N\}})_n$ which converges uniformly on $B(0,N)$ to a function f, and we have $\sigma = \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_N$ with $\sigma_i : \mathbb{N} \nearrow \mathbb{N}$. Moreover f satisfies $f(0) = 0$ by pointwise convergence and $|f(x)-f(y)| \leq 2\|f-f_n\|_\infty + |f_n(x)-f_n(y)|$, which implies that f is a M-Lipschitzian function. ■

Definition A.5.3

We say that a family $\mathscr{F}$ of probability measures on $\mathbb{R}^d$ is **tight** if for any $\varepsilon > 0$, there exists a compact K_ε such that $\mu(\mathbb{R}^d \backslash K_\varepsilon) < \varepsilon$ for all $\mu \in \mathscr{F}$. And we say that $\mathscr{F}$ is **tight for moment of order** p if for any $\varepsilon > 0$, there exists a compact K_ε such that

$$\int_{\mathbb{R}^d \backslash K_\varepsilon} |x|^p d\mu(x) < \varepsilon.$$

"Tightness" is a concept from measure theory, in which the intuitive idea is that a collection of probability measures cannot escape to infinity. The notion of "Tightness for moment of order p" is more stronger and we immediately notice that "Tight for moment of order $p = 0$" is "Tight". In the particular case where $\mathscr{F} = \{\mu_n,\, n \in \mathbb{N}\}$, then the family $\mathscr{F}$ is tight if $\sup_n \mu_n(\{|x| > R\})$ goes to 0 when R goes to infinity. And $\mathscr{F}$ is tight for moment of order p if

$$\lim_{R\to+\infty} \sup_n \int_{|x|>R} |x|^p d\mu_n(x) = 0.$$

We also notice that if $\mathscr{F}$ is tight for moment of order p, then the following inequality

$$\int_{|x|>R} |x|^q d\mu_n(x) \leq \frac{R^q}{R^p} \int_{|x|>R} |x|^p d\mu_n(x), \qquad q < p$$

immediately results in $\mathscr{F}$ is tight for moment of order q for all $q < p$. The tightness of the measures is related to the relative compactness in the space of the probability measures [71].

Theorem A.5.2 (Prokhorov Theorem)

Let $\mathcal{X}$ be a separable metric space. Then a family $\mathscr{F}$ of probability measures on $\mathcal{X}$ is relatively compact for the weak topology if and only if it is tight.

The following result makes the link between the convergence of the moments of order p of a sequence of probability measures $(\mu_n)_n$ and the tightness for the moments of order p of the family $\{\mu_n,\, n \in \mathbb{N}\}$.

Lemma A.5.4

Let $(\mu_n)_n$ be a sequence in $\mathcal{P}_p(\mathbb{R}^d)$ which converges weakly to $\mu \in \mathcal{P}_p(\mathbb{R}^d)$. Then

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} |x|^p d\mu_n(x) = \int_{\mathbb{R}^d} |x|^p d\mu(x)$$

if and only if $(\mu_n)_n$ is tight for moments of order p.

Proof. Let $R > 0$. We set $\chi_R : \mathbb{R}^d \longrightarrow [0,1]$ a continuous function such that $\chi_R(x) = 1$ for $|x| \leq R$ and $\chi_R(x) = 0$ for $|x| \geq R+1$. We assume that the moments of order p of μ_n converge to the moment of order p of μ. We have

$$\begin{aligned}\int_{|x|>R+1} |x|^p d\mu_n(x) &= \int_{\mathbb{R}^d} |x|^p d\mu_n(x) - \int_{|x|\leq R+1} |x|^p d\mu_n(x) \\ &= \int_{\mathbb{R}^d} |x|^p d\mu_n(x) - \int_{\mathbb{R}^d} |x|^p d\mu(x) + \int_{\mathbb{R}^d} |x|^p d\mu(x) - \int_{\mathbb{R}^d} |x|^p \mathbb{1}_{\{|x|\leq R+1\}}(x) d\mu_n(x) \\ &- \int_{\mathbb{R}^d} \chi_R(x)|x|^p d\mu(x) + \int_{\mathbb{R}^d} \chi_R(x)|x|^p d\mu(x).\end{aligned}$$

Since $-\mathbb{1}_{\{|x|\leq R+1\}} \leq -\chi_R$, we therefore have

$$\begin{aligned}\int_{|x|>R+1} |x|^p d\mu_n(x) \leq &\left| \int_{\mathbb{R}^d} |x|^p d\mu_n(x) - \int_{\mathbb{R}^d} |x|^p d\mu(x) \right| + \left| \int_{\mathbb{R}^d} \chi_R(x)|x|^p d\mu(x) - \int_{\mathbb{R}^d} \chi_R(x)|x|^p d\mu_n(x) \right| \\ &+ \int_{\mathbb{R}^d} (1 - \chi_R(x))|x|^p d\mu(x).\end{aligned}$$

The first term tends to 0 by convergence of moments, the second term tends to 0 by weak convergence and the third term tends to 0 when R goes to infinity by Lebesgue's dominated Theorem.

Conversely, we assume that $(\mu_n)_n$ is tight for moments of order p. We have

$$\begin{aligned}\left| \int_{\mathbb{R}^d} |x|^p d\mu_n(x) - \int_{\mathbb{R}^d} |x|^p d\mu(x) \right| &\leq \left| \int_{\mathbb{R}^d} \chi_R(x)|x|^p d\mu_n(x) - \int_{\mathbb{R}^d} \chi_R(x)|x|^p d\mu(x) \right| \\ &+ \left| \int_{\mathbb{R}^d} (1 - \chi_R(x))|x|^p d\mu_n(x) - \int_{\mathbb{R}^d} (1 - \chi_R(x))|x|^p d\mu(x) \right| \\ &\leq \left| \int_{\mathbb{R}^d} \chi_R(x)|x|^p d\mu_n(x) - \int_{\mathbb{R}^d} \chi_R(x)|x|^p d\mu(x) \right| \\ &+ \sup_n \int_{|x|>R} |x|^p d\mu_n(x) + \int_{\mathbb{R}^d} (1 - \chi_R(x))|x|^p d\mu(x).\end{aligned}$$

The first term tends to 0 by weak convergence, the second tends to 0 when R goes to infinity since $(\mu_n)_n$ is tight for moments of order p, and the third term tends to 0 when R goes to infinity by Lebesgue's dominated Theorem. ■

If the moments of order p of μ_n converge to the moment of order p of μ and that $(\mu_n)_n$ converge weakly to μ, then $(\mu_n)_n$ is tight. This comes from the inequality

$$\int_{|x|>R} d\mu_n(x) \leq \frac{1}{R^p} \int_{|x|>R} |x|^p d\mu_n(x).$$

Proof of Theorem A.5.1. (1) $\Rightarrow$ (2) We assume that $W_p(\mu_n, \mu) \longrightarrow 0$ when n goes to infinity. Let us show to begin with the convergence of moments of order p. By (A.4.7), we have

$$\left| \left(\int_{\mathbb{R}^d} |x|^p d\mu_n(x) \right)^{1/p} - \left(\int_{\mathbb{R}^d} |x|^p d\mu(x) \right)^{1/p} \right| \leq W_p(\mu_n, \mu)$$

which tends to 0 by hypothesis. We prove now weak convergence, let $\varphi \in \mathcal{C}_c(\mathbb{R}^d)$. By Lemma A.5.2,

there exists a sequence of functions $(\psi_k)_k \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ which converges uniformly to φ. Which implies

$$\begin{aligned}\left|\int_{\mathbb{R}^d}\varphi(x)d\mu_n(x)-\int_{\mathbb{R}^d}\varphi(x)d\mu(x)\right| &\leq \int_{\mathbb{R}^d}|\varphi(x)-\psi_k(x)|d\mu_n(x)+\int_{\mathbb{R}^d}|\varphi(x)-\psi_k(x)|d\mu(x)\\ &\quad+\left|\int_{\mathbb{R}^d}\psi_k(x)d\mu_n(x)-\int_{\mathbb{R}^d}\psi_k(x)d\mu(x)\right|\\ &\leq 2\|\varphi-\psi_k\|_\infty+\left|\int_{\mathbb{R}^d}\psi_k(x)d\mu_n(x)-\int_{\mathbb{R}^d}\psi_k(x)d\mu(x)\right|.\end{aligned}$$

Let $\pi_n \in \Pi(\mu_n,\mu)$ be an optimal coupling. For fixed k, we have

$$\begin{aligned}\left|\int_{\mathbb{R}^d}\psi_k(x)d\mu_n(x)-\int_{\mathbb{R}^d}\psi_k(x)d\mu(x)\right| &\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d}|\psi_k(x)-\psi_k(y)|d\pi_n(x,y)\\ &\leq \|D\psi_k\|_\infty\iint_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|d\pi_n(x,y)\\ &\leq \|D\psi_k\|_\infty\left(\iint_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^p d\pi_n(x,y)\right)^{1/p}\\ &= \|D\psi_k\|_\infty W_p(\mu_n,\mu)\end{aligned}$$

which goes to 0 when n goes to infinity. So for fixed k, we have

$$\limsup_{n\to+\infty}\left|\int_{\mathbb{R}^d}\varphi(x)d\mu_n(x)-\int_{\mathbb{R}^d}\varphi(x)d\mu(x)\right| \leq 2\|\varphi-\psi_k\|_\infty$$

which can be made as small as we want since $(\psi_k)_k$ converges uniformly to φ.

(2) $\Rightarrow$ (1) We assume that the moments of order p of μ_n converge to the moment of order p of μ and that $(\mu_n)_n$ converge weakly to μ. We first show that $W_1(\mu_n,\mu)$ goes to 0 when n goes to infinity. It is enough to show that for every subsequence of $(\mu_{\sigma(n)})_n$, $\sigma:\mathbb{N}\nearrow\mathbb{N}$, there exists a subsubsequence $\sigma':\mathbb{N}\nearrow\mathbb{N}$ such that $W_1(\mu_{\sigma\circ\sigma'(n)},\mu)$ converges to 0. To simplify the notations, we assume that $\sigma(n)=n$. By the dual representation of W_1 (A.3.10), there exists Lipschitzians functions φ_n such that

$$W_1(\mu_n,\mu) \leq \left|\int_{\mathbb{R}^d}\varphi_n(x)d\mu_n(x)-\int_{\mathbb{R}^d}\varphi_n(x)d\mu(x)\right|+\frac{1}{n}$$

with $\|\varphi_n\|_{\text{Lip}} \leq 1$. We may assume that $\varphi_n(0)=0$ (by considering $\varphi_n-\varphi_n(0)$ if necessary). By Lemma A.5.3, there exists a subsequence $(\varphi_{\sigma(n)})_n$ which converges uniformly on compact set to a 1-Lipschitzian function φ. Let $\chi_R:\mathbb{R}^d\longrightarrow[0,1]$, $R>0$ be a continuous function such that $\chi_R(x)=1$ for $|x|\leq R$ and $\chi_R(x)=0$ for $|x|\geq R+1$. Then

$$\int_{\mathbb{R}^d}\varphi_{\sigma(n)}(x)d\mu_{\sigma(n)}(x)=\int_{|x|>R}(1-\chi_R(x))\varphi_{\sigma(n)}(x)d\mu_{\sigma(n)}(x)+\int_{|x|\leq R+1}\chi_R(x)\varphi_{\sigma(n)}(x)d\mu_{\sigma(n)}(x).$$

Using the fact that $|\varphi_n(x)|\leq|x|$, we obtain

$$\begin{aligned}&\left|\int_{\mathbb{R}^d}\varphi_{\sigma(n)}(x)d\mu_{\sigma(n)}(x)-\int_{\mathbb{R}^d}\varphi_{\sigma(n)}(x)d\mu(x)\right| \leq \int_{|x|>R}|x|d\mu_{\sigma(n)}(x)+\int_{|x|>R}|x|d\mu(x)\\ &+\left|\int_{|x|\leq R+1}\chi_R(x)\varphi(x)d\mu_{\sigma(n)}(x)-\int_{|x|\leq R+1}\chi_R(x)\varphi(x)d\mu(x)\right|\\ &+\int_{|x|\leq R+1}|\chi_R(x)|\cdot|\varphi_{\sigma(n)}(x)-\varphi(x)|d\mu_{\sigma(n)}(x)+\int_{|x|\leq R+1}|\chi_R(x)|\cdot|\varphi_{\sigma(n)}(x)-\varphi(x)|d\mu(x).\end{aligned}$$

