Acknowledgments

My gratitude goes first to my directors Hasnaa Zidani and Nicolas Forcadel. Nicolas, for your unalterable stability and efficiency, and your patience in listening to me despite your countless burdens. I take model on your way of training messy yearlings, I owe you much of my background since... well, Hilbert spaces. Hasnaa, for sharing with me your unbounded knowledge and your experience. Your liking of optimal control is contagious, as seen by the successive generations of your students. It is a privilege to benefit from your open-mindedness and confidence, which really meant something at times. Thank you both for these last years, for all that you asked and gave me.

To Nicola Gigli and Quentin Mérigot, thank you for the time that you devoted to reading this manuscript. To Yves Achdou, Pierre Cardaliaguet, Chloé Jimenez and Filippo Santambrogio, for your presence as examinators. This thesis relies on some of your contributions, it is a much-felt honor to count each of you in my jury.

To Cristopher Hermosilla, who welcomed me in the Universidad Santa María during six fruitful months, and Pedro Gajardo, who made this possible. Cristopher, thank you for your guidance and precious collaboration during my stay in Chile and afterwards – and I wish a long life to CMM visiting programs.

A special thanks to François Delarue and Yves Achdou, who accepted to belong to my follow-up committee and fulfilled this duty very seriously; in the same vein, to Laurent Vercouter for our conversations.

To Olivier, for your continuous support; Giulia, your sharp sagacity; of course, to my comrade Othmane, your uncompromising maths and our heated debates, it has been a pleasure to be your first student.

To the professors turned colleagues, your energy as teachers and our lengthy discussions; Witold for the maths and its complement, Antoine for so many projects and the Von Kármán alley – *it worked!*, Carole for your exactedness, Christian for trapping me into doing a PhD, Ioana for the probas and fairness, Bruno for the argumented coffees, Lionel for joining the Julia side, and the rest of the math department.

Aux indispensables personnels de l'INSA, Brigitte, Momo et ses précieuses léocartes d'urgence, Élodie et Catherine que j'ai tant fait travailler, Moktar, Assié, Pascal pour les rails de poster, et beaucoup d'autres.

Fernanda y Piero. Saben[1]. ♡

Fellows from Rouen and Valparaíso: Augustin – thank you for listening to my (at least ten) rantings, Emma, Zoé, Théau and Nathan the GM veterans, my HJ siblings Rita and Ali – go on Ali, you're a good PhD student, the next gen Sara, Diane, Guillaume and Boris, the previous gen Étienne and Cyril – don't worry, the student seminar is in good hands. Audrey, Nessim, Sam e il egregio Silvio from the other side of the road, thank you for opening the doors of the university, hihi. Bryan el compañero, Javier y sus atenciones, Abraham y Tobias, Sergio – nunca olvidaré tu reactividad, Carolina, John y Enzo, he visto con placer que siguen kafeminariando. Ernesto, mereces lo que vives, disfruta y espero verte siempre tan apasionado.

Aux compagnons hérités de vies antérieures : Ben, Sousou et la Loukoumotive au grand complet, Daou (matheux yi ay doff legn), Clarisse, Marie, Théo et Aymeric pour vos histoires qui me sortent des miennes.

À ma large famille, son entrain communicatif et la légendaire distraction Aussedat. Promis papa, ce n'est pas un canular parodiant les publications scientifiques au moyen d'une débauche typographique jamais vue. J'espère.

Lastly, to all the researchers that are supporting the community with Sisyphean efforts, from young ambitious writing good lecture notes to old crocodiles smiling in seeing them. Perhaps the generations to come will not perceive the involvement behind the names written on standard references. To us, math apprentices of these days, the name of Santambrogio means grandiloquent presentations with knocking and scribbling on all blackboards simultaneously, storming and waving in laudable attempts to pass something to the audience. As that of Pierre Cardaliaguet, Guillaume Carlier, Gabriel Peyré, Jean-Michel Roquejoffre, ... and the many others spending a considerable energy on organising meetings and courses. *It helped me immensely.* To you, bored PhD student that will not read anything else than these acknowledgments[2], I suggest to take example on the common denominator, *la substantifique moelle* of their attitude: ingenuity for mathematics, dedication for people.

[1] Las confs, cuchuflis, la camioneta, the fat Cantor, Santiago, este weón de gato, Cahors, en verdad por compartir temas y tiempo.

[2] Here's how to lose more time. follow linearly | shameful jewels | read miscs | 8:38-8:55 | heavy | oldy | now | get | back | to | work

Abstract

The thesis studies optimal control problems in some spaces that are not vector spaces, with a focus on the link with Hamilton-Jacobi-Bellman equations understood in the viscosity sense. The red wire is the control of a population of drivers in a traffic network. At first, we focus on a single driver, addressing the difficulty of the lack of regularity of the ambient space. We propose a framework for Cauchy-Lipschitz control problems in CAT(0) spaces, in which we are able to give sufficient conditions for the existence of an optimal control, and characterize the value function as the unique viscosity solution to a Hamilton-Jacobi-Bellman equation. Secondly, we consider a probability measure evolving on the Euclidean space, representing a population of drivers. We obtain a comparison principle that is applicable to the control of such a population, that we prove in more generality in spaces with curvature bounded from below. Thirdly, we provide a first step towards the treatment of populations evolving on networks, by proving that the squared Wasserstein distance over a network is directionally differentiable. In formulating the Hamilton-Jacobi-Bellman equation in the Wasserstein space using solely the metric structure, one needs some technical argument to use continuity equations as characteristics; this is developed in a last chapter, focussing in more details about the geometry induced by optimal transport on measures.

Keywords: Optimal control, Hamilton-Jacobi equations, Viscosity solutions, Wasserstein distances, Helmholtz-Hodge decomposition, Metric spaces, Nonlinear partial differential equations.

Mots-clefs : Théorie de la commande, Équations de Hamilton-Jacobi, Solutions de viscosité, Distances de Wasserstein, Théorème de Helmholtz-Hodge, Espaces métriques, Équations aux dérivées partielles non linéaires. Contrôle optimal, Décomposition de Helmholtz-Hodge.

Résumé

Cette thèse a pour objet principal l'étude de problèmes de contrôle optimal sur des espaces qui ne sont pas vectoriels. Nous décrivons ici le contenu du manuscrit, qui se divise en quatre chapitres.

Un problème de contrôle optimal consiste à se donner une famille de courbes, toutes partant d'un même point et définies sur un même intervalle de temps, solutions d'une même équation différentielle dont la dynamique est supposée paramétrée par une variable dite de contrôle. L'objectif est de déterminer quels sont les contrôles permettant d'optimiser un certain critère dépendant de la trajectoire. Cet énoncé général couvre une gamme de situations courantes, et a été largement étudié dans le cadre euclidien où la position est donnée par des coordonnées cartésiennes. Le but de cette étude est d'étendre les techniques connues pour certains cas d'intérêt où la paramétrisation de la position par des coordonnées euclidiennes est inadaptée.

Plus précisément, la situation qui motive ce travail est celle du trafic routier. Les routes s'organisent en réseau, que l'on peut raisonnablement modéliser par des segments de droite reliés entre eux en certains points de jonction. Cet espace peut être représenté comme un sous-espace d'un plan euclidien, via une carte routière, et les problèmes de contrôle peuvent se formuler directement sur la carte avec la contrainte additionnelle de rester sur la route. Cette idée présente deux inconvénients : premièrement, la contrainte est très irrégulière, forçant les trajectoires à se maintenir dans un ensemble d'intérieur vide ; deuxièmement, une méthode numérique sur la carte devra mailler des zones inutiles, non parcourues par des routes, engendrant un coût additionnel. Si l'on considère maintenant le réseau comme un espace muni de sa propre métrique, la formulation du problème n'est plus contrainte, et les maillages sont en général réduits. Cependant, l'espace n'est plus homogène, au sens où la structure du voisinage d'un point change radicalement entre l'intérieur d'un segment de droite ou une zone de jonction. Au niveau théorique, ceci créé des problèmes pour adapter les techniques de résolution auxquelles nous nous intéresserons.

Dans une première partie, nous proposons un cadre de résolution de ces problèmes, formulé sur la famille des espaces dits CAT(0), qui contient les réseaux sans boucles, leurs équivalents à plusieurs dimensions, certaines variétés à courbure négative, et les espaces euclidiens. Ces espaces ne peuvent pas représenter un système routier à l'échelle macroscopique, puisqu'ils ne permettent pas d'avoir plusieurs chemins de longueur minimale reliant deux points donnés, mais sont parfaitement adaptés à leur étude

locale. Ils jouissent en outre d'une grande popularité dans la littérature des flots de gradient, ce qui donne accès à des résultats d'existence, d'unicité et de stabilité de trajectoires guidées par ce type précis d'équations différentielles. En nous appuyant sur ces résultats, ainsi que sur une théorie abstraite des équations différentielles dans les espaces métriques, nous obtenons un formalisme pour les équations contrôlées qui généralise le cadre euclidien, et permet de formuler un problème de contrôle.

Le premier apport de notre travail est de fournir une condition suffisante sur la dynamique sous laquelle le problème de contrôle admet une solution. Pour ceci, il nous faut montrer que l'ensemble des courbes qui définissent notre problème de contrôle est compact, de manière à ce qu'une suite de courbes de plus en plus proches de réaliser l'optimum permette de construire une courbe optimale par extraction. L'essence du problème est de montrer qu'une limite de solutions de l'équation contrôlée restera elle-même solution. Dans le cadre hilbertien, ce passage est donné par une limite faible de la dynamique, ce qui n'est pas bien défini pour notre situation. Par contre, sous certaines hypothèses de régularité, il est possible de reformuler le système contrôlé via des inégalités variationnelles évolutives, qui sont elles-mêmes linéaires par rapport à la dynamique en un certain sens. En exploitant cette linéarité, et sous une hypothèse naturelle de convexité des valeurs de la dynamique, nous pouvons passer à la limite faible, et conclure.

Le second apport sur ce sujet est la mise en lumière du lien avec les équations de Hamilton-Jacobi-Bellman, déjà abordées dans ces espaces sous l'angle des solutions de viscosité. Dans le cas euclidien, ces équations sont très bien connues, et il est établi qu'une fonction auxiliaire du problème de contrôle, appelée fonction valeur, est solution au sens de viscosité. Ce dernier terme fait référence à une théorie d'existence et d'unicité prenant ses racines dans les limites d'approximations visqueuses, d'où son nom, mais dont la formulation aboutie est indépendante de toute approximation. Cette théorie s'adapte bien à des cadres plus larges, et nous montrons que la fonction valeur du problème de contrôle que nous considérons est également l'unique solution d'une équation de Hamilton-Jacobi-Bellman. Pour conclure cette première partie, nous proposons une résolution numérique sur certains exemples académiques.

La seconde partie de notre travail considère une population d'agents, modélisée d'une manière qui sera reprise dans toutes les parties suivantes, et que nous détaillons un peu. L'objectif est de pouvoir traiter simultanément le cas d'un seul conducteur, d'un nombre fini de conducteurs, ou d'une population tellement dense qu'elle est représentée par un fluide dont les particules sont indistinctes. Cette population évolue au cours du temps, et pour mesurer la distance parcourue entre deux états différents, nous utilisons une distance dite de Wasserstein. Pour la calculer, on considère les mesures comme des populations d'agents anonymes, et on cherche à identifier chaque agent de la population initiale à un agent de la population terminale de manière à minimiser la moyenne de la distance parcourue par chacun des conducteurs. La distance elle-même sera donnée par cette moyenne optimale. Ce choix permet de travailler dans un espace de mesures muni de sa distance, et de formuler des problèmes de contrôle directement sur les courbes de populations.

Dans cet espace métrique, nous proposons une extension des solutions de viscosité qui nous permet d'obtenir un principe de comparaison pour les équations de Hamilton-Jacobi. Ce principe implique en particulier l'unicité des solutions. Comparativement à la littérature existante, nous utilisons principalement des arguments métriques, ce qui permet de donner le principe de comparaison sur une classe plus large d'espaces partageant les mêmes propriétés métriques de courbure, appelés espaces CBB(0). En contrepartie de cette plus grande généralité, il n'est pas clair que cette notion soit stable.

Dans un second temps, nous revenons à l'espace des mesures pour étudier les problèmes de contrôle optimal, et établissons le lien entre la fonction valeur et les équations de Hamilton-Jacobi-Bellman. Ici, les agents eux-mêmes évoluent dans un espace euclidien. Ce travail diffère du chapitre précédent au niveau des trajectoires, qui ne sont pas définies en relation avec la structure d'espace courbe, mais via des solutions au sens des distributions apparemment sans lien avec la distance de Wasserstein. En conséquence, nous devons prêter une attention particulière aux vitesses des trajectoires, pour résoudre un point qui n'a de sens que dans cet espace. Ce chapitre se termine sur certaines extensions du problème de contrôle.

La troisième partie vise à entamer l'étude des problèmes de contrôle sur des populations évoluant sur des réseaux. Un des points fondamentaux des deux chapitres précédents est la possibilité de construire facilement des fonctions directionnellement différentiables, et d'utiliser ces dérivées directionnelles en lieu et place des habituels gradients. Si les conducteurs évoluent maintenant sur un réseau, ces constructions ne sont plus aussi directes. Nous montrons que le carré de la distance de Wasserstein est bien directionnellement différentiable, et nous donnons l'expression explicite de la dérivée directionnelle. Les réseaux que nous considérons peuvent admettre des boucles, à la différence du premier chapitre. Ces boucles créent des points de discontinuité de la dérivée directionnelle de la distance sous-jacente, qui doivent être considérés à part. Il n'est pas clair que les arguments employés puissent être réutilisés pour des espaces plus généraux, par exemple en dimension supérieure. Ceci conclut nos contributions liées aux problèmes de contrôle.

La quatrième et dernière partie de la thèse est plus spécifiquement centrée sur l'espace des mesures muni de la distance de Wasserstein, dans le cas où les agents évoluent dans un domaine euclidien. Cet espace est muni de plusieurs structures superposées : c'est un espace métrique courbe, avec ses géodésiques et son espace tangent ; c'est un sous-ensemble convexe de l'espace vectoriel des mesures signées ; c'est enfin une fermeture possible de l'ensemble des fonctions positives d'intégrale 1, ce qui mène à y formuler des équations différentielles. Ces différentes propriétés ont permis de construire des opérations semblables aux opérations euclidiennes d'addition de champs de vecteurs. En utilisant ce calcul, nous donnons les arguments techniques employés précédemment pour traiter les problèmes de contrôle sur l'espace des mesures.

Il se trouve que ces arguments ne s'étendent pas à toutes les équations différentielles ordinaires que l'on pourrait considérer dans l'espace des mesures, mais seulement à celles pour lesquelles la dynamique est de la forme précise venant des densités. Nous donnons un contre-exemple, et quelques cas particuliers dans lesquels ce comportement irrégulier n'apparaît pas. La caractérisation la plus précise que nous proposons n'est valide qu'en dimension un, et s'appuie sur les résultats du transport optimal, qui sous-tend la définition de la distance de Wasserstein. En dimension supérieure, nous avons obtenu certains résultats plus faibles, ainsi qu'une décomposition générale permettant de relier les espaces tangents à la structure de la mesure sous-jacente. Ce chapitre plus exploratoire appelle à être complété dans de futurs travaux.

Contents

All chapters of this manuscript are independent, to the exception of some results of Chapter 5 used in Chapter 3. Each chapter starts by a short descriptive introduction and a detailed table of contents.

Submitted and published articles

The thesis gave rise to the following contributions:

[A1] with H. Zidani, "A Cauchy-Lipschitz setting for control problems in complete CAT(0) spaces". 2025, submitted;

[A2] with O. Jerhaoui and H. Zidani, "Viscosity solutions of centralized control problems in measure spaces". In *ESAIM Control Optimisation and Calculus of Variations* 30 (Oct. 2024), pp. 1–37;

[A3] with C. Hermosilla, "A minimality property of the value function in optimal control over the Wasserstein space". Preprint, available at https://hal.science/hal-04427139. Jan. 2024;

[A4] "On the structure of the geometric tangent cone to the Wasserstein space". Preprint, available at https://hal.science/hal-04672554. Jan. 2025.

Chapter 2 is directly derived from **[A1]**, Chapter 3 is derived from both **[A2]** and **[A3]**, and Chapter 5 is partly derived from **[A4]**.

Notations and conventions

- Given a set A, the function $\mathbb{1}_A$ is the indicator function, that takes the value 1 over A and 0 elsewhere.
- A *modulus of continuity* is a function $m : \mathbb{R}^+ \to \mathbb{R}^+$ that is continuous, nondecreasing and such that $\lim_{r \to 0} m(r) = 0$.
- The *domain* of a function $u : X \to \mathbb{R} \cup \{-\infty, +\infty\}$ is the set dom u of points $x \in X$ such that $u(x) \notin \{-\infty, +\infty\}$.
- Function spaces are denoted $F(X;Y)$, where F is the notation in use ($\mathcal{C}$ for continuous, L^p for Lebesgue spaces...), X is the domain and Y the images. The subscript F_b indicates bounded functions, while F_c denotes compactly supported ones.
- The set of Lipschitz functions from A to B is denoted $\mathbb{L}(A;B)$.
- If $f : X \to Y$ and $A \subset X$, the notation $f(A)$ denotes the set of $f(x)$ for $x \in A$.
- The notation $a_n \searrow_n b$ means that the sequence $(a_n)_n$ converges to b monotonically from above.
- A *vanishing sequence* is a strictly decreasing sequence of positive terms that converges to 0.
- An *empirical measure* is a totally atomic Borel probability measure with finitely many atoms.
- A *submeasure* ν of a nonnegative measure μ is a nonnegative measure such that $\nu \leqslant \mu$, or equivalently, $\int \varphi d\nu \leqslant \int \varphi d\mu$ for all continuous, bounded and nonnegative φ.
- The conditional measure $\mu|_A$ of a measure over a measurable set A is defined as the null measure if $\mu(A) = 0$, and $\mu(\cdot \cap A)/\mu(A)$ otherwise.

Introduction

Maybe you have witnessed, while waiting for the tram, a middle school teacher trying to herd a multitude of careless children safely through the jungle of traffic lights and crossing lanes. Then checked your emails and received the official announcement of a new prize, that you read leaning on an advertisement for the latest version of the original recipe of some popular drink. These are instances of a uniform control applied to all members of a population, in hope to reach some given objective.

This thesis is concerned with simplified models of such problems. The long-term ambition is to control a population of agents, for instance drivers of vehicles, evolving on a traffic network. This has pragmatic applications of immediate interest: prevent or fluidify traffic jams by signalization or influence through autonomous vehicles, assist the design of new infrastructures by solving inverse problems, control information or energy networks, and so on. The mathematical treatment is, however, not direct.

The simplest nontrivial model that one can consider is a population of one driver, evolving on a finite-dimensional domain assumed to be $\mathbb{R}^d$. The computation of an optimal control in this case led to the development of optimal control theory and the tightly related fields of nonlinear Partial Differential Equation (PDEs) of Hamilton-Jacobi-Bellman equations [Bel54]. The most efficient methods are developed in the linear-quadratic case, in which the computation can be transposed into the resolution of the Ricatti equation [Abo+03]. In certain cases, one can also exploit first-order optimality conditions, and base numerical methods on Pontryagin's Maximum Principle [Pon+62; Vin10; Cla13]. Thirdly, one can shift the focus towards the so-called *value function*, compute it with dedicated methods, and use it to recover solutions of the original problem; this is the Hamilton-Jacobi-Bellman approach, that we now develop.

Let us introduce some notations. The general form of Hamilton-Jacobi equations that will be considered in this manuscript reads as

$$\begin{aligned} -\partial_t v(t,x) + H(x, D_x v(t,x)) &= 0 \qquad && (t,x) \in [0,T) \times \Omega, \\ V(T,\cdot) &= \mathfrak{J} && x \in \Omega. \end{aligned} \tag{1}$$

The time horizon $T > 0$ is fixed, the domain Ω is to be understood as a space variable, the function $v : [0,T] \times \Omega \to \mathbb{R}$ is the unknown, and the function $\mathfrak{J} : \Omega \to \mathbb{R}$ is a boundary condition that we refer to as the *terminal cost* for reasons that will soon be clear. The real-valued function H is the Hamiltonian of the system, and its domain of definition is precisely one of the point of discussion. For the sake of the introduction, we may consider that H depends on the space variable x, and a formal object $D_x v$ encoding the variation in space of the function v at the point (t,x).

A common framework: viscosity solutions. The prototype of Hamiltonian in (1) is given by $H(x,p) = \frac{1}{2}|p|^2$, and in general, $H(x,p) = \sup_v -\langle p, v\rangle - L(x,v)$ for a Lagrangian L. In this case and assuming that everything is smooth, (1) is satisfied with $D_x v = \nabla_x v$ by the value function

$$V : [0,T] \times \Omega \to \mathbb{R}, \qquad V(t,x) = \inf_{\substack{\gamma \in AC([t,T];\Omega) \\ \gamma(t) = x}} \int_{s=t}^{T} L(\gamma_s, \dot\gamma_s) ds + \mathfrak{J}(\gamma_T). \tag{2}$$

The Lagrangian L encodes the geometry of the space, and $\mathfrak{J}$ is indeed a terminal cost, that can be tweaked to encode various physical problems. In the extreme case where $\mathfrak{J}(x) = 0$ if $x = \overline{x}$, and $+\infty$ otherwise, the value $V(t,x)$ is the minimal length between x and $\overline{x}$ of a curve running for a time $T - t$. In less extreme cases, $\mathfrak{J}$ can be chosen very smooth but penalizing strongly certain regions. However, as smooth as $\mathfrak{J}$ may be, the function V may develop singularities at the points x for which there are several optimal curves γ. Typically, V will resemble the crest of a mountain, that decreases on two hillsides joined at x. This lack of regularity is not an abstraction of theoretical mathematics: the reader experiences it every time that she has to choose between equally far doors of the metro. Mathematically speaking, the gradient of the value function does not exist, and one has to understand the PDE (1) in another, weaker way.

The problem is that allowing for weaker solutions can result in infinitely many solutions, as it is often the case if one requires the equation to be satisfied at almost every point. The theory of *viscosity solutions* provides a local criterion to sort out the "physical one", corresponding to (2). We detail the core ideas in Section 1.2; for this introduction, it is sufficient to consider that a viscosity solution of $H(x,u(x),D_x u(x))$ is asked to satisfy both inequalities

$$H(x,u(x),p) \leqslant 0 \qquad \text{for all } p \text{ in the superdifferential of } u \text{ at } x \text{ (subsolution condition)}, \tag{3a}$$

$$H(x,u(x),p) \geqslant 0 \qquad \text{for all } p \text{ in the subdifferential of } u \text{ at } x \text{ (supersolution condition)}. \tag{3b}$$

To be complete, this definition should be complemented with regularity, boundary conditions understood in the correct way, and the precise definition of semidifferentials. A crucial point is that the semidifferentials may be empty, and a nonsmooth function can still satisfy (3). When applied to (1), the strength of the definition (3) is that monotonicity of V with respect to $\mathfrak{J}$ translates in a monotonicity of the viscosity solutions of the equation, proved directly from the equation, and that generalizes outside of the control framework. This is the content of *comparison principles*.

Viscosity solutions were introduced by Crandall, Lions and Evans [Lio82; CL83; Eva98] using successive formulations issued from limits of viscous approximations. Shortly after, Ishii [Ish85] introduced viscosity solutions for discontinuous Hamiltonians, and discontinuous viscosity solutions in [Ish89], by imposing conditions on the upper and lower semicontinuous envelopes. The second-order case was first reached by Jensen [Jen88] using sup-convolutions: it was seen from the beginning that this operation is the analogue for viscosity solutions of the convolution for linear equations, as pushed to the extreme by Kolokoltsov and Maslov [KM97]. The presentation of the so-called User's Guide [CIL92] serves as a reference for the Euclidean case, with clear objectives of generalizations to infinite-dimensional spaces.

Convex Hamiltonians have a particular status in viscosity theory, as the main provider of examples by the Hopf-Lax semigroups [Lio82]. Barron and Jensen [BJ90] proved that in this case, it is only necessary to test the solution from one side, the one pushing the graph of the solution in the right direction (towards the positives for a minimization problem, and the negatives for a maximization problem). In parabolic problems, the two inequalities of the definition of a viscosity solution were proved by Frankowska [Fra93] to be equivalent to viability and invariance of the domains enclosed by the graph of the solution, which essentially means that the solution must increase (resp. decrease) along all (resp. one) trajectory of the underlying controlled system.

One particularity of the viscosity solution theory is its adaptability: strict viscosity solutions are allowing $\varepsilon's$ in inequalities [CIL92], much appreciated in infinite-dimension to employ Ekeland principles [Eke74]. Caffarelli [Caf89] considered $\mathcal{C}^2-$, $\mathcal{C}^{1,1}-$ and $W^{2,p}-$viscosity solutions of elliptic equations, proving different regularity estimates for each of them. This was later integrated in the L^p-viscosity solutions exposed by Caffarelli, Crandall, Kocan, and Swiech [Caf+96] to treat equations with a measurable dependence in space, also addressed by Camilli and Siconolfi [CS03; CS05]. The case of the Eikonal equation, and in general the class of equations where the unknown is constrained only through the norm of its gradient, has its dedicated definitions relying on the metric slope (Ambrosio and Feng [AF14], Gangbo and Święch [GŚ14; GŚ15b]) or growth conditions along paths (Giga, Hamamuki, and Nakayasu [GHN15]). Just for measure spaces, Table 3.1 p. 71 counts 9 different definitions of semidifferentials employed to define viscosity solutions, and does not cover the use of negative Sobolev spaces [FN12; Bur+20] or test functions [PW18; CKT23a; MZ24; SY24; BEZ24]. We refer the reader to the series [CL85; CL86b; CL86a; CL90; CL91; CL94] and the monographs [Bar94; BC97; CS04; FS06; BC24] for extensive development of the viscosity tree, and we now concentrate on the ramifications in relation with this thesis.

Towards more realistic models: crowds of drivers. Consider again our problem of controlling a population of drivers. One can allow a finite quantity of drivers in the model, and study the resulting system in large dimension, as in particle simulations. Viewing the particles as Dirac masses, one can also let their number grow to infinity, and derive an equation on the resulting density – or in general, the resulting measure. In statistical physics, this dates back to Gibbs [Gib02], with Morrey [Mor55] introducing hydrodynamic limits, formalized by Dobrushin and Siegmund-Schultze [DS82] – see also the monograph by Kipnis and Landim [KL99]. So far, systems were described by particles or empirical measures, evolving in a state space. At the turn of the millenary, Otto [Ott01] proposed to endow this state space with the distance of optimal transportation[1], and interpreted the porous medium equation as a gradient flow in this metric space. Simultaneously, optimal transport was used by Brenier [Bre01] as a way to reformulate the equations of physics. The idea to consider a *geometry* on measures, with geodesics, differentials and therefore PDEs, was developed by Ambrosio, Gigli, Savaré [AGS05; Gig08], Gangbo, Nguyen, and Tudorascu [GNT08]; at the same time, Lions [Lio06] developed a differential calculus on measures, launching the field of mean-field games. At the time of writing of this thesis, gradient flows on one side, and mean-field games on the other side, are the two main active areas in relation with differentiation in the space of measures.

Let us focus more specifically on the evolution of a measure through time. On the one hand, *densities* can be subject to conservation laws $\partial_t \rho_t + \partial_x [\varphi \circ \rho_t] = 0$, which are solved by Kružkov solutions in asking for the satisfaction of entropy inequalities [Kru70; GR91]. Kružkov solutions are in particular the limits of viscous approximations, and the methods developed for conservation laws were the basis of viscosity solutions, with a correspondence between the derivatives of the HJB flow and the solution of the conservation law fully clarified by Colombo, Perrollaz, and Sylla [CPS23]. This is currently an active theme of research, see for instance Cardaliaguet, Forcadel, Girard, and Monneau [Car+24] and references therein.

On the other hand, *measures* can be evolving under continuity equations $\partial_t \mu_t + \operatorname{div}\left(b[\mu_t]\#\mu_t\right) = 0$, solved in the sense of distributions for sufficiently regular dynamics (see [AGS05], and the recent work of Bonnet and Frankowska [BF21; BF24]). In the formal case where $\mu_t = \rho_t d\mathcal{L}$ in dimension one, the density ρ_t satisfies $\partial_t \rho_t + \partial_x (b[\rho_s d\mathcal{L}]\rho_s)$. Continuity equations are closer to ODEs in that they satisfy *superposition theorems*, usually stating that the solution at time t is the propagation of the initial measure through the flow of the underlying ODE, or a superposition of possible solutions if uniqueness does not hold, as proved by Ambrosio and Crippa [AC14]. In particular, the initial condition can be taken as a Dirac mass to recover classical ODEs. They generalize in Measure Differential Equations (MDEs) introduced by Piccoli [Pic19], in which the velocity is allowed to split mass. These are the most geometric extensions of ODEs, with currently quite few results on uniqueness and stability besides the cases that are linked to continuity equations.

The different ways of moving through the space of measures determine the relevant *tangent cone* to consider. This is especially important in the Wasserstein space, in which a "smooth" function can be defined in essentially two ways: either by following the metric definitions of Alexandrov geometry, and ask quite low regularity, or by employing the specific theories in the space of measures, resulting in much nicer functions. In the first case, the squared distance furnishes an example of a smooth function, that admits directional derivatives along the elements of the geometric tangent cone, but which may not be approximated by a "linear" function in any way. In the second case, the function is forced to vary only along certain "regular" directions, but – precisely because of this – it can be considered $\mathcal{C}^1$ in a reasonable sense. One can construct edge cases of equations with discontinuous Hamiltonian that depend non-trivially on the directional derivatives along non-regular directions, for instance as done by Ambrosio and Feng [AF14] with the metric slope. In these cases, the larger class of functions allows to get a comparison principle with relatively few efforts. It would be quite hasty to conclude that $\mathcal{C}^1$ functions are not sufficient to characterize uniqueness as well, but it is not trivial to draw the limit between the class of problems for which $\mathcal{C}^1$ functions are suited, and the others.

[1] Whereas the introduction of the optimal transport problem is undisputedly attributed to Monge, it seems that besides Kantorovich [Kan42], one could credit Graev, Arens, Eells and Hutchinson for introducing the distance that L. Vaserstein only mentioned in [Vas69] (as reported on the personal page of the latter). Half a century later, the overwhelming majority of mathematical publications on the topic opt for the (inaccurate, ill-written) name of Wasserstein distance. This is unfair but too late, as the term *Wasserstein* now carries a meaning and mental pictures by itself. We adopt the terminology of Monge-Kantorovich distance for a general cost, p−Wasserstein for the distance to the power p, omitting the prefix for $p = 2$ (see Definition 1.1.16).

Towards more realistic models: the underlying space. Back to our driving problem, we may want to get closer to reality by taking into account that cities and traffic lanes are not Euclidean spaces. A network with one-dimensional branches, connected at junctions, and endowed with the shortest path distance, is far more reasonable [GP06; D'A+10; Bre+14], and cheaper in numerical simulations. The geometry here is distorted in two ways: first, geodesics between arbitrary points may not be unique, exactly as navigation apps proposing several optimal paths. However, at small scales, this problem disappears. The second distortion cannot be removed by restriction to a sufficiently small ball: the tangent cones change in a discontinuous way. For instance, in a small ball around a junction o, all points but o have exactly two directions in their tangent cones (towards o or away from it), but o has more. This is a problem when defining $\mathcal{C}^1$ functions, since by essence, one would like to enforce a continuity of the local approximations of the function, but the local approximations are defined on spaces that are themselves not continuous with respect to the base point. These technical problems do not prevent real-life drivers to choose an itinerary, so dedicated adaptation of viscosity solutions have been introduced in networks.

The first comparison principle on a 1-dimensional network was obtained by Achdou, Camilli, Cutrì, and Tchou [Ach+13] with test functions that are piecewise smooth. The point of view is that of constrained problems, with an underlying dynamical system defined on an embedding Euclidean space, but that allows to stay on the network. The key point is the clever design of a penalization function, which grows as fast as needed along each branch as to obtain the correct sign in the critical step. Similar classes of test functions are used by Camilli, Schieborn, and Marchi [SC13; CSM13] to treat the multi-dimensional case, and the control point of view that leads to the clever design of the penalization was clarified by Imbert, Monneau, and Zidani [IMZ13]. Lions and Souganidis [LS17] emphasize the role of the Kirchhoff junction condition in the multi-junction cases. A heavy but systematic approach in the quasi-convex case is proposed by Imbert and Monneau [IM17], in which it clearly appears that this function should compensate the irregularity of the Hamiltonian, that may be discontinuous, and the space, that does not allow for $\mathcal{C}^1$ functions. This is possible – although not trivial – in a network, in which all the problems are concentrated at one point: in higher dimension, Barles, Briani, and Chasseigne [BBC13; BBC14] consider Whitney stratifications, the use of which is surveyed in [BC24]. The stratified setting has the advantage that a trajectory evolving *around* an interface can be approximated by pieces of trajectories evolving *on* the different strata, shall it be for very short times. This becomes increasingly complex as the structure of the space features more and more subdomains, and could degenerate. Motivated by this difficulty, Jerhaoui and Zidani [JZ23b] treated Hamilton-Jacobi equations in a canonical way in large classes of spaces, with test functions that are merely semiconvex or semiconcave. By "large class", we mean the so-called complete CAT(0) spaces, containing networks of one or higher dimension, metric trees, Hadamard manifolds, Euclidean buildings, and general non-positively curved geodesic spaces. The techniques of [JZ23b] are not related to optimal control, and it was an open problem to study Hamilton-Jacobi-*Bellman* equations in such spaces.

To come back to our driving example, the extension of continuity equations on a network is not exempt of its own difficulties. The main approach is a divide-and-conquer strategy, that treats each regular subdomain independently and glues them to form the network. Camilli, De Maio, and Tosin [CDMT17] obtained well-posedness for continuity equation by this way, under the assumption that the velocity orients the network, so that the gluing step can be made by induction. The difficulty lies in the proper definition of the traces at the gluing points, and how to match them. In relation with control theory however, one can consider at first that the curves of measures are "pushed forward" by the flow of an equation satisfied by a single point: this is the representation obtained for the solutions of continuity equations in $\mathbb{R}^d$, and it allows to define reasonable control problems.

However, transposing the idea of a metric space over measures, when these measures are constrained to remain on a network, is a quite steep step. The approach of $\mathcal{C}^1$ functions *à la Lions* does not apply, since the underlying space is not a vector space. The space of measures itself is geodesic, but does not have a curvature bound, and semiconcave or semiconvex test functions are not easy to construct. This manuscript brings a modest contribution in this direction, but leaves many doors open for future research.

Structure of the manuscript. Our contributions are distributed gradually in three main chapters.

– Chapter 2 focuses on networks, and in general CAT(0) spaces, to develop optimal control problems.

– Chapter 3 is concerned with spaces of measures over $\mathbb{R}^d$. Some of the results apply to the larger class of geodesic CBB(0) spaces, and are formulated at this level.
– Chapter 4 opens the way towards measures over networks.

In addition, some side questions about the geometry of optimal transport retained our attention, and grew up as to form Chapter 5. This chapter deviates from the initial topic of the PhD, even if some sections are linked to Chapter 3. We consider nevertheless that its last sections are part of the main results of this thesis.

We now describe more precisely the content of the manuscript.

I Optimal control in CAT(0) spaces

Let (Ω, d) be a complete CAT(0) space, that is, geodesic and non-positively curved in the sense of Alexandrov. The definition of curvature will be discussed in Section 1.1; for this introduction, it is sufficient to consider that CAT(0) spaces are the ones in which the squared distance is geodesically 2-convex. It implies in particular that there cannot be any loop, so this setting is adapted to small-scale study of traffic networks.

We are interested into the class of Mayer problems, that can formally be written as

$$\begin{aligned} &\text{Minimize} \quad \mathfrak{J}(y_T^{x,u}) \quad \text{over } u \in L^1(0,T;U), \\ &\text{where } \dot{y}_t^{x,u} = f(y_t^{x,u}, u(t)) \;\text{ for } t \text{ in } (0,T), \quad \text{and} \quad y_0 = x. \end{aligned} \tag{4}$$

Here U is a set of controls, $\mathfrak{J} : \Omega \to \mathbb{R}$ is again a terminal cost, and $x \in \Omega$ is fixed. The problem (4) can be written in the form of (2) by letting $L(x, v) = 0$ if $v \in f(x,U)$, and ∞ otherwise. Our motivation is to answer the following questions:

– how to define correctly the controlled dynamical system in (2)?
– Can one give conditions on the dynamic f under which the trajectories of this system form a closed subset of $\mathrm{AC}([0,T];\Omega)$?
– Can one characterize the value function as a viscosity solution of an HJB equation?

We start by the ODEs. The reader is familiar with the standard theorems of existence and uniqueness in $\mathbb{R}^d$; the essential assumption, that is not trivial to generalize in CAT(0) spaces, is the Lipschitz character of the dynamic, since there is no direct way to define a distance between elements of tangent cones attached to different points. To the knowledge of the author, at least two theories exist to generalize well-posedness of dynamical systems, both extending beyond CAT(0) spaces, and both requiring some machinery: gradient flows, and mutational analysis.

The ascension towards metric gradient flows goes back to the Hille-Yosida theorem of existence, uniqueness and contraction of flows of Cauchy problems of the form $\dot{x}_t + Ax_t = 0$, where A is accretive [Bré10]. One generalizes the implicit Euler scheme by iterations of the resolvent operator $(I + \Delta t A)^{-1}$, and proves that the limit converges to a unique curve which, when differentiable, is a strong solution. Hille-Yosida is only concerned with linear operators (although not only gradients), but the method was extended to nonlinear operators – changing name in the process – to become the Crandall-Liggett generation theorem in Banach spaces [CL71]. In the latter paper, there is no ambiguity on the definition of a strong solution, which should be a curve given by the integral of its derivative and satisfying $\dot{x}_t = -A(x_t)$ for almost every time. This makes sense in Banach spaces, since composition of translations is given by the addition of vectors, and one can *integrate*.

Such an integral representation is not available in general. However, the resolvent equation solves a minimization problem, which can be written in any metric space. Hence, at least formally, one can define a scheme by solving iteratively the optimization problem, and obtain limiting curves, as done by Mayer [May98] in CAT(0) spaces. Knowing that a curve is the limit of the scheme is not quite practical to manipulate it; the good idea is to characterize these curves by Evolutionary Variational Inequalities (EVI), that are global inequalities imposing the fastest decrease rate on the solution curve. A typical EVI for a gradient curve $y \in \mathrm{AC}([0,T];\Omega)$ of a concave function $\varphi : \Omega \to \mathbb{R}$ writes as

$$\frac{d}{dt}\frac{d^2(y(\cdot), z)}{2} \leqslant \varphi(y_t) - \varphi(z) \qquad \text{for all } z \in \Omega, \text{ for almost all } t \in [0,T].$$

The *variational inequalities* of Kinderlehrer and Stampacchia [KS87] are general first-order conditions for minimization problems when either the space, or the operator, does not allow for the classical gradient equality. *Evolutionary variational inequalities* are their transposition to Cauchy problems that are iteratively solving minimizations – gradient flows. Ambrosio, Gigli, and Savaré [AGS05] introduced them in general metric spaces, with gradually stronger properties when the space or the operators are assumed to satisfy some curvature conditions.

EVIs are quite weak, and do not provide a pointwise representation of the derivative – since they are formulated in such generality that "derivatives" are not defined. In spaces with defined angles however, one can introduce definitions of the *right derivative* y_t^+ of a curve at time t, which belongs to the tangent cone of the point y_t, the *metric gradient* $\nabla_x \varphi$ of a concave function φ, and of gradient flows – detailed in Chapter 1. In spaces with curvature bounded from below, Perel'man and Petrunin [Pet95] proved that gradient curves exist by constructing limits of broken geodesics, and using the semicontinuity of the norm of the metric gradient to obtain that the limit is still a gradient curve. This has been extended to both signs of curvature, and the monograph of Alexander, Kapovitch, and Petrunin [AKP23] contains global existence and uniqueness results for gradient flows in this "stronger" sense. In the part of this manuscript that is concerned with CAT(0) spaces, EVIs and "strong" gradient flows are both used and complementary.

Mutational analysis does not focus on gradient flows, but on the generalization of the Cauchy-Lipschitz setting. It was developed by Aubin [Aub99] as a general framework for morphological analysis, devoted to the motion of sets. The results are stated in general metric spaces, in which one *chooses* or assumes the existence of a family of relatively tame semigroups, replacing the translations $x \mapsto x + hv$ of $\mathbb{R}^d$. These *transitions* are used as surrogate of derivatives, and the main interest of it is that the local condition of "θ is a derivative of a curve y at the time t" is imposed on a *globally-defined* object. Precisely, a transition is a semigroup $\theta : \mathbb{R}^+ \times \Omega \to \Omega$, that must satisfy some contractivity and Lipschitz bounds. The mutation $\mathring{y}_t$ of a curve $y \in \mathrm{AC}([0,T];\Omega)$ is defined as the set of transitions θ such that

$$\lim_{h \searrow 0} \frac{d(y_{t+h}, \vartheta(h, y_t))}{h} = 0. \tag{5}$$

This condition depends only on the values of ϑ around y_t, so that many transitions can belong to $\mathring{y}_t$. The dynamic of an autonomous ODE is defined from Ω to a set of transitions. If ϑ and ϑ' are different transitions, the application $\mathbb{D}(\vartheta, \vartheta') := \sup_{x \in \Omega} \limsup_{h \searrow 0} d(\vartheta(h,x), \vartheta'(h,x))/h$ provides a distance that can be used to impose a Lipschitz variation of the dynamic. Note that in $\mathbb{R}^d$, taking $\vartheta(h,x) = x + hv$ for $v \in \mathbb{R}^d$, there holds $\mathbb{D}(\vartheta, \vartheta') = |v - v'|$, and the definition reduces to the classical setting.

Starting from this abstract structure, [Aub99] provides generalizations of Cauchy-Lipschitz theorems and classical estimates, later developed by Lorenz [Lor10] with a gallery of applications. The recent work of Frankowska and Lorenz [FL23] extends the scope of results to a Filippov theorem with very light assumptions. One has to be aware that a definition with mutations is (a) not geometric, and not *canonic* in that one chooses the sets of transitions, (b) quite strong, in that one imposes (5) instead of a weak definition. However, for our purpose of studying control problems with regular dynamics, it provides readily-usable tools.

To define dynamical systems in CAT(0) spaces even without a gradient flow structure, we combine both theories, defining the transitions (of mutational analysis) as the gradient flows of simple functions, referred to as energies. Let us fix a set $\mathcal{E}$ of such functions, that has to satisfy some technical conditions detailed in Section 2.1.1.1. For instance, one could consider the set of $-\alpha d(\cdot, x_0)$ for $\alpha \geqslant 0$ and $x_0 \in \Omega$. The flows of these energies are the reparametrized geodesics towards x_0, extended by a constant.

Definition (Controlled ODE in CAT(0) spaces, from Definition 2.1.7). *Let f be an application from Ω to the set of energies $\mathcal{E}$; for each $x \in \Omega$, $f(x)$ is a function that has a globally well-defined gradient flow. A curve $y \in AC([0,T];\Omega)$ is a solution of the formal ODE $\mathring{y}_t \ni f(y_t)$ if for almost any time $t \in [0,T]$, the flow of $f(y_t)$ starting from y_t approaches $h \mapsto y_{t+h}$ around $h = 0$ at order 1.*

Using this definition, we can benefit from the well-posedness results of [FL23], adapted with gradient flows of simple concave functions in Proposition 2.1.9. This solves the first point, and we can turn to the properties of such systems.

In $\mathbb{R}^d$, a classical result of Filippov [Fil63] and Aumann [Aum65] states that if the right hand-side of a differential inclusion has convex values, then the set of trajectories issued from a given point is closed under uniform convergence. In essence, one extracts a weakly converging subsequence of dynamics in $L^1(0,T;\mathbb{R}^d)$, and uses the weak closedness provided by the convexity to show that the limit is still a solution of the inclusion. In the abstract theory of [Aub99], there is no way (to our knowledge) to take convex combinations of dynamics, since one starts with semigroups. With our formulation based on gradient flows of *functions*, it is mathematically possible to take a convex combination at the level of energies, and to declare the result as the analogue of a convex hull of dynamics. Our first result is that this allows to extend the Aumann-Filippov theorem.

Theorem (Characterization of the closure of trajectories, from Theorem 2.2.4). *Assume that the dynamic f satisfies regularity conditions of Carathéodory type, and U is compact. The closure in $AC([0,T];\Omega)$ of the set*

$$\left\{ y \in AC([0,T];\Omega) \;\middle|\; y_0 = x, \text{ and } \mathring{y}_s \cap f(y_s,U) \neq \emptyset \text{ for a.e. } s \in [0,T] \right\}$$

is given by the set of solutions of the controlled system with convexified dynamics, i.e.

$$\left\{ y \in AC([0,T];\Omega) \;\middle|\; y_0 = x, \text{ and } \mathring{y}_s \cap \overline{\operatorname{conv}} f(y_s,U) \neq \emptyset \text{ for a.e. } s \in [0,T] \right\}.$$

To get this, we reformulates the ODE system with EVIs in which the potential depends on the current point of the trajectory (see Proposition 2.1.13). The main point is that EVIs are *linear* with respect to the potential, so one can work with convex combinations there. Hence one can provide conditions on the dynamic to ensure closedness of trajectories: it should have compact convex images, as in Euclidean spaces, save for the detail that the images are not of the same nature. This result does not compete with the Filippov theorem of [FL23], on which it partially relies, but gives a stronger characterization since we are able to define dynamics with convex images.

The convexification at the level of energies – instead of dynamics – may seem quite abstract, but under a reasonable assumption, one can compute the right derivative y_t^+ of the solution of the convexified equation. To formulate it, consider a solution y of the convexified system, and a curve of measures $t \mapsto \omega_t$ on the admissible energies, such that for almost each t, the gradient flow of $x \mapsto \int_\varphi \varphi(x) d\omega_t(\varphi)$ lies in the mutation of y_t. The right derivative being defined in the tangent cone, in which one can define *barycenters* $\operatorname{Bary}_{\mathrm{T}\Omega}(\varpi)$ as the unique minimizers of a compromise cost, we can state the following.

Proposition (Barycenter at the differential level, from Proposition 2.2.12). *There holds $y_t^+ = \mathrm{Bary}_{\mathrm{T}\Omega}(\nabla_x \# \omega_t)$ for almost any $t \in [0,T]$.*

In $\mathbb{R}^d$, this equality reduces to a classical integral. This justifies that the convexification taken at the integral level is connected to the usual convex hulls. This gives an answer to the second question, and we turn to the HJB equation.

Our precise aim is to make the connection with the existing theory of viscosity solutions for general Hamilton-Jacobi equations developed in Hadamard spaces by Jerhaoui and Zidani [Jer22; JZ23b]. This theory already obtained general comparison results, stability of viscosity solutions and extended Perron's method in the setting of complete CAT(0) spaces, but so far, the link with control theory was not clear. This is answered in the following theorem.

Theorem (Characterization of the value by an HJB equation, from Proposition 2.3.8). *Assume Carathéodory assumptions on the dynamic f, and $\mathfrak{J}$ to be Lipschitz-continuous. The value of the control problem*

$$V(t,x) := \inf_{u(\cdot)\in L^0(t,T;U)} \mathfrak{J}(y_T^{t,x,u})$$

is the unique viscosity solution of (1) for the control Hamiltonian $H(x,p) := \sup_{u\in U} -p\left(\nabla_x f(x,u)(x)\right)$.

The term $\nabla_x f(x,u)(x)$ in the Hamiltonian should be read "the metric gradient at the point $x \in \Omega$ of the function $y \mapsto f(x,u)(y)$". The argument are quite classical, but facilitated by the definition of viscosity with semiconvex and semiconcave functions, proposed in [JZ23b]. The nontrivial part is to obtain the necessary regularity of the Hamiltonian from the definition of ODE in the mutational sense, and is done in Lemma 2.3.6.

We mention that viscosity solutions in metric spaces have been studied by Conforti, Kraaij, Tamanini and Tonon [CKT23b; CKT23a; Con+24], focussing on controlled gradient flows. The idea in these works is to approximate the Hamiltonian by "upper and lower Hamiltonians", that do not depend on the gradient of test functions, since the latter are not defined, but on zero-order information – in the spirit of the EVI inequalities. In principle, it may be possible to adapt this strategy in our setting by allowing the potential to change with respect to the space variable in a smooth way. The definitions of [JZ23b] seem to us quite minimalist in comparison, but the general setting of [CKT23b] allows for discontinuous test functions, as the entropy in measure spaces.

The work that is closest in spirit would be the results on viability and invariance developed with mutational analysis by Badreddine and Frankowska [BF22b]. In this reference, a notion of viscosity solution is given for control problems, that characterises the value function as the unique solution of a suitable HJB PDE. We use mutational analysis as a building block for control problems, and it is not surprising that we obtain a similar characterization. However, our concern is the specific case of CAT(0) spaces, in which the theory of [JZ23b] applies and is not unrelated to [BF22b].

To conclude the study in the specific case of Hadamard spaces, Section 2.4 provides some numerical schemes based on the formulation of ODE by approximation with gradient flows, and some numerical experiments in julia. The schemes are based on the control formulation, hopefully to be compared with PDE-based schemes in the future.