And so we have

$$\left|\int_{\mathbb{R}^d}\varphi_{\sigma(n)}(x)d\mu_{\sigma(n)}(x)-\int_{\mathbb{R}^d}\varphi_{\sigma(n)}(x)d\mu(x)\right|\leq\int_{|x|>R}|x|d\mu_{\sigma(n)}(x)+\int_{|x|>R}|x|d\mu(x)$$
$$+\left|\int_{|x|\leq R+1}\chi_R(x)\varphi(x)d\mu_{\sigma(n)}(x)-\int_{|x|\leq R+1}\chi_R(x)\varphi(x)d\mu(x)\right|+2\sup_{|x|\leq R+1}|\varphi_{\sigma(n)}(x)-\varphi(x)|.$$

The third term goes to 0 since $(\mu_{\sigma(n)})_n$ converges weakly to μ, and the fourth term goes to 0 since $(\varphi_{\sigma(n)})_n$ converges uniformly to φ on the closed ball $B(0,R+1)$. One obtains

$$\begin{aligned}\limsup_{n\to+\infty}W_1(\mu_{\sigma(n)},\mu)&\leq\sup_n\int_{|x|>R}|x|d\mu_{\sigma(n)}(x)+\int_{|x|>R}|x|d\mu(x)\\&\leq\sup_n\frac{1}{R^{p-1}}\int_{|x|>R}|x|^pd\mu_{\sigma(n)}(x)+\int_{|x|>R}|x|d\mu(x).\end{aligned}$$

The first term goes to 0 when R goes to infinity by the Lemma A.5.4, and the second term goes to 0 when R goes to infinity by the Lebesgue's dominated Theorem. By letting $R\longrightarrow+\infty$, one obtains $W_1(\mu_{\sigma(n)},\mu)\longrightarrow 0$. We now prove that $W_p(\mu_n,\mu)\longrightarrow 0$ when n goes to infinity. Let $\pi_n\in\Pi(\mu_n,\mu)$ be an optimal coupling for W_1, and $R>0$. Then

$$\begin{aligned}W_p(\mu_n,\mu)^p&\leq\iint_{\{|x|\leq R\}\cap\{|y|\leq R\}}|x-y|^pd\pi_n(x,y)+\iint_{|y|\leq R\leq|x|}|x-y|^pd\pi_n(x,y)\\&+\iint_{|x|\leq R\leq|y|}|x-y|^pd\pi_n(x,y)+\iint_{\{R\leq|x|\}\cap\{R\leq|y|\}}|x-y|^pd\pi_n(x,y)\\&\leq(2R)^{p-1}\iint_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|d\pi_n(x,y)+2^p\iint_{R\leq|x|}|x|^pd\pi_n(x,y)+2^p\iint_{R\leq|y|}|y|^pd\pi_n(x,y)\\&+2^{p-1}\iint_{\{R\leq|x|\}\cap\{R\leq|y|\}}(|x|^p+|y|^p)d\pi_n(x,y)\\&\leq(2R)^{p-1}W_1(\mu_n,\mu)+2^{p+1}\int_{R\leq|x|}|x|^pd\mu_n(x)+2^{p+1}\int_{R\leq|y|}|y|^pd\mu(y).\end{aligned}$$

So we have

$$\limsup_{n\to+\infty}W_p(\mu_n,\mu)^p\leq 2^{p+1}\left(\sup_n\int_{|x|\geq R}|x|^pd\mu_n(x)+\int_{|x|\geq R}|x|^pd\mu(x)\right).$$

By letting $R\longrightarrow+\infty$, the first term goes to 0 by the Lemma A.5.4 and the second term goes to 0 by Lebesgue's dominated Theorem. And so we have

$$\lim_{n\to+\infty}W_p(\mu_n,\mu)=0.$$

■

From this Theorem, we deduce immediately that if $(\mu_n)_n$ is a sequence in $\mathcal{P}_p(\mathbb{R}^d)$ which converges weakly to $\mu\in\mathcal{P}_p(\mathbb{R}^d)$. And if $(\nu_n)_n$ is a sequence in $\mathcal{P}_p(\mathbb{R}^d)$ which converges weakly to $\nu\in\mathcal{P}_p(\mathbb{R}^d)$, such that

$$\lim_{n\to+\infty}\int_{\mathbb{R}^d}|x|^pd\mu_n(x)=\int_{\mathbb{R}^d}|x|^pd\mu(x)\qquad\text{and}\qquad\lim_{n\to+\infty}\int_{\mathbb{R}^d}|x|^pd\nu_n(x)=\int_{\mathbb{R}^d}|x|^pd\nu(x).$$

Then we have

$$\lim_{n\to+\infty}W_p(\mu_n,\nu_n)=W_p(\mu,\nu).\tag{A.5.2}$$

In other words, if we consider that to converge in $\mathcal{P}_p(\mathbb{R}^d)$ means to converge weakly and to converge for moments of order p, then W_p is continuous on $\mathcal{P}_p(\mathbb{R}^d)$ for convergence in $\mathcal{P}_p(\mathbb{R}^d)$. Thus, W_p metrizes the weak topology on $\mathcal{P}_p(\mathbb{R}^d)$ (Definition 6.8 and Theorem 6.9 in Ref. [88]).

A.6 Topological properties of the Wasserstein space

The Wasserstein space of order p is the space $\mathcal{P}_p(\mathbb{R}^d)$ equipped with the Wasserstein metric W_p. It is therefore a metric space. In this section, we will see that the topological properties of the Wasserstein space are very similar to the topological properties of the space $L^p(\mathbb{R}^d)$. The first property concerns the completeness and the separability.

Proposition A.6.1

The Wasserstein space $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ is a complete separable metric space.

Another way to state this Proposition is to say that the Wasserstein space $\mathcal{P}_p(\mathbb{R}^d)$ is a Polish space. In general, if $\mathcal{X}$ is a Polish space, then the Wasserstein space $\mathcal{P}_p(\mathcal{X})$ is also a Polish space (Proposition 7.1.5 in Ref. [3] and Theorem 6.18 in Ref. [88]). We will only focus on the case $\mathcal{X} = \mathbb{R}^d$. The proof of separability uses the fact that one can control the Wasserstein metric by the weighted total variation (Theorem 6.15 in Ref [88]).

Proposition A.6.2 (Control of W_p by weighted total variation)

Let μ, $\nu \in \mathcal{P}_p(\mathbb{R}^d)$. Then

$$W_p(\mu,\nu) \leq 2^{1/q} \left(\int_{\mathbb{R}^d} |x|^p d|\mu-\nu|(x) \right)^{1/p} \tag{A.6.1}$$

with $1/p + 1/q = 1$.

Proof. Let $a := (\mu-\nu)_+(\mathbb{R}^d)$ (see Definition A.3.3 for the positive part of a signed measure). Is it easy to check that $a = (\mu-\nu)_-(\mathbb{R}^d)$. We define the coupling π by $\pi := (\mathrm{id},\mathrm{id})\#\min(\mu,\nu) + ((\mu-\nu)_+ \otimes (\mu-\nu)_-)/a$ where $\min(\mu,\nu)$ is defined by (A.3.8). In other words we have for any test function φ

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x,y) d\pi(x,y) := \int_{\mathbb{R}^d} \varphi(x,x) d(\min(\mu,\nu))(x) + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \frac{\varphi(x,y)}{a} d((\mu-\nu)_+)(x) d((\mu-\nu)_-)(y).$$

By easily checking that $\min(\mu,\nu) = \nu - (\mu-\nu)_-$, we have that π is a coupling of (μ,ν) and by the classical convexity inequality $(x+y)^p \leq 2^{p-1}(x^p+y^p)$ we have

$$\begin{aligned}
W_p(\mu,\nu)^p &\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\pi(x,y) \\
&\leq \frac{2^{p-1}}{a} \iint_{\mathbb{R}^d\times\mathbb{R}^d} (|x|^p + |y|^p) d((\mu-\nu)_+)(x) d((\mu-\nu)_-)(y) \\
&= 2^{p-1} \left(\int_{\mathbb{R}^d} |x|^p d((\mu-\nu)_+)(x) + \int_{\mathbb{R}^d} |x|^p d((\mu-\nu)_-)(x) \right) \\
&= 2^{p-1} \int_{\mathbb{R}^d} |x|^p d|\mu-\nu|(x)
\end{aligned}$$

and (A.6.1) follows. ■

And to prove the completeness, we will use the following result taken from [36].

Theorem A.6.1 (The Kolmogorov extension Theorem)

Let $(X_n)_n$ be a sequence of Polish spaces. For each $n \in \mathbb{N}$, let $Y_n := \prod_{i=1}^n X_i = Y_{n-1} \times X_n$ and let $\mu_n \in \mathcal{P}(Y_n)$. If we have for all $n \in \mathbb{N}$, $(\text{proj}_1, \cdots, \text{proj}_n)\#\mu_{n+1} = \mu_n$, then there exists $\mu \in \mathcal{P}(Y^{\mathbb{N}})$ such that for any $n \in \mathbb{N}$, $(\text{proj}_1, \cdots, \text{proj}_n)\#\mu = \mu_n$.

Proof of Proposition A.6.1. We start by showing the separability. Let $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ and let $\varepsilon > 0$. There exists a positive integer N such that

$$\int_{|x|>N} |x|^p d\mu(x) < \varepsilon.$$

We define the projection π_N on the closed ball $\overline{B(0,N)}$, $\pi_N : x \longmapsto x$ if $|x| \leq N$ and $Nx/|x|$ otherwise. By setting $\mu_N := \pi_N\#\mu$, then $(\text{id}, \pi_N)\#\mu$ is a coupling of (μ, μ_N) and we have

$$W_p(\mu, \mu_N)^p \leq \int_{\mathbb{R}^d} |x - \pi_N(x)|^p d\mu(x) = \int_{|x|>N} |x - \pi_N(x)|^p d\mu(x) \leq \int_{|x|>N} |x|^p d\mu(x) \leq \varepsilon$$

since $N/|x| < 1$. As $\overline{B(0,N)}$ is compact, then there exists a finite family of open balls $B(x_i, \varepsilon)$, $x_i \in \mathbb{R}^d$ covering $\overline{B(0,N)}$ in other words

$$\overline{B(0,N)} \subseteq \bigcup_{i=1}^n B(x_i, \varepsilon).$$

We define by induction the family C_i by $C_1 := B(x_1, \varepsilon) \cap \overline{B(0,N)}$ and for every $2 \leq i \leq n$,

$$C_i := \left(B(x_i, \varepsilon) \cap \overline{B(0,N)}\right) \setminus \left(\bigcup_{k=1}^{i-1} C_k\right)$$

in other words $x \in \overline{B(0,N)}$ is in C_i if it is in the ball $B(x_i, \varepsilon)$ but is not in any of the balls $B(x_j, \varepsilon)$, $j < i$. It is clear that the family $(C_i)_{1 \leq i \leq n}$ is a partition of $\overline{B(0,N)}$. We define the measure $\gamma_1 := \sum a_i \delta_{x_i}$ where $a_i := \mu_N(C_i)$. Let f be the function to which x associates x_i if $x \in C_i$, then $(\text{id}, f)\#\mu_N$ is a coupling of (μ_N, γ_1) and we have

$$W_p(\mu_N, \gamma_1)^p \leq \int_{|x|\leq N} |x - f(x)|^p d\mu_N(x) = \sum_{i=1}^n \int_{C_i} |x - f(x)|^p d\mu_N(x) = \sum_{i=1}^n \int_{C_i} |x - x_i|^p d\mu_N(x) < \varepsilon^p$$

since $x \in B(x_i, \varepsilon)$ if $x \in C_i$ and $\sum a_i = 1$. Let $(y_i)_{1 \leq i \leq n} \in \mathbb{Q}^d$ be a family such that for every i, $|x_i - y_i| < \varepsilon$. We define the measure $\gamma_2 := \sum a_i \delta_{y_i}$. Let g be the function which x_i associates y_i, then $(\text{id}, g)\#\gamma_1$ is a coupling of (γ_1, γ_2) and we have

$$W_p(\gamma_1, \gamma_2)^p \leq \int_{\mathbb{R}^d} |x - g(x)|^p d\gamma_1(x) = \sum_{i=1}^n a_i |x_i - y_i|^p < \varepsilon^p.$$

Let $(b_i)_{1 \leq i \leq n} \in \mathbb{Q}^d$ be a family of rational numbers. We define the measure $\nu := \sum b_i \delta_{y_i}$. By definition of the total variation for a complex measure (A.3.6), we have $|\gamma_2 - \nu| = \sum |a_i - b_i| \delta_{y_i}$ and consequently we have by (A.6.1)

$$\begin{aligned} W_p(\gamma_2, \nu) &\leq 2^{1/q} \left(\int_{\mathbb{R}^d} |x|^p d|\gamma_2 - \nu|(x)\right)^{1/p} = 2^{1/q} \left(\sum_{i=1}^n |a_i - b_i||y_i|^p\right)^{1/p} \\ &\leq 2^{1/q} \max_{1 \leq i \leq n} |x_i| \sum_{i=1}^n |a_i - b_i|^{1/p} \end{aligned}$$

By choosing the rationals b_i so that $2^{1/q}\max|x_i|\sum|a_i-b_i|^{1/p}<\varepsilon$, we finally have

$$\begin{aligned}W_p\left(\mu,\sum_{i=1}^n b_i\delta_{y_i}\right)&\leq W_p(\mu,\mu_N)+W_p(\mu_N,\gamma_1)+W_p(\gamma_1,\gamma_2)+W_p(\gamma_2,\nu)\\&\leq \varepsilon^{1/p}+\varepsilon+\varepsilon+\varepsilon.\end{aligned}$$

We have shown that we can approximate μ by ν and we can identify ν by an element of $\bigcup_n \mathbb{Q}^n\times(\mathbb{Q}^d)^n$. This space is countable and is dense in $\mathcal{P}_p(\mathbb{R}^d)$ for the Wasserstein metric so consequently the Wasserstein space $\mathcal{P}_p(\mathbb{R}^d)$ is separable.

Let us now show the completeness. Let $(\mu_n)_n$ be a Cauchy sequence in $\mathcal{P}_p(\mathbb{R}^d)$. Then there exists a subsequence $\sigma:\mathbb{N}\nearrow\mathbb{N}$ such that

$$\sum_{n=1}^{+\infty}W_p(\mu_{\sigma(n)},\mu_{\sigma(n+1)})<+\infty.$$

We will show that the subsequence $(\mu_{\sigma(n)})_n$ is convergent, which will imply the convergence of the sequence $(\mu_n)_n$ since it is Cauchy. To simplifie the notations, we assume that $\sigma(n)=n$. For each $n\in\mathbb{N}$, we take $\pi_n\in\Pi(\mu_n,\mu_{n+1})$ an optimal coupling. By successively applying the Gluing Lemma (Lemma A.2.1), we can construct for each $n\in\mathbb{N}$ the measure $\gamma_n\in\mathcal{P}((\mathbb{R}^d)^n)$ defined by

$$\begin{cases}\gamma_1=\mu_1,\\ \gamma_2=\pi_1,\\ \gamma_n=\mu_{n-1}\otimes P_{\gamma_{n-1}}(x_{n-1},dx_1\cdots dx_{n-2})\otimes P_{\pi_{n-1}}(x_{n-1},dx_n)\qquad n\geq 3\end{cases}$$

with $\pi_n(dx_ndx_{n+1})=\mu_n(dx_n)P_{\pi_n}(x_n,dx_{n+1})$ and $\gamma_n(dx_1\cdots dx_n)=\mu_n(dx_n)P_{\gamma_n}(x_n,dx_1\cdots dx_{n-1})$. By construction, it is easy to check that for all $n\in\mathbb{N}$, $(\mathrm{proj}_1,\cdots,\mathrm{proj}_n)\#\gamma_{n+1}=\gamma_n$ since $P_{\gamma_n}(x_n,\cdot)\in\mathcal{P}((\mathbb{R}^d)^{n-1})$. So by the Kolmogorow extension Theorem (Theorem A.6.1), there exists $\gamma\in\mathcal{P}((\mathbb{R}^d)^{\mathbb{N}})$ such that for all $n\in\mathbb{N}$, $(\mathrm{proj}_1,\cdots,\mathrm{proj}_n)\#\gamma=\gamma_n$. So we have

$$\begin{aligned}\sum_{n=1}^{+\infty}W_p(\mu_n,\mu_{n+1})&=\sum_{n=1}^{+\infty}\left(\iint_{\mathbb{R}^d\times\mathbb{R}^d}|x_n-x_{n+1}|^p d\pi_n(x_n,x_{n+1})\right)^{1/p}\\&=\sum_{n=1}^{+\infty}\left(\int_{(\mathbb{R}^d)^{\mathbb{N}}}|\mathrm{proj}_n(x)-\mathrm{proj}_{n+1}(x)|^p d\gamma(x)\right)^{1/p}\\&=\sum_{n=1}^{+\infty}\|\mathrm{proj}_n-\mathrm{proj}_{n+1}\|_{L^p((\mathbb{R}^d)^{\mathbb{N}},\gamma)}<+\infty.\end{aligned}$$

So the sequence of functions $(\mathrm{proj}_n)_n$ is a Cauchy sequence in the space $L^p((\mathbb{R}^d)^{\mathbb{N}},\gamma)$ which is complete. So there exists $f\in L^p((\mathbb{R}^d)^{\mathbb{N}},\gamma)$ such that $(\mathrm{proj}_n)_n$ converges to f in $L^p((\mathbb{R}^d)^{\mathbb{N}},\gamma)$. We set $\mu:=f\#\gamma\in\mathcal{P}_p(\mathbb{R}^d)$, then by (A.4.6) we have

$$\limsup_{n\to+\infty}W_p(\mu_n,\mu)=\limsup_{n\to+\infty}W_p(\mathrm{proj}_n\#\gamma,f\#\gamma)\leq\limsup_{n\to+\infty}\|\mathrm{proj}_n-f\|_{L^p((\mathbb{R}^d)^{\mathbb{N}},\gamma)}=0.$$

■

For a measured space $\mathcal{X}$, we can identify a measure μ in the Wasserstein space $\mathcal{P}_p(\mathcal{X})$ by a random variable X of law μ in $L^p(\mathcal{X})$. The topological properties of $(\mathcal{P}_p(\mathbb{R}^d),W_p)$ and of $L^p(\mathcal{X})$ are therefore linked since as for $L^p(\mathcal{X})$, the Wasserstein space $\mathcal{P}_p(\mathcal{X})$ is separable if $\mathcal{X}$ is separable and the completeness of $L^p(\mathcal{X})$ results in the completeness of $\mathcal{P}_p(\mathcal{X})$ for the Wasserstein metric. However, the Wasserstein space $(\mathcal{P}_p(\mathcal{X}),W_p)$ has the same negative topological properties as $L^p(\mathcal{X})$. Indeed, just like the space $L^p(\mathcal{X})$, $(\mathcal{P}_p(\mathcal{X}),W_p)$ is neither σ-compact, nor locally compact (see § 2.2.3 in Ref. [68]).

Proposition A.6.3 (Convexity with respect to transition kernel)

Given $P : x \in \mathbb{R}^d \longmapsto P(x, \cdot) \in \mathcal{P}_p(\mathbb{R}^d)$, $Q : x \in \mathbb{R}^d \longmapsto Q(x, \cdot) \in \mathcal{P}_p(\mathbb{R}^d)$ two continuous transition kernels and $\mu \in \mathcal{P}(\mathbb{R}^d)$, then

$$W_p\left(\int_{\mathbb{R}^d} P(x, \cdot)d\mu(x), \int_{\mathbb{R}^d} Q(x, \cdot)d\mu(x)\right)^p \leq \int_{\mathbb{R}^d} W_p(P(x, \cdot), Q(x, \cdot))^p d\mu(x). \quad \text{(A.6.2)}$$

The proof of Proposition A.6.3 uses the following result (Theorem 1 in Ref. [20], Theorem 6.3 in Ref. [74]).

Theorem A.6.2 (The Borel selection Theorem)

Let X, Y be two Polish spaces and $E \subset X \times Y$. If for every $x \in X$, the slice $E_x := \{y \in Y, (x, y) \in E\}$ is σ-compact (i.e a countable union of compacts), or possibly empty, then the first projection $F := \{x \in X, E_x \neq \emptyset\}$ of E is a Borel set and there exists a Borel map $s : F \longrightarrow Y$ such that for all $x \in F$, $(x, s(x)) \in E$.

Proof of Proposition A.6.3. For $x \in \mathbb{R}^d$ and for $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, we define the space E as the set of $(x, \pi) \in \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ satisfying the three following conditions:

(1) $\forall\ \varphi, \psi \in \mathcal{C}_c(\mathbb{R}^d)$, $\displaystyle\iint_{\mathbb{R}^d\times\mathbb{R}^d} (\varphi(u) + \psi(v))d\pi(u, v) = \int_{\mathbb{R}^d} \varphi(u)P(x, du) + \int_{\mathbb{R}^d} \psi(v)Q(x, dv).$

(2) $\displaystyle\iint_{\mathbb{R}^d\times\mathbb{R}^d} |u - v|^p d\pi(u, v) \leq W_p(P(x, \cdot), Q(x, \cdot)).$

(3) $\displaystyle\iint_{\mathbb{R}^d\times\mathbb{R}^d} |u|^p d\pi(u, v) \leq \int_{\mathbb{R}^d} |u|^p P(x, du)$ and $\displaystyle\iint_{\mathbb{R}^d\times\mathbb{R}^d} |v|^p d\pi(u, v) \leq \int_{\mathbb{R}^d} |v|^p Q(x, dv).$

We will show that E is closed. Let $(x_n, \pi_n)_n$ be a sequence in E such that $(\pi_n)_n$ converges weakly to π and $(x_n)_n$ converges to x. By hypothesis the maps P and Q are continuous. So $W_p(P(x_n, \cdot), P(x, \cdot))$ and $W_p(Q(x_n, \cdot), Q(x, \cdot))$ tend to 0 when n tends to infinity which implies by Theorem A.5.1 that $P(x_n, \cdot)$ converges weakly to $P(x, \cdot)$ and that $Q(x_n, \cdot)$ converges weakly to $Q(x, \cdot)$. For $\varphi, \psi \in \mathcal{C}_c(\mathbb{R}^d)$ we have

$$\begin{aligned}
\iint_{\mathbb{R}^d\times\mathbb{R}^d} (\varphi(u) + \psi(v))d\pi(u, v) &= \lim_{n\to+\infty} \iint_{\mathbb{R}^d\times\mathbb{R}^d} (\varphi(u) + \psi(v))d\pi_n(u, v) \\
&= \lim_{n\to+\infty} \left(\int_{\mathbb{R}^d} \varphi(u)P(x_n, du) + \int_{\mathbb{R}^d} \psi(v)Q(x_n, dv)\right) \\
&= \int_{\mathbb{R}^d} \varphi(u)P(x, du) + \int_{\mathbb{R}^d} \psi(v)Q(x, dv).
\end{aligned}$$

So (x, π) satisfies (1). Let $\chi_R : \mathbb{R}^d \longrightarrow [0, 1]$, $R > 0$ be a continuous function such that $\chi_R(x) = 1$ for $|x| \leq R$ and $\chi_R(x) = 0$ for $|x| \geq R + 1$. Then the function $(u, v) \longmapsto \chi_R(u)\chi_R(v)|u - v|^p$ is continuous with compact support and by (A.5.2) we have

$$\begin{aligned}
\iint_{\mathbb{R}^d\times\mathbb{R}^d} \chi_R(u)\chi_R(v)|u - v|^p d\pi(u, v) &= \lim_{n\to+\infty} \iint_{\mathbb{R}^d\times\mathbb{R}^d} \chi_R(u)\chi_R(v)|u - v|^p d\pi_n(u, v) \\
&\leq \lim_{n\to+\infty} W_p(P(x_n, \cdot), Q(x_n, \cdot)) \\
&= W_p(P(x, \cdot), Q(x, \cdot)).
\end{aligned}$$

And by letting $R \longrightarrow +\infty$, we obtain (2) for (x, π). By Theorem A.5.1, the moments of order p of $P(x_n, \cdot)$ converge to the moment of order p of $P(x, \cdot)$ and the moments of order p of $Q(x_n, \cdot)$ converge to the moment of order p of $Q(x, \cdot)$. And since the two functions $(u, v) \longmapsto \chi_R(u)\chi_R(v)|u|^p$ and $(u, v) \longmapsto \chi_R(u)\chi_R(v)|v|^p$ are continuous with compact support, we have