II Viscosity solutions in CBB(0) spaces

This chapter transposes to spaces with Curvature Bounded Below by 0 the idea of defining viscosity solutions with test functions that share the regularity of the squared distance. In complete geodesic CBB(0) spaces, the squared distance is 2-semiconcave along geodesics, hence directionally differentiable. Consequently, one can define first-order Hamilton-Jacobi equations by relying on the directional derivatives along a suitable tangent cone. The questions that arise are the following.

- Does this definition of viscosity solutions imply a comparison principle?
- In the Wasserstein case, does the corresponding Hamilton-Jacobi-Bellman equation characterize the value function?

Let us be more precise on the definition of viscosity solutions, formulated for the equation (1) in which Ω is complete geodesic and CBB(0). Consider sets of test functions $\mathscr{T}_\pm \subset \mathcal{C}((0,T)\times\Omega;\mathbb{R})$ made of locally Lipschitz functions that are $\mathcal{C}^{1,\infty}$ with respect to the time variable, and locally semiconcave (for $\mathscr{T}_+$) or locally semiconvex (for $\mathscr{T}_-$) in the space variable.

Definition (Viscosity solution, from Definition 3.1.6). *A function $u:[0,T]\times\Omega\to\mathbb{R}$ is a subsolution (resp. supersolution) of (1) if it is locally uniformly usc (resp. lsc), and if for any $\varphi\in\mathscr{T}_+$ (resp. $\varphi\in\mathscr{T}_-$) such that $u-\varphi$ reaches a maximum (resp. minimum) at $(t,x)\in[0,T)\times\Omega$, there holds*

$$-\partial_t\varphi(t,x)+H\big(x,D_x\varphi(t,x)\big)\leqslant 0 \quad (\textit{resp.} \geqslant 0).$$

It is a viscosity solution if it is a subsolution, supersolution and satisfies the terminal condition $u(T,\cdot)=\mathfrak{J}$.

The condition of "locally uniform upper semicontinuity" is adapted to spaces with non-compact balls, as the Wasserstein space. It requires the set function $B\mapsto\sup_B u$ to be upper semicontinuous in the Hausdorff topology over nonempty closed *bounded* sets, see Definition 3.1.3. A similar condition is imposed at the terminal time by Fabbri, Gozzi, and Święch [FGŚ17]. With this definition, we are able to show the following.

Theorem (Comparison principle, from Theorem 3.1.12). *Assume that the Hamiltonian H is Lipschitz in the second argument, and satisfies*

$$H\big(y,-aD_yd^2(x,\cdot)\big)-H\big(x,aD_xd^2(\cdot,y)\big)\leqslant 2aC_Hd(x,y)\big(1+d(x,y)\big) \qquad \forall a\geqslant 0,\ (x,y)\in\Omega^2. \tag{6}$$

Then any pair $u, v : [0,T] \times \Omega \to \mathbb{R}$ of locally bounded sub and supersolutions that are ordered on the parabolic boundary, i.e. $u(T,x) \leqslant v(T,x)$ for all $x \in \Omega$, are ordered on the whole domain, i.e. $u \leqslant v$ on $[0,T] \times \Omega$.

The strategy of the proof is mostly classical, employing the Kruškov method of doubling of variables, and an adequate penalization inspired from [FGŚ17] to allow any growth of the semisolutions. Note that this is specific to parabolic equations. In Hilbert spaces, the assumption (6) is implied by the classical assumption that $H(y,p) - H(x,p) \leqslant C|p|(1+|x-y|)$. Our contributions lie in the particular design of test functions, that belong to the domain of the Hamiltonian, and the treatment of the lack of local compactness of the space. This absence of local compactness is a problem in the argument, since one first constructs a sequence of points by maximizing a given function, then extracts a converging subsequence of these points. The extraction is avoided thanks to the assumption of locally uniform upper semicontinuity, which allows in a sense to take the limit of sets and not sequences; on the other hand, the existence of maximizers is obtained by a smooth Ekeland principle. The original version of the Ekeland variational principle states that a lower semicontinuous and lower bounded function $U : X \to \mathbb{R}$, with X a complete metric space, can be "almost-maximized", in that for all $\varepsilon > 0$, there exists $x_\varepsilon \in X$ such that the function $U + \varepsilon d(\cdot, x_\varepsilon)$ admits a maximum. This version does not suit our needs, since $\varepsilon d(\cdot, x_\varepsilon)$ is not semiconcave, and cannot be embedded in the test functions. Hence we use a modified version due to Borwein, Preiss, and Zhu [BZ05], replacing the perturbation by a series of squared distances, that can be taken as test functions.

The comparison principle is such a fundamental tool that it is illusory to claim any breakthrough. However, in this relatively simple setting, we are able to impose quite few assumptions on the semisolutions, to get a so-called *strong comparison principle*, instead of imposing uniform continuity of one or both semisolutions. Since we are looking closely at the squared distance, our definition is comparable to the metric viscosity solutions of Ambrosio, Feng, Gangbo and Święch [AF14; GŚ14; GŚ15b] when the Hamiltonian is restricted to depend only on the metric slope. Importantly, the latter can be computed as $\sup_{v \in \mathrm{T}_x\Omega, |v|_x \leqslant 1} |D_x\varphi(v)|$ for our class of test functions. It is less trivial to compare it with the pathwise definition of Giga, Hamamuki, and Nakayasu [GHN15]; it seems to us that their definition is more restrictive in general, see Remark 3.1.9. The works that are closest in spirit are that of Cardaliaguet, Quicampoix [CQ08], Jimenez, Marigonda [JMQ20; Jim24], Hynd and Kim [HK15], formulated for uniformly continuous semisolutions with semidifferentials, and in the Wasserstein space. Since the rest of our contribution is written in this space as well, let us say a few words of context.

The space of Borel probability measures can be endowed with the Wasserstein distance of optimal transportation. We refer the reader to Section 1.1.4 for precise definitions; for the present discussion, it suffices to have in mind that the Wasserstein distance between μ and ν is the minimal energy that is needed to drag μ onto ν, the energy being computed with respect to a cost that is the squared distance in the Wasserstein case. It is particularly interesting for limits of particle systems, since the distance between empirical measures (a) can always be computed, unlike entropy-based distances, and (b) behaves as a travel distance; if μ is an empirical measure, i.e. a finite combination of Dirac masses, and ν a perturbation of μ by moving the supports of the Diracs by a distance of order ε, then $d_{\mathcal{W}}(\mu,\nu)$ behaves as ε if the latter is sufficiently small. Optimal transport in general furnishes a variety of distances with different costs, and extends to the cases of measure with different masses – we refer to Chizat [Chi17] for references. Regardless of its intrinsic beauties, it is a quite natural metric space to formulate the equations of physics.

And indeed, PDEs in the Wasserstein space are a most active field of research, with different branches: the Otto calculus, founded by Otto [Ott01]; the Lions differentiability, developed during courses at the Collège de France by Lions [Lio06]; geometric differentiability, taking roots in [AGS05] and the subsequent work of Gangbo, Nguyen, and Tudorascu [GNT08]; diffusion semigroups on $\mathscr{P}_2(\Omega)$, pioneered by Von Renesse and Sturm [VRS09], to mention only the main approaches. After the completion of the first part of this chapter, and in particular with the prepublication of Bertucci and Lions [BL24], the author came to the conviction that L-differentiability does provide better test functions in $\mathscr{P}_2(\Omega)$, and is not trivially equivalent to the notion of viscosity solutions presented earlier. L-differentiable functions are defined by coming back to a Hilbert space, in which one may impose as much regularity as needed. Consequently, viscosity solutions defined with L-differentiable test functions can be proved to be stable by locally uniform convergence, under a mere continuity assumption on the Hamiltonian, with the same arguments as in $\mathbb{R}^d$

[CIL92, §6].

Nevertheless, it is interesting to consider optimal control problems with the notion of viscosity that we introduce; it is more restrictive than its L-differentiability analogue, so that the satisfaction of the HJB equation (1) by the value function is, in comparison, harder to obtain. More specifically, we consider the Mayer optimal control problem aiming to

$$\begin{aligned}&\text{Minimize} \quad \mathfrak{J}(\mu_T^{\nu,u}) \quad \text{over } u \in L^1(0,T;U),\\ &\text{where } \partial_t \mu_t^{\nu,u} + \operatorname{div}\left(f[\mu_t^{\nu,u}, u(t)]\#\mu_t^{\nu,u}\right) = 0 \quad \text{for } t \in (0,T), \quad \text{and} \quad \mu_0^{\nu,u} = \nu.\end{aligned} \tag{7}$$

As opposite to Chapter 2, there is no ambiguity on the definition of the trajectories, which solve a controlled continuity equation using results of [AGS05; BF21; BF24]. The value function $V : [0,T] \times \mathscr{P}_2(\Omega) \to \mathbb{R}$ of the problem (7) is given by

$$V(t,\nu) := \inf\left\{ \mathfrak{J}\left(\mu_T^{t,\nu,u(\cdot)}\right) \;\middle|\; u(\cdot) \in L^1(t,T;U),\ \partial_s \mu_s + \operatorname{div}\left(f[\mu_s, u(s)]\#\mu_s\right) = 0 \text{ for } s \in (t,T),\ \mu_t = \nu \right\}.$$

The dynamic f is defined from $\mathscr{P}_2(\Omega) \times U$ towards the set of Lipschitz vector fields with linear growth, endowed with the topology of local uniform convergence. The main result is the following one.

Theorem (Characterization of the value function, from Theorem 3.2.16). *Assume that the dynamic f satisfies Carathéodory-type regularity conditions, has convex images, and that the terminal cost $\mathfrak{J}$ is locally uniformly continuous. Then the value function is locally uniformly continuous, and is the unique viscosity solution of the HJB equation (1) with Hamiltonian $H(\mu, p) := \sup_{u\in U} -p\left(\pi_T^\mu f[\mu,u]\#\mu\right)$.*

This type of result is very classical [CQ08; HK15; Cav+18; BF22a; Dau23; JMQ20; JMQ23; Ber24], and the arguments are not new. However, since we want to keep the coherence with the metric structure, we are faced with a difficulty that does not appear in the other works – save maybe in [AF14] for Eikonal equations, in which the computation of the metric slope forces the use of the geometric tangent cone instead of its regular subset. The Hamiltonian is defined on the elements (μ, p) of the metric cotangent bundle, in which $\mu \in \mathscr{P}_2(\Omega)$ is a point, and $p : \mathbf{Tan}_\mu \to \mathbb{R}$ a Lipschitz map from the (geometric) tangent cone to $\mathbb{R}$. Consequently, to evaluate p, we have to compose the dynamic of the continuity equation with a projection on the geometric tangent cone, producing the term $\pi_T^\mu f$ in the definition of the Hamiltonian. Therefore, when using the Bellman principle, we have to show that a solution of the continuity equation can be approximated by a curve with derivative $\pi_T^\mu f$. In other words, we have to show that for $\xi = f[\mu,u]\#\mu$, there holds

$$\lim_{h \searrow 0} \frac{d_{\mathcal{W}}\left((\pi_x + hf(x))\#\mu, (\pi_x + h\pi_T^\mu f(x))\#\mu\right)}{h} = 0. \tag{8}$$

This question will be discussed in detail in Chapter 5. In short, (8) holds *owing to the fact that f is a map*, and we are able to complete the argument. There might be other ways to conclude: if $(\mu_s)_s$ solves the continuity equation for a dynamic $(f_s)_s$, then the same curve is also a solution for the velocity fields $(\pi_T^{\mu_s} f_s)_s$. However, the projection π_T^μ is discontinuous as a function of μ, and we do not know of any regularity of the curve $s \mapsto \pi_T^{\mu_s} f_s$. This prevents us to apply the arguments in use in the current strategy.

Technically, the comparison principle stands for a slightly larger class of Hamiltonian than needed to treat (7). Such Hamiltonians are described in Proposition 3.2.6: in the vocabulary of control problems, they allow for MDEs instead of continuity equations, with a driving field that may split mass. However, the said driving field should still be locally Lipschitz-continuous with respect to a certain similarity $W_{\mu,\nu}$, and we do not know examples of such regularity for measure fields that would not be induced by a map.

The problem (7) is of first-order, so that the Hamiltonian is defined at all measures. In the case where one considers higher order derivatives of the unknown, several strategies are in use. Bayraktar, Ekren, and Zhang [BEZ24] replaces the squared distance in the doubling function by a smoother function built from series of penalized Fourier moments, and is able to extends Ishii's Lemma for the particular case considered. Daudin and Seeger [DS24] test the solution only at points where the Hamiltonian is defined, here in the domain of the Fisher information. The definition of viscosity solutions imposes a constraint on the semidifferential up to some entropy penalization, so that one has to prove that the value function does satisfy this strict definition. This is done relying on the smoothing properties of the diffusion term in the Fokker-Planck equation. Bertucci and Lions [BL24] rely on a similar strategy, using in addition a regularization of the Wasserstein distance by a sup-convolution of its lift; another regularization has been proposed by Cosso and Martini [CM23] by local averages. One could also consider directly test functions that are $\mathcal{C}^2$−regular [PW18; CD18b; CD18a; CP20; Cos+24]. We stress again that this is not needed in our case, since the HJB equation is of first order only.

This answers the second question motivating this chapter. Section 3.2.4 provides some extensions of the type of control problems.

- If the dynamic f does not have convex images, the value function of (7) is characterized as the unique viscosity solution of the Hamilton-Jacobi-Bellman equation associated to the convexified dynamics, with Hamiltonian $H^{\text{relax}}(\mu, p) := \sup_{b\in\overline{\text{conv}}\, f[\mu,U]} -p(\pi_T^\mu b\#\mu)$.
- We point at the necessary modifications in the case of a Bolza problem, with a running cost in addition of the terminal cost.
- Some arguments of the comparison principle are simplified if one considers a weaker topology on $\mathscr{P}_2(\Omega)$, namely the inductive topology constructed from traces of the narrow one on Wasserstein balls, which makes $\mathscr{P}_2(\Omega)$ ball-compact. However, one has to prove that the value function stays continuous with respect to this weaker topology. This is done in two steps: first, the closedness of trajectories with respect to this weaker topology is proved in Lemmata 3.2.23 and 3.2.24, and Proposition 3.2.26 deduces the desired regularity. The inductive topology is not quite commonly used, and in the part which is concerned with the trajectories, the extraction with respect to $d_{\mathcal{W},p}$ for some $p < 2$ may appear for the first time.
- If the terminal cost $\mathfrak{J}$ is taken equal to ∞ on some part of $\mathscr{P}_2(\Omega)$, then V cannot be proved to be a subsolution. Extending an argument of Lions and Souganidis [LS84], we show the following.

Theorem (Minimality, from Theorem 3.2.28). *Assume that f satisfies Carathéodory-type regularity conditions, and that $\mathfrak{J}$ is lower bounded and lower semicontinuous. The value function is the smallest viscosity supersolution of (1).*

The first two points are quite classical in the literature, and the specificities in our case lie in the fact that the test functions are not locally linear, and neither have the semiconvexity postulated in Chapter 2. The two last points are more directly concerned with the Wasserstein space, being formulated with topologies that are built for measures. This concludes our focus on control problems in the Wasserstein space over $\mathbb{R}^d$.

III Wasserstein space over a network

Chapter 4 is the most technical part of the manuscript, albeit the shortest. It shows that the Wasserstein distance over a one-dimensional network is directionally differentiable in the following sense. Let (Ω, d) be a network, which may admit loops, endowed with the shortest path distance. Denote $\mathscr{G}$ the set of its geodesics, with $e_h(\gamma) = \gamma_h$ the evaluation of the geodesic γ at time $h \in [0,1]$.

Theorem (Differentiability, from Theorem 4.3.2). *Let $\mu, \nu \in \mathscr{P}_2(\Omega)$, and $\xi \in \mathscr{P}_2(\mathscr{G})$ such that $e_0\#\xi = \mu$. Then the directional derivative of $h \mapsto d^2_{\mathcal{W}}(e_h\#\xi, \nu)$ at $h = 0$ exists, and is given by*

$$\lim_{h\searrow 0} \frac{d^2_{\mathcal{W}}(e_h\#\xi,\nu) - d^2_{\mathcal{W}}(\mu,\nu)}{h} = \inf_{\substack{\alpha=\alpha(d\gamma,dz)\in\Gamma(\xi,\nu)\\ (e_0(\pi_\gamma),\pi_z)\#\alpha\in\Gamma_o(\mu,\nu)}} \int_{(\gamma,z)\in\mathscr{G}\times\Omega} \frac{d}{dh}_{|h=0} d^2(\gamma(\cdot), z)\, d\alpha(\gamma, z).$$

In Euclidean spaces, the directional derivative of the Wasserstein distance is known [Gig08, §4.2], and the argument is provided in the case of a $\mathcal{C}^1$ cost with sufficiently smooth derivative for comparison with Ω. The case of a manifold with curvature bounded from below is treated in [Gig11, Theorem 4.2], relying on the curvature properties. In our case, $\mathscr{P}_2(\Omega)$ is neither CBB nor CAT, and there is no gain in restricting to Wasserstein geodesics. To our knowledge, there is no further literature on this question.

The proof goes by bounding the limit inf and limit sup by the same term. One inequality is easy, the other is not. The problem comes from the fact that the application $(\gamma, z) \mapsto \frac{d}{dh_{|h=0}} d^2(\gamma(\cdot), z)$ is not continuous, and has the wrong semicontinuity to apply classical arguments. In the networks that we consider, discontinuities appear only at the junctions, which we know how to treat, and on the "cut locus bundle" − the pairs (x, z) such that x is in the cut locus of z. One has to obtain a control on the mass that is put around the points of discontinuity, prior to passing to the limit in h.

Let us give the core ideas in an informal way. Before the limit in h, the points at which we do not have a uniform estimate are the pairs (γ, z) such that $\gamma(s)$ crosses the cut locus of z for some $s \in]0, h[$. Pick α_h optimal at time h, and restrict it to these problematic sets. Consider a narrow limit point β. The key is to use that β is optimal between its marginals, thus has a monotone support. Vaguely speaking, Lemma 4.2.7 shows that the intersection between a monotone set and the cut locus bundle must be very thin, that is, with $x-$projection finite in any compact. So one can work as if the limit was a single Dirac mass. We use it with the following argument: if $(\mu_n)_{n\in\mathbb{N}}$ is a sequence of measures converging with respect to the Wasserstein distance to δ_0, with $\mu_n \leqslant \nu$ for some fixed $\nu \in \mathscr{P}_2(\Omega)$, then actually $\mu_n(\{0\}) \to_n 1$, otherwise ν would have infinite mass. Up to detail, this implies that β actually has null mass, and the "bad set" disappears.

What is showed is that the Wasserstein distance can be used as a test function if the Hamiltonian is defined on directional derivatives. We did not repeat the study of viscosity solutions for two reasons; first, because the arguments would not change as soon as the assumptions allow it, and secondly, because the assumptions on the Hamiltonian are very restrictive. Thus Theorem 4.3.2 is, at best, a first step towards a sound treatment of crowds on networks by these tools.

IV Aspects of Wasserstein geometry

This chapter starts from the following question, generalizing (8):

$$\text{Does it hold for any } \xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \text{ that} \quad \lim_{h\searrow 0} \frac{d_{\mathcal{W}}\left(\exp_\mu(h\cdot\pi_T^\mu\xi), \exp_\mu(h\cdot\xi)\right)}{h} = 0 \quad ? \tag{9}$$

To clarify the notations, we need some context. One main point in [AGS05; Gig08] is that the Wasserstein space is a geodesic CBB space, endowed with geometric tangent cones, themselves isometric to a subset $\mathbf{Tan}_\mu$ of the larger set $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ of measures on $\mathrm{T}\mathbb{R}^d = \{(x, v) \mid x \in \mathbb{R}^d, v \in \mathrm{T}_x\mathbb{R}^d\}$ with $x-$marginal μ. The elements of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, that we refer to as *measure fields*, can be manipulated analogously as vectors: one can multiply them by $s\cdot\xi := (\pi_x, s\pi_v)\#\xi$, take sums or interpolation along transport plans, and follow an "exponential" $s \mapsto \exp_\mu(s\cdot\xi) := (\pi_x + s\pi_v)\#\xi$. In this simili-Hilbertian space, the tangent cone is a closed convex subset, and any $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ has a unique metric projection $\pi_T^\mu\xi \in \mathbf{Tan}_\mu$ minimizing a certain "cone distance" W_μ. The latter is an infimum on a particular set of plans $\Gamma_\mu(\xi,\zeta)$ that forbid mass transfer between (x, v) and (y, w) if $x \neq y$.

The larger set $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, and its subset of map-induced fields $L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)\#\mu$, are the usual spaces in which the dynamics of the continuity equation take values. Some authors directly consider these as the tangent cone to $\mathscr{P}_2(\mathbb{R}^d)$, without much consequences. If (9) were to hold, then the geometric tangent cone $\mathbf{Tan}_\mu$ would be established as the quotient of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ by the same equivalence relation as in manifolds, and one could pass from the flow of a continuity equation to the exponential of the projected dynamics without problems. Unfortunately, we show that (9) does not hold in general. This requires some preliminaries.

Orthogonal decompositions. At any $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, the CBB theory provides a metric scalar product defined on the tangent cone $\mathbf{Tan}_\mu$. The expression of the metric scalar product can be extended, *a priori* without justification, to all $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. We denote $\langle\cdot,\cdot\rangle^+_\mu$ this extension, with a superscript + to recall that it is defined by a supremum over transport plans, and $\langle\cdot,\cdot\rangle^-_\mu$ its counterpart with an inf. The applications $\langle\cdot,\cdot\rangle^\pm_\mu$ have mixed properties of convexity along the two types of interpolations that can be considered in $\mathscr{P}_2(\mathbb{R}^d)$; the "vertical interpolation", in the Banach sense of measures, and the "horizontal interpolation", along *any* transport plan. We stress that the latter provides more interpolating curves as geodesic/displacement interpolation, since the transport plans are not constrained to be optimal.

Section 5.1 contains a quantity of small algebraic results of constant use on these interpolations, horizontally convex subsets of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ and the metric projection on them. We provide two applications as showcases of the "algebra" in $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$; the exact formula of the superdifferential of the squared distance[1] in Section 5.1.3, and some bounds on Bertucci's regularization by sup-convolution in Section 5.1.4.

The interest of the maps $\langle\cdot,\cdot\rangle_\mu^\pm$ is that they provide orthogonal decompositions that actually have some meaning. The barycentric/centred decomposition is one of them, detailed in Section 5.2.1.1, with the particularity that barycentric (or map-induced) fields reduce $\langle\cdot,\cdot\rangle_\mu^\pm$ to the scalar product of L^2_μ. This simplification brings additional properties that are not representative of a generic orthogonal decomposition. However, introducing *solenoidal measure fields* as

$$\mathbf{Sol}_\mu := \left\{ \zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \;\middle|\; \langle\zeta,\xi\rangle_\mu^+ = \langle\zeta,\xi\rangle_\mu^- = 0 \quad \forall \xi \in \mathbf{Tan}_\mu \right\},$$

we obtain a pair $\mathbf{Tan}_\mu$ / $\mathbf{Sol}_\mu$ that might be taken as a canonical example. The definition of $\mathbf{Sol}_\mu$ itself is not sufficient to support this claim, and the precise result is the following.

Theorem (Helmholtz-Hodge decomposition, from Theorem 5.2.12). *Let $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. Then*

- *ξ has unique metric projections $\pi_T^\mu\xi$ on $\mathbf{Tan}_\mu$ and $\pi_S^\mu\xi$ on $\mathbf{Sol}_\mu$ with respect to W_μ;*
- *The sets $\Gamma_{\mu,o}(\xi,\pi_T^\mu\xi)$ and $\Gamma_{\mu,o}(\xi,\pi_S^\mu\xi)$ are reduced to singletons $\{\alpha_T\}$ and $\{\alpha_S\}$, with*

$$\pi_S^\mu\xi = (\pi_x,\pi_v-\pi_w)\#\alpha_T, \qquad \pi_T^\mu\xi = (\pi_x,\pi_v-\pi_w)\#\alpha_S, \tag{10}$$

  and there exists a plan $\alpha \in \Gamma_\mu(\pi_T^\mu\xi,\pi_S^\mu\xi)$ such that $\xi = (\pi_x,\pi_v+\pi_w)\#\alpha$;
- *Some partial Pythagoras identities hold, and*

$$\langle\xi,\eta\rangle_\mu^\pm = \langle\pi_T^\mu\xi,\eta\rangle_\mu^\pm \quad \forall \eta \in \mathbf{Tan}_\mu, \qquad \langle\xi,\zeta\rangle_\mu^\pm = \langle\pi_S^\mu\xi,\zeta\rangle_\mu^\pm \quad \forall \zeta \in \mathbf{Sol}_\mu.$$

We clarify immediately that the part of this statement concerned with $\mathbf{Tan}_\mu$ comes from [Gig08, Chap 4]. Our contribution is most of all (10), and the extension to $\mathbf{Sol}_\mu$ by relying solely on algebraic properties. It turns out that *all* the arguments are based on these properties, and the result extends to any orthogonal decomposition with respect to $\langle\cdot,\cdot\rangle_\mu^\pm$. However, in targetting (9), we focus primarily on $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$.

Directional derivatives of $d_\mathcal{W}$ and the classification problem. The question (9) contains several particular cases. If the measure field ζ is taken in $\mathbf{Sol}_\mu$, then $\pi_T^\mu\zeta = 0_\mu$, and it writes as

$$\lim_{h\searrow 0} \frac{d_\mathcal{W}\left(\mu, \exp_\mu(h\cdot\zeta)\right)}{h} = 0. \tag{P$_{\min}$}$$

If this were true, a measure field ξ with a non-zero solenoidal component (given by the Helmholtz-Hodge decomposition) would "loose speed" when following the exponential $s \mapsto \exp_\mu(h\cdot\xi)$. One could then expect that if the maximal speed is reached, that is, if

$$\lim_{h\searrow 0} \frac{d_\mathcal{W}\left(\mu, \exp_\mu(h\cdot\xi)\right)}{h} = \|\xi\|_\mu, \tag{P$_{\max}$}$$

then $\xi \in \mathbf{Tan}_\mu$. To advocate for the latter claim, note that the reparametrized geodesics of $\mathscr{P}_2(\mathbb{R}^d)$ are induced by the measure fields on which equality in (P$_{\max}$) holds for some $h > 0$, and that $\mathbf{Tan}_\mu$ is defined as the closure of these fields with respect to a quite strong topology. However, intuition fails there.

Theorem (Classification of $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$, from (5.24) and below). *There always holds that $\xi \in \mathbf{Tan}_\mu$ implies (P$_{\max}$), and that (P$_{\min}$) implies $\zeta \in \mathbf{Sol}_\mu$. Moreover,*

[1] The acute reader will ask: which superdifferential? The most restrictive one in the geometric tangent cone, see (5.6) p. 107.

- *by Proposition 5.3.8, the converse holds if ξ and ζ are induced by maps. Then Corollary 5.3.9 shows that (9) also holds.*
- *Section 5.3.4 provides a counterexample in the general case: there exists $\mu \in \mathscr{P}_2(\mathbb{R})$ and $\zeta \in \mathbf{Sol}_\mu \setminus \{0_\mu\}$ satisfying ($P_{\max}$).*
- *In dimension one, if μ is purely atomic or absolutely continuous with respect to the Lebesgue measure, then the converse holds. This is proved by Lemma 5.4.7 for solenoidal fields on absolutely continuous measures, and directly implied by Theorem 5.2.23 in all other cases.*

In addition, we show – in dimension one – that if ξ is tangent, then for some vanishing sequence $(h_n)_{n\in\mathbb{N}}$, all choices of $\xi_n \in \frac{1}{h_n}\exp_\mu^{-1}(\exp_\mu(h_n \cdot \xi_n))$ converge towards ξ with respect to W_μ (see Proposition 5.3.6). This is stronger than the definition of $\mathbf{Tan}_\mu$, which only requires the approximation of ξ by a sequence of reparametrized geodesics, not specifically passing through the exponential of ξ.

All statements in dimension one rely on the fact that an optimal plan splits mass on *at most countably many points*. This extends in higher dimension with c−c hypersurfaces [Gig11], and we conjecture that all statements do generalize as well. We detail a first step in this direction in the last part of this introduction.

The counterexample given in Section 5.3.4 builds a skinny Cantor measure with the property that along a given vanishing sequence h_n, it is approximated with an error in $o(h_n)$ by a sum of Dirac masses. On such measures, every measure field is tangent, and the limit in ($P_{\max}$) is close to being attained. To the best of our intuition, the set of measures μ that admit a nontrivial nontangent measure field satisfying ($P_{\max}$) could have measure 0 with respect to the measures $\mathbb{P}_\beta$ constructed by Von Renesse and Sturm [VRS09]. The set of measures on which equality is reached has $\mathbb{P}_\beta$−measure zero, and the problem appear precisely when the measure μ is approximated at scale h by elements of this set with an error in $o(h)$. This point is not supported by strong evidence, and at best a curiosity, but might be worth looking at.

A closer look at solenoidal fields. The set $\mathbf{Sol}_\mu$ has quite interesting properties, some symmetric to $\mathbf{Tan}_\mu$, and some not. For instance, the tangent cone is built as a closed cone over a regular subset of measure fields (reparametrized geodesics). In some cases, $\mathbf{Sol}_\mu$ can be obtained in the same way as the cone over velocities that are coming back to μ on a short time, as proved in Lemma 5.4.6. This has a clear geometric meaning: in these “regular” cases, solenoidal measure fields are the velocities of the curves that spiral around μ, approximated with respect to W_μ by velocities of loops. The existence of this representation is a condition on μ; the exact characterization is not clear even in dimension one, where we know that it is satisfied for purely atomic measures and absolutely continuous measures (see Lemma 5.4.7), and not satisfied for at least one Cantor measure (see Section 5.3.4).

In the general case, one can at least say that solenoidal fields are stable with respect to the convex combinations in the classical sense of measures.

Proposition (Vertical convexity, from Proposition 5.4.5). *If $\zeta, \zeta' \in \mathbf{Sol}_\mu$, $\lambda \in [0,1]$, then $(1-\lambda)\zeta + \lambda\zeta' \in \mathbf{Sol}_\mu$.*

This is not satisfied by tangent fields, although a counterpoint is given in Proposition 5.4.4. This property hides two phenomena: the subset of *centred* solenoidal measure fields is vertically convex, which is also true for centred tangent measure fields, and actually any centred, W_μ−closed, horizontally convex cone (see Proposition 5.2.20). Moreover, the barycentric solenoidal fields are supported “on the same vectors” as their centred cousins, in a way made precise in Proposition 5.4.5. This is the part that fails for tangent measure fields.

The picture is clearer in dimension one. In this case, we are able to characterize completely tangent and solenoidal measure fields, by discussing the atom part and the atomless part of μ.

Theorem (Dim 1, from Theorem 5.2.23). *Let $\mu = m_a\mu^a + m_d\mu^d$, with μ^a atomic and μ^d atomless. Then*

- *$\xi \in \mathbf{Tan}_\mu$ iff it is a map on μ^d, i.e. $\xi = m_a\xi^a + m_d f^d\#\mu^d$ for some $\xi^a \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_{\mu^a}$, $f^d \in L^2_{\mu^d}(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$;*
- *$\zeta \in \mathbf{Sol}_\mu$ iff it is centred and 0 on atoms of μ, i.e. $\zeta = m_a 0_{\mu^a} + m_d\zeta^d$ for some centred $\zeta^d \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_{\mu^d}$.*

Here vertical convexity of solenoidal measure fields appears clearly: the vertical sum of centred fields that are concentrated on 0 on atoms stays so. This decomposition contains two types of information: the fact that *centred* solenoidal and measure fields each vanish on some set on which the other does not, and the fact that these sets are determined by the support of μ, one being the countable set of atoms, and the other its complementary. At this point, we might say that the characterization of $\mathbf{Tan}_\mu$ is not surprising in view of the fact that μ^d is the transport-regular component of μ [Gig11], hence all optimal plans are induced by maps, and this passes to the limit. We do not know of a reference considering the set $\mathbf{Sol}_\mu$ besides the solenoidal vector fields of fluid dynamics, which are our *barycentric* solenoidal fields, indeed much studied.

The above characterization implies that solenoidal fields are more stable than tangent ones.

Theorem (Closedness in a weaker topology in dimension 1, from Theorem 5.2.24). *The set $\mathbf{Sol}_\mu$ is closed with respect to the topology on $\mathscr{P}_2(\mathrm{T}\mathbb{R})$ induced by the Wasserstein distance $d_{\mathcal{W},\mathrm{T}\mathbb{R}}(\cdot,\cdot)$.*

This result is reminiscent of the weak closure results of [Gig08, Chap. 5] on "strongly" (horizontally) convex sets, but cannot be deduced from them. One should be careful with intuition coming from Hilbert spaces, since $\mathbf{Tan}_\mu$ has the same properties of horizontal convexity, closure and decomposition, but is not $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}-$closed. Once again, the argument relies on the countable support of centred optimal plans. The generalization to higher dimension is open, but motivates our (final) section.

Centred fields give local information. We come back to any dimension $d \in \mathbb{N}_*$. Let $\mathbf{Tan}^0_\mu$ and $\mathbf{Sol}^0_\mu$ denote the subsets of $\mathbf{Tan}_\mu, \mathbf{Sol}_\mu$ made of fields with barycenter 0 for $\mu-$almost each point x. These subsets satisfy much richer algebraic properties than $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$, for the simple next reason.

The metric scalar product satisfies $\langle \xi,\zeta\rangle^+_\mu = \int_{x\in\mathbb{R}^d} \langle \xi_x,\zeta_x\rangle^+_{\delta_x} \, d\mu(x)$ for any disintegrations $\xi = \xi_x \otimes \mu$ and $\zeta = \zeta_x \otimes \mu$. If ξ and ζ are centred, the disintegrations can be chosen so. The quantity $\langle \xi_x,\zeta_x\rangle^+_{\delta_x}$ is a supremum over the set of transport plans, so greater than the value of $\int_{(v,w)\in\mathrm{T}_x\mathbb{R}^d} \langle v, w\rangle \, d\xi_x \otimes \zeta_x$, which is 0 since ξ_x and ζ_x are centred. *Consequently, if two centred measure fields are orthogonal, their disintegrations are orthogonal $\mu-$almost everywhere.*

This very strong property allows to restrict and glue centred solenoidal fields and still obtain solenoidal measure fields, to rescale them by an element $\lambda \in L^\infty_\mu(\mathbb{R}^d;\mathbb{R})$ by $\lambda\cdot\zeta := (\pi_x, \lambda(\pi_x)\pi_v)\#\zeta$ instead of a scalar, and to "orthogonalize" them. The same holds for centred tangent fields. This is detailed in Proposition 5.2.20, Lemma 5.5.1, Proposition 5.5.4 and Corollary 5.5.5. Interestingly, the introduction of $\mathbf{Sol}_\mu$ really helps there. The fact that the restriction of a tangent field stays tangent is deduced from the classical restriction of optimality [Vil09, Theorem 4.6]. However, to go the other way, one would have to *extend* a tangent field, which is not trivial. Our strategy is to first extend optimal plans under the assumption that the target is compactly supported (Lemma 5.5.2), then prove that orthogonality with respect to this subset is sufficient to characterize solenoidal fields (Lemma 5.5.3), deduce the desired result on $\mathbf{Sol}^0_\mu$, and finally on $\mathbf{Tan}^0_\mu$ by orthogonality.

Bear with us for our last statement: we say that a subset $\mathcal{A}$ of centred fields is of "dimension k" if there exist $f_1,\cdots,f_k \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ such that $\mu-$almost everywhere, $f_1(x),\cdots,f_k(x)$ is an orthonormal family, and $\xi \in \mathcal{A}$ if and only if ξ is centred and $v \in \mathrm{span}\{f_1(x),\cdots,f_k(x)\}$ for $\xi-$a.e. $(x,v)\in\mathrm{T}\mathbb{R}^d$ (see Definition 5.5.8).

Theorem (Decomposition by dimension of $\mathbf{Tan}^0_\mu$, from Theorem 5.5.9). *For any $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, there exists a measurable partition of $\mathbb{R}^d$ by sets $(A_k)_{k\in[\![0,d]\!]}$ with the following properties. For $k \in [\![0,d]\!]$, denote $m_k := \mu(A_k)$ and $m_k\mu_k := \mu(\cdot\cap A_k)$. Then $\mathbf{Sol}^0_{\mu_k}$ is of dimension k, $\mathbf{Tan}^0_{\mu_k}$ is of dimension $d-k$, and $\zeta \in \mathbf{Sol}^0_\mu$ (resp. $\xi \in \mathbf{Tan}^0_\mu$) if and only if it writes as $\sum_{k=0}^d m_k\zeta_k$ for some $\zeta_k \in \mathbf{Sol}^0_{\mu_k}$ (resp. $\sum_{k=0}^d m_k\xi_k$ for some $\xi_k \in \mathbf{Tan}^0_{\mu_k}$).*

Let us detail the case of dimension one, relying on the exact characterization that we already obtained. For any $\mu \in \mathscr{P}_2(\mathbb{R})$, the above theorem constructs a set A_0 and a set A_1. Using the decomposition for $d = 1$, we may further say that A_0 is the set of atoms of μ, and A_1 its complementary. Then $\mu = m_0\mu_0 + m_1\mu_1$ is the classical atomic/diffuse decomposition. The Cantor part is not seen there: however, as we discussed previously, its seems to be detected by the structure of solenoidal fields – equivalently, the satisfaction of ($\mathrm{P}_{\min}$) by all $\zeta \in \mathbf{Sol}_\mu$. On A_0, we know that solenoidal fields are concentrated on 0, hence of "dimension 0" by convention. At the opposite, centred tangent fields are not constrained. They are of “dimension 1” since they put mass on the span of the unit section $f_1 : x \mapsto (x, 1)$ almost everywhere. The situation reverses on A_1: solenoidal fields are any centred fields, and centred tangent fields reduce to 0_μ.

In dimension $d > 1$, we do not have a similar characterization of the sets A_k. The intuition arising from [Gig11] is that the k^{th} submeasure has a support covered by countably many graphs of differences of convex functions. Lott [Lot16] computed the geometric tangent cone to the Hausdorff measure supported on a $k-$dimensional submanifold; in the above notations, this would yield a single non $\mu-$negligible set A_k. The directions of the basis $g_1, \cdots, g_k$ for solenoidal fields are the tangent one to the manifold, and the basis $f_{k+1}, \cdots, f_d$ for tangent measure fields spans the normal directions to the manifold – with a flip of terminology. Our result applies to any measure, but still does not subsume [Lot16], because we lack the description of the support of μ_k.

To conclude, we point that the above decomposition could be related to the body of results on sharp extensions of the Rademacher theorem (see Alberti and Marchese [AM16] and references therein), although the links are not yet clear. In particular, Alberti, Csörnyei, and Preiss [ACP11] state in Theorem 2.13 that “Every measure μ on $\mathbb{R}^n$ can be uniquely decomposed as $\mu = \mu_n + \mu_{n-1} + \cdots + \mu_0$, where each μ_k is a $k-$rectifiably representable measure supported on a $(k+1)-$purely unrectifiable set.” The similarities are immediate, so let us point at the differences: the whole theory is concerned with non-differentiability sets of *Lipschitz functions*, whereas our results would concern the (smoother) class of semiconcave/semiconvex functions. Our approach is entirely based on optimal transportation with $p = 2$, whereas the mentioned results are a branch of geometric measure theory. Nevertheless, it seems promising to turn to the case of $p = 1$ and see if a connection can be made there.

Chapter 1

Background material

This chapter contains the mandatory definitions and properties for the following ones. It may serve as an introduction, but is strongly biased by the necessities of the sequel. To the reader that is interested into metric spaces in the Alexandrov sense, we recommend [AKP23; BH99]; [San15; ABS21] as entry points in the literature of optimal transport, and [AGS05; Vil09] as references; [Lio82] and the excellent [Bar94] for viscosity solutions, with [CIL92] still being up to its name.

Contents of Chapter 1

1.1 Metric spaces and curvature

> *"Il ne faut pas vingt années accomplies pour voir changer les hommes d'opinion sur les choses les plus sérieuses [...]. J'assurerai aussi peu qu'une ligne droite tombant sur une autre ligne droite, fait deux angles droits, ou égaux à deux droits, de peur que les hommes venant à y découvrir quelque chose de plus ou de moins, je ne sois raillé de ma proposition [...]"*
>
> La Bruyère, Caractères/Des Jugements, 94 [VI], 1691.

1.1.1 Shared definitions

This section introduces elementary definitions in use in the sequel, as well as adaptations of classical results. In all the manuscript, (Ω, d) is a *Polish space* (complete separable metric), sometimes assumed *proper* (with compact closed balls).

Definition 1.1.1 (Geodesic). *A curve $\gamma : [0,1] \to \Omega$ is a* unit-speed geodesic*, or simply a* geodesic*, if it satisfies*

$$d(\gamma(t), \gamma(s)) = |t-s|\, d(\gamma(0), \gamma(1)) \qquad \forall (s,t) \in [0,1]^2.$$

The set of geodesics is denoted $\mathcal{G}$, and $\mathcal{G}_x$ is the subset of $\gamma \in \mathcal{G}$ such that $\gamma_0 = x$.

A *geodesic space* is a space in which any two points are joined by at least one geodesic. We are concerned with metric spaces that satisfy a curvature condition, either nonpositive as developed in Section 1.1.2, or nonnegative in Section 1.1.3. Let us provide here a brief nontechnical introduction. Consider three points $x, y, z \in \Omega$ and a geodesic $\gamma_{y,z}$ linking y to z. In the Euclidean space,

$$d^2\left(x, \gamma_{y,z}(s)\right) = (1-s)d^2(x,y) + s d^2(x,z) - s(1-s)d^2(y,z). \tag{1.1}$$

Equality can be weakened in two ways: *non-positively spaces*, in which (1.1) stands with an $\leqslant$ sign, and *non-negatively curved spaces*, in which (1.1) stands with an $\geqslant$ sign. The inequality can be reformulated as a monotonicity of $s \mapsto d(\gamma_s, \gamma'_s)/s$ for each pair of geodesics γ, γ' issued from the same point; in non-positively curved spaces, these functions are nondecreasing, meaning that geodesics have a tendency to diverge. At the opposite, in non-negatively curved spaces, geodesics have a tendency to converge.

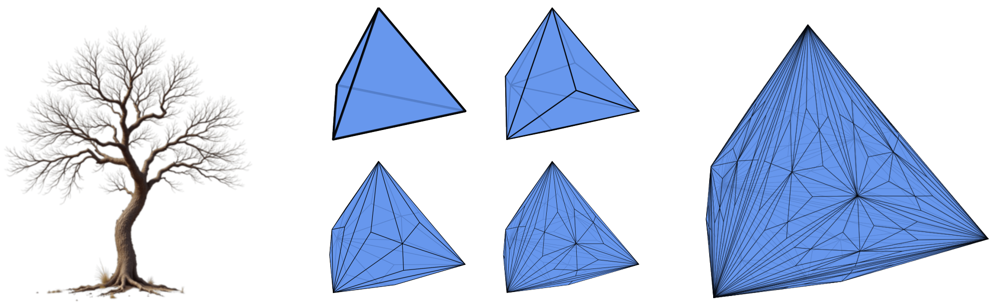


Figure 1.1: A non-positively curved space, and non-negatively curved spaces constructed in [OS94].

The monotonicity of $s \mapsto d(\gamma_s, \gamma'_s)/s$ allows to define angles, then tangent cones, then a first-order differential calculus. Since the definitions are shared between curved spaces, we factor them in a dedicated section.

1.1.1.1 Construction of the tangent cone

The tangent cone is often introduced as an "abstract completion", whose precise meaning is given in each example. The completion can actually be given a tangible meaning as follows. Consider the set of possibly discontinuous curves escaping from $x \in \Omega$ at finite speed, i.e.

$$\mathcal{A}_x := \left\{ \gamma : \mathbb{R}^+ \to \Omega \mid \gamma(0) = x \text{ and } \limsup_{s \searrow 0} d(x, \gamma(s))/s < \infty. \right\}$$

Define *reparametrization* as the operation $\cdot : \mathbb{R}^+ \times \mathcal{A}_x \to \mathcal{A}_x$ given by $(\lambda \cdot \gamma)(s) := \gamma(\lambda s)$, and

$$d_x : \mathcal{A}_x \times \mathcal{A}_x \to \mathbb{R}, \qquad d_x(\gamma,\gamma') := \limsup_{s \searrow 0} \frac{d(\gamma(s),\gamma'(s))}{s}. \tag{1.2}$$

The static curve $\gamma(s) \equiv x$ is denoted 0_x, so that one can write $0 \cdot \gamma = 0_x$ for any $\gamma \in \mathcal{A}_x$. Owing to the triangle inequality, d_x is finite-valued, and the limit sup is always defined.

Remark 1.1.2. *The set $\mathcal{A}_x$ is complete with respect to the pseudo-distance d_x. This can be seen by the following construction, alluded to in [AKP23, Proposition 6.11]: consider $(\gamma^n)_{n\in\mathbb{N}}$ a d_x-Cauchy sequence. For any fixed $m \in \mathbb{N}$, let $\varepsilon_m > 0$ be small enough so that*

$$\sup_{s\in(0,\varepsilon_m]} \frac{d(\gamma^m(s),\gamma^n(s))}{s} \leqslant d_x(\gamma^m,\gamma^n) + 2^{-m}$$

for all $n \in [\![0,m]\!]$. We may assume that $(\varepsilon_m)_m$ goes to 0 when m goes to ∞. Define then a curve γ by $\gamma(s) := \gamma^m(s)$ if $s \in (\varepsilon_{m+1},\varepsilon_m]$, and $\gamma(0) = x$. Although possibly discontinuous, γ satisfies

$$d_x\left(\gamma,\gamma^n\right) = \limsup_{s\searrow 0} \frac{d(\gamma_s,\gamma^n_s)}{s} \leqslant \sup_{m\geqslant n} \sup_{s\in(\varepsilon_{m+1},\varepsilon_m]} \frac{d(\gamma^m_s,\gamma^n_s)}{s} \leqslant \sup_{m\geqslant n} \sup_{s\in(0,\varepsilon_m]} \frac{d(\gamma^m_s,\gamma^n_s)}{s} \leqslant \sup_{m\geqslant n} d_x(\gamma^m,\gamma^n) + 2^{-m} \underset{n\to\infty}{\longrightarrow} 0.$$

In addition, $d_x(0,\gamma) \leqslant d_x(0,\gamma^0) + d_x(\gamma^0,\gamma)$ is bounded, so that $\gamma \in \mathcal{A}_x$ is a limit of the sequence $(\gamma^n)_n$.

One checks that the relation $\sim_x$ on $\mathcal{A}_x$ defined by $\gamma \sim_x \gamma'$ if $d_x(\gamma,\gamma') = 0$ is an equivalence relation. Thanks to Remark 1.1.2, the quotient $\mathcal{A}_x/\sim_x$ is complete with respect to the distance induced by d_x on equivalence classes. The quotient topology is induced by d_x. The tangent cone is built as the subset of $\mathcal{A}_x/\sim_x$ of elements that can be approximated by reparametrized geodesics.

Definition 1.1.3 (Tangent cone). *Let $x \in \Omega$, and $\mathscr{G}_x \subset \mathcal{A}_x$ be the subset of curves whose restriction on* $[0,1]$ *is a unit-speed geodesic. One introduces*

$$\begin{cases} \text{the regular tangent cone:} & \mathrm{T}'_x\Omega := \mathbb{R}^+ \cdot \mathscr{G}_x = \left\{\lambda\cdot\gamma \mid \lambda \in \mathbb{R}^+ \text{ and } \gamma \in \mathscr{G}_x\right\}, & \text{(1.3a)}\\ \text{the tangent cone:} & \mathrm{T}_x\Omega := \overline{\mathrm{T}'_x\Omega/\sim_x}^{d_x}, & \text{(1.3b)}\end{cases}$$

where the completion is taken in the (complete, metric) space $\mathcal{A}_x/\sim_x$.

1.1.1.2 The metric scalar product

To avoid repetition in the sequel, we define the metric scalar product in a quite general setting.

Definition 1.1.4 (Metric scalar product). *Let $x \in \Omega$. The metric scalar product attached to x is the unique extension by continuity of the application $\langle\cdot,\cdot\rangle_x : (\mathrm{T}'_x\Omega)^2 \to \mathbb{R}$ defined by*

$$\langle\gamma,\gamma'\rangle_x := \liminf_{h\searrow 0} \frac{d^2(x,\gamma(h)) + d^2(x,\gamma'(h)) - d^2(\gamma(h),\gamma'(h))}{2h^2} \qquad \forall \gamma,\gamma' \in \mathrm{T}'_x\Omega. \tag{1.4}$$

By the triangular inequality and the definition of $d_x(\cdot,\cdot)$ in (1.2), $\langle\cdot,\cdot\rangle_x$ has finite values, and the extension is indeed well-defined. Since $d(x,\gamma(h)) = hd(x,\gamma(1))$ admits a limit for any geodesic, passing to reparametrization and closure, one gets the simpler expression

$$\langle v, w\rangle_x = \frac{|v|_x^2 + |w|_x^2 - d_x^2(v,w)}{2} \tag{1.5}$$

for any $v, w \in \mathrm{T}_x\Omega$, with $|v|_x = d_x(v,0_x)$. It is clear that $\langle\gamma,\gamma'\rangle_x = \langle\gamma',\gamma\rangle_x$, that $\langle\gamma,0_x\rangle_x = 0$ for any γ, and that $\langle\lambda\cdot\gamma,\lambda\cdot\gamma'\rangle_x = \lambda^2\langle\gamma,\gamma'\rangle_x$ for any $\lambda \geqslant 0$. In the flat space $\mathbb{R}^d$, the definition (1.4) reduces to the classical scalar product. The metric scalar product furnishes an indication of the angle between the velocities of geodesics, as long as the said angle is defined.