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d} \chi_R(u)\chi_R(v)|u|^p d\pi(u,v) &= \lim_{n\to+\infty} \iint_{\mathbb{R}^d\times\mathbb{R}^d} \chi_R(u)\chi_R(v)|u|^p d\pi_n(u,v)\\ &\le \lim_{n\to+\infty} \int_{\mathbb{R}^d} |u|^p P(x_n, du)\\ &= \int_{\mathbb{R}^d} |u|^p P(x, du)\end{aligned}$$

and we also have

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d} \chi_R(u)\chi_R(v)|v|^p d\pi(u,v) &= \lim_{n\to+\infty} \iint_{\mathbb{R}^d\times\mathbb{R}^d} \chi_R(u)\chi_R(v)|v|^p d\pi_n(u,v)\\ &\le \lim_{n\to+\infty} \int_{\mathbb{R}^d} |v|^p Q(x_n, dv)\\ &= \int_{\mathbb{R}^d} |v|^p Q(x, dv).\end{aligned}$$

So by letting $R \longrightarrow +\infty$, we obtain (3) for (x, π) so E is closed. We now show that the slice $E_x := \{\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d),\ (x, \pi) \in E\}$ is compact. Let $R > 0$ and $x \in \mathbb{R}^d$. Then we have for $\pi \in E_x$

$$\begin{aligned}\int_{|u|^2+|v|^2>R^2} d\pi(u,v) &\le \int_{|u|^2>R^2/2} d\pi(u,v) + \int_{|v|^2>R^2} d\pi(u,v)\\ &\le \left(\frac{2}{R}\right)^p \int_{\mathbb{R}^d} |u|^p d\pi(u,v) + \left(\frac{2}{R}\right)^p \int_{\mathbb{R}^d} |v|^p d\pi(u,v)\\ &\le \left(\frac{2}{R}\right)^p \left(\int_{\mathbb{R}^d} |u|^p P(x, du) + \int_{\mathbb{R}^d} |v|^p Q(x, dv)\right).\end{aligned}$$

For any $\varepsilon > 0$, we can choose R large enough such that $\pi(\{(u, v) \in \mathbb{R}^d \times \mathbb{R}^d,\ |u|^2 + |v|^2 > R^2\}) < \varepsilon$ so the slice E_x is tight which leads by Prokhorov Theorem (Theorem A.5.2) that E_x is relatively compact for the weak topology. Moreover, E_x is closed since E is closed so E_x is compact. We deduce by the Borel Selection Theorem (Theorem A.6.2) that the set F of $x \in \mathbb{R}^d$ such that $E_x \neq \emptyset$ and such that there is a coupling $\pi \in \Pi(P(x, \cdot), Q(x, \cdot))$ of moment bounded by $M_p(P(x, \cdot))$, $M_p(Q(x, \cdot))$ is Borelian ($F = \mathbb{R}^d$). And furthermore, there exists a measurable section $s : \mathbb{R}^d \longrightarrow \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ such that for all $x \in \mathbb{R}^d$, $(x, s(x)) \in E$. We can therefore consider for $x \in \mathbb{R}^d$ an optimal coupling $\pi_x \in \Pi(P(x, \cdot), Q(x, \cdot))$ such that the map $x \longmapsto \pi_x$ is measurable. We set $\pi := \int \pi_x d\mu(x)$. It is easy to check that $\pi \in \Pi(\int P(x, \cdot)d\mu(x), \int Q(x, \cdot)d\mu(x))$ and therefore

$$\begin{aligned}W_p\left(\int_{\mathbb{R}^d} P(x, \cdot)d\mu(x), \int_{\mathbb{R}^d} Q(x, \cdot)d\mu(x)\right)^p &\le \iint_{\mathbb{R}^d\times\mathbb{R}^d} |u - v|^p d\pi(u,v)\\ &= \iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} |u - v|^p d\pi_x(u,v)d\mu(x)\\ &= \int_{\mathbb{R}^d} W_p(P(x, \cdot), Q(x, \cdot))^p d\mu(x).\end{aligned}$$

■

The following result is Proposition 14.3 in Ref. [40].

Proposition A.6.4 (Lipschizianity with respect to transition kernel)

Let $P : x \in \mathbb{R}^d \longmapsto P(x, \cdot) \in (\mathcal{P}_p(\mathbb{R}^d), W_p)$ be a transition kernel such that $\|P\|_{\text{Lip}}$ is finite. Then for all $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, we have

$$W_p\left(\int_{\mathbb{R}^d} P(x, \cdot)d\mu(x), \int_{\mathbb{R}^d} P(x, \cdot)d\nu(x)\right) \leq \|P\|_{\text{Lip}} W_p(\mu, \nu). \tag{A.6.3}$$

Proof. Let $\pi_* \in \Pi(\mu, \nu)$ and let $\pi_{x_0,y_0} \in \Pi(P(x_0, \cdot), P(y_0, \cdot))$ be two optimal couplings. We have

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p d\pi_{x_0,y_0}(x, y) \leq \|P\|^p_{\text{Lip}} |x_0 - y_0|^p.$$

For $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ and for $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, we define the space E as the set of $(x, y, \pi) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ satisfying the three following conditions:

(1) $\forall\ \varphi,\ \psi \in \mathcal{C}_c(\mathbb{R}^d)$, $\displaystyle\iint_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi(u) + \psi(v))d\pi(u, v) = \int_{\mathbb{R}^d} \varphi(u)P(x, du) + \int_{\mathbb{R}^d} \psi(v)P(y, dv).$

(2) $\displaystyle\iint_{\mathbb{R}^d \times \mathbb{R}^d} |u - v|^p d\pi(u, v) \leq W_p(P(x, \cdot), P(y, \cdot)).$

(3) $\displaystyle\iint_{\mathbb{R}^d \times \mathbb{R}^d} |u|^p d\pi(u, v) \leq \int_{\mathbb{R}^d} |u|^p P(x, du)$ and $\displaystyle\iint_{\mathbb{R}^d \times \mathbb{R}^d} |v|^p d\pi(u, v) \leq \int_{\mathbb{R}^d} |v|^p P(y, dv).$

We will show that E is closed. Let $(x_n, y_n, \pi_n)_n$ be a sequence in E such that $(\pi_n)_n$ converges weakly to π, $(x_n)_n$ converges to x and $(y_n)_n$ converges to y. By hypothesis the map P is Lipschitzian and therefore continuous. So $W_p(P(x_n, \cdot), P(x, \cdot))$ and $W_p(P(y_n, \cdot), P(y, \cdot))$ tend to 0 when n tends to infinity which implies by Theorem A.5.1 that $P(x_n, \cdot)$ converges weakly to $P(x, \cdot)$ and that $P(y_n, \cdot)$ converges weakly to $P(y, \cdot)$. For $\varphi, \psi \in \mathcal{C}_c(\mathbb{R}^d)$ we have

$$\begin{aligned}
\iint_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi(u) + \psi(v))d\pi(u, v) &= \lim_{n\to+\infty} \iint_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi(u) + \psi(v))d\pi_n(u, v) \\
&= \lim_{n\to+\infty} \left(\int_{\mathbb{R}^d} \varphi(u)P(x_n, du) + \int_{\mathbb{R}^d} \psi(v)P(y_n, dv)\right) \\
&= \int_{\mathbb{R}^d} \varphi(u)P(x, du) + \int_{\mathbb{R}^d} \psi(v)P(y, dv).
\end{aligned}$$

So (x, y, π) satisfies (1). Let $\chi_R : \mathbb{R}^d \longrightarrow [0, 1]$, $R > 0$ be a continuous function such that $\chi_R(x) = 1$ for $|x| \leq R$ and $\chi_R(x) = 0$ for $|x| \geq R + 1$. Then the function $(u, v) \longmapsto \chi_R(u)\chi_R(v)|u - v|^p$ is continuous with compact support and by (A.5.2) we have

$$\begin{aligned}
\iint_{\mathbb{R}^d \times \mathbb{R}^d} \chi_R(u)\chi_R(v)|u - v|^p d\pi(u, v) &= \lim_{n\to+\infty} \iint_{\mathbb{R}^d \times \mathbb{R}^d} \chi_R(u)\chi_R(v)|u - v|^p d\pi_n(u, v) \\
&\leq \lim_{n\to+\infty} W_p(P(x_n, \cdot), P(y_n, \cdot)) \\
&= W_p(P(x, \cdot), P(y, \cdot)).
\end{aligned}$$

And by letting $R \longrightarrow +\infty$, we obtain (2) for (x, y, π). By Theorem A.5.1, the moments of order p of $P(x_n, \cdot)$ converge to the moment of order p of $P(x, \cdot)$ and the moments of order p of $P(y_n, \cdot)$ converge to the moment of order p of $P(y, \cdot)$. And since the two functions $(u, v) \longmapsto \chi_R(u)\chi_R(v)|u|^p$

and $(u,v) \longmapsto \chi_R(u)\chi_R(v)|v|^p$ are continuous with compact support, we have

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d} \chi_R(u)\chi_R(v)|u|^p d\pi(u,v) &= \lim_{n\to+\infty} \iint_{\mathbb{R}^d\times\mathbb{R}^d} \chi_R(u)\chi_R(v)|u|^p d\pi_n(u,v)\\ &\leq \lim_{n\to+\infty} \int_{\mathbb{R}^d} |u|^p P(x_n, du)\\ &= \int_{\mathbb{R}^d} |u|^p P(x, du)\end{aligned}$$

and we also have

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d} \chi_R(u)\chi_R(v)|v|^p d\pi(u,v) &= \lim_{n\to+\infty} \iint_{\mathbb{R}^d\times\mathbb{R}^d} \chi_R(u)\chi_R(v)|v|^p d\pi_n(u,v)\\ &\leq \lim_{n\to+\infty} \int_{\mathbb{R}^d} |v|^p P(y_n, dv)\\ &= \int_{\mathbb{R}^d} |v|^p P(y, dv).\end{aligned}$$

So by letting $R \longrightarrow +\infty$, we obtain (3) for (x,y,π) so E is closed. We now show that the slice $E_{(x,y)} := \{\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d),\ (x,y,\pi) \in E\}$ is compact. Let $R > 0$ and $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$. Then we have for $\pi \in E_{(x,y)}$

$$\begin{aligned}\int_{|u|^2+|v|^2>R^2} d\pi(u,v) &\leq \int_{|u|^2>R^2/2} d\pi(u,v) + \int_{|v|^2>R^2} d\pi(u,v)\\ &\leq \left(\frac{2}{R}\right)^p \int_{\mathbb{R}^d} |u|^p d\pi(u,v) + \left(\frac{2}{R}\right)^p \int_{\mathbb{R}^d} |v|^p d\pi(u,v)\\ &\leq \left(\frac{2}{R}\right)^p \left(\int_{\mathbb{R}^d} |u|^p P(x,du) + \int_{\mathbb{R}^d} |v|^p P(y,dv)\right).\end{aligned}$$

For any $\varepsilon > 0$, we can choose R large enough such that $\pi(\{(u,v) \in \mathbb{R}^d \times \mathbb{R}^d,\ |u|^2 + |v|^2 > R^2\}) < \varepsilon$ so the slice $E_{(x,y)}$ is tight which leads by Prokhorov Theorem (Theorem A.5.2) that $E_{(x,y)}$ is relatively compact for the weak topology. Moreover, $E_{(x,y)}$ is closed since E is closed so $E_{(x,y)}$ is compact. We deduce by the Borel Selection Theorem (Theorem A.6.2) that the set F of $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ such that $E_{(x,y)} \neq \emptyset$ and such that there is a coupling $\pi \in \Pi(P(x,\cdot), P(y,\cdot))$ of moment bounded by $M_p(P(x,\cdot))$, $M_p(P(y,\cdot))$ is Borelian ($F = \mathbb{R}^d \times \mathbb{R}^d$). And furthermore, there exists a measurable section $s : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ such that for all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$, $(x,y,s(x,y)) \in E$. We can therefore define the four-variable coupling π defined for any test function φ by

$$\iiiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \varphi(x,y,u,v) d\pi(x,y,u,v) = \iiiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \varphi(x,y,u,v) d\pi_{x,y}(u,v) d\pi_*(x,y).$$

We set $\pi_{3,4} := (\mathrm{proj}_3, \mathrm{proj}_4)\#\pi$. Since $\pi_* \in \Pi(\mu,\nu)$ and $\pi_{x,y} \in \Pi(P(x,\cdot), P(y,\cdot))$, it is easy to check that $\pi_{3,4} \in \Pi(\int P(x,\cdot)d\mu(x), \int P(x,\cdot)d\nu(x))$. And so we have

$$\begin{aligned}W_p\left(\int_{\mathbb{R}^d} P(x,\cdot)d\mu(x), \int_{\mathbb{R}^d} P(x,\cdot)d\nu(x)\right)^p &\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} |u-v|^p d\pi_{3,4}(u,v)\\ &= \iint_{\mathbb{R}^d\times\mathbb{R}^d} \left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} |u-v|^p d\pi_{x,y}(u,v)\right) d\pi_*(x,y)\\ &\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} \|P\|_{\mathrm{Lip}} |x-y|^p d\pi_*(x,y)\\ &= \|P\|_{\mathrm{Lip}}^p W_p(\mu,\nu)^p.\end{aligned}$$