Definition 1.1.5 (Defined angles). *A geodesic space* (Ω, d) *has defined angles if for all* $x \in \Omega$ *and* $\gamma, \gamma' \in \mathscr{G}_x$, *the limit*

$$\cos(\alpha) := \lim_{\substack{h,s\to 0\\ h,s>0}} \frac{d^2(x,\gamma_s) + d^2(x,\gamma'_h) - d^2(\gamma_s,\gamma'_h)}{2d(x,\gamma_s)d(x,\gamma'_h)}$$

exists (independently of the joint behaviour of h and s).

This condition is reminiscent of a Hilbertian behaviour, and fails even in quite simple spaces. In spaces with defined angles, the metric scalar product satisfies

$$\langle \alpha\cdot\gamma, \alpha^{-1}\cdot\gamma'\rangle_x = \langle\gamma,\gamma'\rangle_x \qquad \forall x,\gamma,\gamma' \text{ and } \alpha > 0.$$

Jointly with the positive homogeneity of the metric scalar product, this gives $\langle\lambda\cdot\gamma,\gamma'\rangle_x = \langle\gamma,\lambda\cdot\gamma'\rangle_x = \lambda\langle\gamma,\gamma'\rangle_x$ for all $\lambda \geqslant 0$.

1.1.2 CAT(0) spaces

We consider the definition of curvature in the Alexandrov sense, by comparison with a flat space.

Definition 1.1.6 (CAT(0) space). *A complete metric space* (Ω, d) *is CAT(0) if it is a geodesic space, and if for any* $x, y, z \in \Omega$, *any* $t \in [0,1]$ *and* $\gamma \in \mathscr{G}_x$ *a geodesic from x to y, there holds*

$$d^2\big(z,\gamma(t)\big) \leqslant (1-t)d^2(z,x) + td^2\big(z,y\big) - t(1-t)d^2(x,y). \tag{1.6}$$

The name CAT stands for Cartan (Élie), Alexandrov, and Topogonov. The 0 stands for a parameter $\kappa \in \mathbb{R}$ bounding the curvature from above. For a general κ, one compares triangles (or simplexes of finitely many points) with respect to the unique 2-dimensional manifold of constant curvature κ, asking triangles to be thinner than in the model space. We will also consider a class of locally CAT(0) spaces in Chapter 4, in which every point admits a CAT(0) neighbourhood.

Since geodesics are unique by the convexity of the squared distance, we allow ourselves the following set of notations.

$[xy]$	$\in \mathrm{AC}([0,1];\Omega)$	the unit-speed geodesic linking x to y	$x, y \in \Omega$
$(1-t)x \oplus ty$	$\in \Omega$	the unique point of $[xy]$ at distance $td(x,y)$ of x	$x, y \in \Omega,\ t \in [0,1]$
$\overrightarrow{xy}$	$\in \mathrm{T}_x\Omega$	the equivalence class of $[xy]$ in $\mathrm{T}_x\Omega$	$x, y \in \Omega$
$\uparrow_x^y$	$\in \mathrm{T}_x\Omega$	the unit direction $d(x,y)^{-1}\cdot\overrightarrow{xy}$	$x, y \in \Omega,\ x \neq y$
$\lambda\cdot v$	$\in \mathrm{T}_x\Omega$	the equivalence class of $\lambda\cdot\gamma$ for $\gamma \in v$	$\lambda \in \mathbb{R}^+,\ v \in \mathrm{T}_x\Omega$

Recall that the CAT(0) inequality (1.6) implies that $h \mapsto d(\gamma(h),\gamma'(h))/h$ is nondecreasing if γ and γ' are geodesics issued from the same point. Using the above notations, the infinitesimal distance is controlled by the global one:

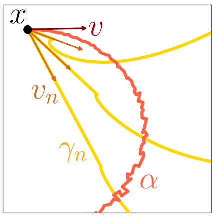
$$d_x(t\cdot\overrightarrow{xy}, t\cdot\overrightarrow{xz}) \leqslant td(y,z) \qquad \forall x,y,z \in \Omega,\ t \in [0,1]. \tag{1.7}$$

As shown in [BH99, Theorem 3.19], each tangent cone $(\mathrm{T}_x\Omega, d_x)$ to a point x of a CAT(0) space is itself CAT(0). It may not be locally compact, even if the space Ω is so[1]. One can correct this by assuming that the space is geodesically extendible, in that any geodesic can be extended in a geodesic ray parametrized on $\mathbb{R}$. This assumption yields impressive results regarding differentiability, as for instance the Alexandrov theorem of Lytchak and Nagano [LN19]. In this manuscript, we do not assume Ω to be geodesically extendible, mainly because our arguments do not seem to get simplified in this setting.

We introduce some definitions and results of differential calculus in CAT(0) spaces, mainly taken from the monograph of Alexander, Kapovitch, and Petrunin [AKP23].

Definition 1.1.7 (Derivative of curve [AKP23, Definition 6.9]). *Let* $a > 0$ *and* $\alpha : [0,a] \to \Omega$ *be a curve issued from* $x \in \Omega$. *The vector* $v \in T_x\Omega$ *is the right derivative of* α *at 0, briefly* $\alpha^+(0) = v$, *if for some sequence of* $(v_n)_n \subset$

[1] For instance, gluing a countable number of shorter and shorter needles around a common point produces a compact CAT(0) space with a tangent cone that is not locally compact, as pointed in https://mathoverflow.net/questions/428545.



$\mathrm{T}'_x\Omega$ *such that* $d_x(v_n, v) \to_n 0$, *and corresponding reparametrized geodesics* γ_n, *we have*

$$\lim_{n\to\infty} \limsup_{\varepsilon\to 0^+} \frac{d\left(\alpha(\varepsilon), \gamma_n(\varepsilon)\right)}{\varepsilon} = 0. \tag{1.8}$$

The left derivative $\alpha^-(t)$ *is the derivative of* $h \mapsto \alpha(t-h)$. *A curve* α *is differentiable at* $t \in (0, a)$ *if it admits a right and a left derivative at* t *such that* $|\alpha_t^+|_{\alpha_t} = |\alpha_t^-|_{\alpha_t} = \frac{1}{2} d_x(\alpha_t^+, \alpha_t^-)$.

If α has a right derivative, and $\beta : [0, a] \to \Omega$ is another curve issued from the same point and such that $d(\alpha(\varepsilon), \beta(\varepsilon)) = o(\varepsilon)$, then β_0^+ is defined and equal to α_0^+. Derivatives of curves allow to define directional derivatives of maps.

Definition 1.1.8 (Differential [AKP23, Definition 6.15]). *Let* $\varphi : \Omega \to \mathbb{R}$ *be a semiconcave and locally Lipschitz function. Its differential at* $x \in \Omega$ *is the unique map* $D_x\varphi : \mathrm{T}_x\Omega \to \mathbb{R}$ *such that* $D_x\varphi(\alpha_0^+) = (\varphi\circ\alpha)^+(0)$ *for any curve* $\alpha : [0, a] \to \Omega$ *with* $a > 0$, $\alpha_0 = x$ *and* α_0^+ *defined.*

By [AKP23, Lemma 13.15], if $\varphi : \Omega \to \mathbb{R}$ is semiconcave and locally Lipschitz, its differential $D_x\varphi$ is positively homogeneous, geodesically concave in $\mathrm{T}_x\Omega$, and Lipschitz with a constant bounded by any local Lipschitz constant of φ in a ball containing x. If $v = \overrightarrow{xy}$ for some point $y \in \Omega$, then $D_x\varphi(v)$ is the directional derivative of φ along the geodesic $[xy]$. As useful examples, let us mention that

$$D_x d(\cdot, y)(v) = \begin{cases} |v|_x & \text{if } y = x, \\ -\langle \uparrow_x^y, v\rangle_x & \text{otherwise,} \end{cases} \qquad \text{and} \qquad D_x d^2(\cdot, y)(v) = -2\,\langle \overrightarrow{xy}, v\rangle_x. \tag{1.8}$$

An important consequence is that the metric scalar product is geodesically concave in $\mathrm{T}_x\Omega$, as the differential of minus the squared distance. By approximation, one gets that for any $x \in \Omega$ and $p, q, v \in \mathrm{T}_x\Omega$,

$$\langle p, v\rangle_x - \langle q, v\rangle_x \leqslant |v|_x\, d_x(p, q). \tag{1.9}$$

The metric gradient of a concave function is defined as the direction of maximal growth.

Definition 1.1.9 (Gradient [AKP23, Definition 13.17]). *Let* $\varphi : \Omega \to \mathbb{R}$. *Suppose that the differential* $D_x\varphi : \mathrm{T}_x\Omega \to \mathbb{R}$ *is defined at a point* $x \in \Omega$. *A tangent vector* $g \in \mathrm{T}_x\Omega$ *is called a* gradient of φ *at* x, *denoted* $g = \nabla_x\varphi$, *if* $D_x\varphi(\omega) \leqslant \langle g, \omega\rangle_x$ *for all* $\omega \in \mathrm{T}_x\Omega$, *with equality at* g.

Any locally Lipschitz and semiconcave function admits a unique gradient at every point [AKP23, Proposition 13.19]. One can follow the direction of these gradients, and we will primarily rely on the following existence result.

Proposition 1.1.10 (Existence and uniqueness of gradient flow [AKP23, Propositions 16.15 and 16.19]). *Let* $\varphi : \Omega \to \mathbb{R}$ *be locally Lipschitz and concave. Then there exists a unique semigroup* $\vartheta : \mathbb{R}^+ \times \Omega \to \Omega$ *such that each* $h \mapsto \vartheta(h, x)$ *is locally Lipschitz, and for all* $s \geqslant 0$,

$$\vartheta(0, x) = x, \qquad \vartheta(h, \vartheta(s, x)) = \vartheta(h+s, x), \qquad \vartheta(\cdot, x)_h^+ = \nabla_{\vartheta(h,x)}\varphi.$$

1.1.3 CBB(0) spaces

Spaces with Curvature Bounded Below (CBB) are defined by comparing triangles. By [AKP23, p. 8.25], this is equivalent to the following statement, that we take as a definition.

Definition 1.1.11 (CBB(0) space). *A complete metric space* (Ω, d) *is CBB(0) if it is a geodesic space, and if for any* $x, y, z \in \Omega$, *any* $t \in [0, 1]$ *and* $\gamma \in \mathscr{G}_x$ *a geodesic from* x *to* y, *there holds*

$$d^2\left(z, \gamma(t)\right) \geqslant (1-t)d^2(z, x) + t d^2\left(z, y\right) - t(1-t)d^2(x, y). \tag{1.10}$$

This manuscript is mainly concerned with the Wasserstein space, whose specificities are gathered in Section 1.1.4. Before this, we point at some properties of CBB spaces constructed by quotients of Hilbert spaces. Let $(E, \langle\cdot,\cdot\rangle_E)$ be a Hilbert space and G be a group of linear bijective isometries on E. Let Ω be the set of equivalence classes for the relation $f^0 \sim f^1$ if there exists $g \in G$ such that $f^1 = g(f^0)$. The equivalence class of $f^0 \in E$ is denoted $[f^0]$. For $x, y \in \Omega$, let

$$d(x,y) := \inf_{f^0\in x, f^1\in y} \|f^0 - f^1\|_E.$$

We assume that each class x is closed in E, and that the infimum is attained for all x and y. This simple setting will serve as a model for the L-differentiability in Wasserstein spaces. In anticipation of this part, we provide the argument of two simple facts, although we do not claim any novelty here.

Lemma 1.1.12. (Ω, d) *is a complete geodesic CBB space.*

Proof. By [AKP23, Corollary 8.35], (Ω, d) is CBB. Let $x, y \in \Omega$ and f^0, f^1 be such that $d(x,y) = \|f^0 - f^1\|_E$. For all $t \in [0,1]$, denote $\gamma_t := [(1-t)f^0 + tf^1] \in \Omega$. Then

$$d(x,y) \leq d(x,\gamma_t) + d(\gamma_t, y) \leq t\|f^0 - f^1\|_E + (1-t)\|f^0 - f^1\|_E = td(x,y) + (1-t)d(x,y) = d(x,y).$$

Hence equality holds everywhere, and $d(x,\gamma_t) = td(x,y) = \|f^0 - [(1-t)f^0 + tf^1]\|_E$ for all $t \in [0,1]$. Repeating the operation between f^0 and $(1-t)f^0 + tf^1$, we get that for all $s \in [0,t]$,

$$\begin{aligned} d(\gamma_s,\gamma_t) &= d\left([(1-s)f^0 + sf^1], \gamma_t\right) = d\left(\left[\left(1-\frac{s}{t}\right)f^0 + \frac{s}{t}\left((1-t)f^0 + tf^1\right)\right], \gamma_t\right) \\ &= \left(1-\frac{s}{t}\right)d(x,\gamma_t) = (t-s)d(x,y). \end{aligned}$$

Therefore γ is a geodesic. Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence. Up to extraction, we may assume that $d(x_n, x_{n+1}) \leq e^{-n}$. Pick $f^0 \in x_0$, and for each n, let $f^{n+1} \in x_{n+1}$ such that $d(x_{n+1}, x_n) = \|f^{n+1} - f^n\|_E$. Then

$$\lim_{n\to\infty}\sup_{m\geq n}\|f^m - f^n\|_E \leq \lim_{n\to\infty}\sup_{m\geq n}\sum_{p=n}^{m}\|f^{p+1} - f^p\|_E \leq \lim_{n\to\infty}\sup_{m\geq n}\sum_{p=n}^{m} e^{-p} \leq \lim_{n\to\infty}\frac{e^{-n}}{1-e^{-1}} = 0.$$

So (Ω, d) is complete. □

Lemma 1.1.13 (Geodesics are image of geodesics). *Let γ be a geodesic of Ω. For any $f^0 \in \gamma_0$, there exists $f^1 \in \gamma_1$ such that $\gamma_t = [(1-t)f^0 + tf^1]$ and $d(\gamma_t,\gamma_s) = |t-s|\,\|f^0 - f^1\|_E$ for all $s,t \in [0,1]$.*

Proof. Let $f^0, f^2 \in E^2$ be such that $f^0 \in \gamma_0$, $f^2 \in \gamma_{1/2}$ and $d(\gamma_0,\gamma_{1/2}) = \|f^0 - f^2\|_E$. Let $t \in [1/2, 1]$ be arbitrary, and $f^3 \in \gamma_{1/2}$, $f^4 \in \gamma_t$ such that $d(\gamma_{1/2},\gamma_t) = \|f^3 - f^4\|_E$. Since $f^3 \sim f^2$, there exists $g \in G$ such that $f^2 = g(f^3)$. Denote $f^5 := g(f^4)$. As g is an isometry, $d(\gamma_{1/2},\gamma_t) = \|f^3 - f^4\|_E = \|f^2 - f^5\|_E$. Consequently

$$\|f^0 - f^5\|_E \leq \|f^0 - f^2\|_E + \|f^2 - f^5\|_E = d(\gamma_0,\gamma_{1/2}) + d(\gamma_{1/2},\gamma_t) = d(\gamma_0,\gamma_t) \leq \|f^0 - f^5\|_E.$$

In the Hilbert space E, equality happens in the triangular inequality if and only if f^2 belongs to the segment joining f^0 to f^5. Checking times, we get $f^5 = f^0 + \frac{t}{1/2}(f^2 - f^0)$. Let $f^1 := f^0 + \frac{1}{1/2}(f^2 - f^0)$. We just proved that for all $t \in [1/2, 1]$, one has $\gamma_t = [(1-t)f^0 + tf^1]$, and $d(\gamma_s,\gamma_t) = |t-s|\,\|f^0 - f^1\|_E$ for all $s \in \{0\} \cup [1/2, 1]$ and $t \in [1/2, 1]$. Applying the same reasoning on $[0, 1/2]$ with $\tilde{f}^0 := f^1$ and $\tilde{f}^2 := 1/2f^0 + 1/2f^1 = f^2$, we get the result on $[0,1]$. □

To justify our interest in this construction, and introduce the following section, consider the Hilbert space $L^2([0,1];\mathbb{R})$. The set of measurable measure-preserving bijections (swapping chunks) is a group. The quotient of $L^2([0,1];\mathbb{R})$ by the closure of its orbits reads as a set of elements that are fully determined by the measure of their level sets, regardless of the shape of these level sets. Here we are not exactly in the setting described above, but this quotient turns out to corresponds to the set of probability measures on $\mathbb{R}$, endowed with the Wasserstein distance, that we now introduce in a more classical way.

1.1.4 Wasserstein spaces

"Le pushforward, c'est comme le vélo"
Gabriel Peyré, SMAI MODE Days 2024

1.1.4.1 First definitions

Let Ω be a Polish space, and $c : \Omega^2 \to \mathbb{R}^+$ be a *cost function*. The Monge-Kantorovich distance associated to the cost c is an application on the set of Borel probability measures over Ω. Let us agree on the terminology.

Definition 1.1.14 (Borel probability measures). *Denote by $\mathcal{B}_\Omega$ the Borel $\sigma-$algebra of Ω. The set $\mathscr{P}(\Omega)$ of Borel probability measures over Ω is the set of applications $\mu : \mathcal{B}_\Omega \to [0,1]$ satisfying $\mu(\Omega) = 1$ and $\mu(\bigsqcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mu(A_n)$ for each countable family of 2-by-2 disjoint sets $(A_n)_{n\in\mathbb{N}} \subset \mathcal{B}_\Omega$.*

If Ω' is another Polish space, one can send the elements of $\mathscr{P}(\Omega)$ on that of $\mathscr{P}(\Omega')$ as follows.

Definition 1.1.15 (Pushforward). *Let $f : \Omega \to \Omega'$ be $(\mathcal{B}_\Omega, \mathcal{B}_{\Omega'})-$measurable. The pushforward by f is the operator $f\# : \mathscr{P}(\Omega) \to \mathscr{P}(\Omega')$ defined by*

$$(f\#\mu)(A') := \mu(f^{-1}(A')) = \mu\left(\left\{x \in \Omega \;\middle|\; f(x) \in A'\right\}\right) \qquad \forall A' \subset \Omega' \textit{ measurable.}$$

The pushforward is the adjoint of composition, in that for any bounded measurable $\varphi : \Omega' \to \mathbb{R}$,

$$\int_{x'\in\Omega'} \varphi(x') d[f\#\mu](x') = \int_{x\in\Omega} \varphi(f(x)) d\mu(x).$$

The reader that feels interested by measures is invited to metabolize this definition, as it provides the basic algebraic tool in all the sequel. As a first example, in a product space $\mathcal{A} \times \mathcal{B}$ with canonical projections π_a, π_b defined by $\pi_a(a,b) = a$ and $\pi_b(a,b) = b$, the pushforward of a measure $\eta \in \mathscr{P}(\mathcal{A}\times\mathcal{B})$ by π_a integrates on the lines $\{(a,b)\}_{b\in\mathcal{B}}$. More precisely, $[\pi_a\#\eta](A) = \eta\{(a,b) \mid a \in A,\ b \in \mathcal{B}\}$ for all $A \subset \mathcal{A}$ measurable. By the socks-shoes formula $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$, pushforwards can be composed in $f\#(g\#\mu) = (f \circ g)\#\mu$.

Definition 1.1.16 (Monge-Kantorovich distance). *Define* transport plans *between $\mu, \nu \in \mathscr{P}(\Omega)$ as*

$$\Gamma(\mu,\nu) := \left\{\eta \in \mathscr{P}(\Omega^2) \;\middle|\; \pi_x\#\eta = \mu \textit{ and } \pi_y\#\eta = \nu\right\}.$$

This set contains at least the product measure $\mu \otimes \nu$. The Monge-Kantorovich distance with cost $c(\cdot,\cdot)$, and the family of $p-$Wasserstein distances, are respectively defined by

$$d_{MK,c}(\mu,\nu) := \inf_{\eta\in\Gamma(\mu,\nu)} \int_{(x,y)\in\Omega^2} c(x,y) d\eta(x,y), \qquad \textit{and} \qquad d_{\mathcal{W},p}(\mu,\nu) := d_{MK,d^p}^{1/p}(\mu,\nu).$$

In the previous definitions, $d_{MK,c}$ and $d_{\mathcal{W},p}$ are not granted to be finite.

Definition 1.1.17 (Monge-Kantorovich space). *The spaces $\mathscr{P}_{MK,c}(\Omega)$ and $\mathscr{P}_p(\Omega)$ are defined as*

$$\mathscr{P}_{MK,c}(\Omega) := \left\{ \mu \in \mathscr{P}(\Omega) \;\middle|\; d_{MK,c}(\mu,\delta_o) < \infty \right\}, \qquad \textit{and} \qquad \mathscr{P}_p(\Omega) := \left\{ \mu \in \mathscr{P}(\Omega) \;\middle|\; d_{\mathcal{W},p}(\mu,\delta_o) < \infty \right\}.$$

On these subsets, it is now true that $d_{\mathcal{W},p}$ is a complete distance, as well as $d_{MK,c}$ if c satisfies the triangular inequality. Moreover, the infimum is attained under very light assumptions on $c(\cdot,\cdot)$. The proof of these facts involves mainly two tools: the possibility to compose (or "glue") plans, and compactness.

1.1.4.2 Gluing plans

A transport plan $\eta \in \Gamma(\mu,\nu)$ provides a way to direct the mass of μ on that of ν: for any pair $A, B \in \mathcal{B}_\Omega$, the quantity $\eta(A \times B)$ models the amount of $\mu(A)$ that is transferred to $\nu(B)$. Given another transport plan γ from ν to some $\omega \in \mathscr{P}(\Omega)$, one expects to be able to chain η and γ to obtain a transport plan from μ to ω. This result, and useful variations, are deduced from the use of disintegration of measures.

Definition 1.1.18 (Disintegration). *Let $\mu \in \mathscr{P}(\Omega)$, and $f : \Omega \to \Omega'$ be measurable. A disintegration of μ with respect to f is a family $(\mu_{x'})_{x'\in\Omega'}$ such that*

- *for each $x' \in \Omega'$, $\mu_{x'}$ belongs to $\mathscr{P}(\Omega)$,*

– *for each* $A \in \mathcal{B}_\Omega$, *the function* $x' \to \mu_{x'}(A)$ *is* $(\mathcal{B}_{\Omega'}, \mathcal{B}_{\mathbb{R}})-$*measurable and* $[f\#\mu]-$*integrable,*

– *for each* $(A, B) \in \mathcal{B}_\Omega \times \mathcal{B}_{\Omega'}$, *there holds*

$$\mu(A \cap f^{-1}(B)) = \int_{x' \in B} \mu_{x'}(A)\, d[f\#\mu](x'). \tag{1.11}$$

If, in addition, $\mu_{x'}$ *is concentrated on* $f^{-1}(\{x'\})$ *for* $[f\#\mu]-$*almost every* $x' \in \Omega'$, *the disintegration is said to be* proper.

This definition corresponds to [Bog07, Definition 10.4.2], where it is called a system of (proper) regular conditional measures. A disintegration complements the pushforward to retrieve the original measure; this is clearer on the equivalent reformulation of (1.11) as

$$\int_{x \in \Omega} \varphi(x)\, d\mu(x) = \int_{x' \in \Omega'} \int_{x \in \Omega} \varphi(x)\, d\mu_{x'}(x)\, d[f\#\mu](x') \qquad \forall \varphi \in \mathcal{C}_b(\Omega; \mathbb{R}). \tag{1.12}$$

Disintegrations are most often used in product spaces, when f is a projection. In this context, it is common to denote $\eta = \eta_x \otimes \mu$ if $\mu = \pi_x \# \eta$. Existence of disintegrations in general settings is by no means an easy matter, and obstructions may appear between the proper condition and the measurability condition. The following result is sufficient in our setting.

Proposition 1.1.19 (Existence and essential uniqueness [Bog07, Lemma 10.4.3 and Theorem 10.4.8]). *Assume that* Ω *and* Ω' *are Polish. Then, for any* $\mu \in \mathscr{P}(\Omega)$ *and* $f : \Omega \to \Omega'$ *measurable, there exists a proper disintegration family* $(\mu_{x'})_{x' \in \Omega'}$ *in the sense of Definition 1.1.18. Moreover, any two disintegrations coincide* $[f\#\mu]-$*almost everywhere.*

The proof of Proposition 1.1.19 combines Lemma 10.4.3 and Theorem 10.4.8 of Bogachev [Bog07], and is detailed in Example 10.4.11 of the same reference. Its most useful consequence is the gluing Lemma.

Lemma 1.1.20 (Gluing Lemma). *Let* Ω_1, Ω_2 *be Polish spaces and* $(\mu^1, \mu^2) \in \mathscr{P}(\Omega_1) \times \mathscr{P}(\Omega_2)$. *Assume that* f_1, f_2 *are measurable applications from* Ω_1, Ω_2 *to a common Polish space* Ω, *and that* $f_1\#\mu^1 = f_2\#\mu^2 =: \nu \in \mathscr{P}(\Omega)$. *Then there exists* $\eta \in \mathscr{P}(\Omega_1 \times \Omega_2)$ *with first marginal* μ^1, *second marginal* μ^2, *and such that* $(f_1, f_2)\#\eta = (id, id)\#\nu \in \mathscr{P}(\Omega^2)$.

Gluing two plans allows to prove that the Monge-Kantorovich distances satisfy the triangular inequality whenever the underlying cost does.

Remark 1.1.21 (On the origin of the gluing lemma)**.** *Lemma 1.1.20 is equivalent to the composition of maps* $f \circ g$ *if* $\alpha = (id, g)\#\mu$ *and* $\beta = (id, f)\#\nu$, *or to the Chapman-Kolmogorov formula to compose Markov kernels. Perhaps under the influence of [Vil03], it is most often cited under the name "Dudley's Lemma", although the given reference concentrates on pairings of finite sets. Richard Dudley himself refers to it as the "pairing lemma" or marriage lemma in [Dud76, Theorem 18.1], and credits previous works of D. König (1931) and P. Hall (1935). The interested reader will find a few historical references in [Pat23], in which the author concludes that "it is probably impossible to determine the origin of the gluing lemma."*

1.1.4.3 Topological matters

From now on, we focus on the $p-$Wasserstein distances. In Polish spaces, all Borel probability measures are *tight*, or *inner regular on compact sets*:

$$\mu(O) = \sup_{K \subset O,\ K \text{ compact}} \mu(K) \qquad \forall O \subset \Omega \text{ open}.$$

In particular, for any $\varepsilon > 0$, there exists a compact set $K \subset \Omega$ such that $\mu(K) \geqslant 1 - \varepsilon$. We refer the reader to [Bil99, Theorem 1.3] for the proof and comments. Tightness provides a very general compactness criterion in a weak topology.

Definition 1.1.22 (Narrow convergence). *A sequence $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}(\Omega)$ converges narrowly to $\mu \in \mathscr{P}(\Omega)$ if*

$$\int_{x\in\Omega} \varphi(x)\,d\mu_n(x) \underset{n\to\infty}{\longrightarrow} \int_{x\in\Omega} \varphi(x)\,d\mu(x) \qquad \forall\, \varphi \in \mathcal{C}_b(\Omega;\mathbb{R}).$$

The narrow topology on $\mathscr{P}(\Omega)$ is generated by the sets whose complementary is narrowly closed.

The narrow topology is the weak-$*$ topology of $\mathcal{C}_b(\Omega;\mathbb{R})$, and often called the weak topology. A set $M \subset \mathscr{P}(\Omega)$ is uniformly tight if for any $\varepsilon > 0$, there exists a compact K such that $\mu(K) \geqslant 1-\varepsilon$ for all $\mu \in M$.

Theorem 1.1.23 (Prokhorov's tightness criterion [Pro56, Theorem 1.12]). *A set $M \subset \mathscr{P}(\Omega)$ is uniformly tight if and only if it is relatively compact in the narrow topology.*

This topology is the one used to prove existence of minimizers in the definitions of $d_{MK,c}$ and $d_{\mathcal{W},p}$, under lower semicontinuity and suitable lower bounds of the cost $c(\cdot,\cdot)$ [Vil09, Theorem 4.1]. In the case where the cost in unbounded, as in most of this manuscript, the compact sets in the $p-$Wasserstein topology are characterized by a stronger criterion. One says that $M \subset \mathscr{P}(\Omega)$ is $p-$uniformly integrable if for any point $x \in \Omega$, there holds

$$\lim_{R\to\infty}\sup \sup_{\mu\in M} \int_{y\in\Omega, d(x,y)\geqslant R} d^p(x,y)\,d\mu(y) = 0.$$

Theorem 1.1.24 (Compactness criterion [AGS05, Theorem 7.1.5]). *A set $M \subset \mathscr{P}_p(\Omega)$ is $p-$uniformly tight if and only if it is relatively compact in $(\mathscr{P}_p(\Omega), d_{\mathcal{W},p})$.*

As a useful example, the set of transport plans between two measures in $\mathscr{P}_p(\Omega)$ is compact. It seems customary to let this as an exercise to the reader, so let us provide the details.

Lemma 1.1.25. *For any $\mu,\nu \in \mathscr{P}_p(\Omega)$, the set $\Gamma(\mu,\nu)$ is compact in $(\mathscr{P}_p(\Omega), d_{\mathcal{W},p})$.*

Proof. For any $R > 0$, denote $I_\mu(R) := \int_{|x|\geqslant R} |x|^2 d\mu$, which goes to 0 when R goes to ∞. For any $\eta \in \Gamma(\mu,\nu)$, we have that

$$\int_{(x,y),|x|^p+|y|^p\geqslant R^p} \left[|x|^p + |y|^p\right] d\eta = \int_{(x,y),|x|^p+|y|^p\geqslant R^p} |x|^p\,d\eta + \int_{(x,y),|x|^p+|y|^p\geqslant R^p} |y|^p\,d\eta. \tag{1.13}$$

One the one hand, dividing the integrals on two subsets and using that ($|x|^p < R^p/2$ and $R^p/2 \leqslant |y|^p$) implies $|x|^p \leqslant |y|^p$,

$$\int_{|x|^p+|y|^p\geqslant R^p} |x|^p\,d\eta \leqslant \int_{|x|^p\geqslant R^p/2} |x|^p\,d\eta + \int_{|x|^p<R^p/2\leqslant |y|^p} |x|^p\,d\eta \leqslant I_\mu(R^p/2) + I_\nu(R^p/2).$$

The same estimate holds for the second term, so that the left hand-side of (1.13) is bounded by $2(I_\mu(R^p/2) + I_\nu(R^p/2))$ independently of $\eta \in \Gamma(\mu,\nu)$, and the latter set is $p-$uniformly tight. It is then relatively compact by Theorem 1.1.24, and as projections are continuous, it is also closed. □

In the particular case of Wasserstein distances, one has the following characterization.

Proposition 1.1.26 (Convergence in $\mathscr{P}_p(\Omega)$ [Vil09, Theorem 6.9]). *Let $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}_p(\Omega)$ and $\mu \in \mathscr{P}_p(\Omega)$. The following are equivalent:*

- *$(\mu_n)_n$ converges to μ with respect to the $p-$Wasserstein distance, i.e. $d_{\mathcal{W},p}(\mu_n,\mu) \underset{n\to\infty}{\longrightarrow} 0$;*
- *$(\mu_n)_n$ converges narrowly towards μ and $d_{\mathcal{W},p}(\mu_n,\delta_o) \longrightarrow_{n\to\infty} d_{\mathcal{W},p}(\mu,\delta_o)$.*

The spaces $\left(\mathscr{P}_p(\Omega), d_{\mathcal{W},p}\right)$ have the inconvenience of not being locally compact. On the other hand, the narrow topology is too weak to ensure boundedness of the moments. This is corrected with an intermediate weak topology in $\mathscr{P}_p(\Omega)$, defined as follows.

Definition 1.1.27 (Inductive topology in $\mathscr{P}_p(\Omega)$). *For any $N \in \mathbb{N}_*$, equip the $p-$Wasserstein ball $\overline{\mathscr{B}}_p(\delta_o, N)$ with the trace topology $\tau_{p,N}$ induced by the narrow topology. Let $\iota_{p,N} : \overline{\mathscr{B}}_p(\delta_o, N) \to \mathscr{P}_p(\Omega)$ be the canonical injection. The inductive topology τ_p in $\mathscr{P}_p(\Omega)$ is defined as the finest topology on $\mathscr{P}_p(\Omega)$ that lets all applications $\iota_{p,N}$ be continuous.*

The difference with convergence with respect to $d_{\mathcal{W},p}$ lies in the fact that the moments are only bounded, and may not converge towards $d_{\mathcal{W},p}(\mu, \delta_0)$. The topology τ_p is interesting for the following reason.

Proposition 1.1.28 (Convergence and compactness with respect to τ_p [Gig08, Def. 2.16 and below]). *A set is closed with respect to τ_p if and only if its intersection with any $\overline{\mathscr{B}}_p(\delta_o, N)$ is narrowly closed. A sequence $(\mu_n)_{n\in\mathbb{N}}$ of $\mathscr{P}_p(\Omega)$ converges towards $\mu \in \mathscr{P}_p(\Omega)$ with respect to τ_p if and only if it converges narrowly and*

$$\sup_{n\in\mathbb{N}} d_{\mathcal{W},p}(\mu_n, \delta_o) < \infty.$$

Moreover, closed $p-$Wasserstein balls in $\mathscr{P}_p(\Omega)$ are τ_p-compact.

In an arbitrary topological space, closed sets are also sequentially closed. The converse is true for metrizable spaces, but not necessarily for any topological space. The spaces that satisfy this property are called *sequential*. Since the narrow topology is metrizable (see [AGS05, Remark 5.1.1]), each topological space $\left(\overline{\mathscr{B}}_p(\delta_o, N), \tau_{p,N}\right)$ is sequential. This property is inherited by the inductive topology τ_p; it is actually a general fact, sequential spaces are stable by inductive limits [Fra65, Corollary 1.7]. The proof in the latter reference being quite elusive, we provide it here.

Lemma 1.1.29. *$\left(\mathscr{P}_p(\Omega), \tau_p\right)$ is a sequential space.*

Proof. Let $A \subset \mathscr{P}_p(\Omega)$ be sequentially τ_p-closed. By Proposition 1.1.28, it is enough to prove that each set $A_N := A \cap \overline{\mathscr{B}}_p(\delta_o, N)$ is narrowly closed, which is equivalent to sequentially narrowly closed. Let $\left(\mu_m\right)_m \subset A_N$ be a sequence converging narrowly to $\mu \in \overline{\mathscr{B}}_p(\delta_o, N)$. By the continuity of the injection $\iota_{p,N}$, the sequence $\left(\iota_{p,N}(\mu_m)\right)_m \subset A$ converges to $\iota_{p,N}(\mu)$, and as A is sequentially τ_p-closed, $\iota_{p,N}(\mu) \in A$. Since $\mu \in \overline{\mathscr{B}}_p(\delta_o, N)$, one has $\iota_{p,N}(\mu) = \mu$, and we conclude. □

In particular, the applications $\mu \mapsto d^p_{\mathcal{W},p}(\mu, \nu)$ are τ_p-lower semicontinuous.

Proposition 1.1.30 (Compact injections). *Let $1 \leqslant p < q$, and $A \subset \mathscr{P}_q(\Omega)$ be bounded with respect to $d_{\mathcal{W},q}$. Then A is relatively compact in $(\mathscr{P}_p, d_{\mathcal{W},p})$. As a consequence, if $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}_q(\Omega)$ converges towards $\mu \in \mathscr{P}_q(\Omega)$ with respect to τ_q, it also converges with respect to $d_{\mathcal{W},p}$.*

However, it is not true that $d_{\mathcal{W},p}(\mu_n, \mu) \to_n 0$ for any $p \in [1, q)$ implies $\mu_n \to \mu$ with respect to τ_q, as illustrated below.

Proof. By Hölder, $d_{\mathcal{W},p}(\mu, \delta_0) \leqslant d_{\mathcal{W},q}(\mu, \delta_0)$, so that $A \subset \mathscr{P}_p(\Omega)$. It then suffices to see that for all $R > 0$,

$$\sup_{\mu\in A} \int_{y\in\Omega, d(x,y)\geqslant R} d^p(x, y) d\mu = \sup_{\mu\in A} \int_{y\in\Omega, d(x,y)\geqslant R} d^{p-q}(x, y) d^q(x, y) d\mu \leqslant R^{p-q} \sup_{\mu\in A} \int_{y\in\Omega} d^q(x, y) d\mu.$$

Since A is bounded with respect to $d_{\mathcal{W},q}$, the last term goes to 0 when $R \to \infty$, and A is relatively compact in $(\mathscr{P}_p, d_{\mathcal{W},p})$ by Theorem 1.1.24. Consequently, if $(\mu_n)_{n\in\mathbb{N}}$ converges with respect to τ_q, the set $\bigcup_n\{\mu_n\}$ is relatively compact with respect to $d_{\mathcal{W},p}$, and some subsequence converges. The limit must coincide with the narrow limit μ, so that the whole sequence converges with respect to $d_{\mathcal{W},p}$. □

Let us collect a few examples to conclude this section. Since the difference in the topologies lies in the behaviour at infinity, it is enough to focus on sequences of measures of the form

$$\mu_n := (1 - \varepsilon_n)\delta_0 + \varepsilon_n \delta_n \qquad \forall n \in \mathbb{N}_*,$$

with their candidate limit being $\mu := \delta_0$, and $(\varepsilon_n)_{n\in\mathbb{N}_*} \subset [0, 1]$ a vanishing sequence.

The sequence μ_n converges with respect to...	value of ε_n	value of the moments
the narrow topology, but not τ_p for $p \geqslant 1$	$1/\ln(n)$	$d_{\mathcal{W},p} = (n^p/\ln(n))^{1/p}$
$d_{\mathcal{W},p}$ for all $p < \overline{p}$, but not $\tau_{\overline{p}}$	$\ln(n)/n^{\overline{p}}$	$d_{\mathcal{W},p}(\mu_n,\mu) = \left(n^{p-\overline{p}}\ln(n)\right)^{1/p}$ if $p < \overline{p}$, $d_{\mathcal{W},\overline{p}}(\mu_n,\mu) = \ln(n)^{1/p}$
τ_p, but not $d_{\mathcal{W},p}$	$1/n^p$	$d_{\mathcal{W},p}(\mu_n,\mu) \equiv 1$
$d_{\mathcal{W},p}$, but not $d_{\mathcal{W},p+1}$	$1/n^{p+1/2}$	$d_{\mathcal{W},p}(\mu_n,\mu) = 1/n^{1/(2p)}$, $d_{\mathcal{W},p+1}(\mu_n,\mu) = n^{1/(2p+2)}$
any $d_{\mathcal{W},p}$ for $p < \infty$, but not $d_{\mathcal{W},\infty}$	e^{-n}	$d_{\mathcal{W},p} = ne^{-n/p}$, and $d_{\mathcal{W},\infty} = n$.

1.1.4.4 Geodesics and the tangent cone to $\mathscr{P}_2(\mathbb{R}^d)$

We now focus on the 2−Wasserstein distance, that we denote $d_{\mathcal{W}}$ instead of $d_{\mathcal{W},2}$. The Wasserstein space $\left(\mathscr{P}_2(\mathbb{R}^d), d_{\mathcal{W}}\right)$ is a geodesic metric space with strong Hilbertian features. It can be studied from two points of views: first as a convenient space in which to define probability densities, with tools growing up from generalizations of vector fields to measure fields; and secondly, as a particular case of non-negatively curved space, with structures coming down from a greater generality. The two points of view are rich and useful, but admittedly produce a lot of definitions and notations, that we try to reduce to a minimum introduced in this section.

Geodesics and measure fields. The first specificity of Wasserstein spaces is that the geodesics are explicitly described by transport plans. Recall that $\Gamma(\mu,\nu) \subset \mathscr{P}_2((\mathbb{R}^d)^2)$ is the set of transport plans between μ and ν; the subset of optimal transport plans for the Wasserstein distance is indexed by o, i.e.

$$\Gamma_o(\mu,\nu) := \left\{ \eta \in \Gamma(\mu,\nu) \;\middle|\; d_{\mathcal{W}}^2(\mu,\nu) = \int_{x,y\in\mathbb{R}^d} |x-y|^2 \, d\eta(x,y) \right\}.$$

The subscript o will always denote optimality with respect to a cost that should be clear from the context. A curve $(\mu_s)_{s\in[0,1]} \subset \mathscr{P}_2(\mathbb{R}^d)$ is a geodesic between its endpoints if and only if it is given by

$$\mu_s = \left((1-s)\pi_x + s\pi_y\right)\#\eta$$

for some $\eta \in \Gamma_o(\mu,\nu)$ [AGS05, Ch. 7]. This parametrization allows to get the following fundamental curvature estimate, proved in [AGS05, Theorem 7.3.2].

Theorem 1.1.31. *$(\mathscr{P}_2(\mathbb{R}^d), d_{\mathcal{W}})$ is CBB(0). Equivalently, for any $\nu \in \mathscr{P}_2(\mathbb{R}^d)$, any $t \in [0,1]$ and $\eta = \eta(dx,dy) \in \mathscr{P}_2((\mathbb{R}^d)^2)$ that is optimal between its marginals, there holds*

$$d_{\mathcal{W}}^2\left(((1-t)\pi_x + t\pi_y)\#\eta, \nu\right) \geqslant (1-t)d_{\mathcal{W}}^2(\pi_x\#\eta,\nu) + td_{\mathcal{W}}^2(\pi_y\#\eta,\nu) - t(1-t)d_{\mathcal{W}}^2(\pi_x\#\eta,\pi_y\#\eta).$$

Note that the plan η sends initial *points* to terminal *points*. We now want to describe the geodesics by their initial point and their initial velocity; to this aim, we perform the change of variable $(x,y) \mapsto (x, y-x)$, or in other words,

$$\xi := (\pi_x, \pi_y - \pi_x)\#\eta.$$

Using the measure $\xi = \xi(x,v)$, the curve $(\mu_s)_{s\in[0,1]}$ rewrites as $\mu_s = (\pi_x + h\pi_v)\#\xi$. If η is induced by a map in the sense that $\eta = (\pi_x, T(\pi_x))\#\mu$, then $\xi = (\pi_x, T(\pi_x) - \pi_x)\#\mu$ is the pushforward of μ by the *vector field* $x \mapsto (x, T(x) - x)$. For this reason, we adopt the terminology of *measure field* to refer to ξ.

Definition 1.1.32 (Measure fields). *Let*

$$\mathrm{T}^n\mathbb{R}^d := \left\{ (x, v_1, \cdots, v_n) \;\middle|\; x \in \mathbb{R}^d,\ v_i \in \mathrm{T}_x\mathbb{R}^d \ \ \forall i \in [\![1,n]\!] \right\},$$

endowed with $\left|(x,v_1,\cdots,v_n)-(y,w_1,\cdots,w_n)\right|^2 := \left|x-y\right|^2+\sum_{i=1}^n |v_i-w_i|^2$. *If* $n=1$*, the tangent bundle* $\mathrm{T}^1\mathbb{R}^d$ *is shortened in* $\mathrm{T}\mathbb{R}^d$. *A* measure field *is an element of* $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)$. *It is* issued from μ *if it belongs to*

$$\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu := \left\{ \xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d) \;\middle|\; \pi_x\#\xi = \mu \right\}.$$

The sets $\mathscr{P}_2(\mathrm{T}^n\mathbb{R}^d)_\mu$ *are defined similarly for* $n \geqslant 2$.

Measure fields are the velocities of curves, exactly as $v \in \mathrm{T}_x\mathbb{R}^d$ is the velocity of the curve $h \mapsto x + hv$. We follow the conventional notations in the definition of a partial inverse of exp.

Definition 1.1.33 (Exponential and partial inverse). *Let* $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. *The exponential*

$$\exp_\mu : \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \to \mathscr{P}_2(\mathbb{R}^d), \qquad \exp_\mu(\xi) := (\pi_x + \pi_v)\#\xi$$

admits the (multivalued, partial) inverse $\exp_\mu^{-1} : \mathscr{P}_2(\mathbb{R}^d) \rightrightarrows \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$*, given by*

$$\exp_\mu^{-1}(\nu) := \left\{ \xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \;\middle|\; \exp_\mu(\xi) = \nu \text{ and } \int_{(x,v)\in\mathrm{T}\mathbb{R}^d} |v|^2\, d\xi = d_{\mathcal{W}}^2(\mu,\nu) \right\}.$$

Tangent cone. The geometric tangent cone is introduced in the thesis of N. Gigli [Gig08], and follows the construction in curved spaces. As opposed to the general case of a non-negatively curved metric space, the tangent cones to $\mathscr{P}_2(\mathbb{R}^d)$ are all embedded into $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)$, which can be equipped with the corresponding Wasserstein, narrow and inductive topologies. However, this is not the natural topology within a cone attached to a given measure.

Definition 1.1.34 (The cone distance). *Let* $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. *Denote by*

$$\Gamma_\mu(\xi,\zeta) := \left\{ \alpha = \alpha(dx,dv,dw) \in \mathscr{P}_2(\mathrm{T}^2\mathbb{R}^d) \;\middle|\; (\pi_x,\pi_v)\#\alpha = \xi,\ (\pi_x,\pi_w)\#\alpha = \zeta \right\}.$$

The cone distance $W_\mu : \left(\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu\right)^2 \to \mathbb{R}^+$ *is defined as*

$$W_\mu^2(\xi,\zeta) := \inf_{\alpha\in\Gamma_\mu(\xi,\zeta)} \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} |v-w|^2\, d\alpha(x,v,w) \qquad \forall \xi,\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu. \tag{1.14}$$

The distance to $0_\mu = (\pi_x, 0)\#\mu$ *is denoted by* $\|\xi\|_\mu := W_\mu(\xi, 0_\mu) = d_{\mathcal{W}}(\xi, 0_\mu) = \sqrt{\int_{(x,v)\in\mathrm{T}\mathbb{R}^d} |v|^2\, d\xi}$.

The cone distance is designed to forbid transfer of mass from (x,v) to (y,w) if $x \neq y$. It satisfies

$$W_\mu^2(\xi,\zeta) = \int_{x\in\mathbb{R}^d} d^2_{\mathcal{W},\mathrm{T}_x\mathbb{R}^d}(\xi_x,\zeta_x)\, d\mu(x),$$

and was originally defined as such in [Gig08, (4.7)], then proved equivalent to (1.14) in Proposition 4.2 of the same reference. In particular, if $\xi = f\#\mu$ and $\zeta = g\#\mu$ for $f,g \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$, one recovers the L^2 norm $W_\mu(\xi,\zeta) = \|f-g\|_\mu$, which may differ greatly from $d_{\mathcal{W}}(\xi,\zeta)$. By the same reasoning as in Theorem 1.1.31, the metric space $(\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu, W_\mu)$ is a complete geodesic CBB(0) space.

Definition 1.1.35 (Geometric tangent cone). *Let* $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. *The tangent cone is defined by*

$$\mathbf{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d) := \overline{\left\{\lambda\cdot\eta \mid \eta \in \exp_\mu^{-1}(\nu),\ \nu\in\mathscr{P}_2(\mathbb{R}^d),\ \lambda\in\mathbb{R}^+\right\}}^{W_\mu} = \overline{\mathbb{R}^+\cdot\exp_\mu^{-1}\left(\mathscr{P}_2(\mathbb{R}^d)\right)}^{W_\mu}.$$

The tangent cone, and in general $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, is endowed with various operations.

Definition 1.1.36 (Operations in $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$). *Given* $\xi,\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, $\lambda\in\mathbb{R}$*, define the*

scalar multiplication	$\lambda\cdot\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$	$\lambda\cdot\xi := (\pi_x, \lambda\pi_v)\#\xi$
set-valued sum	$\xi\oplus\zeta \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$	$\xi\oplus\zeta := \left\{(\pi_x, \pi_v+\pi_w)\#\beta \mid \beta\in\Gamma_\mu(\xi,\zeta)\right\}$
barycenter	$Bary_{\mathrm{T}\mathbb{R}^d}(\xi) \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$	$Bary_{\mathrm{T}\mathbb{R}^d}(\xi)(x) := \left(x, \int_{v\in\mathrm{T}_x\mathbb{R}^d} v\,d\xi_x(v)\right)$.

In the case where ξ or ζ is induced by a map, $\xi \oplus \zeta$ is reduced to one element. Note that there always holds $W_\mu(\lambda\cdot\xi,\lambda\cdot\zeta) \leqslant |\lambda| W_\mu(\xi,\zeta)$ for $\xi,\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ and $\lambda \in \mathbb{R}$. Similarly, one can construct a transport plan between the barycenters as the mean of an optimal transport plan between the measure fields: working with disintegrations, this yields the following estimate.

Lemma 1.1.37 (1-Lipschitz barycenter). *For any $\xi,\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, there holds*

$$W_\mu(Bary_{\mathrm{T}\mathbb{R}^d}(\xi)\#\mu, Bary_{\mathrm{T}\mathbb{R}^d}(\zeta)\#\mu) \leqslant W_\mu(\xi,\zeta).$$

The following result is a modification of the gluing Lemma which is simple, but so fundamental that we prefer to provide a proof.