■

A.7 Implementation of the Wasserstein metric on the real line

Let's come back to Monge's problem but in the particular case $d = 1$. We want to minimize the quantity $\iint c(x,y)d\pi(x,y)$ on the set $\Pi(\mu,\nu)$. We assume that μ and ν are linear combinations of Dirac masses, that is to say that

$$\mu := \sum_{i=1}^{n} a_i \delta_{x_i} \qquad \text{and} \qquad \nu := \sum_{j=1}^{n} b_j \delta_{y_j} \tag{A.7.1}$$

with $\sum a_i = \sum b_j = 1$ and where $(x_i)_{1\leq i\leq n}$, $(y_j)_{1\leq j\leq n}$ are two ordered families of points of $\mathbb{R}$: $x_1 < x_2 < \cdots < x_n$ and $y_1 < y_2 < \cdots < y_n$. Constructing a coupling $\pi \in \Pi(\mu,\nu)$ consists in creating a configuration that sends the entire mass a_i for each point x_i on the points $(y_j)_{1\leq j\leq n}$ so that each point y_j receives all of its mass b_j. If the cost function c satisfies the following property

$$\forall i \in [\![1, n-1]\!], \qquad c(x_i,y_i) + c(x_{i+1},y_{i+1}) \leq c(x_i,y_{i+1}) + c(x_{i+1},y_i), \tag{A.7.2}$$

then it is easy to construct an optimal coupling. We start from $i = 1$ and start by sending the mass of x_1 to y_1. If y_1 receives all of its mass, then we send mass at x_1 to y_2 and so on until x_1 has sent all of its mass. Let y_{i_1} be the point reached when x_1 sent all its mass, then we send the mass of x_2 to y_{i_1}. If y_{i_1} receives all of its mass, then we send mass at x_2 to y_{i_1+1} and so on until x_2 has sent all of its mass. Let y_{i_2} be the point reached when x_2 sent all its mass, we repeat the process with x_3 and we do this with all the points x_i. If μ and ν are empirical measures, in other words $a_i = b_i = 1/n$, then construct an optimal coupling for a cost function c satisfying (A.7.2) is easy. It is enough to send all the mass in x_i on y_i. We can check that if a cost function c is of the form $c(x,y) = h(x-y)$ with $h : \mathbb{R} \longrightarrow \mathbb{R}$ a convex function, then it satisfies (A.7.2). Indeed, for all $i \in [\![1, n-1]\!]$, we set $u_i := x_i - y_{i+1}$, $v_i := x_i - y_i$ and $w_i := x_{i+1} - x_i$. Then (A.7.2) becomes

$$h(v_i) + h(u_i + w_i) \leq h(u_i) + h(w_i + v_i).$$

By dividing on each side by $w_i > 0$, we obtain the following inequality

$$\frac{h(u_i + w_i) - h(u_i)}{w_i} \leq \frac{h(v_i + w_i) - h(v_i)}{w_i}$$

which is true by the chordal slope Lemma. Thus, in dimension 1, we can compute $W_p(\mu,\nu)$ explicitly when μ and ν are defined by (A.7.1) since the function $x \longmapsto |x|^p$ is convex for all $p \geq 1$.

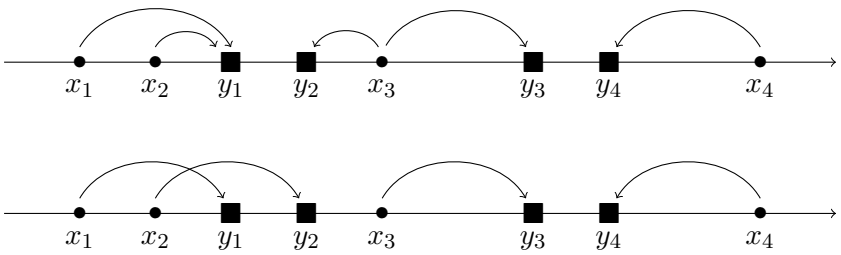


Figure A.5: Example of construction of an optimal coupling $\pi \in \Pi(\mu,\nu)$ which minimizes $\iint c(x,y)d\pi(x,y)$ when the cost function satisfies (A.7.2). Above, case where μ and ν satisfy (A.7.1) with $a_1 + a_2 = b_1$, $a_3 = b_2 + b_3$ and $a_4 = b_4$. Below, case where μ and ν are empirical measures.

Definition A.7.1

For $\mu \in \mathcal{P}(\mathbb{R})$, we define the **generalized inverse** of F_μ the function

$$F_\mu^{-1}(t) := \inf\{x \in \mathbb{R},\, F_\mu(x) > t\}. \tag{A.7.3}$$

The generalized inverse of F_μ is a function that defines quantiles, it is sometimes called quantile function. Since F_μ takes its values in $[0,1]$ and is always non-decreasing and right-continuous, then the generalized inverse of F_μ is defined on $[0,1[$ and is always non-decreasing and right-continuous. And we have the inequalities $F_\mu^{-1}(F_\mu(x)) \geq x$ for all $x \in \mathbb{R}$ and $F_\mu(F_\mu^{-1}(t)) \geq t$ for all $t \in [0,1[$. These two inequalities become equalities if F_μ is increasing, which means in this case that the generalized inverse of F_μ is the reciprocal function of the cumulative distribution function of μ. We have also $\lim_{t\to 1} F_\mu^{-1}(t) = +\infty$. To compute $W_p(\mu,\nu)$ in $\mathbb{R}$, we use the following result.

Theorem A.7.1 (Wasserstein metric on $\mathbb{R}$)

Let μ, $\nu \in \mathcal{P}_p(\mathbb{R})$. Then we have

$$W_p(\mu,\nu) = \left(\int_0^1 |F_\mu^{-1}(t) - F_\nu^{-1}(t)|^p dt\right)^{1/p}. \tag{A.7.4}$$

The generalized inverse provides a closed form of the Wasserstein metric W_p on $\mathbb{R}$. For μ, $\nu \in \mathcal{P}_p(\mathbb{R})$, $W_p(\mu,\nu)$ is the $L^p([0,1[)$ norm of the function $t \longmapsto F_\mu^{-1}(t) - F_\nu^{-1}(t)$. We will give a proof of this Theorem only in the case that interests us, that is to say for the case $p = 2$ only. The proof uses the following Lemma.

Lemma A.7.1

For $\mu \in \mathcal{P}(\mathbb{R})$, we have

$$\mathrm{Vol}(\{t \in]0,1[,\ F_\mu^{-1}(t) \leq x\}) = F_\mu(x) \tag{A.7.5}$$

Proof. We set $\overline{F_\mu^{-1}}(t) := \inf\{x \in \mathbb{R},\ F_\mu(x) \geq t\}$. If $\overline{F_\mu^{-1}}(t) \leq x_0$, then for all $\varepsilon > 0$ we have $\inf\{x \in \mathbb{R},\ F_\mu(x) \geq t\} < x_0 + \varepsilon$ and therefore there exists $y_\varepsilon < x_0 + \varepsilon$ such that $F_\mu(y_\varepsilon) \geq t$. Since F_μ is non-decreasing and right-continuous, we have $t \leq F_\mu(x_0 + \varepsilon)$ and $t \leq \lim_{\varepsilon\to 0^+} F_\mu(x_0 + \varepsilon) = F_\mu(x_0)$ therefore $t \leq F_\mu(x_0)$. Then, if $t \leq F_\mu(x_0)$, then $[x_0, +\infty[\subseteq \{x \in \mathbb{R},\ F_\mu(x) \geq t\}$ and taking the infimum on each side, we have $\overline{F_\mu^{-1}}(t) \leq x_0$. We thus have the equivalence

$$\overline{F_\mu^{-1}}(t) \leq x \qquad \Longleftrightarrow \qquad t \leq F_\mu(x). \tag{A.7.6}$$

We can deduce that

$$\mathrm{Vol}(\{t \in]0,1[,\ \overline{F_\mu^{-1}}(t) \leq x\}) = \mathrm{Vol}(\{t \in]0,1[,\ t \leq F_\mu(x)\}) = \mathrm{Vol}([0, F_\mu(x)]) = F_\mu(x).$$

Then, if at first x is such that $F_\mu(x) > t$, then $F_\mu(x) \geq t$ in other words $\{x \in \mathbb{R},\ F_\mu(x) > t\} \subseteq \{x \in \mathbb{R},\ F_\mu(x) \geq t\}$. Taking the infimum, we have $\overline{F_\mu^{-1}}(t) \leq F_\mu^{-1}(t)$. Hence if $t \in]0,1[$ is such that $F_\mu^{-1}(t) \leq x$, then $\overline{F_\mu^{-1}}(t) \leq x$ in other words $\{t \in]0,1[,\ F_\mu^{-1}(t) \leq x\} \subseteq \{t \in]0,1[,\ \overline{F_\mu^{-1}}(t) \leq x\}$. And if in a second time we have $F_\mu^{-1}(t) > x$ and $\overline{F_\mu^{-1}}(t) \leq x$, then we have by (A.7.6) that $F_\mu(x) \geq t$, and $x \notin \{x \in \mathbb{R},\ F_\mu(x) > t\}$ therefore $F_\mu(x) \leq t$ and therefore $F_\mu(x) = t$. We finally deduce that

$$\begin{aligned}\mathrm{Vol}(\{t \in]0,1[,\ \overline{F_\mu^{-1}}(t) \leq x\}) &= \mathrm{Vol}(\{t \in]0,1[,\ \overline{F_\mu^{-1}}(t) \leq x\} \cap \{t \in]0,1[,\ F_\mu^{-1}(t) \leq x\}) \\ &\quad + \mathrm{Vol}(\{t \in]0,1[,\ \overline{F_\mu^{-1}}(t) \leq x\} \cap \{t \in]0,1[,\ F_\mu^{-1}(t) > x\}) \\ &= \mathrm{Vol}(\{t \in]0,1[,\ F_\mu^{-1}(t) \leq x\}) + \mathrm{Vol}(\{F_\mu(x)\}).\end{aligned}$$

We obtain (A.7.5) since $\mathrm{Vol}(\{F_\mu(x)\}) = 0$ and $\mathrm{Vol}(\{t \in]0,1[,\ \overline{F_\mu^{-1}}(t) \leq x\}) = F_\mu(x)$. ■

Let $x \in \mathbb{R}$. Thanks to (A.7.5), we deduce that $F_\mu^{-1}\#\lambda(]-\infty, x]) = \mathrm{Vol}(\{t \in]0,1[,\, F_\mu^{-1}(t) \leq x\}) = \mu(]-\infty, x])$. We therefore have by the monotone class Theorem that

$$F_\mu^{-1}\#\lambda = \mu. \tag{A.7.7}$$

We also have that $F_\mu^{-1}(t) \leq x$ if and only if $t < F_\mu(x)$. It is enough to make a reasoning analogous to that which we made to obtain (A.7.6).