Lemma 1.1.38. *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ and $\gamma \in \mathscr{P}_2(\mathrm{T}^n\mathbb{R}^d)_\mu$ for $n \geqslant 1$. Denote (x,v) the variables of ξ and $(x,v_1,\cdots,v_n)$ that of γ. Let $T \in L^2_\gamma(\mathrm{T}^n\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ be such that $\pi_x(T(x,v_1,\cdots,v_n)) = x$ for $\gamma-$a.e. point in $\mathrm{T}^n\mathbb{R}^d$. Then*

$$\Gamma_\mu(\xi, T\#\gamma) = \left\{(\pi_x,\pi_v,\pi_v(T(\pi_x,\pi_{v_1},\cdots,\pi_{v_n})))\#\alpha \;\middle|\; \alpha = \alpha(dx,dv,dv_1,\cdots,dv_n) \in \Gamma_\mu(\xi,\gamma)\right\}.$$

Proof. The inclusion $\supset$ is trivial. On the other hand, let $\beta = \beta(dx,du,dw) \in \Gamma_\mu(\xi,T\#\gamma)$. Disintegrate $\beta = \beta_{(x,w)} \otimes (T\#\gamma)$, where the family $(\beta_{(x,w)})_{(x,w)\in\mathrm{T}\mathbb{R}^d} \subset \mathscr{P}_2(\mathrm{T}^2\mathbb{R}^d)$ is unique up to a $T\#\gamma-$negligible set and for $T\#\gamma$–a.e. $(x,w) \in \mathrm{T}\mathbb{R}^d$, $\beta_{(x,w)}$ is concentrated on the set of (x',v,w') such that $x' = x$ and $w' = w$. Disintegrate also $\gamma = \gamma_{(x,w)} \otimes T\#\gamma$ with respect to the measurable map T, in which $\gamma_{(x,w)}$ is concentrated on the set of $(x'',v_1,\cdots,v_n)$ such that $x'' = x$ and $\pi_v(T(x,v_1,\cdots,v_n)) = w$. Let $\alpha = \alpha(dx,dv,dv_1,\cdots,dv_n)$ be given by

$$\int_{\mathrm{T}^{n+1}\mathbb{R}^d} \varphi d\alpha = \int_{(x,w)\in\mathrm{T}\mathbb{R}^d} \int_{\substack{(x',u,w')\in\mathrm{T}^2\mathbb{R}^d \\ (x'',v_1,\cdots,v_n)\in\mathrm{T}^n\mathbb{R}^d}} \varphi(x,v,v_1,\cdots,v_n) d[\beta_{(x,w)} \otimes \gamma_{(x,w)}](x',v,w',x'',v_1,\cdots,v_n) d[T\#\gamma]$$

for all $\varphi \in C_b(\mathrm{T}^{n+1}\mathbb{R}^d;\mathbb{R})$. Then, for all $\varphi \in C_b(\mathrm{T}^3\mathbb{R}^d;\mathbb{R})$ and $\psi \in C_b(\mathrm{T}^n\mathbb{R}^d;\mathbb{R})$,

$$\int_{\mathrm{T}^{n+1}\mathbb{R}^d} \varphi(x,v,\pi_v(T(v_1,\cdots,v_n)))d\alpha = \int_{(x,w)} \int_{(x',v,w')} \varphi(x,v,w) d\beta_{(x,w)}(x',v,w') d[T\#\gamma] = \int_{\mathrm{T}^2\mathbb{R}^d} \varphi d\beta,$$
$$\int_{\mathrm{T}^{n+1}\mathbb{R}^d} \psi(x,v_1,\cdots,v_n)d\alpha = \int_{(x,w)} \int_{(x'',v_1,\cdots,v_n)} \psi(x,v_1,\cdots,v_n) d\gamma_{(x,w)}(x'',v_1,\cdots,v_n) d[T\#\gamma] = \int_{\mathrm{T}^n\mathbb{R}^d} \psi d\gamma.$$

In pushforward notations, $(\pi_x,\pi_v,\pi_v(T(\pi_{v_1},\cdots,\pi_{v_n}))\#\alpha = \beta$ and $(\pi_x,\pi_{v_1},\cdots,\pi_{v_n})\#\alpha = \gamma$. □

In the course of the manuscript, we will need to measure a similarity between measure fields that belong to distinct tangent spaces. This can be done by computing the Wasserstein distance over the tangent bundle $\mathrm{T}\mathbb{R}^d$, with distance $(x,v),(y,w) \mapsto \sqrt{|x-y|^2 + |v-w|^2}$. However, this distance can be too weak in applications. Following Piccoli [Pic19], we consider the following extension of W_μ to couples of measures $\mu,\nu \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)$.

Definition 1.1.39 (Extension of W_μ). *Define $W_{\mu,\nu} : \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \times \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\nu \to \mathbb{R}^+$ by*

$$W^2_{\mu,\nu}(\xi,\zeta) := \inf\left\{ \int_{(x,v),(y,w)\in\mathrm{T}\mathbb{R}^d} |v-w|^2\, d\alpha \;\middle|\; \alpha \in \Gamma(\xi,\zeta) \text{ and } (\pi_x,\pi_y)\#\alpha \in \Gamma_o(\mu,\nu) \right\}. \tag{1.15}$$

The application $W_{\mu,\nu}(\cdot,\cdot)$ does not induce a distance, but naturally appears in estimates. The infimum is attained by classical arguments. If $\mu = \nu$, it reduces to the cone distance W_μ. We refer the reader to [Pic19] for further details and comments on (1.15); let us point that $W_{\mu,\nu}$ is linked to definitions considered in [CSS23a; CSS23b] to formulate assumptions on dissipative measure fields. As opposed to W_μ, the definition in (1.15) is quite coarse, since it simply translates the mass of ξ on that of ζ without following the deformation of $\mathscr{P}_2(\mathbb{R}^d)$. A more geometric attempt has been proposed in [Gig08, Chapter 6, Section 6], resulting in a “distance” with possibly infinite values.

Differential of the squared distance. We define the metric scalar product directly on $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. Note that its restriction to the geometric tangent cone satisfies Definition 1.1.4.

Definition 1.1.40 (Metric scalar product). *To any $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, associate the maps $\langle\cdot,\cdot\rangle_\mu^\pm : \left(\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu\right)^2 \to \mathbb{R}$ given by*

$$\langle \xi,\zeta\rangle_\mu^+ := \frac{\|\xi\|_\mu^2 + \|\zeta\|_\mu^2 - W_\mu^2(\xi,\zeta)}{2} = \sup_{\alpha\in\Gamma_\mu(\xi,\zeta)} \int_{(x,v,w)\in \mathrm{T}^2\mathbb{R}^d} \langle v,w\rangle \, d\alpha(x,v,w),$$

$$\langle \xi,\zeta\rangle_\mu^- := -\langle -\xi,\zeta\rangle_\mu^+ = -\langle \xi,-\zeta\rangle_\mu^+ = \inf_{\alpha\in\Gamma_\mu(\xi,\zeta)} \int_{(x,v,w)\in \mathrm{T}^2\mathbb{R}^d} \langle v,w\rangle \, d\alpha(x,v,w).$$

The application $\langle\cdot,\cdot\rangle_\mu^-$ bears less intuition than the metric scalar product $\langle\cdot,\cdot\rangle_\mu^+$, but appears at least as frequently in the computations. Both applications coincide whenever ξ or ζ is induced by an application.

Theorem 1.1.41 (Directional derivative [Gig08, Proposition 4.10]). *For any $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$,*

$$D_\mu d_{\mathcal{W}}^2(\mu,\nu)(\xi) := \lim_{h\searrow 0} \frac{d_{\mathcal{W}}^2(\exp_\mu(h\cdot\xi),\nu) - d_{\mathcal{W}}^2(\mu,\nu)}{h} = \inf_{\eta\in\exp_\mu^{-1}(\nu)} \langle -2\cdot\eta,\xi\rangle_\mu^- = \inf_{\eta\in\exp_\mu^{-1}(\nu)} -2\langle \eta,\xi\rangle_\mu^+. \tag{1.16}$$

The following estimate justifies our interest for the map $W_{\mu,\nu}(\cdot,\cdot)$. We could not find a reference for it.

Lemma 1.1.42 (First-order estimate). *For all $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$, $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ and $\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\nu$, there holds*

$$D_\mu d_{\mathcal{W}}^2(\cdot,\nu)(\xi) + D_\nu d_{\mathcal{W}}^2(\mu,\cdot)(\zeta) \leqslant 2 d_{\mathcal{W}}(\mu,\nu) W_{\mu,\nu}(\xi,\zeta). \tag{1.17}$$

Proof. Let $\alpha = \alpha(dx,dv,dy,dw) \in \Gamma(\xi,\zeta)$ realize the infimum in $W_{\mu,\nu}(\xi,\zeta)$. Since $(\pi_x,\pi_y)\#\alpha$ is optimal between μ and ν,

$$d_{\mathcal{W}}(\mu,\nu)W_{\mu,\nu}(\xi,\zeta) = \sqrt{\int |x-y|^2 d(\pi_x,\pi_y)\#\alpha \int |v-w|^2\, d\alpha} \geqslant \int |x-y||v-w| d\alpha \geqslant -\int \langle y-x, v-w\rangle \, d\alpha.$$

Using that the metric scalar product writes as a supremum over transport plans and the formula (1.16),

$$\int_{(x,v),(y,w)} \langle y-x, v\rangle \, d\alpha \leqslant \langle (\pi_x,\pi_y-\pi_x)\#\alpha,\xi\rangle_\mu \leqslant \sup_{\eta\in\exp_\mu^{-1}(\nu)} \langle \eta,\xi\rangle_\mu = -\frac{1}{2} D_\mu d_{\mathcal{W}}^2(\cdot,\nu)(\xi).$$

The same reasoning yields $\int_{(x,v),(y,w)} \langle x-y, w\rangle \, d\alpha \leqslant -\frac{1}{2} D_\nu d_{\mathcal{W}}^2(\mu,\cdot)(\zeta)$, and we conclude. □

1.1.4.5 One-dimensional specificities

In dimension one, the optimal transport plan between μ and ν is always unique for the strictly convex cost $(x,y) \to |x-y|^2$, and can be explicitly computed by distributing the mass of μ, gathered from $-\infty$ to ∞, on the mass of ν, spread from $-\infty$ to ∞. More precisely, let $F_\mu : \mathbb{R} \to [0,1]$ be the distribution function of $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, defined as $F_\mu(x) := \mu((-\infty,x])$. Define its pseudo-inverse $F_\mu^{[-1]} : [0,1] \to \mathbb{R}$ by the implicit requirement that

$$F_\mu(y) \geqslant r \quad \Leftrightarrow \quad y \geqslant F_\mu^{[-1]}(r) \qquad \forall y\in\mathbb{R} \text{ and } r\in[0,1], \qquad \text{resulting in} \qquad F_\mu^{[-1]}(r) := \inf\left\{y\in\mathbb{R} \;\middle|\; F_\mu(y)\geqslant r\right\}.$$

In words, $F_\mu^{[-1]}(r)$ is the first $y \in \mathbb{R}$ such that the set $(-\infty,y]$ is of μ−mass at least r. The plan $(F_\mu^{[-1]}, F_\nu^{[-1]})\#\mathcal{L}_{[0,1]}$ is the unique optimal transport plan between μ and ν [San15, Theorem 2.9]. Using this representation, one can show the following simple separation result of classical aspect.

Lemma 1.1.43 (Optimality within cells). *Let $\mu,\nu \in \mathscr{P}_2(\mathbb{R})$, and consider a countable family $(A_i)_{i\in I} \subset \mathbb{R}$ of closed intervals covering* $\operatorname{supp}\mu$ *and* $\operatorname{supp}\nu$*, and such that* $\max A_i < \min A_j$ *for all $i \leqslant j \in I$. Then*

$$d_{\mathcal{W}}^2(\mu,\nu) = \sum_{i\in I} \mu(A_i)\, d_{\mathcal{W}}^2\left(\mu|_{A_i}, \nu|_{A_i}\right). \tag{1.18}$$

Proof. Let $i \in \mathbb{N}$, and denote $a := \max A_i$ and $b := \min A_{i+1}$, with $a < b$. By assumption, $r := F_\mu(a) = F_\nu(a)$. Since $\mu(]a, b-\varepsilon[) = \nu(]a, b-\varepsilon[) = 0$ for all $\varepsilon > 0$ such that $a < b-\varepsilon$, there also holds $r = F_\mu(b-\varepsilon) = F_\nu(b-\varepsilon)$. Then $F_\mu^{[-1]}(r) \leqslant a$ and $F_\nu^{[-1]}(r) \leqslant a$. On the other hand, for any $r' > r$, one must have $F_\mu^{[-1]}(r') \geqslant b$ and $F^{[-1]}(r') \geqslant b$. In addition, both $F_\mu^{[-1]}$ and $F_\nu^{[-1]}$ are nondecreasing, so the restricted plan $(F_\mu^{[-1]}, F_\nu^{[-1]})\#\mathcal{L}|_{[0,r]}$ sends $\mu|_{(-\infty,a]}$ to $\nu|_{(-\infty,a]}$, while still being optimal by restriction of optimality [Vil09, Theorem 4.6]. By the same argument, $(F_\mu^{[-1]}, F_\nu^{[-1]})\#\mathcal{L}|_{]r,1]}$ is optimal between $\mu|_{[b,\infty)}$ and $\nu|_{[b,\infty)}$. So

$$\begin{aligned}
d_{\mathcal{W}}^2(\mu,\nu) &= \int_{s\in[0,r]} |F_\mu^{[-1]}(s) - F_\nu^{[-1]}(s)|^2 ds + \int_{s\in]r,1]} |F_\mu^{[-1]}(s) - F_\nu^{[-1]}(s)|^2 ds \\
&= \mu((-\infty,a])\, d_{\mathcal{W}}^2\left(\mu|_{(-\infty,a]}, \nu|_{(-\infty,a]}\right) + \mu([b,\infty))\, d_{\mathcal{W}}^2\left(\mu|_{[b,\infty)}, \nu|_{[b,\infty)}\right).
\end{aligned}$$

This proves the result for one interval. Since $\mu((-\infty,a])\mu|_{(-\infty,a)}((-\infty,a')) = \mu((-\infty,a])\frac{\mu((-\infty,a'])}{\mu((-\infty,a])} = \mu((-\infty,a'])$ for $a' < a$, we may proceed by induction (masses compensate in the expected way) and obtain (1.18). □

One could also prove (1.18) directly by constructing optimal plans between $\mu|_{A_i}$ and $\nu|_{A_i}$, then summing them with the correct masses, and use the order on $\mathbb{R}$ to show that the resulting plan is monotone. A higher-dimensional equivalent of this result could be formulated with Kantorovich potentials, with the condition of equality of masses becoming a condition on the c-transform of the concatenation of potentials being the concatenation of c-transforms.

To conclude, let us mention some geometric points for the pleasure of the eye. $\mathscr{P}_2(\mathbb{R}^d)$ contains an isometric copy of $\mathbb{R}^d$ given by Dirac masses; Takatsu and Yokota [TY12] showed that in addition, Dirac masses are the only measures that behave like apexes of cones, in that the Wasserstein distance satisfies

$$d_{\mathcal{W}}^2(\mu,\nu) = d_{\mathcal{W}}^2(\mu,\delta_x) + d_{\mathcal{W}}^2(\nu,\delta_x) - 2d_{\mathcal{W}}(\mu,\delta_x)d_{\mathcal{W}}(\nu,\delta_x)\cos\left(\delta_x\angle_\nu^\mu\right) \qquad \forall \mu,\nu \in \mathscr{P}_2(\mathbb{R}^d) \text{ and } x \in \mathbb{R}^d.$$

The rest of $\mathscr{P}_2(\mathbb{R}^d)$ is not that nice; the geodesic subset of centred Gaussian measures is “stratified” in $d+1$ topological manifolds, depending on the rank of the covariance matrix [Tak11]. The stratification is also alluded to in [GKP11], with a hand-wavy identification of the regular tangent cone as the tangent space *to the strata*. Analogies with finite-dimensional geometry reaches some limits in $\mathscr{P}_2(\mathbb{R}^d)$; for instance, the interior points of a geodesic all share isometric tangent cones in a finite-dimensional CBB space [BGP92, p. 7.16], but Juillet [Jui11] constructs a geodesic in $\mathscr{P}_2(\mathbb{R}^2)$ in which all interior points but one have a tangent cone isometric to a Hilbert space. In summary, despite being often compared to an infinite-dimensional manifold – and for good reasons, $\mathscr{P}_2(\mathbb{R}^d)$ is really rough [ANN18].

1.2 Viscosity solutions

“Et tout cela, parce qu’on veut une solution. *Oh! orgueil humain. Une solution! [...] Et à mesure que nous irons, elle se reculera indéfiniment, parce que notre horizon s’élargira. Plus les télescopes seront parfaits et plus les étoiles seront nombreuses. Nous sommes condamnés à rouler dans les ténèbres et dans les larmes.”*

Gustave Flaubert, Lettre à Mme de Chantepie, 1857.

In this short section, we gather some intuition about the notion of solution to a PDE used in Chapters 2 and 3, namely the viscosity solutions introduced by Crandall and Lions [CL83]. For a first introduction, we refer the reader to [CIL92; Bar94]. The aim of viscosity solutions is to solve PDEs of the form

$$H(x, u(x), Du(x), D^2u(x), \cdots) = 0, \qquad x \in \overline{\Omega}, \tag{1.19}$$

where the unknown is a function $u : \overline{\Omega} \to \mathbb{R}$. PDEs are concerned with flow of information, and the domain $\overline{\Omega}$ is split in two parts: one on which *a priori* information must be given, usually the boundary $\partial\Omega$, and one innervated by characteristics reaching the boundary, usually the interior Ω. Such characteristics can be deterministic or stochastic, and the value of u at a point $x \in \Omega$ can be determined by any mean of aggregation of the information brought from the boundary to x: viscosity solutions are concerned with cases in which this aggregation is monotone, by taking maxima, minima, barycenters, or compositions thereof. In this way, the solution u computed in Ω is itself monotone with respect to the value imposed on the boundary. The purpose of viscosity solutions is to provide an infinitesimal characterization of u.

Monotone Hamiltonians. Denote dom $H \subset \mathcal{C}(\overline{\Omega};\mathbb{R})$ a subset of functions possessing all the classical derivatives entering in H. (The set dom H is central in this thesis, but not in this introduction.) For readability, denote by $(H\varphi)(x)$ the composition $H(x,\varphi(x),D\varphi(x),D^2\varphi(x),\cdots)$. Viscosity solutions apply best when H is monotone in two contravariant ways:

- for any $\varphi,\psi \in$ dom H such that ψ "touches φ from above at x", i.e. $\varphi \leqslant \psi$ in a neighbourhood of x with equality at x, there holds $(H\varphi)(x) \geqslant (H\psi)(x)$. In the sequel, we denote

  $$\varphi \leqslant_x \psi \quad \text{whenever } \varphi(x)=\psi(x) \text{ and } \varphi(y)\leqslant\psi(y) \text{ for all } y \text{ in a neighbourhood of } x.$$

  With this notation, the nondecreasing monotonicity imposes that $\varphi \leqslant_x \psi \implies (H\varphi)(x) \geqslant (H\psi)(x)$.

- There exists a perturbation of $\mathcal{C}(\overline{\Omega};\mathbb{R})$ along which the Hamiltonian H is strictly increasing. Precisely, there exists a one-parameter family $(\Lambda_\alpha)_{\alpha\in\mathbb{R}}$ such that each Λ_α is a bijection from $\mathcal{C}(\overline{\Omega};\mathbb{R})$ to itself, $\sup_{u\in\mathcal{C}(\overline{\Omega};\mathbb{R})}\|u-\Lambda_\alpha u\|_\infty = O(\alpha)$, each Λ_α sends dom H on itself, and the following monotonicities hold for small α:

$$
\begin{array}{lcl}
\forall u \in \mathcal{C}(\overline{\Omega};\mathbb{R}),\quad \forall \alpha\leqslant\alpha', & \Lambda_\alpha u \leqslant \Lambda_{\alpha'} u & \Lambda_\alpha \text{ nondecreases,}\\
\forall u,v \in \mathcal{C}(\overline{\Omega};\mathbb{R}),\ \forall x\in\Omega,\ \forall\alpha, & u\leqslant_x v \iff \Lambda_\alpha u \leqslant_x \Lambda_\alpha v & \text{propagation of } \leqslant_x,\\
\forall \varphi\in \text{dom } H,\ \forall x\in\Omega,\ \forall\alpha, & \begin{cases}\alpha\leqslant 0,\ (H\varphi)(x)\leqslant 0 \implies (H\Lambda_\alpha\varphi)(x)\leqslant\alpha\\ \alpha\geqslant 0,\ (H\varphi)(x)\geqslant 0 \implies (H\Lambda_\alpha\varphi)(x)\geqslant\alpha\end{cases} & H \text{ increases in } \alpha.
\end{array}
$$

As intricate as the second condition may seem, it is still a simplification with respect to the properties of real-world Hamiltonians, in which one could perturb the space variable as well, the family Λ_α could be constructed in a neighbourhood of a given x, differ if one wants to increase or decrease the Hamiltonian, and so on. Note that the Hamiltonian $H\varphi(x) := \langle\nabla\varphi(x), Rx\rangle$ for a R rotation matrix in $\Omega = \mathscr{B}_{\mathbb{R}^2}(0,1)$ does satisfy the first monotonicity condition on dom $H = \mathcal{C}^1$ (since there is no high-order term) but not the second. The characteristics loop, and one cannot add a $\mathcal{C}^1$ perturbation that increases strictly along them. This is fortunate, since any smooth radial function is a solution to $H\varphi = 0$, and there is no contradiction with the uniqueness results of viscosity theory.

Examples. Let us give three examples to illustrate the diversity of situations covered by this definition.

In $\Omega = (-1,1)$, consider $H(\nabla_x u) := |\nabla_x u| - 1$ the Hamiltonian of the Eikonal equation. If φ and ψ are $\mathcal{C}^1(\mathbb{R};\mathbb{R})$ functions such that $\varphi-\psi$ reaches a maximum at $x\in\Omega$, then $\nabla_x\varphi = \nabla_x\psi$, and $H(\nabla_x\varphi) = H(\nabla_x\psi)$: this shows the first monotonicity. On the other hand, for $u\in\mathcal{C}(\Omega;\mathbb{R})$, consider the perturbation $\Lambda_\alpha(u)(x) := (1+\alpha)u(x)$ with $\alpha\in\mathbb{R}$. If $\alpha\leqslant\alpha'$, then $\Lambda_\alpha u \leqslant \Lambda_{\alpha'} u$ as functions. The equivalence $u\leqslant_x v$ iff $\Lambda_\alpha u \leqslant_x \Lambda_\alpha v$ for small α is direct. Finally, for $\varphi\in\mathcal{C}^1$,

$$(H\Lambda_\alpha\varphi)(x) = (1+\alpha)|\nabla_x\varphi| - 1 = (1+\alpha)\big(|\nabla_x\varphi|-1\big)+\alpha = (1+\alpha)(H\varphi)(x)+\alpha,$$

from which one gets the increasing monotonicity for small α.

Consider now $H(u(x),D^2u(x)) = u(x) - \text{Trace}\,(D^2u(x))$ and dom $H = \mathcal{C}^2(\overline{\Omega};\mathbb{R})$. If φ,ψ are two $\mathcal{C}^2$ functions such that $\varphi\leqslant\psi$ around x with $\varphi(x)=\psi(x)$, then $D^2\varphi(x)\leqslant D^2\psi(x)$ in the sense of symmetric matrices, and $(H\varphi)(x)\geqslant(H\psi)(x)$. On the other hand, the perturbation $\Lambda_\alpha(u) := u+\alpha$ is monotone with respect to the order on functions, preserves the "touching conditions", and $(H\Lambda_\alpha\varphi)(x) = (H\varphi)(x)+\alpha$ for any $\varphi\in$ dom H.

Thirdly, consider $H(x,p) = -p_1 + \sup_{b\in B(x)} -\langle b,p_2\rangle$, where $p=(p_1,p_2)$ is a vector with first coordinate $p_1\in\mathbb{R}$ and second coordinate $p_2\in\mathbb{R}^d$, and $B(x)\subset\mathbb{R}^d$ for all $x\in[a,b]\times\mathbb{R}^d$. Then $H(x,\nabla_x u)$ only depends on first-order derivatives, so that the non-decreasing monotonicity is trivial in dom $H=\mathcal{C}^1$. On the other hand, the perturbation $(\Lambda_\alpha u)(x) := u(x) - \alpha(x_1-a)$ is monotone in α, preserves $\leqslant_x$, and $(H\Lambda_\alpha\varphi)(x) = \alpha + (H\varphi)(x)$. This Hamiltonian corresponds to parabolic equations, in which x_1 is a time variable.

In these examples, the expression of the *ad hoc* transformation Λ_α does not really matter. Ideally, one could hope to deduce it from the Hamiltonian in a canonical way.

Viscosity solutions. Modulo details, a viscosity solution of the equation $Hu = 0$ is defined as follows:

- [subsolution] for any $\varphi \in \operatorname{dom} H$ such that $u \leqslant_x \varphi$, there holds $(H\varphi)(x) \leqslant 0$,
- [supersolution] for any $\psi \in \operatorname{dom} H$ such that $\psi \leqslant_x u$, there holds $0 \leqslant (H\psi)(x)$.

Morally, the first monotonicity required on the Hamiltonian is *necessary for existence:* if ψ touches φ from above at x, then by transitivity, it also touches u from above, and should satisfy $(H\psi)(x) \leqslant 0$. In favourable cases, the second monotonicity is *sufficient for uniqueness.* This translates in a comparison principle, usually given by a variation on the following theme.

Algorithm 1: Outline of a comparison principle

1 Assume that v is a subsolution, w a supersolution, and $v(x_0) > w(x_0)$ at some point of $x_0 \in \Omega$. For small $\alpha > 0$, one still has $v(x_0) > \Lambda_\alpha w(x_0)$, and $w^\alpha := \Lambda_\alpha w$ is a strict supersolution − for which $H\Lambda_\alpha w \geqslant \alpha$ in the viscosity sense.

2 Construct maximum points (x^*, y^*) of the doubling function $\Phi : (x, y) \mapsto v(x) - \Lambda_\alpha w(y) - G_\varepsilon(x, y)$, with G_ε a smooth map that vanishes if $x = y$, penalizes $x \neq y$ for small ε, and satisfies

$$H\left[\Lambda_\alpha w(y) + G_\varepsilon(\cdot, y)\right](x) \geqslant H\left[v(x) - G_\varepsilon(x, \cdot)\right](y) + O\big(G_\varepsilon(x, y)\big). \tag{1.20}$$

3 By assumption, $\Phi(x_0, x_0) > 0$ and $\Phi < 0$ on the boundary, so x^*, y^* lie in Ω for small ε. Apply the viscosity inequalities on the test functions φ, ψ such that $\Phi(\cdot, y^*) = v - \varphi$ and $\Phi(x^*, \cdot) = -(w - \psi)$.

4 If H is sufficiently regular, deduce that $H\left[\Lambda_\alpha w(y) + G_\varepsilon(\cdot, y)\right](x) \leqslant 0 < \alpha \leqslant H\left[v(x) - G_\varepsilon(x, \cdot)\right](y)$ locally uniformly around (x_0, x_0), independently of ε. This contradicts (1.20) for small ε.

This program is very flexible, and the literature of viscosity solutions is a fantastic collection of tricks and workarounds. Notably, the maximization step usually requires penalizations, Ekeland-type principles or slight modifications of the definition in order to obtain existence, boundedness, convergence and/or regularity of maximizers. The function G_ε can be taken simple in simple cases: most of this thesis is devoted to the exploration of a notion of viscosity where one *wants* to take $G_\varepsilon(x, y) = d^2(x, y)/\varepsilon$. A universal construction stays to be provided, although [IM17] opened the way. The definition of $\leqslant_x$ can also be modified: one can ask for maximization only in certain directions [CM13], or along paths [ETZ16], or up to an $\varepsilon d(\cdot, x_0)$ with strict viscosity solutions [MQ18], or up to a more intricate penalization [DS24], etc.

The point of view that we presented completely hides a geometric meaning of viscosity solutions, that is best seen when the Hamiltonian H is convex in the first-order variable. In this case, viscosity inequalities can be understood respectively as a nondecreasing condition along all characteristics, and a nonincreasing condition along one of them. The generalization of these conditions is the viability and invariance theory of Aubin and Cellina [AC86], Aubin and Frankowska [AF90], with Nagumo's theorem as the bridge between the PDE side and control side of these equations. We conclude with a mention of the extreme theory of Kolokoltsov and Maslov [KM97], pushing monotonicity as far as defining "scalar products" for the $(\min, +)$ semialgebra, with associated "duality", and retrieve viscosity solutions as the limit of Hopf-Cole transforms − converted into changes of the operations from the linear $(+, \cdot)$ to the idempotent $(\min, +)$.

Chapter 2

Optimal control problems in CAT(0) spaces

The aim of this chapter is to develop the Hamilton-Jacobi-Bellman counterpart of HJ equations presented in [JZ23b]. We first introduce controlled dynamical systems by implementing the results of mutational analysis, and obtain a convenient setting for control problems. We provide sufficient conditions for the existence of an optimal control with a suitable notion of convex hull of the dynamic. Relying on the definition of test functions, we are able to prove that the control Hamiltonian satisfies the assumptions of [JZ23b], and characterize the value function as the unique solution of the associated HJB equation. We conclude with some numerical experiments. The content of this chapter is derived from [AZ25], in collaboration with Hasnaa Zidani.

Contents of Chapter 2

Let (Ω, d) be a complete geodesic proper CAT(0) space, and $(\mathrm{T}_x\Omega, d_x)$ be the tangent cone at $x \in \Omega$. Let again

$[xy]$	$\in \mathrm{AC}([0,1];\Omega)$	the unit-speed geodesic linking x to y	$x, y \in \Omega$
$(1-t)x \oplus ty$	$\in \Omega$	the unique point of $[xy]$ at distance $td(x,y)$ of x	$x, y \in \Omega,\ t \in [0,1]$
$\overrightarrow{xy}$	$\in \mathrm{T}_x\Omega$	the equivalence class of $[xy]$ in $\mathrm{T}_x\Omega$	$x, y \in \Omega$
$\uparrow_x^y$	$\in \mathrm{T}_x\Omega$	the unit direction $d(x,y)^{-1} \cdot \overrightarrow{xy}$	$x, y \in \Omega,\ x \neq y$
$\lambda \cdot v$	$\in \mathrm{T}_x\Omega$	the equivalence class of $\lambda \cdot \gamma$ for $\gamma \in v$	$\lambda \in \mathbb{R}^+,\ v \in \mathrm{T}_x\Omega$

2.1 Ordinary differential equations

Our strategy to define ODEs is the following. In CAT(0) spaces, we can produce a lot of semigroups that are solutions of "simple ODEs", by taking gradient flows of concave energies. For instance, geodesics can be recovered as gradient flows of the energies $-d(\cdot, x)$ for $x \in \Omega$. So, if f is so that $f(x) = \alpha \cdot \uparrow_x^y \in \mathrm{T}_x\Omega$ for some $\alpha \in \mathbb{R}^+$ and $y \in \Omega$ that may depend on x, it makes sense to consider that the solution of the formal ODE $\frac{d}{dt}z = f(z)$ should be locally approximated around x by the gradient flow of $-\alpha d(\cdot, y)$.

The definition of ODEs goes in this direction: a dynamic is defined as an application $f : \Omega \to \mathcal{C}(\Omega; \mathbb{R})$, where each $f(x)$ is the energy of the gradient flow that should approximate the solution passing through x.

For this definition to stand, one needs f to be valued in sufficiently regular functions, and itself sufficiently regular with respect to a suitable distance between energies. Fortunately, this procedure is axiomatized and developed at a formal level by the theory of mutations [Aub99; Lor10]. The original motivation of mutational analysis was the so-called morphological analysis, concerned with motions of sets. However, in the meantime, the authors developed a general theory, best suited for Cauchy-Lipschitz equations, with the aim to factor the common arguments to diverse settings. In this chapter, we rely on the results of [FL23].

2.1.1 Definition

2.1.1.1 Choice of energies and transition semigroups

We first define the set of energies in which the dynamic will take values. Given $E : \Omega \to \mathbb{R}$ a concave Lipschitz function, denote by $|\cdot|_{1,\infty}$ the supremum of the metric slope, i.e.

$$|E|_{1,\infty} := \sup_{(x,v) \in \mathrm{T}\Omega, |v|_x = 1} |D_x E(v)|. \tag{2.1}$$

Since we work with gradient flows, we systematically identify energies $E \sim E'$ such that $\left|E - E'\right|_{1,\infty} = 0$. Additionally, E is said to be Fréchet if

$$D_x E(v) = \langle \nabla_x E, v \rangle_x \qquad \forall x \in \Omega \text{ and } v \in \mathrm{T}_x\Omega.$$

This is a restrictive assumption in general, but not for our purposes. For instance, the squared distance is Fréchet at all point, but the distance itself is not around the base point. Note that the differential of a Fréchet map needs to be geodesically linear, since the metric scalar product is not so.

Definition 2.1.1 (Admissible set of energies). *An admissible set of energies is a subset $\mathcal{E}$ of (equivalence classes of) Lipschitz and concave functions from Ω to $\mathbb{R}$, that is closed and separable with respect to the topology induced by $|\cdot|_{1,\infty}$, and includes the equivalence class 0 of constant functions. It is said to be Fréchet if it contains only Fréchet functions.*

The dynamics of the ODEs will be valued in such a set $\mathcal{E}$. The assumptions of closedness and separability are technical requirements for the theory of mutations. The latter does not consider energies, but semigroups, referred to as transitions. In our context, transitions are defined as follows.

Definition 2.1.2 (Set of transitions). *Let $\mathcal{E}$ be an admissible set of energies in the sense of Definition 2.1.1. By Proposition 1.1.10, each $E \in \mathcal{E}$ admits a unique gradient flow $\vartheta \in C(\mathbb{R}^+ \times \Omega; \Omega)$, denoted $\mathcal{G}_{\mathcal{F}}(E)$. The set of transitions is*

$$\Theta := \{\mathcal{G}_{\mathcal{F}}(E) \mid E \in \mathcal{E}\}.$$

Additionally, we may need to impose the following property of uniform approximation by geodesics.

Assumption [A2.1.3]. For each compacts $A \subset \left(\mathcal{E}, |\cdot|_{1,\infty}\right)$ and $B \subset \Omega$, there exist reparameterized geodesics $(\gamma_{E,x})_{(E,x)\in A\times B}$ defined over a nontrivial common interval $[0, t]$ satisfying

$$\gamma_{E,x}(0) = x \qquad \text{and} \qquad \lim_{h\searrow 0} \sup_{(E,x)\in A\times B} \frac{d\left(\mathcal{G}_{\mathcal{F}}(E)(h,x), \gamma_{E,x}(h)\right)}{h} = 0.$$

Remark 2.1.4 (Characterizations). *Under our assumptions, the gradient flow $\vartheta = \mathcal{G}_{\mathcal{F}}(E)$ is characterized as the unique solution of the Evolutionary Variational Inequality (EVI)*

$$\frac{d}{dt}\frac{d^2(\vartheta(t,x), z)}{2} \leqslant (-E)(z) - (-E)(\vartheta(t,x)) = E(\vartheta(t,x)) - E(z) \qquad \forall z \in \Omega, \quad \textit{for a.e. } t \geqslant 0. \tag{2.2}$$

Indeed, let $x \in \Omega$. By Proposition 1.1.10, the curve $t \mapsto \vartheta(t,x)$ is locally Lipschitz, its right derivative is defined for all $t \geqslant 0$, and equal to $\nabla_{\vartheta(h,x)} E$. The composition $t \mapsto d^2(\vartheta(t,x), z)$ is locally Lipschitz, hence admits a derivative for almost all time $t \in \mathbb{R}^+$ by the Rademacher theorem. The squared distance is 2-convex, hence admits a differential in the sense of Definition 1.1.8 by [AKP23, Lemma 13.15], whose very definition is that

$$D_{\vartheta(t,x)} d^2(\cdot, z)(\vartheta(\cdot,x)^+_t) = \frac{d}{dt^+}\frac{d^2(\vartheta(t,x), z)}{2} = \frac{d}{dt}\frac{d^2(\vartheta(t,x), z)}{2}$$

if the derivative is defined. Using that $\vartheta(\cdot,x)^+_t = \nabla_{\vartheta(t,x)} E$, the Definition 1.1.9 of a metric gradient for the concave function E, and the property of concave functions that their directional derivative along the geodesic $[ab]$ is superior to $E(b) - E(a)$,

$$D_{\vartheta(t,x)} d^2(\cdot, z)(\vartheta(\cdot,x)^+_t) = -2\,\langle \nabla_{\vartheta(t,x)} E, \overrightarrow{\vartheta(t,x)z}\rangle_{\vartheta(t,x)} \leqslant -2 D_{\vartheta(t,x)} E\left(\overrightarrow{\vartheta(t,x)z}\right) \leqslant -2\left(E(z) - E\left(\vartheta(t,x)\right)\right).$$

This is valid for all $z \in \Omega$, for almost all $t \in \mathbb{R}^+$, hence $t \mapsto \vartheta(t,x)$ satisfies the EVI (2.2). The solution of the EVI is known to be unique under our assumptions, and the unique limit of the Crandall-Liggett scheme [AGS05, Theorem 4.0.4]. Hence both notions coincide.

Transitions are naturally endowed with the distance $\mathbb{D}$ defined by

$$\mathbb{D}(\vartheta, \vartheta') := \sup_{x\in\Omega} \limsup_{h\searrow 0} \frac{d(\vartheta(h,x), \vartheta'(h,x))}{h} \qquad \forall \vartheta, \vartheta' \in \Theta. \tag{2.3}$$

On the other hand, energies can be embedded in a Banach space as follows. Let $\mathrm{DC}_{\mathrm{Lip}}$ be the set of Lipschitz functions that write as the difference of concave functions. The space

$$\mathbb{E} := \overline{\mathrm{DC}_{\mathrm{Lip}} / \sim}^{|\cdot|_{1,\infty}}$$

is a vector space, since the operations of addition and scalar multiplication are preserved in the quotient. It is also a subset of Lipschitz functions, as a Cauchy sequence $(E_n)_n$ of Lipschitz functions with respect to $|\cdot|_{1,\infty}$ has a pointwise limit that is also Lipschitz, with constant $\lim_n |E_n|_{1,\infty}$, and it is closed with respect to the norm $|\cdot|_{1,\infty}$ by definition. Hence, it forms a Banach space that contains $\mathcal{E}$ as a closed subset, enabling us to apply results from the theory of Banach-valued L^p spaces.

Lemma 2.1.5. *Let $\mathcal{E}$ be admissible according to Definition 2.1.1. Then $\mathcal{G}_{\mathcal{F}} : \mathcal{E} \to \Theta$ is injective, and 1-Lipschitz as a consequence of the stronger inequality*

$$d_x(\nabla_x E, \nabla_x E') \leqslant \left|E - E'\right|_{1,\infty} \qquad \forall E, E' \in \mathcal{E} \quad \textit{and } x \in \Omega. \tag{2.4}$$

It follows that the restriction of $\mathcal{G}_{\mathcal{F}}$ to any compact subset of $\mathcal{E}$ has a continuous inverse on its range. Moreover, if $\mathcal{E}$ is made of Fréchet functions, then $\mathcal{G}_{\mathcal{F}}$ is an isometry.

Proof. We first show that $\mathcal{G}_\mathcal{F}$ is injective on its image. Let $\vartheta \in \Theta = \mathcal{G}_\mathcal{F}(\mathcal{E})$. Consider first the case where there exists at least one $x \in \Omega$ such that $\vartheta(h,x) = x$ for all $h \geqslant 0$. Equivalently, any $E \in {\mathcal{G}_\mathcal{F}}^{-1}(\vartheta)$ admits a maximum at x, and there holds $\nabla_x E = 0_x$. If such a point exist, then, as ϑ is the gradient flow of a concave function, x belongs to the unique connected component of maximum points. Pick any $E \in {\mathcal{G}_\mathcal{F}}^{-1}(\vartheta)$. For any $y \in \Omega$, the curve $s \mapsto E(\vartheta(s,y))$ is locally Lipschitz and there holds

$$E(\vartheta(t,y)) - E(y) = \int_{s=0}^t \frac{d}{ds} E(\vartheta(\cdot,y))ds = \int_{s=0}^t D_{\vartheta(s,y)} E(\vartheta(\cdot,y)_s^+) ds = \int_{s=0}^t \left|\vartheta(\cdot,y)_s^+\right|^2_{\vartheta(t,y)} ds.$$

Hence $t \mapsto E(\vartheta(t,y))$ increases as long as $\vartheta(t,y)$ does not belong to the set of maximum points, on which E is constant, and may be set to 0. Letting t go to ∞, we obtain an expression of $E(y)$ that depends only on ϑ, so that ${\mathcal{G}_\mathcal{F}}^{-1}(\vartheta)$ reduces to a singleton.

Assume now that ϑ does not admit equilibrium points. We would like to reduce to the first case by adding a sufficiently decreasing function to the candidate energies. As the elements of $\mathcal{E}$ are Lipschitz, the concave and locally Lipschitz function $E - d^2(\cdot,o)$ admits maximum points on the locally compact space Ω, and generates a unique gradient flow. However, one has to show that the said gradient flow is independent of the choice of $E \in {\mathcal{G}_\mathcal{F}}^{-1}(\vartheta)$: this is provided by the Trotter-Kato scheme for CAT(0) spaces [Sto12, Theorem 4.5]. As per this result, the gradient flow (in the sense of the Crandall-Liggett scheme, which is equivalent to our definition by Remark 2.1.4) of the sum $E - d^2(\cdot,o)$ is the uniform limit, when h goes to 0, of the curve that follows alternatively ϑ and the flow of $-d^2(\cdot,o)$ for a duration of h. Therefore, it does not depend on the choice of E. As in the first step, $E - d^2(\cdot,o)$ can be computed by integration, hence $E \in {\mathcal{G}_\mathcal{F}}^{-1}(\vartheta)$ is uniquely defined.

We turn to the 1-Lipschitz estimate. One always has $\mathbb{D}(\mathcal{G}_\mathcal{F}(E), \mathcal{G}_\mathcal{F}(E')) = \sup_{x\in\Omega} d_x\left(\nabla_x E, \nabla_x E'\right)$ for any $E, E' \in \mathcal{E}$ by definition of gradient flows and the derivative of curves. From the definition of metric gradient, $D_x E(\nabla_x E') \leqslant \langle \nabla_x E, \nabla_x E'\rangle_x$ and $D_x E'(\nabla_x E) \leqslant \langle \nabla_x E', \nabla_x E\rangle_x$. Hence, by the definition of the metric scalar product ($2\langle v,w\rangle_x = |v|_x^2 + |w|_x^2 - d_x^2(v,w)$) and the fact that $d_x E$ is Lipschitz in $(\mathrm{T}_x\Omega, d_x)$ with constant $|E|_{1,\infty}$,

$$\begin{aligned} d_x^2\left(\nabla_x E, \nabla_x E'\right) &= |\nabla_x E|_x^2 + \left|\nabla_x E'\right|_x^2 - 2\langle \nabla_x E, \nabla_x E'\rangle_x \leqslant D_x E(\nabla_x E) + D_x E'(\nabla_x E') - D_x E(\nabla_x E') - D_x E'(\nabla_x E) \\ &= D_x\left(E' - E\right)(\nabla_x E) - D_x\left(E'-E\right)(\nabla_x E') \leqslant \left|E' - E\right|_{1,\infty} d_x\left(\nabla_x E, \nabla_x E'\right). \end{aligned}$$

This implies $d_x\left(\nabla_x E, \nabla_x E'\right) \leqslant \left|E'-E\right|_{1,\infty}$ for any $x \in \Omega$, and taking the supremum over $x \in \Omega$, the desired estimate. Now, if the set $\mathcal{E}$ is made of Fréchet functions, then for any $(x,v) \in \mathrm{T}\Omega$ with $|v|_x = 1$,

$$\left|D_x E(v) - D_x E'(v)\right| = \left|\langle \nabla_x E, v\rangle - \langle \nabla_x E', v\rangle\right| \leqslant 1 \times d_x(\nabla_x E, \nabla_x E'),$$

and the sets $\left(\mathcal{E}, |\cdot|_{1,\infty}\right)$ and $(\Theta, \mathbb{D})$ are isometric. □

2.1.1.2 Well-posedness results of mutational analysis

We now fix a set of energies $\mathcal{E}$ satisfying Definition 2.1.1, and its corresponding set of transitions $\Theta := \mathcal{G}_\mathcal{F}(\mathcal{E})$. Here, $\mathcal{E}$ is not assumed to be Fréchet. The *mutation* $\mathring{y}_t$ of a curve $y : [0,T] \to \Omega$ at time $t \in [0,T)$ is defined as

$$\mathring{y}_t := \left\{ \vartheta \in \Theta \;\middle|\; \limsup_{h \searrow 0} \frac{d\left(\vartheta(h,y_t), y_{t+h}\right)}{h} = 0 \right\}.$$

The mutations replace derivatives, and conversely, one can "integrate" a curve of elements of Θ. In the theory developed by Aubin and Frankowska, the transitions are the main object of study. In this manuscript, we mainly work with energies instead of transitions, and we state the assumptions and results directly at this level. Let us first gather [FL23, Corollary 2.21 and Proposition 2.22]. In the sequel, $\mathcal{E}$ is an admissible set of energies in the sense of Definition 2.1.1.

Proposition 2.1.6 (Existence and uniqueness of mutational primitives). *Let $(E_s)_{s\in[0,T]} \in L^1(0,T;\mathcal{E})$. Then for any $x \in \Omega$, there exists a unique curve $y \in AC([0,T];\Omega)$ that is mutable almost everywhere and such that $\mathcal{G}_\mathcal{F}(E_s) \in \mathring{y}_s$ for a.e. $s \in [0,T]$.*

Explicitly, denoting $\vartheta := \mathcal{G}_{\mathcal{F}}(E_s)$ the gradient flow semigroup of the concave function E_s, there holds

$$\limsup_{h\searrow 0} \frac{d\left(\vartheta(h,y_s), y_{s+h}\right)}{h} = 0 \qquad \text{for a.e. } s \in [0,T].$$

For the application to controlled systems that we have in mind, the correct setting is that of mutational inclusions.

Definition 2.1.7 (Solution in the sense of mutations). *Let $F : \Omega \rightrightarrows \mathcal{E}$ be a multivalued map in the set of energies. A curve $(y_s)_{s\in[0,T]} \subset \Omega$ is a solution of the mutational inclusion*

$$\mathring{y}_s \cap \mathcal{G}_{\mathcal{F}}(F(y_s)) \neq \emptyset \qquad s \in [0,T). \tag{2.5}$$

if it is absolutely continuous, and if $\mathcal{G}_{\mathcal{F}}(F(y_s)) \cap \mathring{y}_s \neq \emptyset$ for almost any $s \in [0,T]$.

In $\Omega = \mathbb{R}^d$, the natural energies are the linear functions $x \mapsto < x, v >$, whose gradient flows are the translations $\vartheta(h,x) = x + hv$. Then $\mathcal{E}$ identifies with a set of vectors, and Definition 2.1.7 reduces to the classical notion of differential inclusion.

Assumption [A2.1.8] (Structure of the dynamic). The dynamic $F : \Omega \rightrightarrows \mathcal{E}$ is assumed

(a) nonempty- and compact-valued in $(\mathcal{E}, |\cdot|_{1,\infty})$,

(b) locally Lipschitz in the Hausdorff distance, i.e. for any closed and bounded set $B \subset \Omega$, there exists a constant $\mathrm{Lip}(F)_B \geqslant 0$ such that for all $x, x' \in B$ and $E \in F(x)$, there exists $E' \in F(x')$ with $\left|E - E'\right|_{1,\infty} \leqslant \mathrm{Lip}(F)_B d(x,x')$,

(c) with sublinear growth, i.e. there exists $C_f \geqslant 0$ and a point $o \in \Omega$ such that $|F(x)|_{1,\infty} \leqslant C_f(1 + d(x,o))$ for all $x \in \Omega$.

Under the assumption **[A2.1.8]**, we may apply the results of [FL23] to recover the following.

Proposition 2.1.9 (Well-posedness of the mutational inclusion). *Assume that F satisfies **[A2.1.8]**. Then for any $x \in \Omega$, the inclusion (2.5) admits solutions issued from x in the sense of Definition 2.1.7. If F is single-valued, the solution is unique. In addition, there exists a measurable selection $(E_s)_{s\in[0,T]} \in L^1(0,T;\mathbb{E})$ such that $\mathcal{G}_{\mathcal{F}}(E_s) \in \mathring{y}_s \cap \mathcal{G}_{\mathcal{F}}(F(y_s))$ for a.e. $s \in [0,T]$.*

Proof. This result is obtained as a consequence of the abstract setting in [FL23]. Indeed, our construction satisfies the requirements of the notion of mutation introduced in this paper, as we now detail.

First, the set $\Theta := \{\mathcal{G}_{\mathcal{F}}(E) \mid E \in \mathcal{E}\}$ is a set of transitions in the sense of [FL23, Definition 2.1]. Indeed, every $\vartheta \in \Theta$ satisfies the semigroup property and $\vartheta(0,\cdot) = id$. Furthermore, ϑ is contractive as the gradient flow of a convex function (see [AKP23, Proposition 16.20]). Besides, ϑ is Lipschitz-continuous with constant $\mathbb{D}(0_\Theta, \vartheta)$, and has sublinear growth at the initial time, i.e. $\sup_{x\in\Omega} d(\vartheta(t,x),x) \leqslant C(1+d(x,o))t$ for some constant $C > 0$ and all $t \in [0,1]$.

By definition of an admissible set of energies $\mathcal{E}$, the set Θ is complete and separable with respect to the topology induced by $(\vartheta,\vartheta') \mapsto \left|\mathcal{G}_{\mathcal{F}}{}^{-1}(\vartheta) - \mathcal{G}_{\mathcal{F}}{}^{-1}(\vartheta')\right|_{1,\infty}$.

If we additionally assume that F, hence f, is globally bounded, then, with the properties outlined above, we can directly apply [FL23, Theorem 3.4] to conclude that the existence and uniqueness stated in the proposition hold. In the general case, one can first consider a truncated version f_R of f, which coincides with f on a sufficiently large ball B, and takes the value of f at the projection on the said ball outside of it. Balls are convex in CAT(0) spaces, and this projection is uniquely defined. By [Lor10, Proposition 2.22], it follows that in finite time $T > 0$, the solutions starting from a given initial point x remain uniformly bounded, and the truncation has no effect on the solutions starting sufficiently within the ball.

The existence of a measurable selection of transitions $\vartheta_s \in \mathring{y}_s \cap \mathcal{G}_{\mathcal{F}}(F(y_s))$ follows from [FL23, Proposition 3.2]. As pointed before, the solutions of the mutational ODE stay uniformly bounded in finite time, so that up to a redefinition on a negligible set, $\vartheta_s \in \mathcal{G}_{\mathcal{F}}(\bigcup_{z\in B} F(z))$ for some compact $B \subset \Omega$. Under our assumptions, the set $\bigcup_{z\in B} F(z)$ is compact, since it writes as the union over a compact set of the compact images of an usc multifunction [Ber59, Theorem 3 p.116]. From Lemma 2.1.5, the restriction of $\mathcal{G}_{\mathcal{F}}$ to this compact has a continuous inverse, and $s \mapsto E_s := \mathcal{G}_{\mathcal{F}}{}^{-1}(\vartheta_s)$ is measurable. The L^1 integrability follows from the growth assumption on the dynamic. □

2.1.1.3 An example of an admissible set of energies

In the Euclidean case, it is natural and convenient to choose linear maps as transitions, as detailed in [Aub99]. In the general case of a complete proper CAT(0) space, one may consider the following.

Definition 2.1.10 (Set of energies). *Let* $\kappa : \mathbb{R}^+ \to \mathbb{R}^+$ *be given by* $r^2/2$ *if* $r \in [0,1]$*, and* $r - 1/2$ *if* $r > 1$*. Define*

$$\mathcal{E} := \left\{ x \mapsto -\alpha\kappa(d(x,x_0)) \;\middle|\; \alpha \in \mathbb{R}^+ \text{ and } x_0 \in \Omega \right\}. \tag{2.6}$$

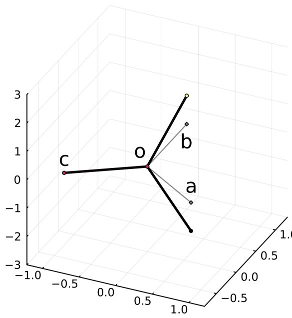
The composition with κ forces the norm of the gradient of E to be continuous, while not breaking the Lipschitz character. In the case where Ω is compact, one could directly consider $\kappa : r \mapsto r^2/2$. Let us verify that $\mathcal{E}$ satisfies Definition 2.1.1; the distance function is convex in CAT(0) spaces [BH99, Proposition 2.2], and κ is convex and nondecreasing, so that the elements of $\mathcal{E}$ are concave and Lipschitz-continuous. The differential can be explicitly computed as

$$D_x E(v) = \langle \alpha\kappa'(d(x,x_0)) \uparrow_x^{x_0}, v\rangle_x \qquad \forall\, v \in \mathrm{T}_x\Omega,$$

with by convention $\kappa'(0) \uparrow_{x_0}^{x_0} = 0_{x_0}$. Thus E is Fréchet.

The rest of this section is devoted to the verification of technical assumptions. Before entering the details, let us point that an extension of convex analysis to CAT(0) spaces is proposed in [BN08; AKA10], based on the following pseudo-scalar product:

$$x \mapsto \ll \overrightarrow{ab}, \overrightarrow{ox} \gg := \frac{1}{2}\left[d^2(a,x) + d^2(b,o) - d^2(a,o) - d^2(b,x)\right]$$



for a fixed reference point $o \in \Omega$ and any pair $(a,b) \in \Omega^2$. Scalar products are the natural energies in $\mathbb{R}^d$, and one might hope to use these maps in the same way in CAT(0) spaces. However, $x \mapsto \ll \overrightarrow{ab}, \overrightarrow{ox} \gg$ is not semiconvex nor semiconcave in general, and the gradient flow is not uniquely defined. The lack of semiconvexity and semiconcavity can be seen on a three-legged network, taking for a and b the endpoints of different branches glued at the junction o.

The parametrization by $\mathbb{R}^+ \times \Omega$ implies the following.

Lemma 2.1.11 (Topological properties). *The set* $\mathcal{E}$ *defined in (2.6) is complete and separable in* $\left(\mathbb{E}, |\cdot|_{1,\infty}\right)$*.*

Proof of Lemma 2.1.11. We begin by separability. Let $\mathcal{A} \subset \mathbb{R}^+, \mathcal{O} \subset \Omega$ be a countable dense subsets. Consider $E := -\alpha\kappa \circ d(\cdot, x_0)$ and $\overline{E} := -\overline{\alpha}\kappa \circ d(\cdot, \overline{x}_0)$. For any $(x,p) \in \mathrm{T}_x\Omega$ such that $|p|_x = 1$, one has

$$\begin{aligned}
\left|d_x E(p) - d_x\overline{E}(p)\right| &= \left|\alpha\kappa'(d(x,x_0))\,\langle\uparrow_x^{x_0}, p\rangle_x - \overline{\alpha}\kappa'(d(x,\overline{x}_0))\,\langle\uparrow_x^{\overline{x}_0}, p\rangle_x\right| \\
&\leqslant \left|\alpha - \overline{\alpha}\right| \left|\kappa'(d(x,x_0))\,\langle\uparrow_x^{x_0}, p\rangle_x\right| + \overline{\alpha}\left|\kappa'(d(x,x_0)) - \kappa'(d(x,\overline{x}_0))\right| \left|\langle\uparrow_x^{x_0}, p\rangle_x\right| \\
&\quad + \overline{\alpha}\kappa'(d(x,\overline{x}_0)) \left|\langle\uparrow_x^{x_0}, p\rangle_x - \langle\uparrow_x^{\overline{x}_0}, p\rangle_x\right| \\
&\leqslant \left|\alpha - \overline{\alpha}\right| + \overline{\alpha} d(x_0,\overline{x}_0) + \overline{\alpha}\left|\langle\uparrow_x^{x_0}, p\rangle_x - \langle\uparrow_x^{\overline{x}_0}, p\rangle_x\right|.
\end{aligned} \tag{2.7}$$

Here we used that $\kappa'(r) = \max(0, \min(1,r))$ is nonnegative, bounded by 1 and 1-Lipschitz over $\mathbb{R}^+$. By (1.9), $|\langle\uparrow_x^{x_0}, p\rangle_x - \langle\uparrow_x^{\overline{x}_0}, p\rangle_x| \leqslant d_x(\uparrow_x^{x_0}, \uparrow_x^{\overline{x}_0}) \leqslant d(x_0, \overline{x}_0)$. If now E is fixed, we may choose $\overline{\alpha} \in \mathcal{A}$ and then $\overline{x}_0 \in \mathcal{O}$ rendering (2.7) arbitrarily small regardless of (x,p), showing that $\mathcal{E}$ is separable in $(\mathcal{E}, |\cdot|_{1,\infty})$.

Let us turn to completeness. Let $(\overline{E}^n)_n$ be a Cauchy sequence in $(\mathcal{E}, |\cdot|_{1,\infty})$. Denote $E^n = -\alpha^n\kappa \circ d(\cdot, x_0^n) - c_n$ the member of the equivalence class of $\overline{E}^n$ that vanishes in $o \in \Omega$. In particular, $\alpha^n = |E^n|_{1,\infty} = \mathrm{Lip}(E^n)$ converges towards a limit $\alpha \in \mathbb{R}^+$. If $\alpha = 0$, then $E^n \to_n 0_{\mathbb{E}}$. Assume by now that $\alpha > 0$, and (up to a shift of indexes) $\alpha^n \geqslant \alpha/2 > 0$ for all $n \in \mathbb{N}$. Let us show that this uniform coercivity forces the sequence $(x_0^n)_n$ to stay in a compact. The sequence $(E^n)_n$ is equiLipschitz by construction, and for each $x \in \Omega$, one has $|E^n(x) - E^m(x)| \leqslant \sup_k \left|E^k\right|_{1,\infty} d(x,o)$. By Arzelà-Ascoli, up to a subsequence, we may assume that E^n converges locally uniformly to some Lipschitz E, that satisfies

$$\lim_{d(x,o)\to\infty} E(x) \leqslant \lim_{d(x,o)\to\infty} E^n(x) + \sup_k \left|E - E^k\right|_\infty \leqslant \lim_{d(x,o)\to\infty} -\frac{\alpha}{2}\kappa(d(x,x_0)) - c_n + \sup_k \left|E - E^k\right|_\infty = -\infty.$$

In consequence, the level sets $E \geqslant c$ are bounded in Ω, thus compact. As x_0^n is the maximum of E^n, we have

$$x_0^n \in \{E^n \geqslant 0\} \subset \left\{ E \geqslant -\sup_k \left| E - E^k \right|_\infty \right\},$$

showing that up to a subsequence, x_0^n admits a limit $x_0 \in \Omega$. Invoking (2.7) again, the associated subsequence $(E^{n_k})_k$ converges towards $E := -\alpha\kappa \circ d(\cdot, x_0)$ with respect to $|\cdot|_{1,\infty}$, and the Cauchy sequence $(E^n)_n$ admits $E \in \mathcal{E}$ as a limit. □

To be complete, we also show that $\mathcal{E}$ satisfies **[A2.1.3]**.

Lemma 2.1.12. *For any compact $A \subset \mathcal{E}$ and $B \subset \Omega$, there exist $t > 0$ and reparametrized geodesics $(\gamma_{E,x})_{E\in A, x\in B}$ all defined on $[0, t]$ such that*

$$\lim_{h\searrow 0} \sup_{E\in A, x\in B} \frac{d\left(\mathcal{G}_{\mathcal{F}}(E)(h,x), \gamma_{E,x}(h)\right)}{h} = 0.$$

Proof. We compute explicitly the expression of the gradient flows, and derive an estimate that is uniform over any compact. Let $A \subset \mathcal{E}$ be compact for the topology induced by $|\cdot|_{1,\infty}$. Each $E \in A$ writes as $-\alpha^E\kappa \circ d(\cdot, x_0^E)$. First notice that $E \mapsto \alpha^E$ is bounded over A: indeed (assuming that diam $\Omega > 0$) there is some $\underline{\kappa} > 0$ such that for all $x_0^E \in \Omega$, there exists $x \in \Omega$ with $\kappa'(d(x, x_0)) \geqslant \underline{\kappa}$. Hence

$$|E|_{1,\infty} = \sup_{(x,v)\in \mathrm{T}\Omega, |v|_x=1} \left| \alpha^E \kappa'(d(x, x_0^E)) \langle \uparrow_x^{x_0^E}, v\rangle_x \right| \geqslant \alpha^E \underline{\kappa},$$

and the boundedness of A implies that of α^E.

Since any energy E is concave and Lipschitz, its gradient flow is uniquely defined, and one checks that it is given by

$$\vartheta(h,x) = (1-\tau(h))\, x \oplus \tau(h) x_0^E,$$

where $\tau(\cdot) \in \mathcal{C}^1(\mathbb{R}^+; [0,1])$ is identically 0 if $x = x_0^E$, and the unique solution of the Cauchy-Lipschitz ODE

$$\dot{\tau}(h) = \frac{\alpha^E \kappa'(d(x, x_0^E)(1-\tau(h)))}{d(x, x_0^E)}, \qquad \tau(0) = 0$$

otherwise. This corresponds to following the geodesic linking x to x_0^E with a speed proportional to α^E if x is sufficiently far from x_0^E, and exponentially decreasing in the region where κ becomes quadratic. Let $t_E := 1/\dot{\tau}(0)$ if $\dot{\tau}(0) > 0$, and 1 if $\dot{\tau}(0) = 0$. The time t_E is lower bounded uniformly on $E \in A$: indeed,

$$\dot{\tau}(0) = \frac{\alpha^E \kappa'(d(x, x_0^E))}{d(x, x_0^E)} = \begin{cases} \frac{\alpha^E}{d(x, x_0^E)} & d(x, x_0^E) \geqslant 1 \\ \alpha^E & d(x, x_0^E) < 1 \end{cases} \leqslant \alpha^E$$

is bounded over the compact A. Hence $t_E \geqslant \max(1, \min_{E\in A} 1/\alpha^E) > 0$.

For any $x \in B$, consider the reparametrized geodesic

$$\gamma_{E,x} : [0,1] \cap [0, t_E] \to \Omega, \qquad \gamma_{E,x}(s) := \left(1 - \frac{s}{t_E}\right) x \oplus \frac{s}{t_E} x_0^E.$$

The initial velocity of $\gamma_{E,x}$ coincides with $\nabla_x E = \vartheta(\cdot, x)_0^+$ by construction. Moreover,

$$d\left(\vartheta(h,x), \gamma_{E,x}(h)\right) \leqslant \left| \tau(h) - \frac{h}{t_E} \right| d(x, x_0^E) = |\tau(h) - h\dot{\tau}(0)|\, d(x, x_0^E) \leqslant \int_{s=0}^h \int_{r=0}^s |\ddot{\tau}(r)|\, d(x, x_0^E)\, dr\, ds.$$

Here $|\ddot{\tau}(r)|$ is defined almost everywhere by

$$|\ddot{\tau}(r)| = \left| \frac{\alpha^E}{d(x, x_0^E)} \kappa''(d(x, x_0^E)(1-\tau(r)))\, d(x, x_0^E)(-\dot{\tau}(r)) \right| = \begin{cases} 0 & \text{if } d(x, x_0^E)(1-\tau(r)) > 1, \\ (\alpha^E)^2(1-\tau(r)) & \text{if } d(x, x_0^E)(1-\tau(r)) < 1. \end{cases}$$

Hence in all cases, $|\ddot{\tau}(r)|\, d(x, x_0^E) \leqslant (\alpha^E)^2$, and the coarse estimate

$$d\left(\vartheta(h,x), \gamma_{E,x}(h)\right) \leqslant \frac{h^2}{2} (\alpha^E)^2$$

suffices to show that the approximation is uniform over any compact set. □

2.1.2 Variational characterization

We now characterize the solutions of the system (2.5) by evolutionary variational inequalities.

Proposition 2.1.13 (Variational characterization of trajectories). *Let $\mathcal{E}$ be an admissible set of energies, and Θ its associated set of transitions. Let $F : \Omega \to \mathcal{E}$ be a dynamic satisfying **[A2.1.8]**. Then a curve $y \in AC([0,T];\Omega)$ is a solution of the mutational inclusion (2.5) if and only if there exists a measurable selection $(E_s)_{s\in[0,T]}$ of $s \mapsto F(y_s)$ such that for almost any $s \in [0,T)$, there holds*

$$\frac{d}{ds}\frac{d^2(y_s,z)}{2} \leqslant E_s(y_s) - E_s(z) \qquad \forall z \in \Omega. \tag{2.8}$$

Note the following particularity of (2.8): the energy E_s varies with y_s, so that the flow defined by these "EVIs" does not have a gradient structure in general. The proof of Proposition 2.1.13 is split in two lemmas: first, it is shown that any solution of (2.5) in the sense of Definition 2.1.7 satisfies (2.8); uniqueness is then obtain by a stability estimate on the curves satisfying (2.8).

Lemma 2.1.14 (Existence). *Let $(y_s)_{s\in[0,T]}$ be mutable almost everywhere, and assume that $(\vartheta_s)_{s\in[0,T]} \subset \Theta$ is a measurable selection of $s \mapsto \mathring{y}_s$ that is contained in the image by $\mathcal{G}_\mathcal{F}$ of a compact of $(\mathcal{E}, |\cdot|_{1,\infty})$. Then $s \mapsto E_s := {\mathcal{G}_\mathcal{F}}^{-1}(\vartheta_s)$ is measurable, and (2.8) holds for this curve.*

Proof. Let $(y_s)_{s\in[0,T]} \in \mathrm{AC}(0,T;\Omega)$ be a mutable a.e., and $\vartheta \in L^1(0,T;\Theta)$ be a measurable selection of $s \mapsto \mathring{y}_s$. Let $s \in [0,T)$ such that $\vartheta_s \in \mathring{y}_s$. Then, for any $z \in \Omega$ and for sufficiently small $h > 0$, using the formula (1.8) of the directional derivative of the squared distance, we obtain

$$\begin{aligned}\frac{d^2(y_{s+h},z) - d^2(y_s,z)}{2h} &\in \frac{d^2(\vartheta_s(h,y_s),z) - d^2(y_s,z)}{2h} \pm \big(d(y_{s+h},z) + d(\vartheta_s(h,y_s),z)\big)\frac{d(\vartheta_s(h,y_s),y_{s+h})}{2h}\\ &\xrightarrow[h\searrow 0]{} -\langle \vartheta_s(\cdot,y_s)^+_0, \overrightarrow{y_s z}\rangle_{y_s}.\end{aligned}$$

Let $K_\mathbb{E} \subset \mathcal{E}$ be a compact for the topology induced by $|\cdot|_{1,\infty}$, containing the curve $\big({\mathcal{G}_\mathcal{F}}^{-1}(\vartheta_s)\big)_{s\in[0,T]}$. As $\mathcal{G}_\mathcal{F}$ in injective and continuous, its restriction to $K_\mathbb{E}$ has a continuous inverse. Hence $(E_s)_{s\in[0,T]} := \big({\mathcal{G}_\mathcal{F}}^{-1}(\vartheta_s)\big)_{s\in[0,T]}$ is measurable. By definition of gradient flow, the last term in the above equation is equal to $-\langle \nabla_{y_s} E_s, \overrightarrow{y_s z}\rangle_{y_s}$. Since E_s is concave and $\langle \nabla_z E_s, v\rangle_z \geqslant D_z E_s(v)$ by Definition 1.1.9 of the gradient, we get

$$\frac{d}{ds}\frac{d^2(y_s,z)}{2} = -\langle \nabla_{y_s} E_s, \overrightarrow{y_s z}\rangle_{y_s} \leqslant -D_{y_s} E_s(\overrightarrow{y_s z}) \leqslant E_s(y_s) - E_s(z).$$

The point z being arbitrary, and $(y_s)_s$ mutable almost everywhere, the claim holds. □

We turn to the dependence of the set of curves satisfying (2.8) with respect to the dynamic. The statement is written in a way so that linearity with respect to the dynamic clearly appears, which will be useful in Section 2.2.

Lemma 2.1.15 (Stability). *Let $\mathcal{E}$ be an admissible set of energies in the sense of Definition 2.1.1. For $i \in \{1,2\}$, let $E^i \in L^1(0,T;\mathbb{E})$ be valued in $\mathcal{E}$ a.e., and $(y^i_s)_{s\in[0,T]} \in AC([0,T];\Omega)$ be such that for almost any $s \in [0,T)$,*

$$\frac{d}{ds}\frac{d^2(y^i_s,z)}{2} \leqslant E^i_s(y^i_s) - E^i_s(z) \qquad \forall z \in \Omega.$$

Then for all $t \in [0,T]$,

$$\frac{d^2(y^1_t,y^2_t)}{2} \leqslant \frac{d^2(y^1_0,y^2_0)}{2} + \int_{s=0}^t \left[D_{y^2_s} E^1_s(\overrightarrow{y^2_s y^1_s}) - D_{y^2_s} E^2_s(\overrightarrow{y^2_s y^1_s})\right] ds. \tag{2.9}$$

Proof. The argument is adapted from [AGS05, Corollary 4.3.3], in which $s \mapsto E^i_s$ are constant in time. Since $E^i \in L^1(0,T;\mathcal{E})$ are Lebesgue-measurable curves valued in the Banach space $\mathbb{E}$, by [Hyt+16, Theorem

2.3.4], there exists $A_i \subset [0,T]$ of full Lebesgue measure such that any $t \in A_i$ is a Lebesgue point for E^i. In particular,

$$\lim_{h\searrow 0} \frac{1}{h} \int_{s=t-\alpha h}^{t+(1-\alpha)h} \left|E_t^i - E_s^i\right|_{1,\infty} ds = 0 \qquad \forall \alpha \in [0,1] \text{ and } t \in A_i.$$

Moreover, by the one-dimensional Rademacher theorem [AKP23, Proposition 13.9], both curves y^i are differentiable in the sense of Definition 1.1.7 at all points of some sets $A_{2+i} \subset [0,1]$ of full Lebesgue measure. Consider then $A_0 \subset [0,T]$ a conegligible subset such that $t \mapsto d^2(y_t^1, y_t^2)$ is differentiable for all $t \in A_0$, and let $t \in (0,T) \cap \bigcap_{i\in[\![0,4]\!]} A_i$.

Integrating the evolutionary variational inequality (2.8) over $[t-h,t]$ and $[t,t+h]$ for a sufficiently small $h>0$, there holds

$$\frac{d^2(y_t^1,z) - d^2(y_{t-h}^1,z)}{2} \leqslant \int_{s=t-h}^{t} \left(E_s^1(y_s^1) - E_s^1(z)\right) ds \qquad \forall z \in \Omega,$$
$$\frac{d^2(z,y_{t+h}^2) - d^2(z,y_t^2)}{2} \leqslant \int_{s=t}^{t+h} \left(E_s^2(y_s^2) - E_s^2(z)\right) ds \qquad \forall z \in \Omega.$$

Recall that by definition of $\mathcal{E}$, E_s^1 identifies with a Lipschitz-continuous function with constant $|E_s^1|_{1,\infty}$. Hence

$$\begin{aligned}\int_{s=t-h}^{t} \left(E_s^1(y_s^1) - E_s^1(z)\right) ds &\leqslant \int_{s=t-h}^{t} \left(E_s^1(y_t^1) - E_s^1(z) + |E_s^1|_{1,\infty}\, d(y_s,y_t)\right) ds \\ &\leqslant h\left(E_t^1(y_t^1) - E_t^1(z)\right) + \int_{s=t-h}^{t} \left(|E_s^1|_{1,\infty}\, d(y_s,y_t) + |E_s^1 - E_t^1|_{1,\infty}\, d(z,y_t^1)\right) ds.\end{aligned}$$

Dividing by $h>0$ and taking the limit sup in $h \searrow 0$, the second term vanishes. Following the same reasoning with $i=2$, we obtain that for all $z \in \Omega$,

$$\limsup_{h\searrow 0} \frac{d^2(y_t^1,z) - d^2(y_{t-h}^1,z)}{2} \leqslant E_t^1(y_t^1) - E_t^1(z), \tag{2.10}$$
$$\limsup_{h\searrow 0} \frac{d^2(z,y_{t+h}^2) - d^2(z,y_t^2)}{2} \leqslant E_t^2(y_t^2) - E_t^2(z). \tag{2.11}$$

As $t \in A_3$, the curve y^1 is differentiable in the sense of Definition 1.1.7 at t, and there holds

$$\limsup_{h\searrow 0} \frac{d^1(y_t^1,z) - d^1(y_{t-h}^1,z)}{2h} = D_{y_t^1}\frac{d^2(\cdot,z)}{2}((y^1)_t^-) = -\langle (y^1)_t^-, \overrightarrow{y_t^1 z}\rangle_{y_t^1}.$$

We may choose $z = (1-\alpha)y_t^1 \oplus \alpha y_t^2$ for $\alpha \in (0,1]$, which implies $\uparrow_{y_t^1}^{z} = \alpha \uparrow_{y_t^1}^{y_t^2}$. Using the positive homogeneity of the metric scalar product, we deduce

$$\limsup_{h\searrow 0} \frac{d^2(y_t^1,(1-\alpha)y_t^1 \oplus \alpha y_t^2) - d^2(y_{t-h}^1,(1-\alpha)y_t^1 \oplus \alpha y_t^2)}{2h} = \alpha \limsup_{h\searrow 0} \frac{d^2(y_t^1,y_t^2) - d^2(y_{t-h}^1,y_t^2)}{2h}.$$

On the other hand, E^1 is directionally differentiable, so that $E_t^1((1-\alpha)y_t^1 \oplus \alpha y_t^2) - E_t^1(y_t^1) = \alpha d_{y_t^1} E^1(\overrightarrow{y_t^1 y_t^2}) + o(\alpha)$. Dividing by $\alpha>0$ and letting $\alpha \searrow 0$, we obtain from (2.10) that

$$\limsup_{h\searrow 0} \frac{d^2(y_t^1,y_t^2) - d^2(y_{t-h}^1,y_t^2)}{2h} \leqslant -D_{y_t^1} E_t^1(\overrightarrow{y_t^1 y_t^2}),$$

and symmetrically

$$\limsup_{h\searrow 0} \frac{d^2(y_{t+h}^2,y_t^1) - d^2(y_t^2,y_t^1)}{2h} \leqslant -D_{y_t^2} E_t^2(\overrightarrow{y_t^2 y_t^1}).$$

By [AGS05, Lemma 4.3.4], there holds that

$$\begin{aligned}\frac{d}{dt}\frac{d^2(y_t^1,y_t^2)}{2} &\leqslant \limsup_{h\searrow 0}\frac{d^2(y_t^1,y_t^2)-d^2(y_{t-h}^1,y_t^2)}{2h}+\limsup_{h\searrow 0}\frac{d^2(y_t^1,y_{t+h}^2)-d^2(y_t^1,y_t^2)}{2h}\\ &\leqslant \underbrace{-D_{y_t^1}E_t^1(\overrightarrow{y_t^1y_t^2})-D_{y_t^2}E_t^1(\overrightarrow{y_t^2y_t^1})}_{\mathcal{A}}+D_{y_t^2}E_t^1(\overrightarrow{y_t^2y_t^1})-D_{y_t^2}E_t^2(\overrightarrow{y_t^2y_t^1}).\end{aligned} \tag{2.12}$$

As E_t^1 is concave, the term $\mathcal{A}$ is nonpositive. Integrating over time yields (2.9). □

Proposition 2.1.13 can now be completed in the following way.

Proof of Proposition 2.1.13. On the one hand, Lemma 2.1.14 shows that any solution is the sense of mutations satisfies the EVI characterization. One the other hand, assume that y^1 satisfies the EVI (2.8) for some $(E_s)_{s\in[0,T]} \in L^1(0,T;\mathbb{E})$ with $E_s \in F(y_s^1)$ a.e.. Applying Lemma 2.1.15 with the constant (mutable) curve $[0,T] \ni s \mapsto x$, one sees that

$$\frac{d^2(y_t^1,x)}{2} \leqslant 0+\int_{s=0}^t D_xE_s(\overrightarrow{xy_s^1})ds \leqslant \int_{s=0}^t |E_s|_{1,\infty}\, d(x,y_s^1)ds \leqslant \int_{s=0}^t C_f\left(1+d(o,y_s^1)\right)d(y_s^1,x)ds,$$

and applying a Grönwall Lemma, the curve y^1 lies in a compact subset B of Ω on the time interval $[0,T]$. Up to redefining $(E_s)_s$ on a negligible subset, it holds that $E_s \in \bigcup_{z\in B} F(z)$. The latter set is compact as the union over a compact of the images of a compact-valued usc multifunction [Ber59, Theorem 3 p.116].

By Proposition 2.1.6, there exists an absolutely continuous curve y^2 issued from $y^1(0)$, that is mutable almost everywhere and such that $\mathcal{G}_{\mathcal{F}}(E_s) \in \mathring{y}_s^2$ for a.e. $s \in [0,T]$. Applying Lemma 2.1.14, y^2 also satisfies (2.8) for the curve $(E_s)_s$. Then, by Lemma 2.1.15,

$$\frac{d^2(y_t^1,y_t^2)}{2} \leqslant \frac{d^2(y_0^1,y_0^2)}{2}+\int_{s=0}^t\left[D_{y_s^2}E_s(\overrightarrow{y_s^2y_s^1})-D_{y_s^2}E_s(\overrightarrow{y_s^2y_s^1})\right]ds=0,$$

and we conclude that y^1 coincides with y^2, hence is a solution of the mutational inclusion (2.5). □

2.2 Relaxation theorem

In this section, we define the relaxed system using convex hulls in the Banach space $\mathbb{E}$. We then prove the relaxation theorem, and give an interpretation of the solutions in terms of barycenter of gradients.

2.2.1 Convexified system

The relaxation of the dynamic is made at the level of energies. Denote by $\overline{\mathrm{conv}}\mathcal{E} \subset \mathbb{E}$ the closed convex hull of $\mathcal{E}$ as a subset of $\left(\mathbb{E},|\cdot|_{1,\infty}\right)$, i.e. the smallest convex closed subset of $(\mathbb{E},|\cdot|_{1,\infty})$ containing $\mathcal{E}$.

Definition 2.2.1 (Relaxed dynamic). *Let $F:\Omega \rightrightarrows \mathcal{E}$ be a dynamic. The relaxed dynamic is defined as*

$$\overline{co}F:\Omega \rightrightarrows \overline{\mathrm{conv}}\mathcal{E}, \qquad \overline{co}F(x) := \overline{\mathrm{conv}}F(x) \qquad \forall x\in\Omega. \tag{2.13}$$

We first check that the relaxed system admits solutions. To this aim, we show that the set $\overline{\mathrm{conv}}\mathcal{E}$ is itself an admissible set of energies, and that the relaxed dynamic $\overline{\mathrm{co}}F$ satisfies the assumptions of Proposition 2.1.9. Both arguments, as well as some in the sequel, are eased by a parametrization of the convex hull through probability measures, and we introduce the necessary tools here.

Monge-Kantorovich tools. Let (X,d_X) be a Polish space. Denote by $\mathscr{P}(X)$ the set of Borel probability measures over X. If $m:(X,d_X)\to(Y,d_Y)$ is a measurable map between Polish spaces, the pushforward $m\#\omega$ of $\omega\in\mathscr{P}(X)$ is the element of $\mathscr{P}(Y)$ given by $(m\#\omega)(A) := \omega\left(m^{-1}(A)\right)$ for all measurable $A\subset Y$. The canonical projections from (X^2,d_{X^2}) are denoted by π_1,π_2, i.e. $\pi_1(x_1,x_2)=x_1$ and $\pi_2(x_1,x_2)=x_2$.

Let $\mathscr{P}_1(X)$ be the subset of $\mu \in \mathscr{P}(X)$ such that $\int_{x\in X} d_X(o,x)d\mu(x) < \infty$ for one (thus all) $o \in X$. $\mathscr{P}_1(X)$ is itself Polish when endowed with the Monge-Kantorovich distance

$$d_{\mathrm{MK}}(\omega,\varpi) := \inf\left\{\int_{(x_1,x_2)\in X^2} d_X(x_1,x_2)d\alpha(x_1,x_2)\ \middle|\ \alpha \in \mathscr{P}(X^2),\ \pi_1\#\alpha = \omega,\ \pi_2\#\alpha = \varpi\right\}. \tag{2.14}$$

Using the separability of the Polish space X, one shows that finite combinations of Dirac masses are dense in $\mathscr{P}_1(X)$ with respect to d_{MK}. For more details, we refer the reader to [AGS05, Chap. 5-7], [Vil09, Chap. 6] or [San15]. For convenience, if ω,ϖ are two nonnegative Borel measures with the same mass $m > 0$, we still denote

$$d_{\mathrm{MK}}(\omega,\varpi) := m d_{\mathrm{MK}}\left(\frac{\omega}{m},\frac{\varpi}{m}\right).$$

Lemma 2.2.2 (Properties of the relaxed system). *The set $\overline{\mathrm{conv}}\mathcal{E}$ is an admissible set of energies in the sense of Definition 2.1.1. Moreover, the relaxed dynamic can be parametrized as*

$$\overline{co}F(x) = Bary_{\mathbb{E}}\left(\mathscr{P}_1(F(x))\right),$$

and satisfies ***[A2.1.8]*** *if so does the original dynamic F.*

Proof. We first parametrize the convex hull by barycenters of measures in $\mathscr{P}_1(\mathbb{E})$. Given $\omega \in \mathscr{P}_1(\mathbb{E})$, denote $\mathrm{Bary}_{\mathbb{E}}(\omega)$ its barycenter, defined as the function $x \mapsto \int_{E\in\mathbb{E}} E(x)d\omega(E)$. The function $\mathrm{Bary}_{\mathbb{E}}(\omega)$ is real-valued and Lipschitz-continuous: indeed, for $x, y \in \Omega$,

$$\left|\mathrm{Bary}_{\mathbb{E}}(\omega)(y) - \mathrm{Bary}_{\mathbb{E}}(\omega)(x)\right| = \left|\int_{E\in\mathcal{E}} (E(y)-E(x))d\omega(E)\right| \leqslant d(x,y)\int_{E\in\mathcal{E}} |E|_{1,\infty}\, d\omega(E) = d(x,y)d_{\mathrm{MK}}(\omega,\delta_{0_{\mathbb{E}}}),$$

where we used that $\mathrm{Lip}(E) = |E|_{1,\infty}$ for any $E \in \mathcal{E}$. Moreover, if ω is supported on concave functions, the function $\mathrm{Bary}_{\mathbb{E}}(\omega)$ is concave as well, and hence belongs to $\mathbb{E}$. Let us show that the barycenter operation is 1-Lipschitz from $(\mathscr{P}_1(\mathbb{E}), d_{\mathrm{MK}})$ to $(\mathbb{E}, |\cdot|_{1,\infty})$. Pick $\omega,\varpi \in \mathscr{P}_1(\mathbb{E})$, and $\alpha \in \mathscr{P}(\mathbb{E}^2)$ satisfying the marginal constraints $\pi_1\#\alpha = \omega$ and $\pi_2\#\alpha = \varpi$. By Jensen inequality, there holds

$$\left|\mathrm{Bary}_{\mathbb{E}}(\omega) - \mathrm{Bary}_{\mathbb{E}}(\varpi)\right|_{1,\infty} = \left|\int_{(E,E')\in\mathbb{E}^2} (E-E')d\alpha(E,E')\right|_{1,\infty} \leqslant \int_{(E,E')\in\mathbb{E}^2} \left|E-E'\right|_{1,\infty} d\alpha(E,E')$$

Taking the infimum over α, we get that $\left|\mathrm{Bary}_{\mathbb{E}}(\omega) - \mathrm{Bary}_{\mathbb{E}}(\varpi)\right|_{1,\infty} \leqslant d_{\mathrm{MK}}(\omega,\varpi)$. Now, one can compute $\overline{\mathrm{conv}}\mathcal{E}$ as

$$\overline{\mathrm{conv}}\mathcal{E} = \overline{\mathrm{Bary}_{\mathbb{E}}(\mathscr{P}_1(\mathcal{E}))}, \quad \text{where} \quad \mathrm{Bary}_{\mathbb{E}}(\mathscr{P}_1(\mathcal{E})) := \left\{\mathrm{Bary}_{\mathbb{E}}(\omega)\ \middle|\ \omega \in \mathscr{P}_1(\mathbb{E}) \text{ is supported on } \mathcal{E}\right\}.$$

Indeed, $\mathrm{Bary}_{\mathbb{E}}(\mathscr{P}_1(\mathcal{E}))$ is a convex subset of $\mathbb{E}$, and contains $\mathcal{E}$ as the barycenters of Dirac masses. Hence $\overline{\mathrm{conv}}\mathcal{E} \subset \overline{\mathrm{Bary}_{\mathbb{E}}(\mathscr{P}_1(\mathcal{E}))}$. On the other hand, using the density of finite combinations of Dirac masses, each $\mathrm{Bary}_{\mathbb{E}}(\omega)$ is a convex combination of elements of $\mathcal{E}$, hence belongs to $\mathrm{conv}\,\mathcal{E}$. Thus $A \subset \mathrm{conv}\,\mathcal{E}$, and $\overline{A} = \overline{\mathrm{conv}}\mathcal{E}$.

As $\overline{\mathrm{conv}}\mathcal{E}$ contains $0_{\mathbb{E}}$, it remains only to prove that it is closed and separable in $(\mathcal{E}, |\cdot|_{1,\infty})$. Closedness follows by definition, and separability from the fact that $\mathscr{P}_1(\mathcal{E})$ is separable in the topology induced by d_{MK} [Vil09, Theorem 6.18], and $\mathrm{Bary}_{\mathbb{E}}(\cdot)$ continuous.

We turn to the properties of the relaxed dynamic. By [Vil09, Remark 6.19], the set of measures $\mu \in \mathscr{P}_1(\mathbb{E})$ that are supported on a compact is itself compact. Hence, for all $x \in \Omega$,

$$\overline{\mathrm{co}}\,F(x) = \mathrm{Bary}_{\mathbb{E}}\left(\mathscr{P}_1(F(x))\right) \tag{2.15}$$

is a compact subset of $\mathbb{E}$. Additionally, the mapping $K \mapsto \mathscr{P}_1(K)$ is 1-Lipschitz from nonempty compact subsets endowed with the Hausdorff distance to $(\mathscr{P}_1(\mathbb{E}), d_{\mathrm{MK}})$. Indeed, let K and K' be nonempty compacts, and $\omega \in \mathscr{P}_1(K)$. By a measurable selection theorem, for instance [AB06], one can build a measurable function $g : K \to K'$ such that $d(k, g(k)) \leqslant d_H(K,K')$ for any $k \in K$. Denote then $\omega' := g\#\omega \in \mathscr{P}(K')$. There holds

$$d_{\mathrm{MK}}(\omega,\omega') \leqslant \int_{x\in K} d(x,g(x))d\omega(x) \leqslant d_H(K,K').$$

As we may switch the role of K and K', the desired 1-Lipschitz continuity holds. As the barycenter is 1-Lipschitz, the composition (2.15) satisfies the compactness of images, growth and local Lipschitz behaviour required by **[A2.1.8]**. □

Remark 2.2.3 (Parametrization by a curve of measures)**.** *Let $y(\cdot) \in AC([0,T];\Omega)$ be a solution of the relaxed dynamical system. By Proposition 2.1.9, there exists $(E_s)_{s\in[0,T]} \in L^1(0,T;\overline{\mathrm{conv}}\mathcal{E})$ such that $\mathcal{G}_{\mathcal{F}}(E_s) \in \mathring{y}_s \cap \mathcal{G}_{\mathcal{F}}(\overline{co}F(y_s))$ for a.e. $s \in [0,T]$. This curve can be parametrized by a curve of measures $\omega(\cdot) \in L^1(0,T;\mathscr{P}(\mathcal{E}))$ as follows: as $\overline{co}F$ has a controlled growth, $y(\cdot)$ stays in a compact. Taking the union over this compact of the usc map $F(\cdot)$ still yields a compact subset of $\mathcal{E}$. The set of probability measures over it is in turn a compact in the Monge-Kantorovich topology, and the restriction of $Bary_{\mathbb{E}}(\cdot)$ to this set is continuous. Hence one can apply a measurable selection theorem to obtain a curve $\omega(\cdot) \in L^1(0,T;\mathscr{P}(\mathcal{E}))$ such that $Bary_{\mathbb{E}}(\omega_s) = E_s$ for a.e. $s \in [0,T]$.*

2.2.2 The relaxation theorem

We turn to the main result of this section.

Theorem 2.2.4 (Relaxation). *Let $\mathcal{E}$ be an admissible set of energies in the sense of Definition 2.1.1, and $F : \Omega \rightrightarrows \mathcal{E}$ satisfy **[A2.1.8]**. Then for any $x \in \Omega$ and $T > 0$, the closure in $AC([0,T];\Omega)$ of the set of solutions of*

$$\mathring{y}_s \cap \mathcal{G}_{\mathcal{F}}(F)(y_s) \neq \emptyset, \qquad y_0 = x$$

is given by the set of solutions of

$$\mathring{y}_s \cap \mathcal{G}_{\mathcal{F}}(\overline{co}F)(y_s) \neq \emptyset, \qquad y_0 = x. \tag{2.16}$$

Moreover, the latter set is compact in $AC([0,T];\Omega)$, endowed with the usual distance $d_{AC}(y,y') = \sup_{t\in[0,T]} d(y_t, y'_t)$.

The proof of Theorem 2.2.4 goes by double inclusion. Lemma 2.2.6 below shows that any Cauchy sequence of solutions of the mutational inclusion converges to a solution of the relaxed system, and Lemma 2.2.9 constructs approximations of a given trajectory of the relaxed system by solutions of the original one. The proof relies on the following Lemma, adapted from [BF24].

Lemma 2.2.5 (Local weak compactness). *Let $F : \Omega \rightrightarrows \mathcal{E}$ satisfy **[A2.1.8]**. For any compact $B \subset \Omega$, the set $L^1(0,T;X)$ with*

$$X := \bigcup_{y\in[0,T]\times B} \overline{co}F(y) \subset \mathbb{E}$$

is weakly compact in $L^1(0,T;\mathbb{E})$.

Proof. Each $\overline{\mathrm{co}}\,F(y)$ is nonempty, convex and compact in the Banach space $\mathbb{E}$, hence weakly closed, and weakly compact by James' Theorem [Jam64, Theorem 5]. The set-valued map $\overline{\mathrm{co}}\,F$ is upper semicontinuous from (Ω, d) to $(\mathbb{E}, |\cdot|_{1,\infty})$: since each weakly open set of $\mathbb{E}$ is also open, it is also upper semicontinuous into $\mathbb{E}$ equipped with its weak topology. We may then apply [Ber59, Theorem 3 p.116] to get that the union X of the images $\overline{\mathrm{co}}\,F(y)$ when y ranges in the compact B is again weakly compact. By Diestel's theorem [DU77, Proposition 7], $L^1(0,T;X)$ is then relatively weakly compact. As it is closed and convex, it is additionally weakly closed, hence weakly compact in $L^1(0,T;\mathbb{E})$. □

Let us begin by the characterization of accumulation points.

Lemma 2.2.6 (Stability of solutions). *Consider the assumptions of Theorem 2.2.4. Let $(y^n)_{n\in\mathbb{N}} \subset AC([0,T];\Omega)$ be a sequence of solutions of the differential inclusion (2.5) in the sense of Definition 2.1.7, that converges uniformly towards some $y \in AC([0,T];\Omega)$. Then y is a solution of the relaxed dynamical system (2.16).*

Proof. Let $B \subset \Omega$ be a closed ball containing the graph of all solutions of the relaxed system (2.16) issued from some y_0^n or y^0. Denote $\mathrm{Lip}(F)_B$ a local Lipschitz constant of F on B. For each $n \in \mathbb{N}$, let $\varphi^n \in L^1(0,T;\mathbb{E})$ be a measurable selection of $s \mapsto F(y_s^n)$ such that $\varphi_s^n \in \mathring{y}_s^n$ for almost every $s \in [0,T]$. As F has sublinear growth, $\|\varphi^n\|_{L^1(0,T;\mathbb{E})} \leqslant C$ for some C that is independent of n. Moreover, each φ^n is valued a.e. in $X := \bigcup_{z\in B} F(z) \subset \mathbb{E}$. Applying Lemma 2.2.5, we obtain that up to a non relabelled subsequence, φ^n converges weakly in the Banach space $L^1(0,T;\mathbb{E})$ towards some φ that is valued in $X \subset \overline{\mathrm{conv}}\mathcal{E}$ for a.e. s.

We now show that $\varphi \in \overline{\text{co}}\, F(y_s)$. Let $L \in \mathbb{E}'$ be a linear continuous form and $p \in \mathcal{C}^\infty([0,T];\mathbb{R}^+)$. For each $n \in \mathbb{N}$,

$$\begin{aligned}\int_{s=0}^T p(s) L\big(\varphi_s^n\big)\, ds &\leqslant \int_{s=0}^T p(s) \max_{b \in F(y_s^n)} L(b)\, ds \\ &\leqslant \int_{s=0}^T p(s) \max_{b \in F(y_s)} L(b)\, ds + \big|p\big|_\infty \|L\|_{\mathbb{E}'} \text{Lip}(F)_B \sup_{s \in [0,T]} d(y_s, y_s^n).\end{aligned}$$

As $\psi \mapsto \int_{s=0}^T p(s) L\big(\psi_s\big)\, ds$ belongs to the dual of $L^1(0,T;\mathbb{E})$, we may pass to the limit in $n \to \infty$ in the above inequality. As $p \in \mathcal{C}^\infty([0,T];\mathbb{R}^+)$ and $L \in \mathbb{E}'$ are arbitrary, there exists a measurable conegligible set $I \subset [0,T]$ such that $\varphi_s \in \overline{\text{conv}}\, F(y_s) = \overline{\text{co}}\, F(y_s)$ for every $s \in I$.

To conclude, we only have to show that $\varphi_s \in \mathring{y}_s$ for a.e. $s \in [0,T]$. By Proposition 2.1.13, this is equivalent to the satisfaction of the evolutionary variational inequalities (2.8). Let $z \in \Omega$ be fixed, and $0 \leqslant s < t \leqslant T$. For any n, there holds

$$\begin{aligned}\frac{d^2(y_t^n, z) - d^2(y_s^n, z)}{2} &= \int_{\tau=s}^t \frac{d^2(y_\tau^n, z)}{2} d\tau \leqslant \int_{\tau=s}^t \big[\varphi_\tau^n(y_\tau^n) - \varphi_\tau^n(z)\big]\, d\tau \\ &\leqslant \int_{\tau=s}^t \big[\varphi_\tau^n(y_\tau) - \varphi_\tau^n(z)\big]\, d\tau + \|\varphi^n\|_{L^1(s,t;\mathbb{E})} \sup_{\tau \in [s,t]} d(y_\tau^n, y_\tau).\end{aligned}$$

The functional $\psi \mapsto \int_{\tau=s}^t \big(\psi_\tau(y_\tau) - \psi_\tau(z)\big)\, d\tau$ is linear and continuous from $L^1(0,T;\mathbb{E})$ to $\mathbb{R}$. Consequently, we may pass to the limit in $n \to \infty$ in the previous inequality to get that

$$\frac{d^2(y_t, z) - d^2(y_s, z)}{2} \leqslant \int_{\tau=s}^t \big[\varphi_\tau(y_\tau) - \varphi_\tau(z)\big]\, d\tau \qquad \text{for all } 0 \leqslant s < t \leqslant T, \text{ and } z \in \Omega. \tag{2.17}$$

We now have to get back to a differential version in time, while paying attention to the fact that for a.e. time, the EVI must be satisfied for all z. Let $I^0 \subset [0,T]$ be the set of Lebesgue points of $\tau \mapsto \varphi_\tau \in \mathbb{E}$, which is of full measure by [Hyt+16, Theorem 2.3.4]. Fix $s \in I^0$ and $z \in \Omega$. Using the Lipschitz-continuity of φ_τ locally uniformly in τ, the local Lipschitz-continuity of the curve y and the fact that the norm of $\mathbb{E}$ induces local uniform convergence, one has

$$\lim_{t \searrow s} \frac{1}{t-s} \int_{\tau=s}^t \big[\varphi_\tau(y_\tau) - \varphi_\tau(z)\big]\, d\tau = \varphi_t(y_s) - \varphi_s(z).$$

On the other hand, by the one-dimensional Rademacher Lemma [AKP23, p. 13.9], there exists $I^1 \subset [0,T]$ of full Lebesgue measure such that $y(\cdot)$ is differentiable in the sense of Definition 1.1.7 for all $s \in I^1$. Consequently, the derivative of $\tau \mapsto d^2(y_\tau, z)$ exists at $s \in I^1$ for any $z \in \Omega$. Dividing by $t-s$ and passing to the limit in $t \searrow s$ for $s \in I^0 \cap I^1$ in (2.17), we obtain the desired EVI.

□

Let $\omega(\cdot) \in L^1(0,T;\mathscr{P}(\mathcal{E}))$ and $I \subset [0,T]$ be a nontrivial interval. For brevity, we denote by $\mathcal{L}_I \otimes \omega$ the measure on $[0,T] \times \mathcal{E}$ given by $\int_{s \in I} \delta_s \otimes \omega_s ds$, i.e.

$$\int_{(s,E) \in [0,T] \times \mathcal{E}} \varphi(s,E)\, d[\mathcal{L}_I \otimes \omega] = \int_{s \in I} \int_{E \in \mathcal{E}} \varphi(s,E)\, d\omega_s(E)\, ds \qquad \forall \varphi \in \mathcal{C}_b([0,T] \times \mathcal{E};\mathbb{R}).$$

The measure $\mathcal{L}_I \otimes \omega$ has mass $\mathcal{L}(I)$. This notation is consistent with the case where $\omega(s) \equiv \varpi$, in which case $\mathcal{L}_I \otimes \omega$ is given by the product measure $\mathcal{L}_I \otimes \varpi$. Moreover, using a measurable selection of optimal plans for each s, one has $d_{\text{MK}}(\mathcal{L}_I \otimes \omega, \mathcal{L}_I \otimes \varpi) \leqslant \int_{s \in I} d_{\text{MK}}(\omega_s, \varpi_s) ds$.