Proof of Theorem A.7.1. We show the case $p = 2$ only. We set $\pi_* := (F_\mu^{-1}, F_\nu^{-1})\#\lambda$. We have by (A.7.7) that $\pi_* \in \Pi(\mu,\nu)$. Then, we have

$$\begin{aligned}
\pi_*(]-\infty, x] \times]-\infty, y]) &= \mathrm{Vol}(\{t \in]0,1[,\, F_\mu^{-1}(t) \leq x,\, F_\nu^{-1}(t) \leq y\}) \\
&= \mathrm{Vol}(\{t \in]0,1[,\, F_\mu^{-1}(t) \leq x\} \cap \{t \in]0,1[,\, F_\nu^{-1}(t) \leq y\}) \\
&= \mathrm{Vol}(\{t \in]0,1[,\, t < F_\mu(x)\} \cap \{t \in]0,1[,\, t < F_\nu(y)\}) \\
&= \mathrm{Vol}([0, F_\mu(x)[\cap [0, F_\nu(y)[) \\
&= \min(F_\mu(x), F_\nu(y)).
\end{aligned}$$

Let us now show that π_* is optimal. We will start by assuming that μ and ν are supported on $[0, +\infty[$ (by restricting μ et ν on $[-R, +\infty[$, $R > 0$, normalizing to a probability measure, and translating by R). We have

$$\begin{aligned}
\int_0^{+\infty}\int_0^{+\infty} xy d\pi_*(x,y) &= \int_0^{+\infty}\int_0^{+\infty}\int_0^x\int_0^y d\pi_*(x,y) dadb \\
&= \int_0^{+\infty}\int_0^{+\infty}\left(\int_a^{+\infty}\int_b^{+\infty} d\pi_*(x,y)\right) dadb \\
&= \int_0^{+\infty}\int_0^{+\infty} \pi_*(]a, +\infty[\times]b, +\infty[) dadb.
\end{aligned}$$

Let $\pi \in \Pi(\mu,\nu)$, we notice that $\pi(]-\infty, x] \times]-\infty, y]) \leq \pi(]-\infty, x] \times \mathbb{R}) = F_\mu(x)$ and therefore $\pi(]-\infty, x] \times]-\infty, y]) \leq \min(F_\mu(x), F_\nu(y))$. Since $\pi_*(]a, +\infty[\times]b, +\infty[) = 1 - F_\mu(a) - F_\mu(b) + \pi_*(]-\infty, a] \times]-\infty, b])$, we can deduce that

$$\begin{aligned}
\int_0^{+\infty}\int_0^{+\infty} xy d\pi_*(x,y) = &\int_0^{+\infty}\int_0^{+\infty} (\pi_*(]a, +\infty[\times]b, +\infty[) - \pi(]a, +\infty[\times]b, +\infty[)) dadb \\
&+ \int_0^{+\infty}\int_0^{+\infty} xy d\pi(x,y).
\end{aligned}$$

The first term in the right-hand side is non-negative. Hence

$$\begin{aligned}
\int_0^{+\infty}\int_0^{+\infty} |x-y|^2 d\pi_*(x,y) &= \int_0^{+\infty} |x|^2 d\mu(x) + \int_0^{+\infty} |y|^2 d\nu(y) - 2\int_0^{+\infty}\int_0^{+\infty} xy d\pi_*(x,y) \\
&\leq \int_0^{+\infty} |x|^2 d\mu(x) + \int_0^{+\infty} |y|^2 d\nu(y) - 2\int_0^{+\infty}\int_0^{+\infty} xy d\pi(x,y) \\
&= \int_0^{+\infty}\int_0^{+\infty} |x-y|^2 d\pi(x,y).
\end{aligned}$$

Suppose now that the support of μ and ν are not lower bound. We define the two sequences of probability measures $(\mu_n)_n$ and $(\nu_n)_n$ by $\mu_n := \mathbb{1}_{]-n,+\infty[}\mu/\mu(]-n, +\infty[)$ and $\nu_n := \mathbb{1}_{]-n,+\infty[}\nu/\nu(]-n, +\infty[)$. In others words, we have for any test function φ

$$\int_{\mathbb{R}} \varphi(x) d\mu_n(x) := \int_{-n}^{+\infty} \frac{\varphi(x) d\mu(x)}{\mu(]-n, +\infty[)} \qquad \text{and} \qquad \int_{\mathbb{R}} \varphi(x) d\nu_n(x) := \int_{-n}^{+\infty} \frac{\varphi(x) d\nu(x)}{\nu(]-n, +\infty[)}.$$

For all n, μ_n and ν_n are supported in $[-n, +\infty[$ so (A.7.4) is satisfied for μ_n and ν_n. Since $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$, it is easy to check that μ_n converge weakly to μ, ν_n converge weakly to ν, and the moment of order 2 of μ_n converge to the moment of order 2 of μ, the moment of order 2 of ν_n converge to the moment of order 2 of ν. Hence, we have by Theorem A.5.1

$$W_2(\mu,\nu) = \lim_{n\to+\infty} W_2(\mu_n,\nu_n) = \lim_{n\to+\infty} \left(\int_0^1 |F_{\mu_n}^{-1}(t) - F_{\nu_n}^{-1}(t)|^2 dt \right)^{1/2}.$$

We will apply the Lebesgue's dominated Theorem. To begin, we have

$$F_{\mu_n}(x) = \int_{-\infty}^{x} \frac{\mathbb{1}_{]-n,+\infty[}(x) d\mu(x)}{\mu(]-n,+\infty[)} = \frac{F_\mu(x) - F_\mu(-n)}{1 - F_\mu(-n)} \mathbb{1}_{]-n,+\infty[}(x).$$

So by definition of the generalized inverse (A.7.3), we have

$$F_{\mu_n}^{-1}(t) = \inf \left\{ x \in \mathbb{R},\ \frac{F_\mu(x) - F_\mu(-n)}{1 - F_\mu(-n)} \mathbb{1}_{]-n,+\infty[}(x) > t \right\}.$$

Since F_μ is non-decreasing and that $t \in [0,1[$,

$$\begin{aligned} F_{\mu_n}^{-1}(t) &= \inf \left\{ x \in \mathbb{R},\ \frac{F_\mu(x) - F_\mu(-n)}{1 - F_\mu(-n)} > t \right\} \\ &= \inf\{x \in \mathbb{R},\ F_\mu(x) > t(1 - F_\mu(-n)) + F_\mu(-n)\} \\ &= F_\mu^{-1}(t(1 - F_\mu(-n)) + F_\mu(-n)). \end{aligned}$$

Since $F_\mu(x) \longrightarrow 0$ when $x \longrightarrow -\infty$ and that for every n, $t(1 - F_\mu(-n)) + F_\mu(-n) \geq t$, then by right-continuity of F_μ we have that $F_\mu^{-1}(t(1 - F_\mu(-n)) + F_\mu(-n))$ converge to $F_\mu^{-1}(t)$ when n goes to infinity. Which entails that

$$\begin{aligned} \lim_{n\to+\infty} |F_{\mu_n}^{-1}(t) - F_{\nu_n}^{-1}(t)|^2 &= \lim_{n\to+\infty} |F_\mu^{-1}(t(1 - F_\mu(-n)) + F_\mu(-n)) - F_\nu^{-1}(t(1 - F_\nu(-n)) + F_\nu(-n))|^2 \\ &= |F_\mu^{-1}(t) - F_\nu^{-1}(t)|^2. \end{aligned}$$

Next, for $t \in [0,1[$ fixed, the function $n \longmapsto F_\mu^{-1}(t(1 - F_\mu(-n)) + F_\mu(-n)) = F_\mu^{-1}(F_\mu(-n)(1-t) + t)$ is a non-increasing function. So we have

$$\begin{aligned} |F_{\mu_n}^{-1}(t) - F_{\nu_n}^{-1}(t)|^2 &\leq 2|F_{\mu_n}^{-1}(t)|^2 + 2|F_{\nu_n}^{-1}(t)| \\ &\leq 2|F_\mu^{-1}(t(1 - F_\mu(0)) + F_\mu(0))|^2 + 2|F_\nu^{-1}(t(1 - F_\mu(0)) + F_\mu(0))|^2. \end{aligned}$$

By making the change of variable $u = t(1 - F_\mu(0)) + F_\mu(0)$, we have

$$\int_0^1 |F_\mu^{-1}(t(1 - F_\mu(0)) + F_\mu(0))|^2 dt = \frac{1}{1 - F_\mu(0)} \int_{F_\mu(0)}^1 |F_\mu^{-1}(x)|^2 dx \leq \frac{1}{1 - F_\mu(0)} \int_0^1 |F_\mu^{-1}(x)|^2 dx.$$

By (A.7.7), the right integral is finite since $\mu \in \mathcal{P}_2(\mathbb{R})$. Same for ν. So we can apply the Lebesgue's dominated Theorem. ■

In the particular case $p = 1$, there is an alternative formula of (A.7.4).

Corollary A.7.1

If $p = 1$, then $W_1(\mu,\nu)$ is the L^1 distance between the cumulative distribution function. In other words

$$W_1(\mu,\nu) = \int_{\mathbb{R}} |F_\mu(x) - F_\nu(x)| dx. \tag{A.7.8}$$

Proof. Let $\mathcal{A} \subset \mathbb{R}^2$ defined by

$$\begin{aligned}\mathcal{A} :=& \{(x,t),\ \min(F_\mu(x), F_\nu(x)) \le t \le \max(F_\mu(x), F_\nu(x)),\ x \in \mathbb{R}\} \\ =& \{(x,t),\ \min(F_\mu^{-1}(t), F_\nu^{-1}(t)) \le x \le \max(F_\mu^{-1}(t), F_\nu^{-1}(t)),\ t \in]0,1[\}.\end{aligned}$$

By Fubini's Theorem, we have

$$\mathrm{Vol}(\mathcal{A}) = \int_{\mathbb{R}} \int_{\min(F_\mu(x),F_\nu(x))}^{\max(F_\mu(x),F_\nu(x))} dtdx = \int_0^1 \int_{\min(F_\mu^{-1}(t),F_\nu^{-1}(t))}^{\max(F_\mu^{-1}(t),F_\nu^{-1}(t))} dxdt.$$

Since $\max(a,b) - \min(a,b) = |a-b|$, we have by (A.7.4)

$$\begin{aligned}W_1(\mu,\nu) &= \int_0^1 |F_\mu^{-1}(t) - F_\nu^{-1}(t)|dt = \int_0^1 (\max(F_\mu^{-1}(t), F_\nu^{-1}(t)) - \min(F_\mu^{-1}(t), F_\nu^{-1}(t)))dt \\ &= \int_0^1 \int_{\min(F_\mu^{-1}(t),F_\nu^{-1}(t))}^{\max(F_\mu^{-1}(t),F_\nu^{-1}(t))} dxdt = \int_{\mathbb{R}} \int_{\min(F_\mu(x),F_\nu(x))}^{\max(F_\mu(x),F_\nu(x))} dtdx \\ &= \int_{\mathbb{R}} (\max(F_\mu(x), F_\nu(x)) - \min(F_\mu(x), F_\nu(x)))dx \\ &= \int_{\mathbb{R}} |F_\mu(x) - F_\nu(x)|dx.\end{aligned}$$

■

To compute numerically the Wasserstein metric W_p between μ and $\nu \in \mathcal{P}_p(\mathbb{R})$, we consider that μ and ν are linear combinations of Dirac masses:

$$\mu := \sum_{i=1}^n a_i \delta_{x_i} \qquad \text{and} \qquad \nu := \sum_{j=1}^m b_j \delta_{y_j}$$

with $\sum a_i = \sum b_j = 1$ and where $(x_i)_{1\le i\le n}$, $(y_j)_{1\le j\le m}$ are two families of points of $\mathbb{R}$ such that $x_1 < x_2 < \cdots < x_n$ and $y_1 < y_2 < \cdots < y_m$. It is so easily to compute the cumulative distribution functions F_μ and F_ν. If we set $s_k := \sum_{i=1}^k a_i$ and $s'_k := \sum_{i=1}^k b_i$, then we have $F_\mu(x) = s_k$ if $x \in [x_k, x_{k+1}[$ and $F_\nu(y) = s'_k$ if $y \in [y_k, y_{k+1}[$ where in both cases $x_{n+1} = y_{n+1} = +\infty$. F_μ and F_ν are therefore step functions, so by the formula which defines the generalized inverse (A.7.3), we have an explicit formula for F_μ^{-1} and F_ν^{-1}. We have $F_\mu^{-1}(t) = x_k$ if $t \in [s_{k-1}, s_k[$ and $F_\mu^{-1}(t) = y_k$ if $t \in [s'_{k-1}, s'_k[$ with for convention $s_0 = s'_0 = 0$. F_μ^{-1} and F_ν^{-1} are also step functions, but on two different subdivisions.