Lemma 2.2.7. *Let $\mathcal{E}$ be an admissible set of energies. Let $I \subset [0,T]$ be a nontrivial closed interval, $K \subset \mathcal{E}$ and $B \subset \Omega$ be compact, and $y \in AC(I;\Omega)$ lie in B. Let $(\omega^n)_{n \in \mathbb{N}} \subset L^1(I;\mathscr{P}_1(K))$ and $\omega \in L^1(I;\mathscr{P}_1(K))$ be such that*

$$\lim_{n \to \infty} d_{MK}\big(\mathcal{L}_I \otimes \omega, \mathcal{L}_I \otimes \omega^n\big) \underset{n \to \infty}{\longrightarrow} 0.$$

Then for any constant $C \geqslant 0$,

$$\sup_{\substack{y' \in AC(I;B) \\ Lip(y') \leqslant C}} \left| \int_{s \in I} D_{y_s} Bary_{\mathbb{E}}(\omega_s) \left(\overrightarrow{y_s y_s'}\right) ds - \int_{s \in I} D_{y_s} Bary_{\mathbb{E}}\left(\omega_s^n\right) \left(\overrightarrow{y_s y_s'}\right) ds \right| \underset{n \to \infty}{\longrightarrow} 0. \tag{2.18}$$

Proof. We change variables to apply the results of integration of Banach-valued functions. The space $\mathcal{C}(K\times B;\mathbb{R})$ is a Banach space when endowed with the sup norm. For each $s\in I$, the function $\rho_s : K\times B\to\mathbb{R}$ defined by $\rho_s(E,z):=D_{y_s}E\left(\overrightarrow{y_s z}\right)$ is continuous in both variables, with in addition

$$\begin{aligned}\left|\rho_s(E,z)-\rho_s(E',z')\right| &\leqslant \left|D_{y_s}E\left(\overrightarrow{y_s z}\right)-D_{y_s}E'\left(\overrightarrow{y_s z}\right)\right|+\mathrm{Lip}(D_{y_s}E')d_{y_s}\left(\overrightarrow{y_s z},\overrightarrow{y_s z'}\right)\\ &\leqslant \left|E-E'\right|_{1,\infty}d(y_s,z)+\left|E'\right|_{1,\infty}d(z,z')\end{aligned}$$

coming from the definition of $|\cdot|_{1,\infty}$ in (2.1), and the estimate (1.7). The function $s\mapsto\rho_s$ is measurable from I to $\mathcal{C}(K\times B;\mathbb{R})$, since it is the pointwise limit as h goes to 0 of the continuous functions $s\mapsto h^{-1}(E\left((1-h)y_s\oplus hz\right)-E(y_s))$. Moreover,

$$\sup_{E\in K,z\in B}\left|D_{y_s}E\left(\overrightarrow{y_s z}\right)\right|\leqslant\sup_{E\in K,z\in B}|E|_{1,\infty}d(y_s,z)\leqslant\sup_{E\in K}|E|_{1,\infty}\mathrm{diam}\,B,$$

so that ρ is in $L^1(I;\mathcal{C}(K\times B;\mathbb{R}))$. Let α^n be an optimal transport plan between $\mathcal{L}_I\otimes\omega$ and $\mathcal{L}_I\otimes\omega^n$. Then

$$\begin{aligned}&\int_{(s,E)\in I\times\mathcal{E}}\rho_s(E,y'_s)d[\mathcal{L}_I\otimes\omega](E,s)-\int_{(s,E)\in I\times\mathcal{E}}\rho_s(E,y'_s)d[\mathcal{L}_I\otimes\omega^n](E,s)\\ &=\int_{((s,E),(s',E'))\in(I\times\mathcal{E})^2}\rho_s(E,y'_s)-\rho_{s'}(E',y'_{s'})d\alpha^n((s,E),(s',E'))\\ &\leqslant\int_{(I\times\mathcal{E})^2}\rho_s(E,y'_s)-\rho_{s'}(E,y'_s)+\left|E-E'\right|_{1,\infty}d(y_{s'},y'_{s'})+\left|E'\right|_{1,\infty}d(y'_{s'},y'_s)d\alpha^n\\ &\leqslant\int_{s,s'\in I}\|\rho_s-\rho_{s'}\|_{\mathcal{C}(K\times B;\mathbb{R})}d(\pi_s,\pi_{s'})\#\alpha^n(s,s')+\widetilde{C}d_{\mathrm{MK}}\left(\mathcal{L}_I\otimes\omega,\mathcal{L}_I\otimes\omega^n\right),\end{aligned}\tag{2.19}$$

with $\widetilde{C}:=\max(\mathrm{diam}\,B,|K|_{1,\infty}C)$. Each plan $\beta^n:=(\pi_s,\pi_{s'})\#\alpha^n(s,s')$ has both marginals equal to $\mathcal{L}_I$, and by [Vil09, Theorem 5.20], converges to the unique optimal transport plan between $\mathcal{L}_I$ and itself, that is, $(id,id)\#\mathcal{L}_I$. This implies that the first summand in (2.19) goes to 0 when n goes to ∞; indeed, by [Hyt+16, Lemma 1.2.31], there exists a family $(\rho^m)_{m\in\mathbb{N}}\subset\mathcal{C}_c(I;\mathcal{C}(K\times B;\mathbb{R}))$ approximating ρ in the $L^1(I;\mathcal{C}(K\times B;\mathbb{R}))$ norm. Using the triangular inequality and the fact that both marginals of β^n are $\mathcal{L}_I$,

$$\int_{s,s'\in I}\|\rho_s-\rho_{s'}\|_{\mathcal{C}(K\times B;\mathbb{R})}d\beta^n\leqslant\int_{s,s'\in I}\|\rho^m_s-\rho^m_{s'}\|_{\mathcal{C}(K\times B;\mathbb{R})}d\beta^n+2\|\rho^m-\rho\|_{L^1(I;\mathcal{C}(K\times B;\mathbb{R}))},$$

and choosing m large enough, then n large enough, the limit ensue. The second summand in (2.19) goes to 0 by assumption, and (2.18) holds. □

Lemma 2.2.8 (Approximation by chattering dynamics). *Let $A\subset\mathcal{E}$ be compact, $I\subset[0,T]$ be a nontrivial closed interval, and $\omega\in L^1(I;\mathscr{P}_1(A))$. For any $\iota>0$, there exists $E^\iota\in L^1(I;A)$ such that in the Monge-Kantorovich distance associated to $\mathscr{P}_1(I\times A)$, there holds*

$$d_{MK}\left(\mathcal{L}_I\otimes\omega,\,(id,E^\iota)\#\mathcal{L}_I\right)\leqslant\iota.\tag{2.20}$$

Proof. Any $\omega(\cdot)\in L^1(0,T;\mathscr{P}_1(A))$ can be approximated in the L^1 distance by simple functions of the form $\omega^{(0)}:=\sum_{j=1}^M\omega_j^{(0)}\mathbb{1}_{J_j}$, where $(J_j)_{j\in[\![1,M]\!]}$ are disjoint nontrivial intervals covering $[0,T]$, and $\omega_j^{(0)}\in\mathscr{P}_1(A)$. Each $\omega_j^{(0)}$ can in turn be approximated with respect to d_{MK} by an empirical measure $\omega_j^{(1)}=\sum_{k=1}^P m_{jk}\delta_{E_{jk}}$, with $E_{jk}\in A$. Partition each J_j in subintervals J_{jk} such that the relative length $|J_{jk}|/|J_j|$ coincides with the mass m_{jk} of the k^{th} atom. Define E^ι to be identically equal to E_{jk} on J_{jk}. With this choice, $E^\iota\#\mathcal{L}_{J_j}=\mathcal{L}(J_j)\omega_j^{(1)}$, and the plan $\eta:=\frac{1}{\mathcal{L}(J_j)}\int_{s\in J_j}\delta_s\otimes(E^\iota,id,E^\iota)\#\mathcal{L}_{J_j}ds$ is a transport plan between $\mathcal{L}_{J_j}\otimes\omega_j^{(0)}$ and $(id,E^\iota)\#\mathcal{L}_{J_j}$. Hence

$$d_{\mathrm{MK}}\left(\mathcal{L}_{J_j}\otimes\omega_j^{(0)},(id,E^\iota)\#\mathcal{L}_{J_j}\right)\leqslant\frac{1}{\mathcal{L}(J_j)}\int_{s\in J_j}\int_{\tau\in J_j}|s-\tau|+\left|E^\iota_\tau-E^\iota_\tau\right|d\mathcal{L}_{J_j}(\tau)ds\leqslant\mathrm{diam}(J_j)\mathcal{L}(J_j).$$

Using the triangular inequality, and the convexity of the Monge-Kantorovich distance with respect to sums in the Banach space of measures,

$$\begin{aligned} d_{\mathrm{MK}}\left(\mathcal{L}_{[0,T]}\otimes\omega,\,(id,E^{\iota})\#\mathcal{L}_{[0,T]}\right) &\leqslant d_{\mathrm{MK}}\left(\mathcal{L}_{[0,T]}\otimes\omega,\mathcal{L}_{[0,T]}\otimes\omega^{(0)}\right)+d_{\mathrm{MK}}\left(\mathcal{L}_{[0,T]}\otimes\omega^{(0)},\,(id,E^{\iota})\#\mathcal{L}_{[0,T]}\right)\\ &\leqslant \int_{s\in[0,T]} d_{\mathrm{MK}}\left(\omega_s,\omega_s^{(0)}\right)ds+\sum_{j=1}^{M} d_{\mathrm{MK}}\left(\mathcal{L}_{J_j}\otimes\omega_j^{(0)},\,(id,E^{\iota})\#\mathcal{L}_{J_j}\right)\\ &\leqslant \int_{s\in[0,T]} d_{\mathrm{MK}}\left(\omega_s,\omega_s^{(0)}\right)ds+T\sup_{j\in[\![1,M]\!]}\mathrm{diam}\,(J_j). \end{aligned}$$

Choosing first $\omega^{(0)}$, then the partitions $(J_j)_j$ fine enough to have both terms inferior to $\iota/2$, we conclude. □

We can now approximate solutions of the relaxed system by solutions of the original system.

Lemma 2.2.9 (Density of trajectories of the original system). *Consider the assumptions of Theorem 2.2.4. Let $y(\cdot)\in AC([0,T];\Omega)$ be a mutational solution of the relaxed dynamical system (2.16). For any $\varepsilon>0$, there exists a solution $z^{\varepsilon}\in AC([0,T];\Omega)$ of the original system (2.5) issued from y_0 and such that* $\sup_{t\in[0,T]} d(y_t,z_t^{\varepsilon})\leqslant\varepsilon$.

Proof. Let $N\in\mathbb{N}_*$ be large enough so that $T/N\leqslant\varepsilon$, and $t_n:=nT/N$ for $n\in[\![0,N]\!]$. Let B be a compact containing all solutions of the relaxed system (2.16) issued from y_0 up to time T. Denote $\mathrm{Lip}(F)_B$ a local Lipschitz constant of F on B, and C_B a bound on F on this set, that provides a common Lipschitz constant for all solutions. Reasoning as before, the union $A:=\bigcup_{x\in B}F(x)$ is a compact subset of $\mathcal{E}$.

Approximation by a chattering dynamic. By Remark 2.2.3, there exists a relaxed control $\omega\in L^1(0,T;\mathscr{P}_1(\mathcal{E}))$ such that $\mathcal{G}_{\mathcal{F}}(\mathrm{Bary}_{\mathbb{E}}(\omega_s))\in\mathring{y}_s\cap\mathcal{G}_{\mathcal{F}}(\overline{\mathrm{co}}\,F(y_s))$ for almost all $s\in[0,T]$. Up to a redefinition on a negligible set, ω_s is concentrated on A. Let $\iota>0$. Applying Lemma 2.2.8 on each interval $[t_n,t_{n+1}]$, there exists $E^{\iota}\in L^1(0,T;A)$ such that

$$\sup_{n\in[\![0,N-1]\!]} d_{\mathrm{MK}}\left(\mathcal{L}_{[t_n,t_{n+1}]}\otimes\omega,(id,E^{\iota})\#\mathcal{L}_{[t_n,t_{n+1}]}\right)\leqslant\iota. \tag{2.21}$$

By the well-posedness result of Proposition 2.1.6, there exists a unique absolutely continuous trajectory $(y_s^{\iota})_{s\in[0,T]}$ such that $\mathcal{G}_{\mathcal{F}}(E_s^{\iota})\in\mathring{y}_s^{\iota}$ for a.e. $s\in[0,T]$, and $y_0^{\iota}=x$. This trajectory is contained in the compact B and Lipschitz with a constant C_B. Moreover, by Lemma 2.1.15,

$$\frac{d^2(y^{\iota}_{t_{n+1}},y_{t_{n+1}})}{2}-\frac{d^2(y^{\iota}_{t_n},y_{t_n})}{2}\leqslant\int_{s=t_n}^{t_{n+1}}\left[D_{y_s}E_s^{\iota}\left(\overrightarrow{y_sy_s^{\iota}}\right)-D_{y_s}\mathrm{Bary}_{\mathbb{E}}(\omega_s)\left(\overrightarrow{y_sy_s^{\iota}}\right)\right]ds\qquad\forall n\in[\![0,N-1]\!].$$

One deduces the coarse estimate

$$d^2\left(y_n^{\iota},y_{t_n}\right)\leqslant 2N\sup_{n\in[\![0,N-1]\!]}\ \sup_{\substack{y'\in\mathrm{AC}([t_n,t_{n+1}];B)\\ \mathrm{Lip}(y')\leqslant C_B}}\left|\int_{s=t_n}^{t_{n+1}}\left[D_{y_s}E_s^{\iota}\left(\overrightarrow{y_sy_s'}\right)-D_{y_s}\mathrm{Bary}_{\mathbb{E}}(\omega_s)\left(\overrightarrow{y_sy_s'}\right)\right]ds\right|.$$

By Lemma 2.2.7, the last term goes to 0 when ι does. We can therefore choose $\iota=\iota_{\varepsilon}>0$ smaller than ε/N and small enough so that $d(y^{\iota}_{t_n},y_{t_n})\leqslant\varepsilon$ for all $n\in[\![1,N]\!]$. To lighten the presentation, we denote E^{ε} and y^{ε} the approximations constructed for this choice of ι, instead of $E^{\iota_{\varepsilon}}$ and $y^{\iota_{\varepsilon}}$.

Approximation by a trajectory of the original system. At this stage, y^{ε} has no reason to be a solution of the original system (2.5), since E_s^{ε} may not belong to $F(y_s^{\varepsilon})$. To recover a solution of (2.5), we apply the results of Frankowska and Lorenz [FL23]. Namely, let $\kappa>1$. By [FL23, Theorem 3.4], there exists $z^{\varepsilon}\in\mathrm{AC}([0,T];\Omega)$, issued from $y_0=y_0^{\varepsilon}$, satisfying $\mathring{z}_s^{\varepsilon}\cap\mathcal{G}_{\mathcal{F}}(F(z_s^{\varepsilon}))\neq\emptyset$ for a.e. $s\in[0,T]$, and the following estimate:

$$d(y^{\varepsilon}_{t_n},z^{\varepsilon}_{t_n})\leqslant\kappa e^{\kappa T\mathrm{Lip}(F)_B}\int_{s=0}^{t_n}\inf_{E''\in F(y_s^{\varepsilon})}\left|E_s^{\varepsilon}-E''\right|_{1,\infty}ds\qquad\forall n\in[\![1,N]\!].$$

Let α^{ε} be an optimal transport plan between $(id,E^{\varepsilon})\#\mathcal{L}_{[0,t_n]}$ and $\mathcal{L}_{[0,t_n]}\otimes\omega$. Recalling that $F(\cdot)$ is Lipschitz with constant $\mathrm{Lip}(F)_B$ over B, there holds

$$\begin{aligned}\int_{s=0}^{t_n}\inf_{E''\in F(y_s^{\varepsilon})}\left|E_s^{\varepsilon}-E''\right|_{1,\infty}ds &=\int_{((s,E),(s',E'))\in([0,t_n]\times\mathcal{E})^2}\inf_{E''\in F(y_s^{\varepsilon})}\left|E-E''\right|_{1,\infty}d\alpha^{\varepsilon}\\ &\leqslant\int_{((s,E),(s',E'))\in([0,t_n]\times\mathcal{E})^2}\left|E-E'\right|_{1,\infty}+\inf_{E''\in F(y_{s'})}\left|E'-E''\right|_{1,\infty}+\mathrm{Lip}(F)_B d(y_s^{\varepsilon},y_{s'})d\alpha^{\varepsilon}.\end{aligned}$$

Here, the term

$$\int_{((s,E),(s',E'))\in([0,t_n]\times\mathcal{E})^2} \inf_{E''\in F(y_{s'})} \left|E'-E''\right|_{1,\infty} d\alpha^\varepsilon = \int_{s'\in[0,t_n]} \int_{E'\in\mathcal{E}} \inf_{E''\in F(y_{s'})} \left|E'-E''\right|_{1,\infty} d\omega_{s'}(E') ds'$$

vanishes, since $\omega_{s'}$ is supported on $F(y_{s'})$ for a.e. $s' \in [0, t_n]$ by definition of a solution of the relaxed system. On the other hand, recall that $y(\cdot)$ is Lipschitz in time with constant C_B. Using the triangular inequality,

$$\begin{aligned}\int_{([0,t_n]\times\mathcal{E})^2} \left|E-E'\right|_{1,\infty} + \mathrm{Lip}(F)_B d(y^\varepsilon_s, y_{s'}) d\alpha^\varepsilon &\leqslant \int_{([0,t_n]\times\mathcal{E})^2} \left|E-E'\right|_{1,\infty} + \mathrm{Lip}(F)_B \left(d(y^\varepsilon_s, y_s) + C_B \left|s-s'\right|\right) d\alpha^\varepsilon \\ &\leqslant \max\left(1, \mathrm{Lip}(F)_B C_B\right) d_{\mathrm{MK}}\left((id, E^\varepsilon)\#\mathcal{L}_{[0,t_n]}, \mathcal{L}_{[0,t_n]}\otimes\omega\right) + T\mathrm{Lip}(F)_B \sup_{s\in[0,T]} d(y^\varepsilon_s, y_s).\end{aligned}$$

By the convexity of d_{MK} with respect to sums in the Banach space of measures, the first term in inferior to $\max\left(1, \mathrm{Lip}(F)_B C_B\right) n\iota$ as a consequence of (2.21), with $n\iota \leqslant \varepsilon$ by the choice of ι. Consequently,

$$d\left(y_{t_n}, z^\varepsilon_{t_n}\right) \leqslant d\left(y_{t_n}, y^\varepsilon_{t_n}\right) + d(y^\varepsilon_{t_n}, z^\varepsilon_{t_n}) \leqslant d\left(y_{t_n}, y^\varepsilon_{t_n}\right) + \kappa e^{\kappa T \mathrm{Lip}(F)_B}\left(\max\left(1, \mathrm{Lip}(F)_B C_B\right)\varepsilon + T\mathrm{Lip}(F)_B d_{\mathrm{AC}}(y^\varepsilon, y)\right).$$

As all the curves $y(\cdot)$, $y^\varepsilon(\cdot)$ and $z(\cdot)$ are Lipschitz with the same constant C_B, the estimates at times $(t_n)_{n\in[\![0,N]\!]}$ imply uniform estimates on $[0, T]$ with an additional error bounded by $C_B\varepsilon$. Sending ε to 0, we conclude. □

2.2.3 Right derivatives of the relaxed flow

In this section, we compute the right derivative of a solution of the relaxed dynamic system (2.16) under an additional regularity assumption on the energies. Recall that if y is a solution of the original system $\mathring{y}_s \cap \mathcal{G}_{\mathcal{F}}(F(y_s)) \neq \emptyset$, then for almost all s, the curve $h \mapsto y_{s+h}$ is approximated at order 1 by the gradient flow of $F(y_s)$ issued from y_s. As gradients flows in the sense of Proposition 1.1.10 admit a right derivative at each point, given by the metric gradient $\nabla_{y_s} E$, there holds that $y^+_s = \nabla_{y_s} F(y_s)$ whenever defined.

In the case where all the energies in $\mathcal{E}$ are Fréchet, we can give a similar expression for the relaxed system. One needs to take barycenters in tangent spaces, defined as follows.

Definition 2.2.10 (Barycenter [Stu03, Prop. 4.3 and 4.4]). *Let Ω be a complete CAT(0) space, and $\omega \in \mathscr{P}_1(\Omega)$. There exists a unique point $b \in \Omega$, called the barycenter of ω and denoted $Bary_\Omega(\omega)$, such that*

$$\int_{x\in\Omega} \left[d^2(z,x) - d^2(y,x)\right] d\omega(x) \geqslant \int_{x\in\Omega} \left[d^2(z,x) - d^2(y,x)\right] d\omega(x) \qquad \forall (y,z)\in\Omega^2. \tag{2.22}$$

Moreover, this point satisfies the variance inequality

$$\int_{x\in\Omega} \left[d^2(x,z) - d^2(x,b)\right] d\omega(x) \geqslant d^2(z,b) \qquad \forall z\in\Omega. \tag{2.23}$$

The barycenter needs not to be associative. As an example, consider the three-legged network with edges $[oa]$, $[ob]$ and $[oc]$ of length one, glued at o. Then $\frac{1}{2}a \oplus \frac{1}{2}b = \frac{1}{2}a \oplus \frac{1}{2}c = \frac{1}{2}b \oplus \frac{1}{2}c = o$. However,

$$\frac{1}{2}\left(\frac{1}{2}a \oplus \frac{1}{2}b\right) \oplus c = \frac{1}{2}o \oplus \frac{1}{2}c \neq o = \frac{1}{2}o \oplus \frac{1}{2}o = \frac{1}{2}\left(\frac{1}{2}a \oplus \frac{1}{2}c\right) \oplus \frac{1}{2}\left(\frac{1}{2}b \oplus \frac{1}{2}c\right).$$

We can anyway use it to give a meaning to the gradient of a convex combination. Recall that a map is Fréchet if $D_x E(v) = \langle \nabla_x E, v\rangle_x$ for all $(x, v) \in \mathrm{T}\Omega$.

Lemma 2.2.11 (Convex combination of Fréchet maps). *Let $\omega \in \mathscr{P}_1(\mathbb{E})$ be concentrated on a set of equiLipschitz, concave and Fréchet maps. The function $Bary_{\mathbb{E}}(\omega) : x \mapsto \int_{E\in\mathbb{E}} E(x) d\omega$ is Lipschitz and concave, so that it admits a gradient everywhere. Moreover, for any $x \in \Omega$,*

$$\nabla_x Bary_{\mathbb{E}}(\omega) = Bary_{\mathrm{T}_x\Omega}(\nabla_x \#\omega). \tag{2.24}$$

Proof. Let $x \in \Omega$ be fixed. The application $E \to \nabla_x E$ is 1-Lipschitz from $\mathbb{E}$ to $\mathrm{T}_x\Omega$ by (2.4), and the pushforward $\nabla_x \# \omega$ belongs to $\mathscr{P}_1(\mathrm{T}_x\Omega)$. The tangent cone $\mathrm{T}_x\Omega$ is itself a CAT(0) space (see [BH99, Theorem 3.19]), so both sides of (2.24) are defined. Denote $p := \mathrm{Bary}_{\mathrm{T}_x\Omega}(\nabla_x \# \omega) \in \mathrm{T}_x\Omega$. On the one hand, recalling that ω is supported on Fréchet maps, there holds for any $v \in \mathrm{T}_x\Omega$ that

$$D_x \mathrm{Bary}_{\mathbb{E}}(\omega)(v) = \int_{E \in \mathbb{E}} D_x E(v) d\omega(E) = \int_{E \in \mathbb{E}} \langle \nabla_x E, v \rangle_x \, d\omega(E) \leqslant \langle p, v \rangle_x$$

by the concavity of the metric scalar product and the Jensen inequality of [Stu03, Theorem 6.2]. On the other hand,

$$D_x \mathrm{Bary}_{\mathbb{E}}(\omega)(p) = \int_{E \in \mathbb{E}} \langle \nabla_x E, p \rangle_x \, d\omega(E) = \int_{E \in \mathbb{E}} \frac{1}{2} \left[|\nabla_x E|_x^2 + |p|_x^2 - d_x^2(\nabla_x E, p) \right] d\omega(E).$$

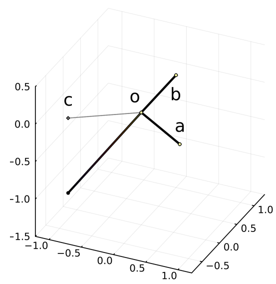
Using the variance inequality (2.23) in the CAT(0) space $\mathrm{T}_x\Omega$,

$$\int_{E \in \mathbb{E}} \left[|\nabla_x E|_x^2 - d_x^2(\nabla_x E, p) \right] d\omega(E) = \int_{v \in \mathrm{T}_x\Omega} \left[d_x^2(0_x, v) - d_x^2(v, p) \right] d\left[\nabla_x \# \omega\right](E) \geqslant d_x^2(0_x, p) = |p|_x^2.$$

Hence $D_x \mathrm{Bary}_{\mathbb{E}}(\omega)(p) \geqslant |p|_x^2$, and equality holds. By uniqueness of the gradient, $\nabla_x \mathrm{Bary}_{\mathbb{E}}(\omega) = p$. □

The assumption of Fréchet character of the elements of $\mathcal{E}$ is not restrictive, since the construction of Section 2.1.1.3 satisfies it. In Euclidean spaces, it reduces to the Fréchet-differentiability of the energies. However, as opposite to the latter case, it may not be preserved by convex combinations.

Indeed, consider again the tripod space $[oa] \cup [ob] \cup [oc]$ endowed with the shortest path distance, and let $d(o,a) = d(o,b) = d(o,c) = 1$. As seen from the explicit formulae (1.8), the applications $x \mapsto 1 - d(x,z)$ are Fréchet at all points but z. Consider $E_a := 1 - d(x,a)$ and $E_b = 1 - d(x,b)$, as well as $E := 1/2 E_a + 1/2 E_b$. Then E is constant on the branches $[oa]$ and $[ob]$, and decreases linearly from $E(o) = 0$ to $E(c) = -1$ over the branch $[oc]$. Consequently it is not Fréchet at o, since if it were the case, its differential at this point would either identically vanish, or change sign.



We deduce the following interpretation.

Proposition 2.2.12 (Equality in the tangent cone). *Assume that* **[A2.1.8]** *holds. Let $y \in AC([0,T];\Omega)$ be a solution of the relaxed ODE (2.16), associated to the relaxed control $\omega \in L^1(0,T;\mathscr{P}_1(\mathbb{E}))$. Then, for almost any time $t \in [0,T]$, the equality $y_t^+ = Bary_{\mathrm{T}\Omega}(\nabla_{y_t} \# \omega_t)$ holds in the tangent cone $\mathrm{T}_{y_t}\Omega$.*

Proof. The curve y is locally Lipschitz-continuous by **[A2.1.8]**, thus is differentiable almost everywhere by the Rademacher theorem in CAT(0) spaces [AKP23, Proposition 13.9]. By definition, for almost any t, it is approximated at order 1 by the gradient flow of $E_t := \mathrm{Bary}_{\mathbb{E}}(\omega_t)$. The set of times t at which both properties hold is still of full measure in $[0,T]$, and on any such t, the right derivative of y coincides with that of the gradient flow of E_t issued from y_t at $h = 0$. By Proposition 1.1.10, the latter is equal to the metric gradient of E_t, and by Lemma 2.2.11, the desired equality holds. □

2.3 Hamilton-Jacobi-Bellman approach

We can now introduce the Mayer control problem that motivates this study. Let $\mathfrak{J} : \Omega \to \mathbb{R}$ be a given terminal cost, U a set of controls, and $f : \Omega \times U \to \mathcal{E}$ a controlled dynamic. The original problem writes

$$\underset{u(\cdot) \in L^1(t,T;U)}{\text{Minimize}} \quad \mathfrak{J}(y_T^{t,x,u}), \quad \text{where } y_t^{t,x,u} = x \text{ and } \mathring{y}_s \ni f(y_s, u(s)). \tag{2.25}$$

Denote again $\overline{\mathrm{co}}\, f$ the relaxed dynamic given by Definition 2.2.1. Using the parametrization by controls, there holds

$$\overline{\mathrm{conv}}\, f(x,U) = \left\{ \mathrm{Bary}_{\mathbb{E}}(\omega) \mid \omega \in \mathscr{P}_1(f(x,U)) \right\} = \left\{ \mathrm{Bary}_{\mathbb{E}}\left(f(x,\cdot) \# \omega\right) \mid \omega \in \mathscr{P}_1(U) \right\},$$

so that we may define *relaxed controls* as measures in $\mathscr{P}_1(U)$, and the *relaxed controlled dynamic* by

$$\overline{\mathrm{co}}\, f : \Omega \times \mathscr{P}_1(U) \to \overline{\mathrm{conv}}\,\mathcal{E}, \qquad \overline{\mathrm{co}}\, f(x,\omega) := \mathrm{Bary}_{\mathbb{E}}\left(f(x,\cdot)\#\omega\right) = \int_{u\in U} f(x,u)\,d\omega(u).$$

The associated control problem writes

$$\underset{\omega(\cdot)\in L^1(t,T;\mathscr{P}_1(U))}{\text{Minimize}} \quad \mathfrak{J}(y_T^{t,x,\omega}), \quad \text{where } y_t^{t,x,\omega} = x \text{ and } \mathring{y}_s \ni \overline{\mathrm{co}}\, f(y_s,\omega(s)). \tag{2.26}$$

We first obtain the existence of an optimal control by applying the results of the previous section, then turn to the link with Hamilton-Jacobi-Bellman equations.

2.3.1 Control problem

Let $\mathcal{E}$ be an admissible set of energies in the sense of Definition 2.1.1.

Assumption [A2.3.1] (Structure of the controlled dynamic). Let (U, d_U) be a compact metric space, and $f : \Omega \times U \to \mathcal{E}$ be the controlled dynamic. It is assumed that

(a) for each fixed $x \in \Omega$, the application $u \mapsto f(x,u)$ is continuous from (U, d_U) to $(\mathcal{E}, |\cdot|_{1,\infty})$,

(b) for each compact $B \subset \Omega$, there exists $\mathrm{Lip}(f)_B \geqslant 0$ such that for any $(x,y) \in B^2$ and $u \in U$,

$$\left|f(x,u) - f(y,u)\right|_{1,\infty} \leqslant \mathrm{Lip}(f)_B\, d(x,y),$$

(c) there exists $C_f \geqslant 0$ and a point $o \in \Omega$ such that $\sup_{u\in U} \left|f(x,u)\right|_{1,\infty} \leqslant C_f(1 + d(x,o))$ for all $x \in \Omega$.

Under **[A2.3.1]**, the set-valued dynamic $x \mapsto f(x,U)$ satisfies the assumption **[A2.1.8]**.

Proposition 2.3.2 (The problem with relaxed dynamic is the relaxed problem). *Assume that $\mathfrak{J} : \Omega \to \mathbb{R}$ is continuous, and that the controlled dynamic f satisfies **[A2.3.1]**. Then (2.26) is the relaxation of (2.25), in the sense that*

$$\inf_{u(\cdot)\in L^1(t,T;U)} \mathfrak{J}(y_T^{t,x,u}) = \min_{\omega(\cdot)\in L^1(t,T;\mathscr{P}_1(U))} \mathfrak{J}(y_T^{t,x,\omega}). \tag{2.27}$$

In the case where the sets $f(\cdot,U)$ are convex, the relaxed dynamic coincides with f, and Proposition 2.3.2 implies the existence of an optimal control. Otherwise, the optimal control may be seen as a Young measure [War72, III.3].

Proof. By Theorem 2.2.4, the set of trajectories of the relaxed control system is the closure in $\mathrm{AC}([0,T];\Omega)$ of the set of trajectories of the original system, and is compact. Consequently, for $(t,x) \in [0,T] \times \Omega$ fixed,

$$\left\{y_T^{t,x,\omega} \;\middle|\; \omega(\cdot) \in L^1(t,T;\mathscr{P}_1(U))\right\} \quad \text{is the closure in } \Omega \text{ of} \quad \left\{y_T^{t,x,u} \;\middle|\; \omega(\cdot) \in L^1(t,T;U)\right\},$$

and the latter set is compact (since $y \mapsto y_T$ is continuous from $\mathrm{AC}([0,T];\Omega)$ to Ω). As $\mathfrak{J}$ is continuous, (2.27) holds. □

To go on, we introduce the *value function* of the control problem (2.25), as

$$V : [0,T] \times \Omega \to \mathbb{R}, \qquad V(t,x) := \inf_{u(\cdot)\in L^1(t,T;U)} \mathfrak{J}(y_T^{t,x,u}) = \min_{\omega(\cdot)\in L^1(t,T;\mathscr{P}_1(U))} \mathfrak{J}(y_T^{t,x,\omega}). \tag{2.28}$$

Here the second equality stands by Proposition 2.3.2. The value function has the interesting property to be constant along optimal trajectories, as a consequence of the Dynamic Programming Principle

$$V(t,x) = \inf_{u(\cdot)\in L^1(t,t+h;U)} V(t+h, y_{t+h}^{t,x,u}) = \min_{\omega(\cdot)\in L^1(t,t+h;\mathscr{P}_1(U))} V(t+h, y_{t+h}^{t,x,\omega}) \tag{2.29}$$

for any $(t,x) \in [0,T) \times \Omega$ and $h \in (0, T-t]$. The equality (2.29) holds in our setting, since trajectories on $[t,T]$ are concatenations of trajectories on $[t,t+h]$ and $[t+h,T]$. Moreover, using the estimates of [FL23] with Cauchy-Lipschitz assumptions, one easily proves that the set-valued map $(t,x) \mapsto \left\{y^{t,x,\omega} \;\middle|\; \omega \in L^1(t,T;\mathscr{P}_1(U))\right\}$ is locally Lipschitz in the Hausdorff distance over Ω. As a consequence, if $\mathfrak{J}$ is locally Lipschitz, then V is locally Lipschitz as well. The proofs of these assertions follow verbatim the proofs of the Euclidean case, and we refer the reader to [Bar94, Section 3.1.3] or [BC97, Section 2.1].

2.3.2 The HJB equation

Denote

$$\mathbb{T}_x := \left\{p : \mathrm{T}_x\Omega \to \mathbb{R} \;\middle|\; p(\cdot) \text{ is positively homogeneous and Lipschitz w.r.t. } d_x(\cdot,\cdot)\right\},$$

and $\mathbb{T} := \bigcup_{x\in\Omega}\{x\} \times \mathbb{T}_x$ the *metric cotangent bundle*. Define the Hamiltonian of the control problem (2.25) as

$$H : \mathbb{T} \to \mathbb{R}, \qquad H(x,p) := \sup_{u\in U}\left[-p\left(\nabla_x f(x,u)(x)\right)\right]. \tag{2.30}$$

Here we recall that for any $(x,u) \in \Omega \times U$, the term $f(x,u)$ is a concave and Lipschitz function from Ω to $\mathbb{R}$, and $\nabla_x f(x,u)(x)$ is the metric gradient of $y \mapsto f(x,u)(y)$ at the point x, defined as an element of $\mathrm{T}_x\Omega$. Our aim is to give a meaning to the following Hamilton-Jacobi-Bellman equation:

$$-\partial_t v(t,x) + H(x, D_x v(t,x)) = 0 \quad (t,x) \in [0,T) \times \Omega, \qquad v(T,\cdot) = \mathfrak{J}. \tag{2.31}$$

2.3.2.1 Viscosity solutions

Following [JZ23b], we introduce viscosity solutions using test functions that are semiconcave/semiconvex. More precisely, consider the sets of test functions

$$\mathscr{T}_\pm := \left\{ \varphi : (0,T) \times \Omega \to \mathbb{R} \;\middle|\; \varphi(t,x) = \psi(t) \pm g(x) \text{ with } \begin{array}{c} \psi \in \mathcal{C}^2((0,T);\mathbb{R}), \\ g \text{ locally Lipschitz and semiconvex} \end{array} \right\}.$$

Any $\varphi \in \mathscr{T}_\pm$ admits a differential $D_x\varphi \in \mathbb{T}_x$ at all point $x \in \Omega$ in the sense of Definition 1.1.8. This allows to define a viscosity solution to (2.31) as follows.

Definition 2.3.3 (Viscosity solution). *An application $v : [0,T] \times \Omega \to \mathbb{R}$ is a viscosity*

- *subsolution of (2.31) if it is upper semicontinuous, satisfies $v(T,\cdot) \leq \mathfrak{J}$ and if for any $\varphi \in \mathscr{T}_+$ such that $v - \varphi$ reaches a maximum in $(t,x) \in (0,T) \times \Omega$, there holds*

$$-\partial_t\varphi(t,x) + H(x, D_x\varphi(t,x)) \leq 0. \tag{2.32}$$

- *supersolution of (2.31) if it is lower semicontinuous, satisfies $v(T,\cdot) \geq \mathfrak{J}$ and if for any $\varphi \in \mathscr{T}_-$ such that $v - \varphi$ reaches a minimum in $(t,x) \in (0,T) \times \Omega$, there holds*

$$-\partial_t\varphi(t,x) + H(x, D_x\varphi(t,x)) \geq 0. \tag{2.33}$$

- *solution of (2.31) if it is both a sub and a supersolution.*

This definition appears in [Jer22; JZ23b], where the authors provide comparison and stability results.

2.3.2.2 Characterization of the value function

In this section, we prove that the value function is the unique viscosity solution of the HJB equation. The proof that V is a subsolution is completely analogous to the classical setting, but for the reader's convenience, we give the details below.

Lemma 2.3.4 (The value function is a subsolution of (2.31)). *Assume that $\mathfrak{J} : \Omega \to \mathbb{R}$ is locally Lipschitz, and that $f : \Omega \times U \to \mathcal{E}$ satisfies **[A2.3.1]**. Then the value function V given by (2.28) is a viscosity subsolution of (2.31) in the sense of Definition 2.3.3.*

Proof. Under the given assumptions, V is locally Lipschitz, and satisfies the terminal condition by definition. Let $\varphi \in \mathscr{T}_+$ such that $V - \varphi$ reaches a maximum at (t,x). For any $u \in U$, the unique solution

$y \in \mathrm{AC}([t,T];\Omega)$ issued from x of the ODE $\mathring{y}_s \ni f(y_s,u)$ is mutable at $s=t$, and admits $\nabla_x f(x,u)$ as a right derivative; indeed, it is the case for the gradient curve $z(\cdot)$ of $\overline{\mathrm{co}}\, f(x,u)$ issued from x, and by Lemma 2.2.6,

$$d^2(y_t,z_t) \leqslant \int_{s=0}^{t} \left[D_{z_s} f(y_s,u)\left(\overrightarrow{z_s y_s}\right) - D_{z_s} f(z_s,u)\left(\overrightarrow{z_s y_s}\right)\right] ds \leqslant \mathrm{Lip}(f) \int_{s=0}^{t} d^2(y_s,z_s) ds.$$

By a Grönwall Lemma, $d^2(y_t,z_t) = o(t)$. Owing to the DPP (2.29), for any $h>0$ such that $t+h \leqslant T$, one has $V(t,x) \leqslant V(t+h,y_{t+h})$. Consequently,

$$V(t+h,y_{t+h}) - \varphi(t+h,y_{t+h}) \leqslant V(t,x) - \varphi(t,x) \;\Rightarrow\; -\frac{\varphi(t+h,y_{t+h}) - \varphi(t,x)}{h} \leqslant -\frac{V(t+h,y_{t+h}) - V(t,x)}{h} \leqslant 0.$$

Recalling that test functions are directionally differentiable everywhere, we may let $h \searrow 0$ and find that

$$-\partial_t \psi(t) - D_x g\left(\nabla_x f(x,u)\right) \leqslant 0.$$

Taking the supremum over $u \in U$ yields the subsolution inequality (2.32). □

The supersolution side is slightly more involved. Here one should use the dynamical programming principle along an optimal solution, which may not satisfy the dynamical system if the latter does not have convex images. Hence, one argues with $\varepsilon-$optimal curves instead. However, to complete the argument, one needs to take a limit in an expression of the form $(t,x) \mapsto H(x, D_x\varphi(t,x))$, where $\varphi \in \mathscr{T}_-$ is a test function. In $\mathbb{R}^d$, test functions can be taken $\mathcal{C}^1$, and this is not a problem. In our setting, we need to rely on the semiconcavity to get the required semicontinuity. This is where the assumption **[A2.1.3]** intervenes, to turn semiconcavity along geodesics into a "sufficient semiconcavity" along the flows of the energies. More precisely, we can show the following preparatory lemmata.

Lemma 2.3.5 (Upper semicontinuity of the differential of semiconvex functions). *Assume that $\mathcal{E}$ satisfies **[A2.1.3]**. Let $(y_i)_{i\in\mathbb{N}} \xrightarrow[i\to\infty]{d(\cdot,\cdot)} y \in \Omega$ and $E_i \xrightarrow[i\to\infty]{|\cdot|_{1,\infty}} E \in \mathcal{E}$. For any locally Lipschitz $\varphi:\Omega\to\mathbb{R}$ that is $\lambda-$semiconvex for some $\lambda \in \mathbb{R}$, there holds*

$$\limsup_{i\to\infty} D_{y_i}\varphi\left(\nabla_{y_i} E_i\right) \leqslant D_y\varphi(\nabla_y E).$$

Proof. The argument is based on the classical proof of [Roc70, Theorem 24.5]. To lighten the notation, denote $z_i(s) := \mathcal{G}_{\mathcal{F}}(E_i)(s,y_i)$, and $z(s) := \mathcal{G}_{\mathcal{F}}(E)(s,y)$. As the sets $\{y_i\}_{i\in\mathbb{N}}$ and $\{E_i\}_{i\in\mathbb{N}}$ are relatively compact (Ω,d) and $(\mathcal{E},|\cdot|_{1,\infty})$, we may use assumption **[A2.1.3]** and approximate each curve $z_i(\cdot)$ at order one by some reparametrized geodesic $\gamma_i = \gamma_{E_i,y_i}$ defined on some interval $[0,t]$ with t independent of i. Let $\varepsilon>0$, and consider $0 < h^\varepsilon \leqslant t$ small enough so that

$$\frac{\varphi(z(h^\varepsilon)) - \varphi(y)}{h^\varepsilon} \leqslant D_y\varphi(\nabla_y E) + \varepsilon.$$

By [Aub99, Proposition 1.3.2] applied with $\vartheta(t) \equiv \mathcal{G}_{\mathcal{F}}(E_i)$ and $\tau(t) \equiv \mathcal{G}_{\mathcal{F}}(E)$, noticing that the contraction constant $\alpha = 0$ in our case, there holds

$$d(z_i(h^\varepsilon), z(h^\varepsilon)) \leqslant d(y_i,y) + h^\varepsilon \mathbb{D}(\mathcal{G}_{\mathcal{F}}(E_i), \mathcal{G}_{\mathcal{F}}(E)) \leqslant d(y_i,y) + h^\varepsilon\, |E_i - E|_{1,\infty} \xrightarrow[i\to\infty]{} 0.$$

Denote $\mathrm{Lip}(\varphi)_{\mathrm{loc}}$ a local Lipschitz constant of φ in a compact containing all $\overline{\mathscr{B}}(y_i, h^\varepsilon \max_i |E_i|_{1,\infty})$. Consider i_ε large enough so that for all $i \geqslant i_\varepsilon$,

$$\left|\varphi(y_i) - \varphi(y)\right| \leqslant \varepsilon h^\varepsilon, \qquad \text{and} \qquad \left|\varphi(z_i(h^\varepsilon)) - \varphi\left(z(h^\varepsilon)\right)\right| \leqslant \varepsilon h^\varepsilon.$$

Then for $i \geqslant i_\varepsilon$,

$$\begin{aligned} D_y\varphi(\nabla_y E) + \varepsilon &\geqslant \frac{\varphi(z(h^\varepsilon)) - \varphi(y)}{h^\varepsilon} \geqslant \frac{\varphi(z_i(h^\varepsilon)) - \varphi(y_i)}{h^\varepsilon} - 2\varepsilon \\ &\geqslant \frac{\varphi\left(\gamma_i(h^\varepsilon)\right) - \varphi(y_i)}{h^\varepsilon} - \mathrm{Lip}(\varphi)_{\mathrm{loc}} \frac{d\left(\gamma_i(h^\varepsilon), z_i(h^\varepsilon)\right)}{h^\varepsilon} - 2\varepsilon \\ &\geqslant D_{y_i}\varphi(\nabla_{y_i} E_i) - \frac{|\lambda|}{2} h^\varepsilon \left|\gamma_i^+\right|_{y_i}^2 - \mathrm{Lip}(\varphi)_{\mathrm{loc}} \frac{d\left(\gamma_i(h^\varepsilon), z_i(h^\varepsilon)\right)}{h^\varepsilon} - 2\varepsilon. \end{aligned}$$

Here the last inequality stands since φ is semiconvex. As $\left|\gamma_i^+\right|_x = \left|\nabla_{y_i} E_i\right|_x \leqslant |E_i|_{1,\infty}$ converges towards $|E|_{1,\infty}$ when i goes to ∞, we may pass to the limit sup in $i \to \infty$ and obtain

$$D_y\varphi(\nabla_y E) + \varepsilon \geqslant \limsup_{i\to\infty} D_{y_i}\varphi(\nabla_{y_i} E_i) - \frac{|\lambda|}{2} h^\varepsilon |E|_{1,\infty}^2 - \mathrm{Lip}(\varphi)_{\mathrm{loc}} \sup_{i\in\mathbb{N}} \frac{d\left(\gamma_i(h^\varepsilon), z_i(h^\varepsilon)\right)}{h^\varepsilon} - 2\varepsilon.$$

We may choose h^ε as a decreasing function of ε, so that sending ε to 0 and using **[A2.1.3]**, we conclude. □

As a consequence, the Hamiltonian is semicontinuous in the following way.

Lemma 2.3.6 (Semicontinuities of the Hamiltonian). *Assume* ***[A2.1.3]*** *and* ***[A2.3.1]*** *hold. Then*

- *if $\varphi : \Omega \to \mathbb{R}$ is locally Lipschitz and semiconvex, then $x \mapsto H(x, D_x\varphi)$ is lower semicontinuous.*
- *if $\varphi : \Omega \to \mathbb{R}$ is locally Lipschitz and semiconcave, then $x \mapsto H(x, D_x\varphi)$ is upper semicontinuous.*

Proof. Assume first that φ is semiconvex. Then the application $(x, E) \mapsto D_x\varphi(\nabla_x E)$ is upper semicontinuous by Lemma 2.3.5, so that $(x, u) \mapsto -D_x\varphi(\nabla_x f(x, u))$ is lower semicontinuous, and so is the supremum over $u \in U$. On the other hand, if φ is semiconcave, let $x_n \to_n x \in \Omega$. Up to extraction, we may assume that

$$\limsup_{n\to\infty} H(x_n, D_{x_n}\varphi) = \lim_{n\to\infty} H(x_n, D_{x_n}\varphi) = \lim_{n\to\infty} \sup_{u\in U} -D_{x_n}\varphi(\nabla_{x_n} f(x_n, u)).$$

The composition $u \mapsto D_{x_n}\varphi\left(\nabla_{x_n} f(x_n, u)\right)$ being Lipschitz, its supremum over the compact U is reached at some point $u_n \in U$. Up to further extraction, we may assume that $(u_n)_{n\in\mathbb{N}}$ converges to some $\overline{u} \in U$. By continuity, $\left|f(x_n, u_n) - f(x, \overline{u})\right|_{1,\infty}$ goes to 0 when n goes to ∞, so that Lemma 2.3.5 applied to the function $-\varphi$ yields

$$\limsup_{n\to\infty} -D_{x_n}\varphi(\nabla_{x_n} f(x_n, u_n)) \leqslant -D_x\varphi(\nabla_x f(x, \overline{u})).$$

Hence

$$\limsup_{n\to\infty} H(x_n, D_{x_n}\varphi) = \lim_{n\to\infty} -D_{x_n}\varphi(\nabla_{x_n} f(x_n, u_n)) \leqslant -D_x\varphi(\nabla_x f(x, \overline{u})) \leqslant \sup_{u\in U} -D_x\varphi(\nabla_x f(x, u)) = H\left(x, D_x\varphi\right),$$

and the desired semicontinuity holds. □

We can now prove the supersolution property.

Lemma 2.3.7 (The value function is a supersolution). *Assume that* ***[A2.1.3]*** *and* ***[A2.3.1]*** *hold, and that $\mathfrak{J} : \Omega \to \mathbb{R}$ is locally Lipschitz. The value function V given by (2.28) is a viscosity supersolution of (2.31) in the sense of Definition 2.3.3.*

Proof. We just have to show the viscosity inequality (2.33). Let $\varphi \in \mathscr{T}_-$ such that $V - \varphi$ reaches a minimum at point $(t, x) \in (0, T) \times \Omega$. For each $\varepsilon > 0$ and $h \in]0, t - T]$, let $y^{\varepsilon,h} : [t, t+h] \mapsto \Omega$ be such that

$$V(t, x) \geqslant V(t+h, y_h^{\varepsilon,h}) - \varepsilon h,$$

and denote $u^{\varepsilon,h} \in L^1(0, h; U)$ a control driving it. There holds

$$V\left(t+h, y_{t+h}^{\varepsilon,h}\right) - \varphi\left(t+h, y_{t+h}^{\varepsilon,h}\right) \geqslant V(t, x) - \varphi(t, x) \quad \Longrightarrow \quad -\frac{\varphi(t+h, y_{t+h}^{\varepsilon,h}) - \varphi(t, x)}{h} \geqslant -\varepsilon.$$

The trajectory $y^{\varepsilon,h}$ is mutable almost everywhere, so that using the separation $\varphi(s, y) = \psi(s) + g(y)$ for some $\psi \in \mathcal{C}^2$ and g semiconcave,

$$-\varepsilon \leqslant -\frac{\varphi(t+h, y_{t+h}^{\varepsilon,h}) - \varphi(t, x)}{h} = -\frac{\psi(t+h) - \psi(t)}{h} - \frac{1}{h}\int_{s=t}^{t+h} D_{y_s^{\varepsilon,h}} g\left(\nabla_{y_s^{\varepsilon,h}} f(y_s^{\varepsilon,h}, u^{\varepsilon,h}(s))\right) ds.$$

Let C_f be a bound on the dynamic f in a sufficiently large ball around x. One has $d\left(x, y_s^{\varepsilon,h}\right) \leqslant hC_f$ for all $s \in [0, h]$, hence

$$-\varepsilon \leqslant -\partial_t\psi + O(h) + \frac{1}{h}\int_{s=t}^{t+h} \sup_{u\in U}\left[-D_{y_s^{\varepsilon,h}} g\left(\nabla_{y_s^{\varepsilon,h}} f(y_s^{\varepsilon,h}, u)\right)\right] ds \leqslant -\partial_t\psi + O(h) + \sup_{z\in\overline{\mathscr{B}}(x, hB)} H\left(z, D_z g\right).$$

Since g is semiconcave, by Lemma 2.3.6, the map $z \mapsto H\big(z, D_z g\big)$ is upper semicontinuous. Hence, taking the limit in $h \searrow 0$, we have that

$$-\varepsilon \leqslant -\partial_t \psi(t) + H\big(x, D_x g\big) = -\partial_t \varphi(t,x) + H\big(x, D_x \varphi(t,x)\big).$$

Letting $\varepsilon \searrow 0$, we conclude to the supersolution property. □

Combining Lemmata 2.3.4 and 2.3.7, we get to the main result of this section.

Proposition 2.3.8 (Characterization of the value function). *Assume that $\mathfrak{J} : \Omega \to \mathbb{R}$ is Lipschitz and bounded, $f : \Omega \times U \to \mathcal{E}$ is globally Lipschitz in both variables and globally bounded, and that $\mathcal{E}$ satisfies **[A2.1.3]**. Then the value function is the unique viscosity solution of (2.31) in the sense of Definition 2.3.3.*

Proof. The assumptions on f imply that **[A2.3.1]** holds, so that V is a viscosity solution by Lemmata 2.3.4 and 2.3.7. To obtain uniqueness, we apply the comparison results of [JZ23b]. As f is bounded, there holds

$$\begin{aligned} H(x,p) - H(x,q) &\leqslant \sup_{u\in U} -p(\nabla_x f(x,u)) + q(\nabla_x f(x,u)) \leqslant \sup_{u\in U} \big|\nabla_x f(x,u)\big|_x \sup_{v \in \mathrm{T}_x\Omega, |v|_x = 1} \big|q(v) - p(v)\big| \\ &\leqslant \|f\|_\infty \sup_{v \in \mathrm{T}_x\Omega, |v|_x=1} \big|q(v) - p(v)\big|. \end{aligned}$$

In the same way, using the explicit expression of the derivative of the squared distance,

$$\begin{aligned} H\big(x, -\alpha D_x d^2(\cdot, y)\big) - H\big(y, \alpha D_y d^2(x,\cdot)\big) &\leqslant \sup_{u\in U} \alpha D_x d^2(x,y)(\nabla_x f(x,u)) + \alpha D_y d^2(x,y)(\nabla_y f(y,u)) \\ &= 2\alpha \sup_{u\in U} -\langle \nabla_x f(x,u), \overrightarrow{xy}\rangle_x - \langle \nabla_y f(y,u), \overrightarrow{yx}\rangle_y. \end{aligned}$$

Using that

$$\Big|\langle \nabla_y f(y,u), \overrightarrow{yx}\rangle_y - \langle \nabla_y f(x,u), \overrightarrow{yx}\rangle_y\Big| \leqslant d_x\big(\nabla_y f(y,u), \nabla_y f(x,u)\big)\,\big|\overrightarrow{yx}\big|_y \leqslant \mathrm{Lip}(f) d^2(x,y),$$

we get that

$$H\big(x, -\alpha D_x d^2(\cdot,y)\big) - H\big(y, \alpha D_y d^2(x,\cdot)\big) \leqslant 2\alpha \sup_{u\in U} \Psi(x,y,u) + 2\alpha \mathrm{Lip}(f) d^2(x,y),$$

where $\Psi(x,y,u) := -\langle \nabla_x f(x,u), \overrightarrow{xy}\rangle_x - \langle \nabla_y f(x,u), \overrightarrow{yx}\rangle_y$. However, the function $z \mapsto f(x,u)(z)$ is concave, so that $\Psi(x,y,u) \leqslant 0$ (by usual directional derivative arguments, or [AKP23, Proposition 13.24]). By [JZ23b, Theorem 4.3], the viscosity solution is unique, and must be V. □

To conclude, we discuss the HJB equation associated to the convexified dynamics.

Remark 2.3.9. *If **[A2.1.8]** holds, then the value function V defined in (2.28) is also a viscosity solution of the HJB equation associated to the relaxed Hamiltonian*

$$H^{relax}(x,p) := \sup_{\omega \in \mathscr{P}_1(U)} -p\big(\nabla_x \overline{co} f(x,\omega)(x)\big). \tag{2.34}$$

Indeed, by Proposition 2.3.2, V coincides with the value function of the control problem with relaxed dynamics f. The latter inherits its Lipschitz-continuity, boundedness and compact images from f, so that by Lemma 2.3.4, it is a subsolution of the HJB equation with H^{relax}. Moreover, if $\varphi \in \mathscr{T}_-$ is such that $V - \varphi$ reaches a minimum at (t,x), then

$$-\partial_t \varphi(t,x) + H^{relax}(x, D_x\varphi(t,x)) \geqslant -\partial_t\varphi(t,x) + H(x, D_x\varphi(t,x)) \geqslant 0,$$

by definition of H and H^{relax}, and V is a viscosity supersolution as well.

Lemma 2.3.7 does not apply directly in the above argument, since it may happen that $\mathcal{E}$ satisfies the assumption **[A2.1.3]**, but not $\overline{\mathrm{conv}}\,\mathcal{E}$.

2.3.3 Numerical approximation

In this section, we provide numerical schemes to approximate the value function V, and an optimal trajectory of the optimal control problem (2.25). For simplicity, the domain Ω is by now assumed to be compact, and the dynamic $f : \Omega \times U \to \mathcal{E}$ is assumed to satisfy **[A2.1.8]**. Consequently, f will be bounded w.r.t. $|\cdot|_{1,\infty}$ by a constant denoted $\|f\|_\infty$. Numerical illustrations are provided in Section 2.4.

2.3.3.1 Approximation of the reachable sets

The first aim is to provide a numerical approximation of the reachable sets of a controlled mutational ODE $\mathring{y}_s \ni \mathcal{G}_{\mathcal{F}}(f(y_s, u(s)))$, that is, the sets

$$R[f]_s^{t,x} := \left\{ y_s^{t,x,u} \;\middle|\; y_t^{t,x,u} = x,\ \mathring{y}_s^{t,x,u} \ni \mathcal{G}_{\mathcal{F}}(f(y_s^{t,x,u}, u(s))),\ u \in L^1(t,s;U) \right\}, \qquad 0 \leqslant t \leqslant s \leqslant T,\ x \in \Omega.$$

We start by restricting to a constant dynamic. If $u(\cdot) \in L^1(t,s;U)$ and $x \in \Omega$ are given, denote $z(\cdot)$ the solution of the mutational ODE $\mathring{z}_s \ni f(x, u(s))$ with a dynamic being constant in space. Then by [Aub99, Proposition 1.3.2],

$$\begin{aligned} d\left(y_s^{t,x,u}, z(t-s)\right) &\leqslant 0 + \int_{\tau=t}^{s} \left| f(y_\tau^{t,x,u}, u(\tau)) - f(x, u(\tau)) \right|_{1,\infty} d\tau \\ &\leqslant \int_{\tau=t}^{s} \mathrm{Lip}(f) d\left(y_\tau^{t,x,u}, x\right) d\tau \leqslant \frac{\mathrm{Lip}(f)\|f\|_\infty (t-s)^2}{2}. \end{aligned} \tag{2.35}$$

One directly deduces that the reachable set $R[f]_s^{t,x}$ is approximated by the trajectories of the constant dynamic $(y,u) \mapsto f_x(y,u) := f(x,u)$ with

$$d_H\left(R[f]_s^{t,x}, R[f_x]_s^{t,x}\right) \leqslant \frac{\mathrm{Lip}(f)\|f\|_\infty (t-s)^2}{2},$$

where d_H denotes the Hausdorff distance.

In some cases, one may have an exact formula for the solution of the mutational ODE driven by the constant dynamic f_x. For instance, in $\Omega = \mathbb{R}^d$, it is known that $R[f_x]_s^{t,x}$ coincides with the convex hull of the points

$$\left\{ \mathcal{G}_{\mathcal{F}}(f(x,u))(s-t, x) = x + (s-t) v_{x,u} \;\middle|\; u \in U, \text{ with } v_{x,u} = \nabla_x f(x,u) \right\}.$$

In the case where such a representation is not available, it is still possible to use simple Euler schemes, as we now detail.

Algorithm 2: Approximation of the reachable sets by an Euler scheme

---

1 Let $\varepsilon > 0$, $N \in \mathbb{N}_*$, $x \in \Omega$, $0 \leqslant t < s$ be given.
2 Denote $\Delta t := (t-s)/N$.
3 Let $\hat{U} \subset U$ be a finite subset such that $U \subset \bigcup_{\hat{u} \in \hat{U}} \overline{\mathscr{B}}(\hat{u}, \varepsilon)$.
4 Set $\hat{R}_0 := \{x\}$.
5 **for** $n \in [\![0, N-1]\!]$ **do**
6   Set $\hat{R}_{n+1} := \left\{ \mathcal{G}_{\mathcal{F}}(f(x,\hat{v}))(\Delta t, \hat{y}) \;\middle|\; \hat{y} \in \hat{R}_n,\ \hat{v} \in \hat{U} \right\}$.
7 Return $\hat{R}_N$.

Lemma 2.3.10 (Convergence of Algorithm 2). *Let $x \in \Omega$ and $0 \leqslant t < s \leqslant T$. Then $d_H\left(R[f]_s^{t,x}, \hat{R}_N\right)$ goes to 0 when N goes to infinity and ε to 0.*

Proof. Let $t_n := t + n\Delta t$ for $n \in [\![0, N]\!]$. Denote $L^1_N(t,s;\hat{U})$ the subset of controls that are piecewise constant over each $[t_n, t_{n+1})$ for $n \in [\![0, N-1]\!]$. By the density of simple functions in L^1, one has that

$$\Delta_{[t,s]}(N,\varepsilon) := \sup_{u(\cdot) \in L^1(t,s;U)} \inf_{v(\cdot) \in L^1_N(t,s;\hat{U})} \int_{\tau=s}^{t} d_U(u(\tau), v(\tau)) d\tau \underset{N \to \infty, \varepsilon \searrow 0}{\longrightarrow} 0.$$

Since the application $u \mapsto (y_s^{0,x,u})_s$ is Lipschitz-continuous by Proposition 2.1.9, the reachable set $R[f_x]_s^{t,x}$ is approximated by the set of trajectories induced by constant controls valued in $\hat{U}$ with an error proportional to $\Delta_{[0,\Delta t]}(N)$. As the application $(y,v)\mapsto f_x(y,v)=f(x,v)$ does not depend on the space variable, the set $\hat{R}_N$ computed by Algorithm 2 is exactly the set of terminal points of these trajectories, completing the proof. □

Remark 2.3.11 (Sub-optimality)**.** *In $\Omega=\mathbb{R}^d$, one can obtain explicit bounds on the error committed by the Euler scheme. Indeed, in this case, one can approximate a given trajectory of the controlled system using a* constant *relaxed control in a clever way. More precisely, denoting $F:\Omega\to \mathrm{T}\Omega$ a classical dynamic, if*

$$y_{\Delta t}=x+\int_{s=0}^{\Delta t}F(y_s,u(s))ds,$$

let $\omega(\cdot):=\frac{1}{\Delta t}\int_{s=0}^{\Delta t}\delta_{u(s)}ds$, and z be the unique solution of the ODE $\dot{z}_s=F(z_s,\omega)$ issued from x. Then

$$\left|y_{\Delta t}-z_{\Delta t}\right|=\left|\int_{s=0}^{\Delta t}F(y_s,u(s))ds-\int_{\tau=0}^{\Delta t}\frac{1}{\Delta t}\int_{s=0}^{\Delta t}F(z_\tau,u(s))dsd\tau\right|$$

by definition of the relaxed control. Using Fubini,

$$\left|\int_{s=0}^{\Delta t}\frac{1}{\Delta t}\int_{\tau=0}^{\Delta t}F(y_s,u(s))-F(y_\tau,u(s))d\tau ds\right|\leqslant\int_{s=0}^{\Delta t}\frac{1}{\Delta t}\int_{\tau=0}^{\Delta t}Lip(F)\|F\|_\infty|\tau-s|\,d\tau ds\leqslant Lip(F)\|F\|_\infty\frac{\Delta t^2}{2}.$$

With this trick, one can replace the term $\Delta_{t_n,t_{n+1}}(1)$ in (2.38) by an error term of order Δt^2. In the setting of CAT(0) spaces, the integral representation is not valid, and Fubini does not make sense. One could hope that using the EVI, a similar result could be obtained; with the present definitions, this seemed out of reach.

2.3.3.2 Approximation of the value function

We now assume that we have a convergent numerical scheme to approximate the reachable sets, with a quadratic error estimate with respect to the time variable.

Assumption [A2.3.12] (Convergence of reachable set). It is assumed that for each $x\in\Omega$ and $0\leqslant t\leqslant s$, one knows how to compute a set $\hat{R}_s^{t,x}$ such that

$$d_H\left(\hat{R}_s^{t,x},R_s^{t,x}\right)\leqslant\varpi|t-s|^2$$

for some constant ϖ that is independent of x,s,t.

One may then restrict the computation of an approximation over a mesh $\hat{\Omega}$, that is, a finite subset of Ω such that for some (morally small) step $\Delta x>0$, there holds $\Omega\subset\bigcup_{\hat{x}\in\hat{\Omega}}\overline{\mathscr{B}}(x,\Delta x)$. Since $\hat{\Omega}$ is a finite set, one can project any point $x\in\Omega$ over $\hat{\Omega}$, to recover the set of $\hat{x}\in\hat{\Omega}$ realizing the minimum of the distance to x: the step controls the error introduced by this operation, since any projection of x over $\hat{\Omega}$ is at distance inferior or equal to Δx of x.

The value function may be approximated by a discrete dynamical programming principle as follows.

Algorithm 3: Approximation of the value function by a semi-Lagrangian scheme

Let $\mathfrak{J}:\Omega\to\mathbb{R}$, $T>0$, $N\in\mathbb{N}_*$ and a mesh $\hat{\Omega}$ be given.
Denote $\Delta t:=T/N$ and $t_n=n\Delta t$ for $n\in[\![0,N]\!]$.
Let $\hat{V}_N:\hat{\Omega}\to\mathbb{R}$ be given by $\hat{V}_N(\hat{x}):=:=\mathfrak{J}(\hat{x})$ for all $\hat{x}\in\hat{\Omega}$.
for $n\in\{N-1,N-2,\cdots,0\}$ **do**
  for $\hat{x}\in\hat{\Omega}$ **do**
    Compute the approximation $\hat{R}_{t_{n+1}}^{t_n,\hat{x}}$ of the reachable set $R_{t_{n+1}}^{t_n,\hat{x}}$.
    Let $\hat{P}_n^{\hat{x}}$ be the set of projections of the elements of $\hat{R}_{t_{n+1}}^{t_n,\hat{x}}$ over the mesh $\hat{\Omega}$.
    Let $\hat{V}_n(\hat{x}):=\min_{\hat{y}\in\hat{P}_n}\hat{V}_{n+1}(\hat{y})$.
Return $(\hat{V}_n)_{n\in[\![0,N]\!]}$.

The convergence estimate involves a play between the time step Δt and the space step Δx, which is classical in semi-Lagrangian approximation, and usually referred as the *inverse CFL condition.* Indeed, to obtain convergence, one should take $\Delta x=o(\Delta t)$, whereas the opposite condition was highlighted by Courant, Friedrich and Lewy in the case of wave equations.

Lemma 2.3.13 (Convergence of Algorithm 3). *Assume that [A2.3.12] holds. Then*

$$\max_{n\in[\![0,N]\!]}\sup_{\hat{y}\in\hat{\Omega}}\left|\hat{V}_n(\hat{y})-V(t_n,\hat{y})\right|\leqslant Lip(V)\,T\left(\frac{\Delta x}{\Delta t}+\varpi\Delta t\right).$$

Proof. Denote $e_n:=\sup_{\hat{y}\in\hat{\Omega}}\left|\hat{V}_n(\hat{z})-V(t_n,\hat{z})\right|$ the error at step n. We proceed by induction: for any $\hat{x}\in\hat{\Omega}$, there holds

$$\begin{aligned}\left|\hat{V}_n(\hat{x})-V(t^n,\hat{x})\right| &\leqslant \left|\inf_{\hat{y}\in\hat{P}_n^{\hat{x}}}\hat{V}_{n+1}(\hat{y})-\inf_{\hat{y}\in\hat{P}_n^{\hat{x}}}V(t^n,\hat{y})\right|+\left|\inf_{\hat{y}\in\hat{P}_n^{\hat{x}}}V(t^n,\hat{y})-\inf_{y\in R_{t_{n+1}}^{t_n,\hat{x}}}V(t^n,y)\right|\\ &\leqslant \sup_{\hat{y}\in\hat{P}_n^{\hat{x}}}\left|\hat{V}_{n+1}(\hat{y})-V(t^n,\hat{y})\right|+\mathrm{Lip}(V)\,d_H\left(\hat{P}_n^{\hat{x}},R_{t_{n+1}}^{t_n,\hat{x}}\right).\end{aligned}$$

As by the construction of the projection and **[A2.3.12]**,

$$d_H\left(\hat{P}_n^{\hat{x}},R_{t_{n+1}}^{t_n,\hat{x}}\right)\leqslant d_H\left(\hat{P}_n^{\hat{x}},\hat{R}_{t_{n+1}}^{t_n,\hat{x}}\right)+d_H\left(\hat{R}_{t_{n+1}}^{t_n,\hat{x}},R_{t_{n+1}}^{t_n,\hat{x}}\right)\leqslant\Delta x+\varpi\Delta t^2,$$

taking the maximum over $\hat{x}\in\hat{\Omega}$, we recover $e_n\leqslant e_{n+1}+\mathrm{Lip}(V)(\Delta x+\varpi\Delta t^2)$. As $e_N=0$ and $e_0\geqslant e_n$ for all n, there holds

$$E_0\leqslant 0+\sum_{n=0}^{N-1}\mathrm{Lip}(V)(\Delta x+\varpi\Delta t^2)=\mathrm{Lip}(V)\,(N\Delta x+T\varpi\Delta t)=\mathrm{Lip}(V)\,T\left(\frac{\Delta x}{\Delta t}+\varpi\Delta t\right).$$

Hence the result. □

Observe that we may extend the function $\hat{V}$ over the domain $[0,T]\times\Omega$ in the following, quite coarse, way: to any (t,x), associate the value $\hat{V}(t_n,\hat{x})$ such that $(t_n,\hat{x})\in\{0,\Delta t,2\Delta t,\cdots,T\}\times\hat{\Omega}$ realize the infimum of the distance to (t,x). Denote $\hat{V}$ such an extension by nearest neighbour. With this choice, there holds

$$\left|\hat{V}(t,x)-V(t,x)\right|\leqslant\left|\hat{V}(t_n,\hat{x})-V(t_n,\hat{x})\right|+|V(t_n,\hat{x})-V(t,x)|\leqslant\mathrm{Lip}(V)\,T\left(\frac{\Delta x}{\Delta t}+\varpi\Delta t\right)+\mathrm{Lip}(V)(\Delta t+\Delta x),$$

so that under the inverse CFL condition $\Delta x=o(\Delta t)$, the order of the error is the same as in Lemma 2.3.13.

2.3.3.3 Numerical resolution of the optimal control problem

We assume by now that we can compute approximations of the value function, in the sense that for any ε, we have access to a function $\hat{V}=\hat{V}_\varepsilon:[0,T]\times\Omega\to\mathbb{R}$ such that

$$\sup_{(t,x)\in[0,T]\times\Omega}\left|\hat{V}(t,x)-V(t,x)\right|\leqslant\varepsilon.$$

The optimal trajectories are characterized by the fact that they let the value function invariant. On the other hand, all the other characteristics of the system see the value function grow over time. At an infinitesimal level and in the case where V is smooth, one can recover an optimal trajectory by computing the gradient of the value function. However, in non-smooth settings, difficulties arise: one has to resort to subgradients, which may form an irregular function. Consequently, numerical schemes appeared (as in [RV91]) that do not rely on differential information. We consider the following version of such schemes.

Algorithm 4: Numerical approximation of the optimal control and trajectory

```
Let x ∈ Ω, N ∈ ℕ_*, T > 0 and V̂ be given.
Define Δt := T/N and t_n := nΔt for any n ∈ ⟦0, N⟧.
Initialize ŷ_0 := x.
for n ∈ ⟦0, N−1⟧ do
    Let û^n ∈ argmin_{v∈U} V̂(t_{n+1}, G_F(f(ŷ_{t_n}, v))(Δt, ŷ_{t_n})).
    Define ŷ_s := G_F(f(ŷ_{t_n}, û^n))(s − t_n, ŷ_{t_n}) for s ∈ ]t_n, t_{n+1}].
Return (û_n)_{n∈⟦0,N−1⟧} ⊂ U and ŷ ∈ AC([0, T]; Ω).
```

Remark 2.3.14 (Factorization)**.** *In practice, the approximations of the reachable sets are computed and stored once and for all before entering the propagation step.*

To give a quantitative estimate of the quality of the scheme, we consider the *loss of optimality*

$$y \in \mathrm{AC}([0,T];\Omega) \mapsto \mathfrak{J}(\hat{y}_T) - V(0, y_0). \tag{2.36}$$

By definition of the value function, y is an optimal trajectory if and only if $s \mapsto V(s, y_s)$ is constant, that is, if the quantity in (2.36) vanishes. Moreover, $y \mapsto \mathfrak{J}(\hat{y}_T) - V(0, y_0)$ converges to 0 when y becomes close to an optimal trajectory. However, as opposite to the distance between y and the nearest optimal solution, the loss of optimality is not affected by the possible non-uniqueness of optimal trajectories.

Lemma 2.3.15 (Convergence of Algorithm 4). *There exists $\Delta_{[0,T]}(M) = O(M)$ depending only on U such that*

$$\mathfrak{J}(\hat{y}_T) - V(0,x) \leqslant Lip(V)\,T\,\frac{Lip(f)\|f\|_\infty \Delta s}{2} + Lip(V)e^{Lip(f)\Delta s}\Delta_{[0,T]}(M) + (2M+1)\varepsilon. \tag{2.37}$$

By Arzelà-Ascoli, the family of trajectories $(\hat{y}^M)_{M\in\mathbb{M}_*}$ admits cluster points in $\mathrm{AC}([0,T];\Omega)$. Choosing $\varepsilon = o(M)$ and passing to the limit in (2.37) along the appropriate subsequence, one gets that the limit points are optimal solutions of the control problem (2.25). By the relaxation result of Theorem 2.2.4, there exists a relaxed control $\omega(\cdot) \in L^1(0,T;\mathscr{P}_1(U))$ generating each limit point; however, the sequence of controls computed by Algorithm 4 has no reason to converge.

Proof. By definition of $\hat{y}$ and $\hat{u}$, there holds $\mathfrak{J}(\hat{y}_T) - \hat{V}(0,x) \leqslant \mathfrak{J}(\hat{y}_T) - V(0,x) + \varepsilon$, where

$$\mathfrak{J}(\hat{y}_T) - \hat{V}(0,x) = \sum_{m=0}^{M-1} \hat{V}(s_{m+1}, \hat{y}_{s_{m+1}}) - \hat{V}(s_m, \hat{y}_{s_m}) = \sum_{m=0}^{M-1} \inf_{v_m\in U} \hat{V}(s_{m+1}, \mathcal{G}_{\mathcal{F}}(f(\hat{y}_{s_m}, v_m))(\Delta s, \hat{y}_{s_m})) - \hat{V}(s_m, \hat{y}_{s_m}).$$

Using the estimates of Proposition 2.1.9,

$$\begin{aligned}\hat{V}(s_m, \hat{y}_{s_m}) + \varepsilon \geqslant V(s_m, \hat{y}_{s_m}) &= \inf_{v(\cdot)\in L^1(s_m,s_{m+1};U)} V(s_{m+1}, y_T^{s_m,\hat{y}_{s_m},v}) \\ &\geqslant \inf_{v_m\in U} V(s_{m+1}, y_{s_{m+1}}^{s_m,\hat{y}_{s_m},v_m}) - \mathrm{Lip}(V)e^{\mathrm{Lip}(f)\Delta s}\Delta_{s_m,s_{m+1}}(1),\end{aligned} \tag{2.38}$$

where $\Delta_{s_m,s_{m+1}}(1) := \sup_{v(\cdot)\in L^1(s_m,s_{m+1};U)} \inf_{v_m\in U} \int_{s=s_m}^{s_{m+1}} d_U(v(s), v_m)ds$. Using $\inf_A \phi - \inf_A \psi \leqslant \sup_A \phi - \psi$, one deduces

$$\begin{aligned}&\mathfrak{J}(\hat{y}_T) - \hat{V}(0,x) \\ &\leqslant \sum_{m=0}^{M-1}\left[\inf_{v_m\in U} V(s_{m+1}, \mathcal{G}_{\mathcal{F}}(f(\hat{y}_{s_m}, v_m))(\Delta s, \hat{y}_{s_m})) - \inf_{v_m\in U} V(s_{m+1}, y_{s_{m+1}}^{s_m,\hat{y}_{s_m},v_m}) + \mathrm{Lip}(V)e^{\mathrm{Lip}(f)\Delta s}\Delta_{s_m,s_{m+1}}(1) + 2\varepsilon\right] \\ &\leqslant \sum_{m=0}^{M-1}\left[\sup_{v_m\in U} V(s_{m+1}, \mathcal{G}_{\mathcal{F}}(f(\hat{y}_{s_m}, v_m))(\Delta s, \hat{y}_{s_m})) - V(s_{m+1}, y_{s_{m+1}}^{s_m,\hat{y}_{s_m},v_m}) + \mathrm{Lip}(V)e^{\mathrm{Lip}(f)\Delta s}\Delta_{s_m,s_{m+1}}(1) + 2\varepsilon\right] \\ &\leqslant \sum_{m=0}^{M-1} \mathrm{Lip}(V)\frac{\mathrm{Lip}(f)\|f\|_\infty \Delta s^2}{2} + \mathrm{Lip}(V)e^{\mathrm{Lip}(f)\Delta s}\sum_{m=0}^{M-1}\Delta_{s_m,s_{m+1}}(1) + 2M\varepsilon.\end{aligned}$$

By Chasles, approximating $L^1(s_m, s_{m+1}; U)$ by constants on M intervals amounts to approximating $L^1(0,T;U)$ by piecewise constant controls on each $[s_m, s_{m+1}]$:

$$\begin{aligned}\sum_{m=0}^{M-1}\Delta_{s_m,s_{m+1}}(1) &= \sum_{m=0}^{M-1} \sup_{v\in L^1(s_m,s_{m+1};U)} \inf_{w\in U^M} \int_{s=s_m}^{s_{m+1}} d_U(v(s), w_m)ds \\ &= \sup_{v\in L^1(0,T;U)} \inf_{w\in U^M} \int_{s=0}^{T} d_U(v(s), w_m)ds =: \Delta_{[0,T]}(M).\end{aligned}$$

By the density of simple functions in L^1, the latter term vanishes when M goes to infinity. Gathering the above, we obtain the desired estimate. □

Once again, the estimate in (2.37) is highly suboptimal with respect to the situation in $\Omega = \mathbb{R}^d$, where the term $\Delta_{s_m,s_{m+1}}(1)$ in (2.38) could be replaced by an error term in Δs^2 (see Remark 2.3.11).

2.4 Numerical illustrations

This part provides numerical illustration of the numerical schemes of Section 2.3.3. The set of transitions is taken as in (2.6), that is,

$$\Theta := \mathcal{G}_{\mathcal{F}}\left\{\alpha\kappa\circ d(\cdot,x_0)\;\middle|\;\alpha\in\mathbb{R}^+,\; x_0\in\Omega\right\}$$

with $\kappa:\mathbb{R}^+\to\mathbb{R}^+$ a smooth Lipschitz function whose derivative at 0 vanishes. Errors on the value function are given as relative infinite error, i.e.

$$\left|V-\hat{V}\right|_\infty := \frac{\max_{n\in[\![0,N]\!]}\max_{\hat{x}\in\hat{\Omega}}\left|V(t_n,\hat{x})-\hat{V}(t_n,\hat{x})\right|}{\max_{n\in[\![0,N]\!]}\max_{\hat{x}\in\hat{\Omega}}|V(t_n,\hat{x})|}.$$

The numerical approximation of the optimal trajectory is assessed through the loss of optimality

$$\mathrm{Error}:\mathrm{AC}([0,T];\Omega)\to\mathbb{R},\qquad \mathrm{Error}(\hat{y}) := \mathfrak{J}\left(\hat{y}_T\right)-V(0,\hat{y}_0). \tag{2.39}$$

By definition of V, this quantity is nonnegative, and vanishes only for optimal solutions. It presents the double advantage of being easy to compute, and blind to the possible non-uniqueness of optimal trajectories. Source files in julia are available at https://github.com/averil-aussedat/FLagHada.jl. Simulations were carried on a Dell Inc. Precision 3561 laptop with 11^{th} Gen. Intel I9 processor and 64GB of RAM (although most of it is not used because the code is not parallelized).

2.4.1 Eikonal equation

Consider the CAT(0) space made of the gluing of three linearly-parametrized segments $[oa]$, $[ob]$ and $[oc]$ at the junction point o, with the shortest path distance. In this example, it is considered that $d(o,a) = d(o,b) = d(o,c) = 1$. Let $U = \{u_0,u_a,u_b,u_c\}$ be the set of controls, and

$$f:\Omega\times U\to\Theta,\qquad f(x,u_0) := \mathcal{G}_{\mathcal{F}}(\kappa\circ d(\cdot,x))\quad\text{and}\quad f(x,u_z) := \mathcal{G}_{\mathcal{F}}(\kappa\circ d(\cdot,z))\qquad\text{for } z\in\{a,b,c\}.$$

The dynamic f is Lipschitz with respect to x, and offers the possibility to stay at a given point or to move towards each of the boundaries a,b,c of the network. Consider the terminal cost

$$\mathfrak{J} := d(\cdot,o).$$

The value function of the optimal control problem

$$\text{Find } u(\cdot)\in L^1(0,T;U) \text{ minimizing } \mathfrak{J}(y_T^{0,x,u})$$

is given by

$$V(t,x) = \max(0, d(x,o)-(T-t)).$$

For any $x\in\Omega$, the optimal control exists and is unique in $L^1(0,T;U)$. It is given by

$$u^*(t) = \begin{cases} u_z & \text{if } x\in(oz] \text{ with } z\in\{a,b,c\}, \text{ and } t\leqslant d(x,o),\\ u_0 & \text{if } x=o \text{ or } t>d(x,o).\end{cases}$$

The optimal trajectory issued from x follows the geodesic $[xo]$ with unit speed up to time $t = d(x,o)$, after which it stops at o. We turn to the numerical results. The approximation of the value function is represented for $T = 0.5$ and varying discretization parameters in Figure 2.1. To assess the order of the scheme, the error is computed as a function of Δt and Δx for a hundred simulations, where Δx varies log-linearly between 5×10^{-2} and 5×10^{-6}, and N is chosen according to the inverse CFL condition as $N = \lceil T/\sqrt{dx}\rceil$. The numerical optimal trajectories issued from o,a,b and c are computed according to Algorithm 4. Results are presented graphically in Figure 2.2, and precise numerical values of the error are shown in Table 2.1.

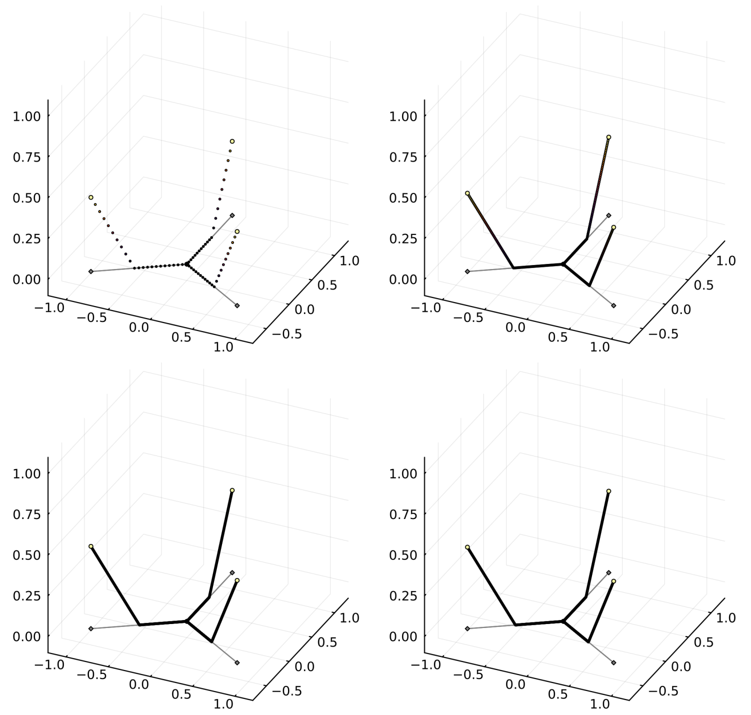


Figure 2.1: Approximation of the value function V at $T = 0.5$ for the Eikonal equation in the tripod network.

From left to right and top to bottom, N_{mesh} is respectively equal to 67, 382, 6 001 and 600 001 points, whereas N takes the values 3, 6, 23, 224 iterations.

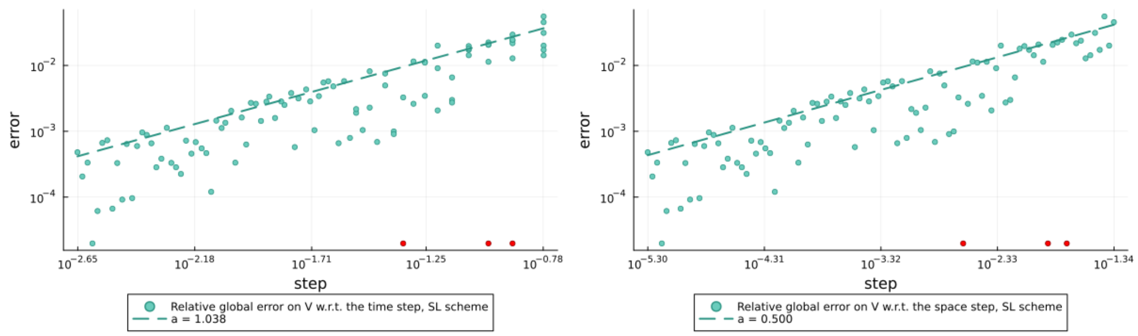


Figure 2.2: Evolution of errors on as a function of the time step Δt (left) and the space step Δx (right) for the Eikonal equation on the tripod network.

Each dot corresponds to a simulation, with a space step Δx ranging in 100 log-uniformly distributed values between 4.56×10^{-2} *and* 5×10^{-6}*, and the time step is computed as to keep the CFL $\Delta x / \Delta t^2 = 1$. The loss of optimality is computed as the maximum* Error*, defined in (2.39), between all numerical optimal trajectories issued from o, a, b and c. The red dots are used to indicate outlier errors, that are significantly smaller than the other and get out of the frame of the graph.*

Parameters		Errors on V		Errors on $\hat{y}$		Time (s)
Δt	Δx	global	at $t=0$	worst Error	mean Error	
1.67e-01	4.56e-02	4.55e-02	9.09e-02	1.11e-16	8.33e-17	0.99
8.33e-02	7.92e-03	1.97e-02	3.94e-02	1.67e-16	1.25e-16	1.02
3.33e-02	1.26e-03	8.16e-03	1.63e-02	1.11e-16	8.33e-17	1.02
1.39e-02	1.99e-04	1.59e-03	3.18e-03	1.78e-15	1.33e-15	1.30
5.56e-03	3.15e-05	2.84e-04	5.68e-04	1.78e-15	1.33e-15	6.36
2.23e-03	5.00e-06	4.80e-04	9.60e-04	1.06e-14	7.95e-15	110.72

Table 2.1: Evolution of error for the Eikonal equation on the tripod network.

Errors are shown for the first, 20^{th}, 40^{th}, 60^{th}, 80^{th} and last simulation appearing in Figure 2.2. The CFL $\Delta x/\Delta t^2$ is kept constant and equal to 1. The errors on the value function V are normalized, and the CPU time is in seconds.

On this example, we observe convergence of order 1 with respect to the time step Δt, and 1/2 with respect to the space step Δx. This is expected from the choice of N as a function of Δx, and in the sequel, we only present the evolution of errors as a function of the space step. In Figure 2.1, one notices high oscillations of the error. To our understanding, this is due to the choice of the nearest neighbour in the definition of the semi-Lagrangian scheme. At each step, the reachable set is projected on the mesh, and this projection is discontinuous. In some very particular cases, the ratio between Δt and Δx makes the approximation of the reachable set fall very close to the mesh, and the error $\| V - \hat{V}\|_\infty$ gets close to machine precision: this explains the red dots appearing in Figure 2.1. In general however, the error is very sensible to $\Delta t/\Delta x$, but stays bounded from above by a decreasing function.

Let us mention that "traditional" semi-Lagrangian uses more elaborate reconstructions of that value of V at the foot of the characteristics, such as interpolation of various order or Galerkin methods. In the context of CAT(0) spaces, even first-order interpolation is not a trivial matter, and we postpone such a refinement of Algorithm 3 for future research. In this example, the trajectories are reconstructed by Algorithm 4 with machine precision, and we do not comment further.

2.4.2 Aestival dilemma

The first example sets in a simplified model of the French fast train network (TGV) in 2000-2010, pictured in Figure 2.3. The control problem under consideration consists in the aestival dilemma of reaching southern locations in minimal time, namely Bordeaux ($\mathrm{B_o}$), Marseille (M) and Barcelona ($\mathrm{B_a}$). It is assumed that the controller do not favour one of the destinations, leading to a terminal cost of

$$\mathfrak{J}(x) := \min\left(d(x,\mathrm{B_o}), d(x,\mathrm{M}), d(x,\mathrm{B_a})\right).$$

The set of controls is chosen as $U = \{u_{\mathrm{L_o}}, u_{\mathrm{N}}, u_{\mathrm{P_u}}, u_{\mathrm{D}}\}$. For each control $u \in U$, the dynamic $x \mapsto f(x,u)$ is given as $\mathcal{G}_{\mathcal{F}}(m_u(x)\kappa \circ d(\cdot, x_{0,u}))$, where

$$x_{0,u_{\mathrm{L_o}}} = \mathrm{L_o}, \quad x_{0,u_{\mathrm{P_u}}} = \mathrm{P_u}, \quad x_{0,u_{\mathrm{N}}} = \mathrm{N}, \quad x_{0,u_{\mathrm{D}}} = \mathrm{D}, \qquad m_{u_{\mathrm{L_o}}} \equiv m_{u_{\mathrm{P_u}}} \equiv 1,$$
$$m_{u_{\mathrm{N}}}(x) = \frac{1}{5} + \min\left(1, \frac{d(x,[\mathrm{M},\mathrm{N}])}{d(\mathrm{L},\mathrm{M})}\right), \quad \text{and} \quad m_{u_{\mathrm{D}}}(x) = \frac{3}{2} - \frac{3}{5}\min\left(1, \frac{d(x,[\mathrm{B_o},\mathrm{D}])}{d(\mathrm{LM},\mathrm{B_o})}\right).$$

Here $d(x,[\mathrm{A},\mathrm{B}]) = \inf_{z\in[\mathrm{A},\mathrm{B}]} d(x,z)$ is the distance to the geodesic $[\mathrm{A},\mathrm{B}]$. In English, f allows to choose between going to Londres or Madrid (Puerta de Atocha) with speed 1, to Nice with lower speed limits in the Provence region, or to Dax with an increasingly efficient network in the Aquitaine region. Each partial dynamic $x \mapsto d(x,u)$ is Lipschitz, and our setting applies.

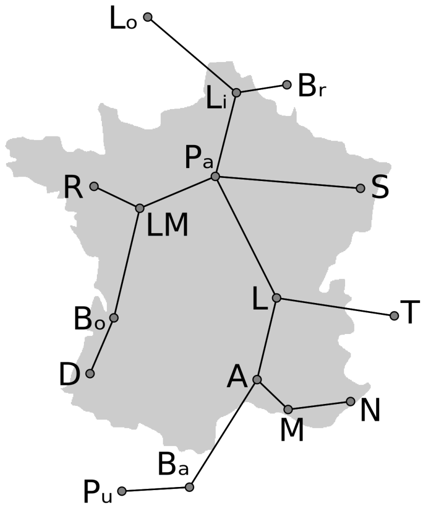


Figure 2.3: Graphical representation of the simplified SNCF network.

The dynamic programming principle implies that the value function V is given as the minimum of the three value functions $V_{\mathrm{B_o}}$, V_{M} and $V_{\mathrm{B_a}}$ respectively associated with the terminal costs $d(\cdot,\mathrm{B_o})$, $d(\cdot,\mathrm{M})$ and $d(\cdot,\mathrm{B_a})$. Each of these value functions may be computed by hand, providing us with a reference solution.

We turn to the numerical results. In this example and the following ones, Algorithms 3 and 4 are run 100 times with a decreasing space step Δx. The time step Δt is computed as indicated in the legends of the figures. The numerical values of the errors are presented for the first, 20^{th}, 40^{th}, 60^{th}, 80^{th} and last simulation.

Figure 2.4 presents the numerical value functions for $T = 6$ and varying Δx. Figure 2.5 shows the evolution of the errors for $T = 6$, and precise numerical values are given in Table 2.2. Figure 2.6 shows the numerical approximation of the optimal control, in the form a feedback law computed on an independent fixed mesh.

The algorithm of approximation of the value function achieves convergence of numerical order 1/2. One observes a convergence of the loss of optimality with numerical order 1/2 as well. Relaxed controls appear in the feedback maps of Figure 2.6: at the destination points $\mathrm{B_o}$, $\mathrm{B_a}$ and M, the optimal dynamic does not vanish, and points upstream with respect to the optimal flow. The induced trajectory will oscillate, or chatter, and eventually stay close to the destination point.

In Figure 2.4, one notices oscillations of the error around their mean. To our understanding, this is due to the choice of the nearest neighbour in the definition of the semi-Lagrangian scheme. At each step, the reachable set is projected on the mesh, and this projection is discontinuous. In some very particular cases, the ratio between Δt and Δx makes the approximation of the reachable set fall very close to the mesh, and the error $\|V - \hat{V}\|_\infty$ decreases. In general however, the error is very sensible to $\Delta t/\Delta x$, but stays bounded from above by a decreasing function.

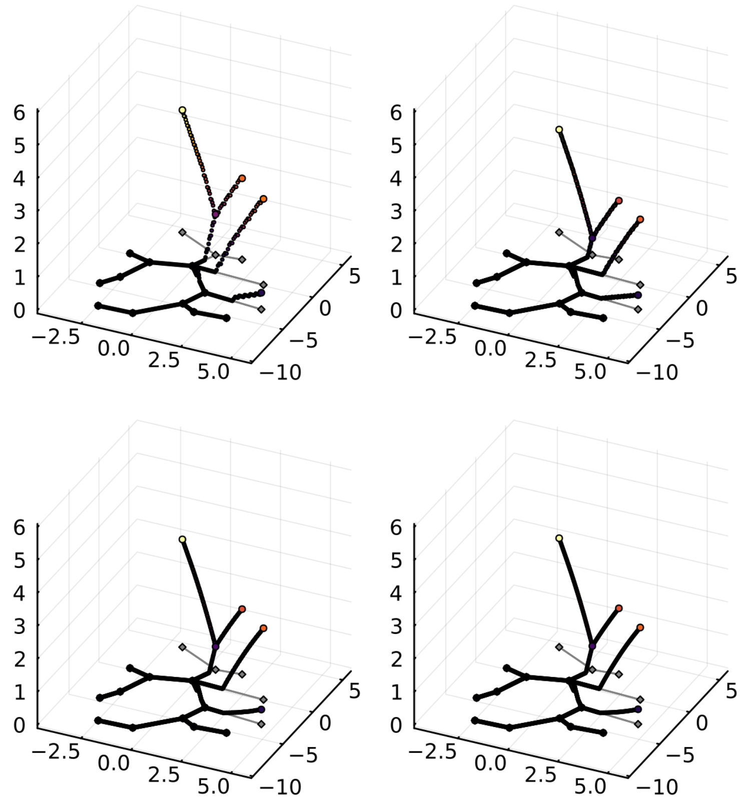


Figure 2.4: Numerical approximation of the value function V at $T = 6$ for the aestival dilemma.

From left to right and top to bottom, N_{mesh} is respectively equal to 398, 1593, 14 505 and 579 943 points, whereas N takes the values 23, 45, 135, 849 iterations.

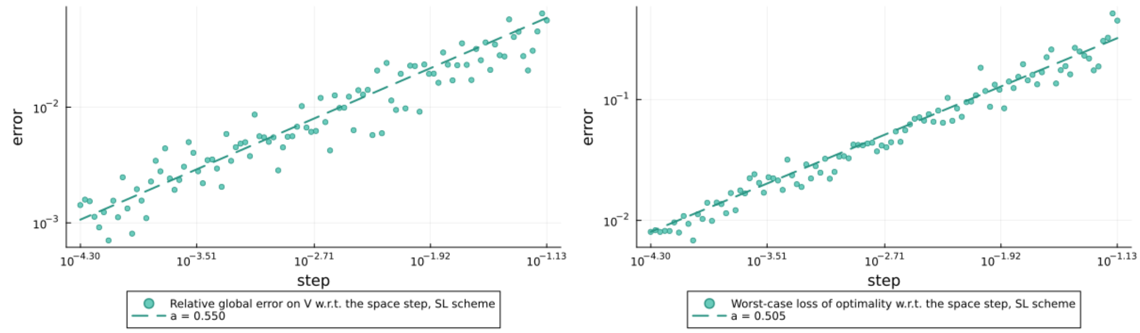


Figure 2.5: Evolution of the error for the aestival dilemma, as a function of the space step Δx for $T = 6$.

Left: Relative error on V in the time-space domain. Right: worst-case loss of optimality. Each dot corresponds to a simulation, with a space step Δx ranging in 100 log-uniformly distributed values between 7.43×10^{-2} and 5×10^{-5}, and the time step is computed as to keep the CFL $\Delta x / \Delta t^2 = 1$. The loss of optimality is computed as the maximum Error*, defined in (2.39), between all numerical optimal trajectories issued from the junctions of the network.*

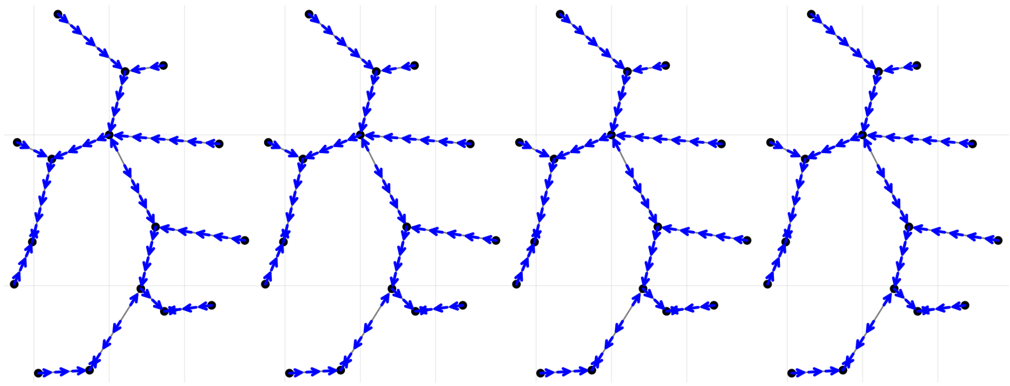


Figure 2.6: Numerical approximation of the feedback control for the aestival dilemma.