Calculating $W_p(\mu,\nu)$ consists of integrating the step function $t \in [0,1[\longmapsto |F_\mu^{-1}(t) - F_\nu^{-1}(t)|^p$. We can represent a step function $f : x \in [a,b[\longmapsto f(x_{i-1})$ if $x \in [x_{i-1}, x_i[$ on a subdivision $a = x_0 < x_1 < x_2 < \cdots < x_n < b$ by an array with two rows and n columns which gives $f(x_i)$ on the first line and x_{i-1} on the second line. Which gives for F_μ^{-1} and F_ν^{-1}

$F_\mu^{-1}(t)$	x_1	x_2	x_3	$\cdots$	x_n
t	$s_0 = 0$	s_1	s_2	$\cdots$	s_{n-1}

and

$F_\nu^{-1}(t)$	y_1	y_2	y_3	$\cdots$	y_m
t	$s'_0 = 0$	s'_1	s'_2	$\cdots$	s'_{m-1}

To obtain the difference $F_\mu^{-1} - F_\nu^{-1}$, we must sort the family of points $(t_i)_{0\le i\le n+m-2}$ where $t_0 = 0$, $t_i = s_i$ for $i \in [\![1, n-1]\!]$ and $t_i = s'_{i-(n-1)}$ for $i \in [\![n, n+m-2]\!]$. The array corresponding to F_μ^{-1} on the new subdivision $t_0 = 0 < t_1 \le t_2 \le \cdots \le t_{n+m-2} < 1$ is

$F_\mu^{-1}(t)$	x_1	x_1	x_2	x_2	x_3	$\cdots$	x_n
t	$t_0 = 0$	$\cdots$	s_1	$\cdots$	s_2	$\cdots$	t_{n+m-2}

and the array corresponding to F_ν^{-1} on the new subdivision $t_0 = 0 < t_1 \leq t_2 \leq \cdots \leq t_{n+m-2} < 1$ is

$F_\nu^{-1}(t)$	y_1	y_1	y_2	y_2	y_3	$\cdots$	y_m
t	$t_0 = 0$	$\cdots$	s'_1	$\cdots$	s'_2	$\cdots$	t_{n+m-2}

We obtain by (A.7.4) that

$$W_p(\mu,\nu) = \left(\sum_{i=1}^{n+m-1} (t_i - t_{i-1}) |F_\mu^{-1}(t_{i-1}) - F_\nu^{-1}(t_{i-1})|^p \right)^{1/p} \tag{A.7.9}$$

with $t_{n+m-1} = 1$. In the particular case where μ and ν are two empirical measures, in other words if $n = m$ and $a_i = b_i = 1/n$ for all i, then we have $s_k = s'_k = k/n$ for all k. So F_μ^{-1} and F_ν^{-1} are two step functions on the same subdivision. Thus, $W_p(\mu,\nu)$ is computed more easily and (A.7.9) becomes

$$W_p(\mu,\nu) = \left(\frac{1}{n} \sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p}.$$

Appendix B

Recalls on Ordinary Differential Equations

Given continuous functions a, $b : I \longrightarrow \mathbb{R}$ where I is an open interval, and constants $t_0 \in I$, $y_0 \in \mathbb{R}$. We consider the initial value problem

$$\begin{cases} y' = a(t)y + b(t), \\ y(t_0) = y_0. \end{cases} \tag{B.1}$$

In this appendix chapter, we give two results on the first order linear ODE (B.1) that we will use as well as their proof. The first result is the Duhamel's formula and the second is the Gronwall's Lemma.

B.1 Duhamel's Formula

The following result gives an explicit formula for the solution of the first order linear ODE (B.1).

Proposition B.1.1 (Duhamel's formula)

Then the initial value problem (B.1) has an unique solution y on I, given by

$$y(t) = y_0 \exp\left(\int_{t_0}^{t} a(s)ds\right) + \int_{t_0}^{t} b(s) \exp\left(\int_{s}^{t} a(u)du\right) ds. \tag{B.1.1}$$

Proof. Let $t_0 \in I$ and $y_0 \in \mathbb{R}$. We define the function y by

$$y : t \in I \longmapsto y(t) := y_0 \exp\left(\int_{t_0}^{t} a(s)ds\right) + \int_{t_0}^{t} b(s) \exp\left(\int_{s}^{t} a(u)du\right) ds.$$

It is clear that $y(t_0) = y_0$ and that y is derivable on I since a and b are continuous. So by derivating y we have

$$\begin{aligned} y'(t) &= a(t)y_0 \exp\left(\int_{t_0}^{t} a(s)ds\right) + b(t) \exp\left(\int_{t}^{t} a(u)du\right) + \int_{t_0}^{t} b(s)a(t) \exp\left(\int_{s}^{t} a(u)du\right) ds \\ &= a(t)\left[y_0 \exp\left(\int_{t_0}^{t} a(s)ds\right) + \int_{t_0}^{t} b(s) \exp\left(\int_{s}^{t} a(u)du\right) ds\right] + b(t) \\ &= a(t)y(t) + b(t). \end{aligned}$$

So y is a solution of (B.1) and the uniqueness follows directly from the Cauchy-Lipschitz Theorem. ■

The formula (B.1.1) is obtained by applying the method of variation of parameter (or variation of constant) to the first order linear ODE (B.1). The name "Duhamel" refers to the mathematician Jean-Marie Duhamel. This formula (B.1.1) is sometimes called Duhamel's principle. Note that the method of the variation of parameters gives an explicit formula for the solution of a linear ODE whatever its order.

B.2 Gronwall's Lemma

The Gronwall's Lemma, or the Gronwall's inequality, allows to bound a function satisfying a certain differential or integral inequality by the solution of the corresponding differential or integral equation. There exist two forms of the Lemma, a differential form and an integral form. We give in this section the integral form with its complete proof.

Lemma B.2.1 (Gronwall's Lemma)

Let $I = [a, b]$, $b \in]a, +\infty[$ an interval on $\mathbb{R}$. Let α, β and y be real valued continuous functions defined on I, and assume that β is non-negative. If y satisfies the integral inequality

$$y(t) \leq \alpha(t) + \int_a^t \beta(s)y(s)ds \qquad \forall t \in I, \tag{B.2.1}$$

then for all $t \in I$,

$$y(t) \leq \alpha(t) + \int_a^t \alpha(s)\beta(s) \exp\left(\int_s^t \beta(u)du\right) ds. \tag{B.2.2}$$

This result is named after Thomas Hakon Gronwall who first showed it in 1919 when α and β are [51] constants. The integral form in the more general case was proved by Richard Bellman in 1943 [7]. The proof of the Gronwall's Lemma (Lemma B.2.1) uses the following result which shows that a function satisfying a first order linear differential inequality can be bounded by the solution of the considered differential equation whose expression is given by the Duhamel formula (B.1.1).

Lemma B.2.2

Let $I = [a, b]$, $b \in]a, +\infty[$ an interval on $\mathbb{R}$. Let α, β and y be real valued continuous functions defined on I. If y is derivable on $]a, b[$ and satisfies the differential inequality

$$y'(t) \leq \alpha(t)y(t) + \beta(t) \qquad \forall t \in]a, b[, \tag{B.2.3}$$

then for all $t \in I$,

$$y(t) \leq y(a) \exp\left(\int_a^t \alpha(s)ds\right) + \int_a^t \beta(s) \exp\left(\int_s^t \alpha(u)du\right) ds. \tag{B.2.4}$$

Proof. For $t \in I$, we set

$$f(t) := y(t) \exp\left(-\int_a^t \alpha(s)ds\right) - \int_a^t \beta(s) \exp\left(-\int_a^s \alpha(u)du\right) ds.$$

Since α and β are continuous on I, and y is derivable on $]a, b[$, then f is derivable on $]a, b[$ and we have by (B.2.3)

$$\begin{aligned} f'(t) &= y'(t) \exp\left(-\int_a^t \alpha(s)ds\right) - \alpha(t)y(t) \exp\left(-\int_a^t \alpha(s)ds\right) - \beta(t) \exp\left(-\int_a^t \alpha(s)ds\right) \\ &\leq (\alpha(t)y(t) + \beta(t)) \exp\left(-\int_a^t \alpha(s)ds\right) - \alpha(t)y(t) \exp\left(-\int_a^t \alpha(s)ds\right) - \beta(t) \exp\left(-\int_a^t \alpha(s)ds\right) \\ &= 0. \end{aligned}$$

So $f'(t) \leq 0$ for all $t \in I$ which implies that $f(t) \leq f(a)$ for all $t \in I$. So using the definition of f, we have

$$y(t) \exp\left(-\int_a^t \alpha(s)ds\right) - \int_a^t \beta(s) \exp\left(-\int_a^s \alpha(u)du\right) ds \leq y(a)$$

and (B.2.4) follows. ■

Proof of Lemma B.2.1. For $t \in I$, we set

$$f(t) := \int_a^t \beta(s)y(s)ds.$$

Since β and y are continuous on I, then f is derivable on $]a, b[$. Moreover, the non-negativity of β and (B.2.1) results in $f'(t) \leq \beta(t)f(t) + \alpha(t)\beta(t)$. So by the Lemma B.2.2, we have

$$\int_a^t \beta(s)y(s)ds \leq \int_a^t \alpha(s)\beta(s) \exp\left(-\int_s^t \beta(u)du\right) ds.$$

And the inequality (B.2.1) results in (B.2.2). ■

We can use the Duhamel's formula (B.1.1) to show that if y is a real-valued continuous function on $I = [a, b]$, $b \in]a, +\infty[$ which satisfies the integral equation for all $t \in I$

$$y(t) = \alpha(t) + \int_a^t \beta(s)y(s)ds,$$

with α, β real-valued continuous functions on I, then for all $t \in I$

$$y(t) = \alpha(t) + \int_a^t \alpha(s)\beta(s) \exp\left(\int_s^t \beta(u)du\right) ds. \tag{B.2.5}$$

This is the particular case of equality in the Gronwall Lemma. Note that in this particular case, the non-negativity of β is not necessary.

Bibliography

[1] Ambrosetti. A. Un teorema di esistenza per le equazioni differenziali negli spazi di Banach. Rendiconti del Seminario Matematico della Università di Padova. tome 39 (1967), p. 349-361.

[2] Ambrosio. L; Gigli. N. A user's guide to optimal transport. CIME summer school, Italy (2009).

[3] Ambrosio. L; Gigli. N; Savaré. G. Gradient flows in metrics spaces and in the space of probability measures. Lectures in Mathematics ETH Zürich. Birkäuser Verlag, Basel (2005).

[4] Aw. A; Klar. A; Rascle. M; Materne. T. Derivation of continuum traffic flow models from microscopic follow-the-leader models. SIAM Journal on Applied Mathematics, 63(1):259-278 (2002).

[5] Aw. A; Rascle. M. Resurrection of second order models of traffic flow. SIAM J. Appl. Math., 60(3):916-938 (2000).

[6] Bando. M; Hasebe. K; Nakayama. A; Shibata. A; Sugiyama. Y. Dynamical model of traffic congestion and numerical simulation. Phys. Rev. E, 51(2):1035-1042 (1995).

[7] Bellman. R. The Stability of Solutions of Linear Differential Equations. Duke Mathematical Journal, 10, 643-647 (1943).

[8] Bernoulli. D. Hydrodynamica, sive de viribus et motibus fluidorum commentarii. Strasbourg (1738). Die Werke van Daniel Bernoulli. Vol 5, Gleb K. Mikhailov (ed.), Basel (2002).

[9] Bertin. E; Droz. M; Gregoire. G. Boltzmann and hydrodynamic description for self-propelled particles. Phys.Rev. E 74, 022101 (2006).

[10] Bertin. E; Droz. M; Gregoire. G. Hydrodynamic equations for self-propelled particles: microscopic derivation and stability analysis. J.Phys. A Math.Theor.42, 445001 (2009).

[11] Bisi. M; Carrillo. J. A; Toscani. G. Contractive metrics for a Boltzmann equation for granular gases: diffusive equilibria. J.Stat.Phys.118 (2005), No.1-2, 301-331.

[12] Bolley. F; Cañizo. J.A; Carrillo. J.A. Mean-field limit for the stochastic Vicsek model. Applied Mathematics Letters, Elsevier (2012), 25 (3), pp. 339-343. hal-00563915.

[13] Boltzmann. L. Lectures on gas theory. University of California Press, Berkley (1964). Translated by Stephen. G. Brush. Reprint of the 1896-1898 Edition. Reprinted by Dover publications, (1995).

[14] Boltzmann. L. Über die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Wärmetheorie und der Wahrschleinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht. Sitzungsberichte der Akademie des Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Classe. Vol 76 (1877), pp. 373-435.

[15] Boltzmann. L. Weitere Studien über das Wärne gleichgenicht unfer Gasmoläkuler. Sitzungsberichte der Akademie der Wissenschaften 66 (1872), 275-370. Translation: Further studies on the thermal equilibrium of gas molecules, in Kinetic Theory 2, 88-174, Ed. S.G. Brush, Pergamon, Oxford (1966).

[16] Bonsal. F.F. Lectures on some Fixed-Point Theorems of Functional Analysis. Bombay (1962).

[17] Braun. W; Hepp. K. The Vlasov Dynamics and its Fluctuations in the 1/N Limit of Interacting Classical Particles. Commun. Math. Phys. 56 (1977), 101-113.

[18] Brenier. Y. Polar factorization and monotone rearrangement of vector-valued functions. In Communications on Pure and Applied Mathematics, Vol.44, pp. 375-417, Wiley (1991).

[19] Brézis. H. Analyse fonctionnelle. Théorie et applications. Masson, Paris (1983).

[20] Brown. L.D; Purves. R. Mesurable selections of extrema. The Annals of Statistics, Vol.1, No.5 (1973), 902–912.

[21] Cañizo. J.A; Carrillo. J.A; Rosado. J. A well-posedness theory in measures for some kinetic models of collective motion. Math.Models Methods Appl.Sci.21 (2011), No.3, 515-539.

[22] Carlen. E; Carvalho. M. C; Degond. P; Wenneberg. B. A Boltzmann model for rod alignment and schooling fish. Nonlinearity 28 (2015), No.6, 1783-1803.

[23] Carlen. E; Gabetta. G; Toscani. G. Propagation of smoothness and the rate of exponential convergence to equilibrium for a spatially homogeneous Maxwellian gas. Comm.Math.Phys.199 (1999), No.3, 521-546.

[24] Carrillo. J.A; Toscani. G. Contractive probability metrics and asymptotic behavior of dissipative kinetic equations. Riv.Mat.Univ.Parma (7) 6 (2007), 75–198.

[25] Carrillo. J.A; D'Orsogna. M.R; Panferov. V. Double milling in self-propelled swarms from kinetic theory. Kinetic and Related Models 2 (2009) 363-378.

[26] Cercignani. C; Illner. R; Pulvirenti. M. The mathematical theory of dilute gases. Applied Mathematical Sciences, Vol 106 (1994), Springer, 139-166.

[27] Clausius. R. Über die Art der Bewegung, welche wir Wärme nennen. Ann. Phys. 176, 353-380 (1857).

[28] Cordier. S; Pareschi. L; Piatecki. C. Mesoscopic modelling of financial markets. J. Stat. Phys. 134 (2009), no.1, 161-184.

[29] Cordier. S; Pareschi. L; Toscani. G. On a kinetic model for a simple market economy. J. Stat. Phys. 120 (2005), 253-277.

[30] Cucker. F; Smale. S. Emergent behavior in flocks. IEEE Trans. Automat. Control, 52 (2007), 852-862.

[31] Cucker. F; Smale. On the mathematics of emergence. Japan. J. Math, 2 (2007), 197-227.

[32] Cuturi. M; Peyré. G. Computational Optimal Transport. Foundations and Trends in machine learning, Vol.11, No.5-6 (2019), pp. 355-607.

[33] Daganzo. C.F. Requiem for second-order fluid approximations of traffic flow. Transport. Res. B: Meth., 29(4), pp. 277-286 (1995).

[34] Degond. P; Frouvelle. A; Liu. J.G; Motsch. S; Navoret. L. Macroscopic models of collective motion and self-organization. Seminaire Laurent Schwartz. EDP et applications. Exp. No. I (2012-2013), 1-27.

[35] Degond. P; Frouvelle. A; Raoul. G. Local stability of perfect alignment for a spatially homogeneous kinetic model. J.Stat.Phys.157 (2014), No.1, 84-112.

[36] Dellacherie. C; Meyer. P.A. Probabilities and potential, vol.29 of North-Holland Mathematics Studies. North-Holland Publishing Co., Amsterdam (1978).

[37] Desvillettes. L; Villani. C. On the trend to global equilibrium for spatially inhomogeneous kinetic systems: the Boltzmann equation. Invent.Math.159 (2005), No.2, 245-316.

[38] Dimarco. G; Tosin. A; Zanella. M. Kinetic derivation of Aw-Rascle-Zhang-type traffic models with driver-assist vehicles. Journal of Statistical Physics (2022).

[39] DiPerna. R.J; Lions. P.L. On the Cauchy problem for Boltzmann equations: Global existence and weak stability. Annals of Mathematics, 130 (1989), 321-366.

[40] Dobrushin. R.L. Perturbation methods of the theory of Gibbsian fields. (1996).

[41] Dobrushin. R.L. Vlasov equations. Funct. Anal. Appl.13 (1979), 115-123.

[42] Dowson. D.C; Landau. B.V. The Fréchet distance between multivariate normal distributions. Journal of multivariate analysis 12 (1982), No.3, 450-455.

[43] Dunford. N; Schwartz. J.T. Linear Operators: Part I. Intersciences Publishers, Inc., New York (1958).

[44] Düring. B; Toscani. G. Hydrodynamics from kinetic models of conservative economies. Phys. A, 384(2):493-506 (2007).

[45] D'Orsogna. M.R; Chuang. Y.L; Bertozzi. A.L; Chayes. L. Self-propelled particles with soft-core interactions: patterns, stability, and collapse. Phys. Rev. Lett. 96 (2006) 104302-1/4.

[46] Enskog. E. Kinetische Theorie des Vorgänge in mässig verdünnten Gasen. Uppsala, Almqvist & Wiksell. Trad. Kinetic Theory of Process in Dilute Gases. In Kinetic Theory, 3, S.G. Brush, Ed. Pergamon Press, Oxford (1972), 125-225.

[47] Gabetta. G; Toscani. G; Wennberg. B. Metrics for probability distributions and the trend to equilibrium for solutions of the Boltzmann equation. J.Statist.Phys.81 (1995), No.5-6, 901-934.

[48] García. F; Oncina. L; Orihuela. J; Troyanski. S. Kuratowski's Index of Non-Compactness and Renorming in Banach Spaces. Journal of Convex Analysis. Volume 11, No. 2, 477-494 (2004).

[49] Gazis. D.C; Herman. R; Rothery. R.W. Nonlinear follow-the-leader models of traffic flow. Oper. Res., 9:545-567 (1961).

[50] Golse. F. The Boltzmann Equation and Its Hydrodynamic Limits. In Handbook of Differential Equations. Evolutionary Equations, Vol 2. C.M. Dafermos and E. Feireisl eds. 161-301, Elsevier B.V. (2005).

[51] Gronwall. T.H. Note on the Derivatives with Respect to a Parameter of Solutions of a System of Differential Equations. Annals of Mathematics, 20, 292-296 (1919).

[52] Ha. S.Y; Lee. K; Levy. D. Emergence of time-asymptotic flocking in stochastic Cucker-Smale system. Comm. Math. Sci., 7:453-469 (2009).

[53] Ha. S.Y; Tadmor. E. From particle to kinetic and hydrodynamic descriptions of flocking. Kinetic and Related Models 1:415-435 (2008).

[54] Hilbert. D. Bergründung der kinetischen Gastheorie. Math. Annalen 72 (1912), 562-577. Trad. Foundations of the Kinetic Theory of Gases. In Kinetic Theory, 3, S.G. Brush, Ed. Pergamon Press, Oxford (1972), 89-101.

[55] Jiang. N; Xiong. L; Zhang. T.F. Hydrodynamic limits of the kinetic self-organized models. SIAM J. Math. Anal. 48 (2016), pp. 3383-3411.

[56] Kantorovich. L. On the translocation of masses. Dokl. Akad. Nauk. USSR 27 (1942), 199-201. English translation in J. Math. Sci. 133, 4 (2006), 1381-1382.

[57] Kirkpatrick. M; Barton. N.H. Evolution of a species' range. Amer. Nat. 150 (1), 1-23 (1997).

[58] Landau. L; Lifshitz. E. Course of Theoretical Physics. Vol 6, Fluid Mechanics, Pergamon Press, Oxford (1987).

[59] Lanford III. O.E. Time evolution of large classical systems. In dynamical systems, theory and applications. Springer, Berlin (1975), pp. 1-111.

[60] Lightill. M.J; Whitham. J.B. On kinematic waves. I: Flow movement in long rivers. II: A Theory of traffic flow on long crowded roads, Proc. Royal Soc. London Ser. A, 229, pp. 281-345 (1955).

[61] Maxwell. J.C. Illustrations of the dynamical theory of gases. Part I. On the motions and collisions of perfectly elastic spheres. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. Vol 19 (1860), 19-32.

[62] Maxwell. J.C. Illustrations of the dynamical theory of gases. Part II. On the process of diffusion of two or more kinds of moving particles among one another. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. Vol 20 (1860), 21-37.

[63] Maxwell. J.C. On the dynamical theory of gases. Philos. Trans. Roy. Soc. London Ser. A 157 (1867), 49-88.

[64] Mirrahimi. S; Raoul. G. Population structured by a space variable and a phenotypical trait. Theor. Popul. Biol. 84, 87-103 (2013).

[65] Monge. G. Mémoire sur la théorie des déblais et des remblais. In Histoire de l'Académie Royale des Sciences de Paris (1781), pp. 666-704.

[66] Neunzert. H. An introduction to the nonlinear Boltzmann-Vlasov equation. Lecture notes in Math. 1048. Springer, Berlin (1984).

[67] Newell. G.F. Nonlinear effects in the dynamics of car-following. Op. Res., 9(2):209-229 (1961).

[68] Panaretos. V; Zemel. Y. An invitation to statistics in Wasserstein space. SpringerBriefs in Probability and Mathematical Statistics (2020).

[69] Payne. H.J. Models of Freeway Traffic and Control. Mathematical Models of Public Systems, Simulation Council Proc. 28, Vol. 1, pp. 51-61 (1971).

[70] Pipes. L.A. An operational analysis of traffic dynamics. J. Appl. Phys., 24(3):274-281 (1953).

[71] Prokhorov. Y.V. Convergence of random processes and limit theorems in probability theory. Theory of Probability & Its Applications. vol. 1, no 2 (1956), p. 157-214.

[72] Raoul. G. Exponential convergence to a study-state for a population genetics model with sexual reproduction and selection. arXiv preprint. arXiv:2104.06089 (2021).

[73] Raoul. G. Macroscopic limit from a structured population model to the Kirkpatrick-Barton model. arXiv preprint. arXiv:1706.04094 (2017).

[74] Repovs. D; Semenov. P.V. Continuous selections of Multivalued Mappings. Mathematics and its Applications. Springer (1998).

[75] Richards. P.I. Shock waves on the highway. Oper. Res. 4, pp. 42-51 (1956).

[76] Rudin. W. Real and complex analysis, third ed. McGraw-Hill Book Co., New York (1987).

[77] Santambrogio. F. Optimal transport for applied mathematicians. Birkhauser (2015).

[78] Schauder. J. Der Fixpunktsatz in Funktionalramen. Studia Math. Vol 2 (1930), 171-180.

[79] Sznitman. A.S. Topics in propagation of chaos. Lecture Notes in Math. 1464, Springer, Berlin (1991).

[80] Tordeux. A; Seyfried. A. Collision-free nonuniform dynamics within continuous optimal velocity models. Phys. Rev. E, 90:042812 (2014).

[81] Toscani. G; Villani. C. Probability metrics and uniqueness of the solution to the Boltzmann equation for a Maxwell gas. J.Statist.Phys.94 (1999), No.3-4, 619-637.

[82] Toscani. G; Villani. C. Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation. Comm.Math.Phys.203 (1999), No.3, 667-706.

[83] Tychonov. A. Ein Fixpunktsatz. Math. Ann. 111 (1935), pp. 767-776.

[84] Ukai. S. On the existence of Global solutions of mixed problem for non-linear Boltzmann equation, 50 (1974), 179-184.

[85] Vasershtein. L. Markov process over denumerable products of spaces describing large system of automata. Problemy Peredači Informacii 5,3 (1969), 64-72.

[86] Vicsek. T; Czirók. A; Cohen. I; Ben-Jacob. E; Shochet. O. Novel type of phase transition in a system of self-driven particles. Phys. Rev. Lett. 75, (1995), 1226-1229.

[87] Villani. C. A review of mathematical topics in collisional kinetic theory. In Handbook of mathematical fluid dynamics. Vol 1. North-Holland, Amsterdam (2002), pp. 71-305.

[88] Villani. C. Optimal transport: Old and new. Grundlehren der Mathematischen Wissenschaften, Vol. 338. Springer, Berlin, (2009).

[89] Villani. C. Topics in optimal transportation. Graduate Studies in Mathematics, Vol.58. American Mathematical Society, Providence (2003).

[90] Villani. C. Transport optimal de mesure: Coup de neuf pour un très vieux problème. Images des mathématiques (2004), pp. 114-119.

[91] Vlasov. A. On the Vibrational Properties of an Electron Gas. Journal of Experimental and Theoretical Physics. Vol 8, p. 291 (1938).

[92] Whitham. G.B. Linear and Nonlinear Waves. Pure and Applied Math., Wiley-Interscience, New York, (1974).

[93] Zhang. M.H. A non-equilibrium traffic model devoid of gas-like behavior. Transportation Res. Part B, 36(3):275-290 (2002).