Algorithm 4 is applied with $\hat{V}$ the approximation of the value function computed at the first, 20^{th}, 50^{th} and last simulations appearing in Figure 2.5. Computations are carried on a dedicated mesh, starting at time $T - \Delta t$, where the time step Δt is taken equal to the time step of the corresponding simulation of $\hat{V}$.

Parameters		Errors on V		Errors on $\hat{y}$		Time (s)
Δt	Δx	global	at $t = 0$	worst Error	mean Error	
2.61e-01	7.43e-02	5.63e-02	1.70e-01	4.50e-01	2.10e-01	1.10
1.33e-01	1.83e-02	2.31e-02	6.95e-02	1.44e-01	4.73e-02	1.04
6.45e-02	4.18e-03	1.28e-02	3.32e-02	6.55e-02	2.67e-02	1.18
3.08e-02	9.56e-04	5.02e-03	1.20e-02	3.37e-02	1.09e-02	3.11
1.48e-02	2.19e-04	1.93e-03	5.14e-03	1.67e-02	5.07e-03	29.27
7.07e-03	5.00e-05	1.42e-03	4.29e-03	8.04e-03	2.17e-03	232.05

Table 2.2: Evolution of errors in the aestival dilemma problem.

Errors are shown for the first, 20^{th}, 40^{th}, 60^{th}, 80^{th} and last simulation appearing in Figure 2.5. The CFL $\Delta x/\Delta t^2$ is kept constant and equal to 1. The errors on the value function V are normalized, and the CPU time is in seconds.

2.4.3 Plane-and-line problem

We turn to CAT(0) spaces made by gluing Euclidean subdomains of different dimensions. First, consider Ω given by the union of an Euclidean 2-dimensional square Ω_1 isometric to $[0,4]^2$, and an unit segment $\Omega_2 = [a, a+4]$, where a is identified with the point $\overline{a} \in \Omega_1$ of coordinates $(4,2)$. Denote e_0, e_1, e_2, e_3 the four sides of Ω_1, with e_3 containing the point a. The set of controls is chosen as $U = \{u_0, u_1, u_2, u_3, u_4\}$, and the dynamic as

$$
\begin{aligned}
f(x, u_i) &= \mathcal{G}_{\mathcal{F}}\left(\kappa \circ d(\cdot, p_{e_i}(x))\right) \text{ for } i \in \{0,1,2\},\\
f(x, u_3) &= \mathcal{G}_{\mathcal{F}}\left(2\kappa \circ d(\cdot, p_{e_3})\right), \qquad \text{and}\\
f(x, u_4) &= \mathcal{G}_{\mathcal{F}}\left(m_4(x)\kappa \circ d(\cdot, b)\right),
\end{aligned}
$$

where $p_{e_i}(x)$ is the projection of x on the side e_i if x belongs to Ω_1, and $p_{e_i}(\overline{a})$ otherwise. In Ω_1, the magnitude $m_4(x) \geqslant 0$ is chosen so that $m_4(x)(\overline{a} - x)$ lies on the boundary of $B := \operatorname{conv}\{(2,0),(0,1),(-1,0),(0,-1)\}$; Ω_2, it is chosen identically equal to 2. In words, the dynamic allows to move up, right and down with unit speed, and left with speed 2.

Consider the terminal cost

$$
\mathfrak{J}(x) := \min\left(d(\cdot, x_1), d(\cdot, x_2)\right),
$$

where $x_1 \in \Omega_1$ identifies with the point of coordinates $(1,3)$, and $x_2 \in \Omega_2$ identifies with 3. The value

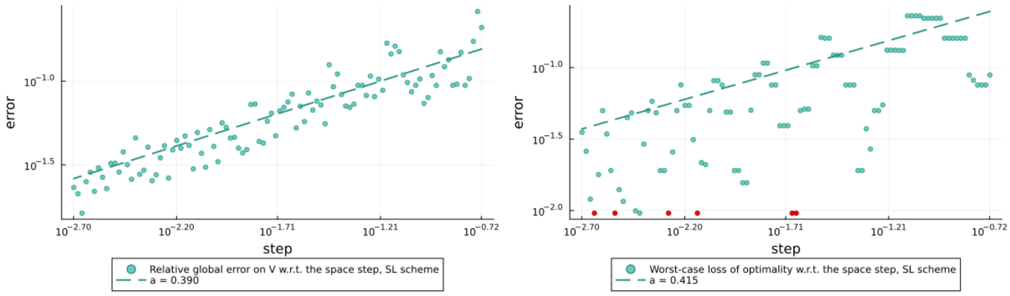


Figure 2.7: Error for the plane-and-line problem, as a function of the space step Δx for $T = 1.8$.

Left: Relative error on V in the time-space domain. Right: worst-case loss of optimality. Each dot corresponds to a simulation, with a space step Δx ranging in 100 log-uniformly distributed values between 1.91×10^{-1} and 2×10^{-3}, and the time step is computed as to keep the CFL $\Delta x/\Delta t^2 = 2.25$. The loss of optimality is computed as the maximum Error, *defined in (2.39), between all numerical optimal trajectories issued from 8 points: the four corners of Ω_1, the centre of Ω_1, the junction between both domains, and the points of Ω_2 identifying respectively with $a+2$ and $a+4$.*

function of the problem is given by $\min(V^1, V^2)$, where

$$V^1_{|\Omega_1}(t,x) = \inf_{v \in B} |x + (T-t)v - x_1|, \qquad V^1_{|\Omega_2}(t,x) = \begin{cases} d(x,x_1) - (T-t) & \text{if } T - t \leqslant d(x,a), \\ V^1_{|\Omega_1}(t + d(x,a), \overline{a}) & \text{otherwise} \end{cases}$$

corresponds to the value function of the cost $d(\cdot, x_1)$, and

$$V^2_{|\Omega_2}(t,x) = (d(x,x_2) - 2(T-t))_+, \qquad V^2_{|\Omega_1}(t,x) = \begin{cases} \inf_{v \in B} \left| x + (T-t)v - \overline{a} \right| + d(a, x_2) & \text{if } T - t \leqslant \frac{d(x,\overline{a})}{|\overline{a} - v|_B}, \\ V^2_{|\Omega_2}(t + d(x,a), \overline{a}) & \text{otherwise} \end{cases}$$

corresponds to the value function of the cost $d(\cdot, x_2)$.

Parameters		Errors on V		Errors on $\hat{y}$		Time (s)
Δt	Δx	global	at $t = 0$	worst Error	mean Error	
2.57e-01	1.91e-01	2.09e-01	2.98e-01	8.88e-02	3.29e-02	1.10
1.80e-01	7.96e-02	1.09e-01	1.56e-01	2.30e-01	4.61e-02	1.10
1.13e-01	3.17e-02	7.20e-02	1.03e-01	1.61e-01	3.53e-02	1.62
7.20e-02	1.26e-02	3.98e-02	4.74e-02	1.56e-02	6.08e-03	6.48
4.62e-02	5.02e-03	2.76e-02	3.94e-02	1.90e-02	4.17e-03	73.49
2.95e-02	2.00e-03	2.31e-02	3.30e-02	3.52e-02	7.62e-03	1803.75

Table 2.3: Evolution of errors in the plane-and-line problem.

Errors are shown for the first, 20^{th}, 40^{th}, 60^{th}, 80^{th} and last simulation appearing in Figure 2.7. The CFL $\Delta x/\Delta t^2$ is kept constant and equal to 2.25. The errors on the value function V are normalized, and the CPU time is in seconds.

The error on the value function exhibits the oscillations induced by the nearest neighbour, while in some particular cases, the relation between Δt and Δx falls precisely on a favourable value, and the error committed by Algorithm 4 jumps down to machine precision. However, the error is bounded above by a decreasing function with a numerical order staying close to $1/2$.

2.4.4 Robot problem

We turn to CAT(0) spaces made by gluing Euclidean subdomains of different dimensions. Consider a gluing of Euclidean subdomains with dimensions in $\{1,2,3\}$, depicted in Figure 2.10. The set of controls is

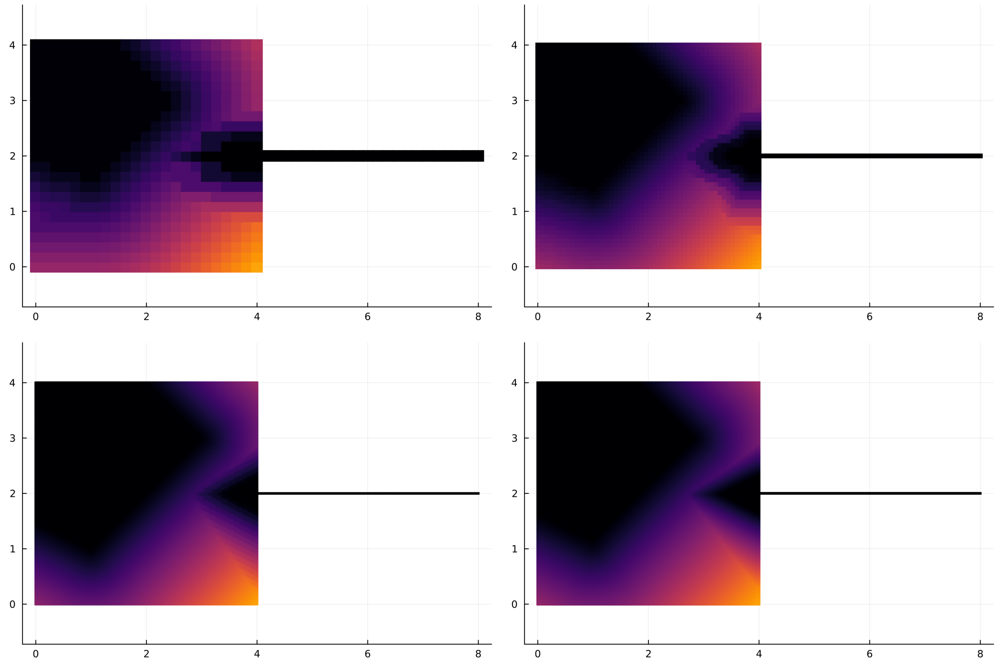


Figure 2.8: Numerical approximation of the value function in the plane-and-line problem.

Results are presented for the first, 20^{th}, 50^{th} and last simulations appearing in Figure 2.7.

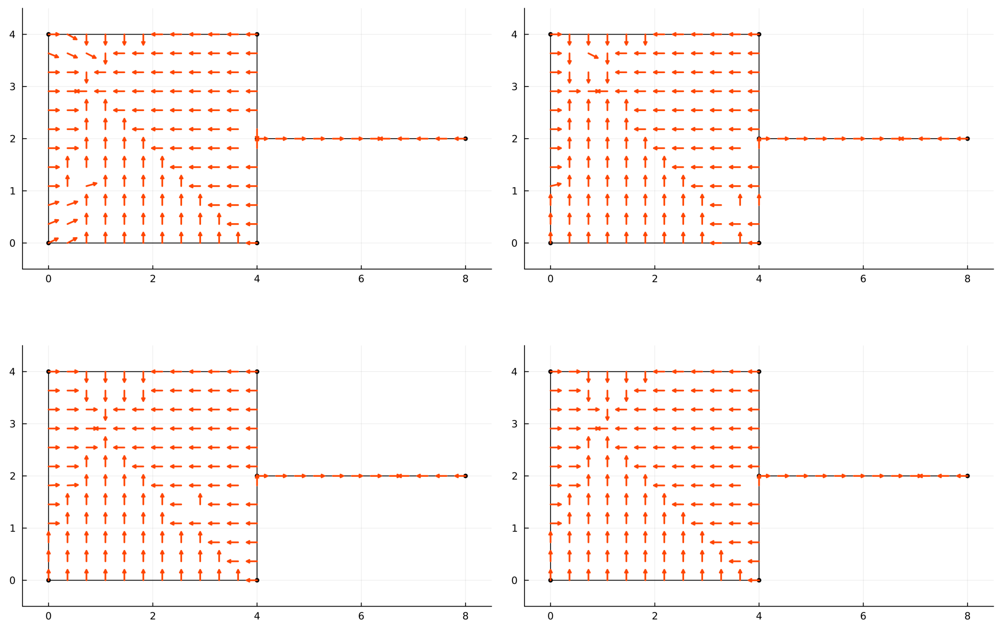


Figure 2.9: Feedback reconstruction in the plane-and-line problem.

Results are presented for the first, 20^{th}, 50^{th} and last simulations appearing in Figure 2.7. The feedback controls are computed on a dedicated mesh by using one step of Algorithm 4 starting from $t = T - \Delta t$, where Δt is the time step used by Algorithm 3 to approximate the corresponding value function.

chosen as $U = \{u_i\}_{i\in[\![0,5]\!]}$, and the dynamic as

$$f(x,u_i) = \mathcal{G}_{\mathcal{F}}\left(m_i(x)\kappa\circ d(\cdot,x_{0,i})\right),$$

where in this example,

$$\kappa:\mathbb{R}^+\to\mathbb{R}^+, \qquad \kappa(r) := \begin{cases} r^2/(2\theta) & r\in[0,\theta], \\ r-\theta/2 & r>\theta, \end{cases} \qquad \theta = 1/4,$$

and the magnitudes are given by

$$m_i \equiv 1 \text{ for } i\in[\![0,4]\!], \qquad m_5(x) = \max(1, 2\min(1, d(x,a), d(x,b))),$$

a and b being points at distance 1 of the central junction o respectively lying in the subdomains Z and H. The target points are chosen as follows:

- $x_{0,i}$ for $i\in[\![0,3]\!]$ are chosen as arbitrary Lipschitz functions such that their restriction to each pane P_j corresponds to the projections on the four sides, and their restriction to the subdomains H and Z is constant and equal to o,
- $x_{0,4}\equiv c$ a point on the leftmost pane, and $x_{0,5}\equiv d$ a point of the rightmost pane.

One checks that the choice of the Lipschitz functions $x_{0,i}$, $i\in[\![0,3]\!]$ does not influence the convex hull of the dynamic, and yield the same solution to the control problem. The dynamic is chosen as to model the motion of solar sailing maintenance robots on the outside of a satellite, with a mechanical engine allowing to move at speed 1, and a favourable drift in the direction of the sunlight. By construction, the scheme does not distinguish between junction points and points in Euclidean subdomains, and the numerical characteristics may change domain in the course of the approximation.

Remark 2.4.1 (Inward pointing condition)**.** *The construction of the dynamic by flows towards point of the space naturally implies an* inward pointing condition. *In classical control theory, this condition ensures that the trajectories of the control system remain in a given region, and allows to treat constrained problems by Hamilton-Jacobi methods (see [Son86] or [Vin10, Section 12.6]). In our case, the trajectories of the control system are build from gradient flows, and forced to lie within Ω by the very choice of the geometry.*

The terminal cost is chosen as

$$\mathfrak{J}(x) = \min_{j\in\{0,1,2\}}\left(d(x,z_j)\right),$$

where the points $(z_j)_{j\in[\![0,2]\!]}$ are depicted in Figure 2.10, and model impacts of meteors on the panes. The explicit solution of the control problem may be computed by hand as the minimum of the value functions associated to each terminal cost $d(\cdot,z_j)$. The optimal trajectories are reconstructed starting from 6 points located at the top of the panes and the central modules, displayed by the green crosses in Figure 2.10.

Simulations are carried for $T = 2.5$. Figure 2.11 presents the errors on the value function and the loss of optimality, with numerical values gathered in Table 2.4. Approximations of the value function for different space steps are displayed in Figure 2.12.

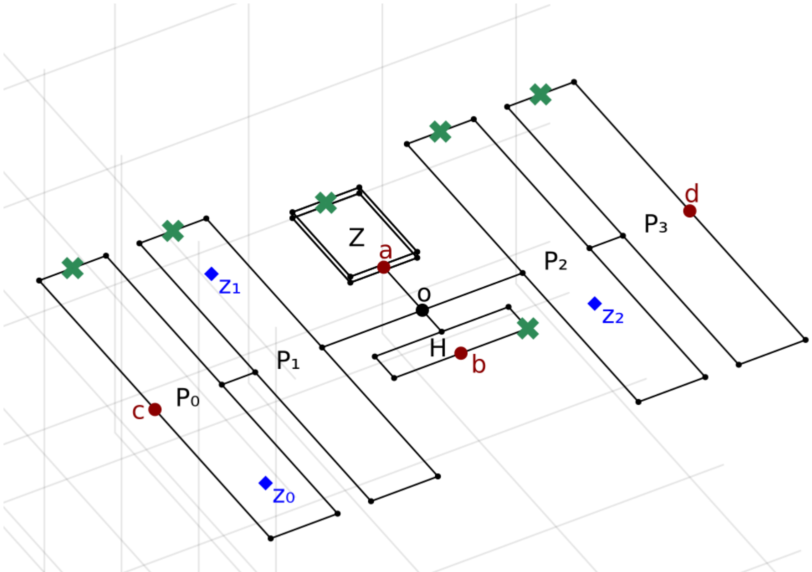


Figure 2.10: CAT(0) domain in the robot problem.

The domain is made of four Euclidean panes $(P_i)_{i\in[\![0,3]\!]}$ *of dimension 2, a 2-dimensional surface H and a 3-dimensional volume Z, linked by linearly parametrized segments. Points* o, a, b, c, d *are references used in the text. Points* $(z_j)_{j\in\{0,1,2\}}$ *are the target points of the dynamic, and the crosses indicate the initial points from which the optimal trajectory is reconstructed.*

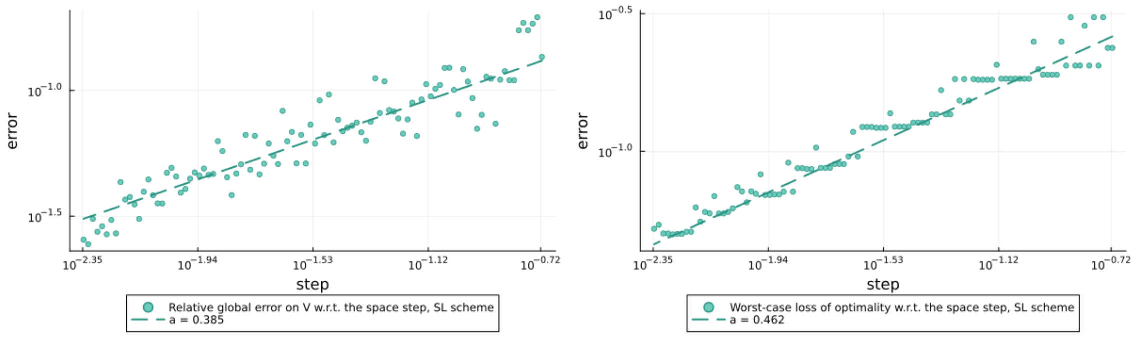


Figure 2.11: Evolution of the error for the solar robot problem, as a function of the space step Δx for $T = 2.5$.

Left: Relative error on V *in the time-space domain. Right: worst-case renormalized loss of optimality. Each dot corresponds to a simulation, with a space step* Δx *ranging in 100 log-uniformly distributed values between* 2×10^{-1} *and* 4.5×10^{-3}, *and the time step is computed as to keep the CFL* $\Delta x/\Delta t^2 = 1$. *The loss of optimality is computed as the maximum* Error, *defined in (2.39), between all numerical optimal trajectories issued from the points indicated by crosses in Figure 2.10.*

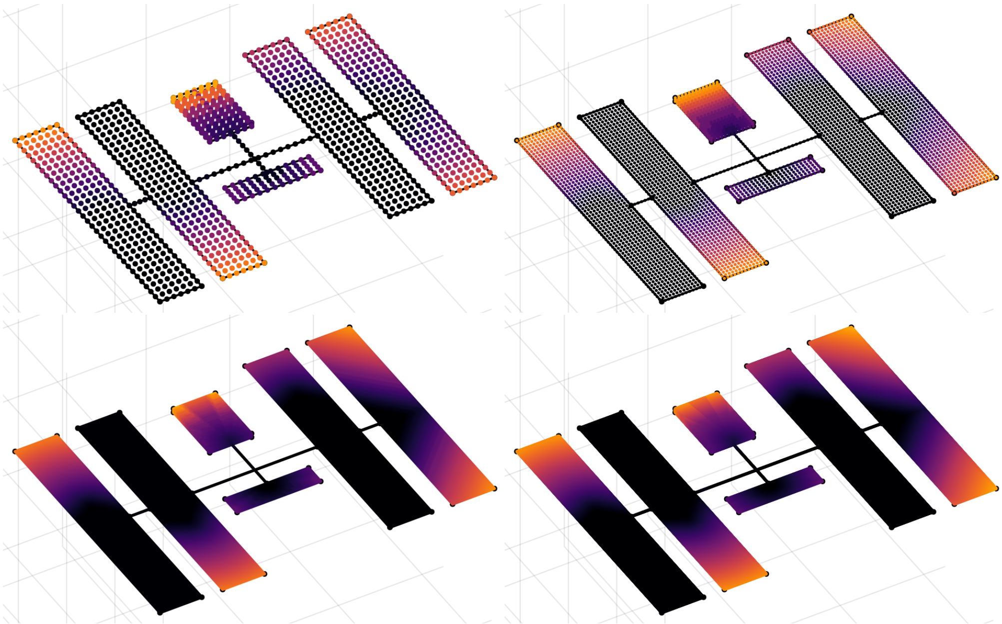


Figure 2.12: Numerical approximation of the solar robot problem.

Results are presented for the first, 20^{th}, 50^{th} and last simulations appearing in Figure 2.11. Lighter colours indicate larger values.

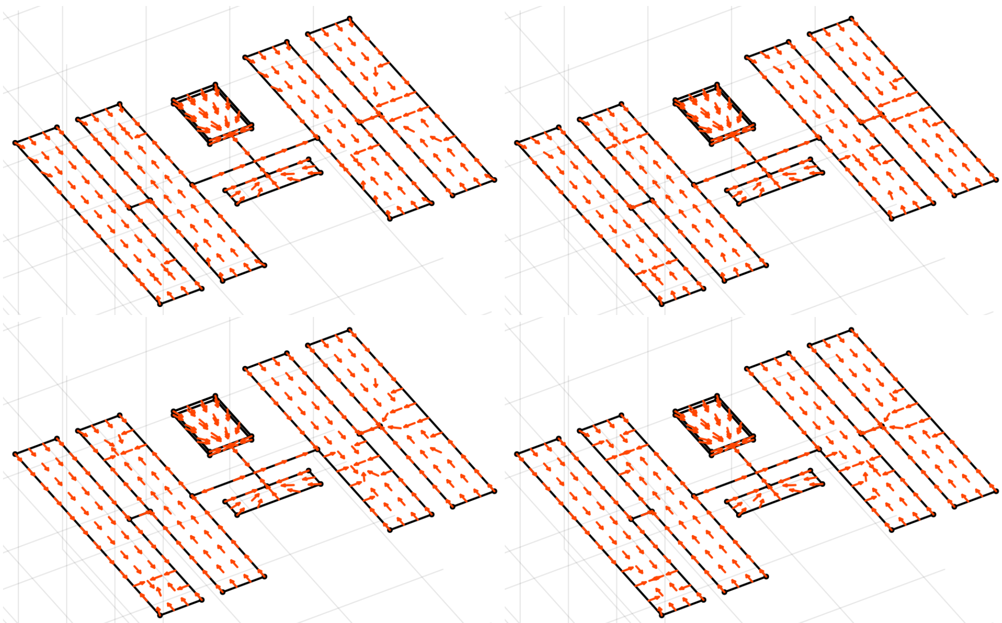


Figure 2.13: Feedback reconstruction in the solar robot problem.

Results are presented for the first, 20^{th}, 50^{th} and last simulations appearing in Figure 2.11. For readability, the controls are represented through the dynamic that they parametrize. The feedback controls are computed on a dedicated mesh by using one step of Algorithm 4 starting from $t = T - \Delta t$.

Parameters		Errors on V		Errors on $\hat{y}$		Time (s)
Δt	Δx	global	at $t=0$	worst Error	mean Error	
4.17e-01	1.93e-01	1.36e-01	2.64e-01	2.37e-01	9.94e-02	1.04
2.78e-01	9.36e-02	1.00e-01	1.61e-01	1.83e-01	6.85e-02	1.15
2.08e-01	4.38e-02	6.82e-02	1.16e-01	1.36e-01	4.23e-02	1.70
1.39e-01	2.05e-02	6.15e-02	1.19e-01	8.99e-02	2.90e-02	5.34
9.62e-02	9.61e-03	4.55e-02	8.83e-02	6.52e-02	2.29e-02	38.81
6.58e-02	4.50e-03	2.55e-02	4.95e-02	5.23e-02	1.45e-02	422.74

Table 2.4: Evolution of errors in the solar robot problem.

Errors are shown for the first, 20^{th}, 40^{th}, 60^{th}, 80^{th} and last simulation appearing in Figure 2.11. The CFL $\Delta x/\Delta t^2$ is kept constant and equal to 1. The errors on the value function V are normalized, and the CPU time is in seconds.

2.4.5 Robust control problem

The last example focuses on the hyperbolic manifold of symmetric positive definite matrices, endowed with the so-called geometric distance

$$d(P,Q) := \sqrt{\text{Trace}\log\left(P^{-1/2}QP^{-1/2}\right)^2}.$$

The interested reader is referred to [Bha07] for an extensive study of this space, including the CAT(0) property and the analytical expression of the geodesics.

Formulation of the problem

Consider the problem

$$\text{Minimize } \max_{x\in\overline{\mathscr{B}}(0,1)} \langle G y_T^{x,u}, y_T^{x,u}\rangle \text{ over all controls } v \in L^1(0,T:U),$$
$$\text{where } (y_s^{x,v})_{s\in[0,T]} \text{ satisfies } y_0^{x,v} = x \text{ and } \dot{y}_s^{x,v} = A(v(s))\,y_s^{x,v}.$$

Here $G \in \mathbb{M}_{2,2}(\mathbb{R})$ is a symmetric positive definite matrix, $A : U \to \mathbb{M}_{2,2}$ is a matrix-valued Lipschitz-continuous dynamic and the ODE is understood in the classical sense. Any control $v \in L^1(0,T;U)$ gives rise to a resolvent operator $R^v : [0,T] \to \mathbb{M}_{2,2}$ such that

$$\dot{R}_s^v = A(v(s))R_s^v, \qquad R_0^v = I_d.$$

The trajectories of the dynamical system are then given by $y_s^{x,v} = R_s^v x$. Consequently,

$$\max_{x\in\overline{\mathscr{B}}(0,1)} \langle G y_T^{x,u}, y_T^{x,u}\rangle = \max_{x\in\overline{\mathscr{B}}(0,1)} \langle G R_T^v x, R_T^v x\rangle = \lambda_{\max}\left((R_T^v)^t G R_T^v\right),$$

where $\lambda_{\max}$ denotes the maximal eigenvalue of the symmetric positive matrix $(R_T^v)^t G R_T^v$. Introduce now $G_s^v := (R_s^v)^t G R_s^v$ for any $s \in [0,T]$. Then the robust optimal control problem amounts to minimizing the maximal eigenvalue over the terminal values of the trajectories $\left(G_s^v\right)_{s\in[0,T]}$ when $v(\cdot)$ ranges in $L^1(0,T;U)$.

Underlying dynamical system

The curve $(G_s)_{s\in[0,T]}$ satisfies the ODE

$$\dot{G}_s = A^t(v(s))G_s + G_s A(v(s)), \qquad G_0 = G$$

in the linear space of matrices. However, in order to approximate the trajectory numerically and still preserve the positive definite character of the numerical solution, one would have to choose a sufficiently small time step. If instead one formulates the problem in the CAT(0) space of symmetric positive definite matrices, the geometry puts non-definite matrices at infinite distance, and naturally forbids degeneracy. Using a Taylor expansion and the explicit expression of the geodesic, one sees that the geodesic approximating $s \mapsto \exp(sA^t(v))G\exp(sA(v))$ has a right derivative equal to $V_G^v := G^{-1/2}A^t(v)G^{1/2} + G^{1/2}A(v)G^{-1/2}$. Then one computes the mutational dynamic as

$$f(G,v) := \mathcal{G}_{\mathcal{F}}(\alpha\kappa \circ d(\cdot, G_0)), \qquad \text{where} \qquad G_0 = \exp_G\left(2 \cdot V_G^v\right), \quad \text{and} \quad \alpha = \left|V_G^v\right|_G.$$

Up to normalization, we may assume that the determinant of the initial condition G equals 1. Taking then $U = \{-1,1\}$ as the set of controls, and $A : u \mapsto \begin{pmatrix} u & 0 \\ 0 & -u \end{pmatrix}$, we observe that $\det(G_s) = \exp(2s \operatorname{Trace} A)\det G = 1$, so that we may restrict the computations to the 2-dimensional convex surface $\det = 1$.

Analytical solution

On the surface $\det = 1$, the maximal eigenvalue is given by

$$\lambda_{\max}(G) = \frac{\operatorname{Trace} G}{2} + \sqrt{\frac{(\operatorname{Trace} G)^2}{4} - 1},$$

which is an increasing function of the trace. The trajectories issued from a point $G = (G_{ij})_{ij}$ and following constant controls write

$$s \mapsto \begin{pmatrix} e^{-2s}G_{11} & G_{12} \\ G_{21} & e^{2s}G_{22} \end{pmatrix} \qquad \text{for } u = -1, \text{ or} \qquad s \mapsto \begin{pmatrix} e^{2s}G_{11} & G_{12} \\ G_{21} & e^{-2s}G_{22} \end{pmatrix} \qquad \text{for } u = 1.$$

In order to decrease the maximal eigenvalue, hence the trace, the optimal choice is first to apply the constant control that lets e^{-2s} be in factor of $\max(G_{11}, G_{22})$. This will decrease the trace until both diagonal terms are equal, a situation reached at time $s^* = \frac{1}{4}\log(\max(G_{11}, G_{22})/\min(G_{11}, G_{22}))$. Once this critical time reached, it becomes optimal to stay frozen by applying the relaxed control $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. Consequently, the value function of the problem reads

$$V(t,G) = \frac{\tau_*}{2} + \sqrt{\frac{\tau_*^2}{4} - 1},$$

where $\tau_* := e^{-2\max(0,\min(T-t,s^*))}\max(G_{11}, G_{22}) + e^{2\max(0,\min(T-t,s^*))}\min(G_{11}, G_{22})$.

Numerical domain

The space Ω of symmetric positive definite matrices is not compact, and we have to restrict the computation to a subdomain of interest. We consider

$$\Omega_{\text{num}} := \left\{G \in \Omega \;\middle|\; \det(G) = 1 \text{ and } \exists v(\cdot) \in L^1(0,T;U),\ s \in [0,T] \text{ such that } \mathfrak{J}(G_s^v) \leqslant \Lambda\right\},$$

for some $\Lambda > 1$ arbitrarily fixed to 2 in the simulations. Consider the truncated cost $\widetilde{\mathfrak{J}} = \min(\Lambda, \mathfrak{J})$. By the DPP (2.29), the value function associated to $\widetilde{\mathfrak{J}}$ is exactly $\widetilde{V} := \min(\Lambda, V)$. Observe that by construction, there holds $\widetilde{V}(t,G) = \Lambda$ for any $(t,G) \in [0,T] \times (\Omega \setminus \Omega_{\text{num}})$. Hence we may restrict the computational domain to any set containing Ω_{num}, and assign the value Λ to any evaluation of $\hat{V}$ outside the computational domain, without lack of information. This trick may be seen as an HJB version of transparent boundary conditions.

Numerically, it is interesting to consider a parametrization of Ω by the coordinates (α, β, γ) as

$$G_{\alpha,\beta,\gamma} = \exp\left(\frac{\beta}{2}V\right)\exp(\alpha U + \gamma W)\exp\left(\frac{\beta}{2}V\right) = \exp\left(\frac{\alpha+\gamma}{2}\right)\cosh\left(\frac{\alpha-\gamma}{2}\right)\begin{pmatrix} c_{2\beta} + \tau & s_{2\beta} \\ s_{2\beta} & c_{2\beta} - \tau \end{pmatrix},$$

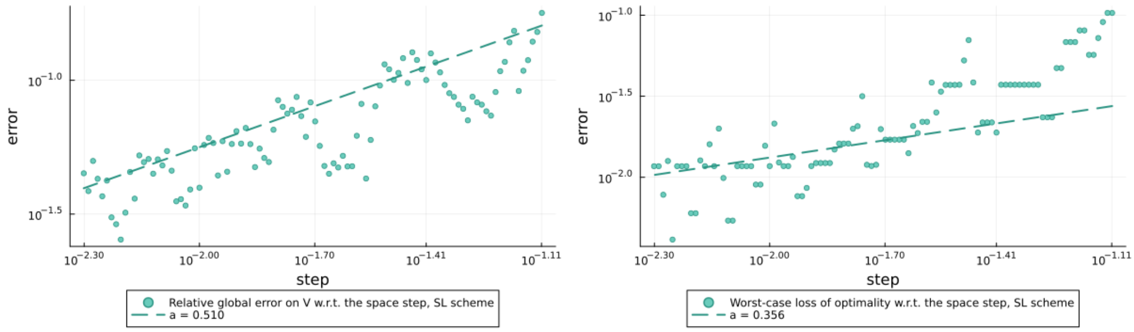


Figure 2.14: Evolution of the error in the robust control problem as a function of the space step Δx for $T = 1$.

Left: Relative error on V in the time-space domain. Right: worst-case loss of optimality. Each dot corresponds to a simulation, with a space step Δx ranging in 100 log-uniformly distributed values between 7.78×10^{-2} and 5×10^{-3}, and the time step is computed as to keep the CFL $\Delta x / \Delta t^2 = 36$. The loss of optimality is computed as the maximum Error *between the trajectories issued from 10 points that are randomly sampled with $(\alpha, \beta) \in [-1/2, 1/2]^2$ at the beginning of the simulation process.*

where $U = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $V = \begin{pmatrix} 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 \end{pmatrix}$, and $W = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ form an orthonormal basis of the tangent space to the identity, and $c_{2\beta} = \cosh(\beta/\sqrt{2})$, $s_{2\beta} = \sinh(\beta/\sqrt{2})$, $\tau = \tanh\left(\frac{\alpha-\gamma}{2}\right)$. The plane $\beta = 0$ is isometric to the Euclidean 2-dimensional plane, and the map $\beta \mapsto G_{\alpha,\beta,\gamma}$ is itself an isometry, so that meshes with nice geometric features are easily constructed in this coordinate system. One proves that Ω_{num} is contained in the image by $(\alpha, \beta) \mapsto G_{\alpha,\beta-\alpha}$ of

$$\left\{(\alpha, \beta) \;\middle|\; 2\cosh(\alpha)\cosh(\beta/\sqrt{2}) = \operatorname{Trace}\left(G_{\alpha,\beta,-\alpha}\right) \leq 2\cosh(\log(\Lambda))e^{2T}\right\},$$

which is compact since cosh is a strictly convex function bounded below by 1. Taking $\Lambda = 2$ and $T = 0.3$, we may then carry the numerical study over the image of the square $[-\overline{\alpha}, \overline{\alpha}] \times [-\overline{\beta}, \overline{\beta}]$ with $\overline{\alpha} = \overline{\beta}/\sqrt{2} = 1.5$.

Results

Figure 2.14 gathers the errors committed by Algorithms 3 and 4, and Table 2.5 collects the precise numerical values. The approximation of the value function is pictured in Figure 2.15, and the reconstruction of the feedback control in Figure 2.15. Any 2-dimensional graphical representation of the curved space Ω has to be distorted; with our choice, the distance between two points sharing the same α coordinate is respected, while matrices sharing the same β look closer than they actually are.

The approximation of the value function exhibits a numerical order of convergence of 1/2. The approximation of the optimal trajectories is more difficult to read, but the maximal errors seem to decrease with a similar order. The reconstruction of the feedback is meaningless in the region where $\hat{V}(t,x) = \hat{V}(t+h, y_{t+h}^{t,x,u}) = \Lambda$ for all controls (in grey in Figure 2.16), and the implementation chooses by default the first control.

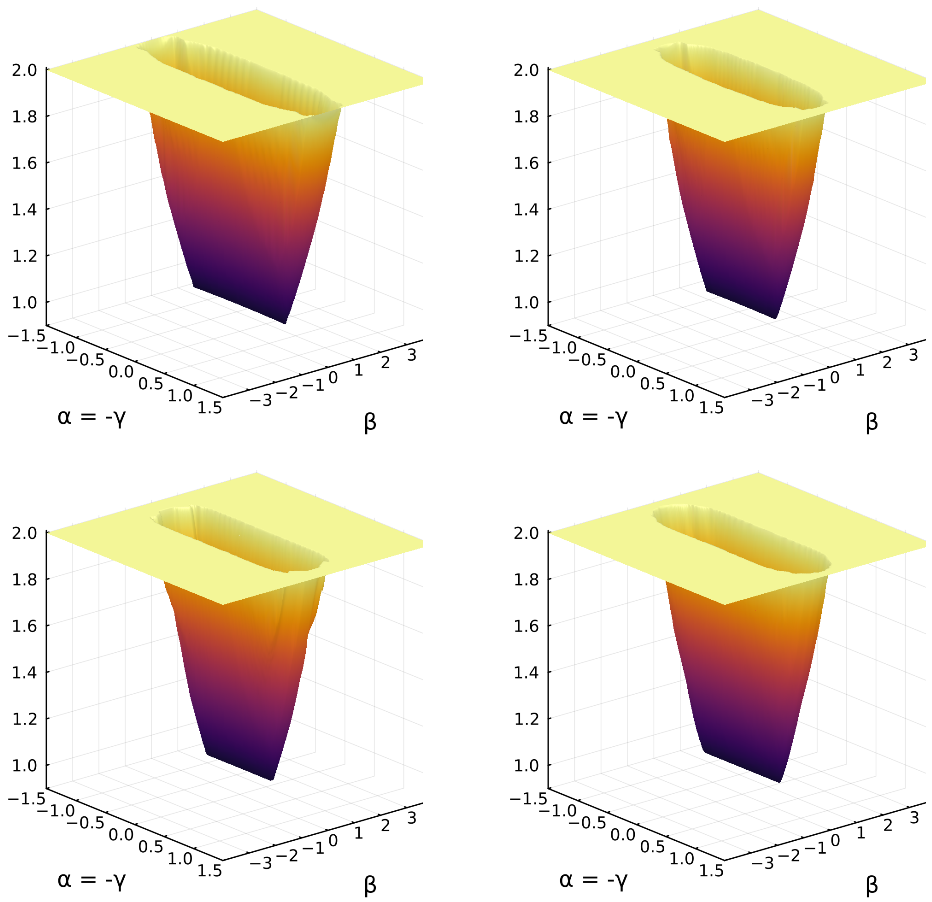


Figure 2.15: Numerical approximation of the value function for the robust control problem.

Results are presented for the first, 20^{th}, 50^{th} and last simulations appearing in Figure 2.14.

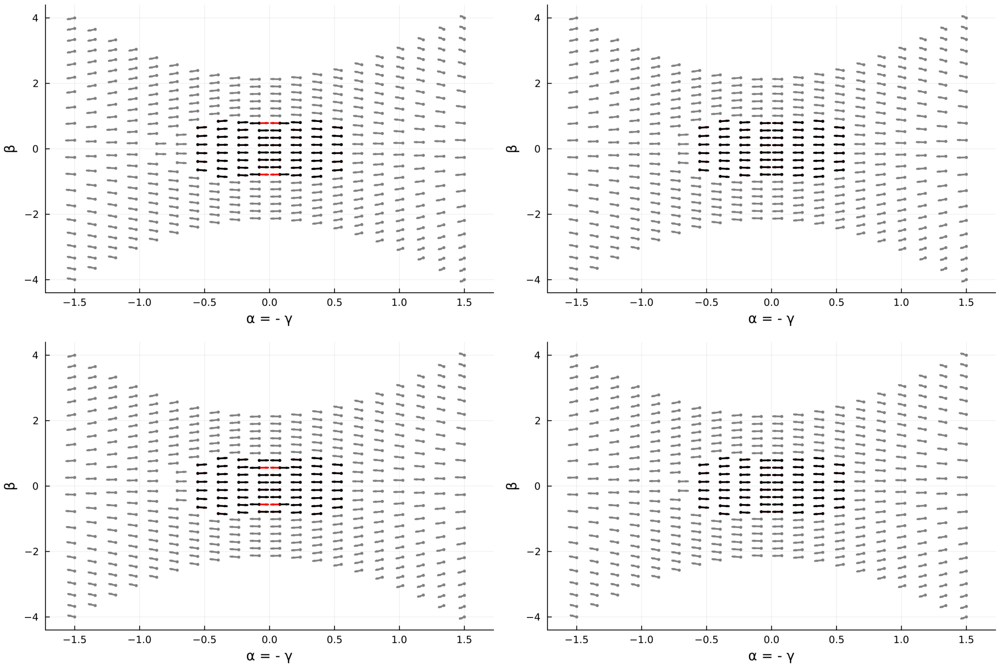


Figure 2.16: Numerical approximation of the feedback control for the robust control problem.

Results are presented for the first, 20^{th}, 50^{th} *and last simulations appearing in Figure 2.14. The feedback controls are computed on a dedicated mesh by using one step of Algorithm 4 starting from* $t = T - \Delta t$. *The arrows indicate the direction of the dynamic associated to the control, and are renormalized for clarity. Points for which the value function reaches the truncating constant are displayed in grey. The numerical feedback is displayed in black, whereas the theoretical optimal feedback is indicated in red.*

Parameters		Errors on V		Errors on $\hat{y}$		Time (s)
Δt	Δx	global	at $t = 0$	worst Error	mean Error	
4.29e-02	7.78e-02	1.79e-01	1.79e-01	1.03e-01	4.17e-02	1.31
3.33e-02	4.59e-02	8.66e-02	8.66e-02	3.72e-02	1.51e-02	2.15
2.50e-02	2.64e-02	8.16e-02	8.16e-02	3.84e-02	1.87e-02	5.57
2.00e-02	1.52e-02	4.95e-02	4.95e-02	1.61e-02	9.02e-03	18.62
1.50e-02	8.71e-03	3.53e-02	3.53e-02	1.17e-02	5.00e-03	77.00
1.15e-02	5.00e-03	4.49e-02	4.49e-02	1.17e-02	4.69e-03	311.79

Table 2.5: Evolution of errors in the robust control problem.

Errors are shown for the first, 20th*, 40*th*, 60*th*, 80*th *and last simulation appearing in Figure 2.14. The CFL* $\Delta x/\Delta t^2$ *is kept constant and equal to* 36. *The errors on the value function V are normalized, and the CPU time is in seconds.*

Chapter 3

Viscosity solutions in non-negatively curved spaces

This chapter focuses on viscosity solutions to Hamilton-Jacobi-Bellman equations in a complete geodesic CBB(0) space (Ω, d). The first part is devoted to a comparison principle for a notion of viscosity solution based on semiconcave/semiconvex test functions, with no restriction on the “testing” points.

The second part of the chapter focuses on optimal control problems, formulated in the Wasserstein space. In this space, it is more advantageous to define viscosity solutions with L-differentiable test functions. However, for the control problems under investigation, the comparison principle for general CBB(0) spaces applies, and we are able to characterize the value function as the unique viscosity solution, in the sense of Definition 3.1.6 below, of a suitable HJB equation. We conclude with some extensions, some of which being treated with adaptations of classical viscosity techniques, and some of which requiring arguments that are specific to the measure setting.

The content of this chapter is derived from [AJZ24], in collaboration with Othmane Jerhaoui and Hasnaa Zidani, and [AH24], in collaboration with Cristopher Hermosilla.

Contents of Chapter 3

In all this chapter, (Ω, d) is a complete geodesic CBB(0) space.

3.1 A comparison principle

Let $T > 0$, $\mathfrak{J} : \Omega \to \mathbb{R}$, and $H : \mathbb{T} \to \mathbb{R}$. Consider the parabolic first-order Hamilton-Jacobi equation

$$\begin{cases} -\partial_t u(t,x) + H(x, D_x u(t,x)) = 0 & (t,x) \in (0,T) \times \Omega, \quad (3.1a) \\ u(T,x) = \mathfrak{J}(x) & x \in \Omega. \quad (3.1b) \end{cases}$$

Our aim is to provide a notion of viscosity solution for (3.1) that supports a strong comparison principle. Our proposition is inspired from the previous chapter, but differs in the sense that one cannot use semiconvex test functions. We also treat the case of a space that is not locally compact.

3.1.1 Viscosity solutions

We introduce some regularity conditions for the test functions and the semisolutions.

Definition 3.1.1 (Locally semiconcave function). *Let (X,d) be a geodesic space. A function $\varphi : X \to \mathbb{R}$ is locally semiconcave if for any $x \in X$, there exists $R > 0$ and a constant λ_R depending on R such that for any geodesic $\gamma \in AC([0,1];X)$ completely contained in $\mathscr{B}_X(x,R)$, there holds*

$$\varphi\left(\gamma(t)\right) \geqslant (1-t)\varphi(\gamma(0)) + t\varphi(\gamma(1)) - \frac{\lambda_R}{2} t(1-t) d^2(\gamma(0),\gamma(1)) \qquad \forall t \in [0,1].$$

A function $\psi : X \to \mathbb{R}$ is locally semiconvex if $-\psi$ is locally semiconcave.

Remark 3.1.2 (Composition rule)**.** *Let $\varphi \in \mathcal{C}^2\left(\mathbb{R}^+;\mathbb{R}^+\right)$ be nondecreasing, and consider the composition $\psi : x \mapsto \varphi(d^2_{\mathcal{W}}(o,x))$. Denote λ_R a local constant of semiconcavity of φ over $[0,3R]$, and Lip$(\varphi)R$ a local constant of Lipschitz-continuity of φ over the same domain. Then ψ is semiconcave with modulus $R\lambda_R +$ Lip$(\varphi)R$ (see [CS04, Proposition 2.1.12]).*

Definition 3.1.3 (Locally uniform semicontinuity). *Let (X,d) be a complete metric space. A locally bounded function $\varphi : X \to \mathbb{R}$ is locally uniformly upper semicontinuous (luusc) if for any decreasing family of closed bounded sets $(B_n)_{n\in\mathbb{N}}$ that converges to its intersection, in the sense that $B := \bigcap_{n\in\mathbb{N}} B_n$ is not empty, and $\lim_{n\to\infty} \sup_{x\in B_n} \inf_{y\in B} d(x,y) = 0$, there holds*

$$\lim_{n\to\infty} \sup_{y\in B_n} \varphi(y) \leqslant \sup_{x\in B} \varphi(x).$$

A function $\psi : X \to \mathbb{R}$ is locally uniformly lower semicontinuous (lulsc) is $-u$ is luusc.

Remark 3.1.4 (On locally uniform upper semicontinuity)**.** *In general, Definition 3.1.3 is stronger than upper semicontinuity. For instance, let $X = \ell^2$ and $\varphi(x) = \mathbb{I}_{\cup_i e_i}$ be the indicator of the canonical basis, i.e. the set of the elements $e_i = (0,\cdots,0,1,0,\cdots)$ with a 1 in the i^{th} position. The latter set is closed, so φ is upper semicontinuous. However, we may take B_n as the subset of those x in the unit ball such that $|\langle u,x\rangle| \leqslant e^{-n}$, where $u = (e^{-n})_{n\in\mathbb{N}}$. Each B_n is closed and bounded, and the family $(B_n)_{n\in\mathbb{N}}$ decreases towards its intersection $\{u\}^\perp$. Now, $\sup_{B_n} \varphi = 1$ for all n, but $\sup_{\{u\}^\perp} \varphi = 0$, so φ is not luusc.*

If (X,d) has compact balls, then both notions coincide. Indeed, if φ is usc, let $(B_n)_n$ be as in the definition. The family $(B_n)_{n\in\mathbb{N}}$ eventually lies in a (compact) large ball, and one can pick a sequence of elements in each B_n which approximate the sup of φ with error $O(1/n)$. Any limit point will belong to B by the approximation property, so using the classical upper semicontinuity of φ is enough to prove locally uniform upper semicontinuity.

The applications which are both luusc and lulsc are the locally uniformly continuous applications. Indeed, if φ is locally uniformly continuous, one can pick a sequence y_n in B_n almost realizing the sup, then find $x_n \in B$ very close to y_n by the assumption of Hausdorff convergence, and conclude by the uniform *continuity of φ in a ball containing B and the B_n. Conversely, if φ is luusc and lulsc, then so is $(x, y) \mapsto \pm(\varphi(x) - \varphi(y))$, and*

$$\lim_{r \searrow 0} \sup_{\substack{x,y \in \overline{\mathscr{B}}(o,R) \\ d(x,y) \leqslant r}} \left|\varphi(x) - \varphi(y)\right| \leqslant \max_{s \in \{-1,1\}} \lim_{n \to \infty} \sup_{\substack{x,y \in \overline{\mathscr{B}}(o,R) \\ d(x,y) \leqslant e^{-n}}} s(\varphi(x) - \varphi(y)) \leqslant 0.$$

The leftmost term, taken as a function of r, furnishes a local modulus of continuity.

Lastly, Definition 3.1.3 is equivalent to the upper semicontinuity of the function $\Phi : \mathcal{S} \to \mathbb{R}$ given by $\Phi(S) = \sup_{x \in S} \varphi(x)$, where $\mathcal{S}$ is the set of nonempty closed and bounded sets of X endowed with the Hausdorff distance. This makes it a natural assumption when dealing with control problems, in which the viscosity solution is expected to behave with respect to sup as solutions of linear PDEs behave with respect to convolution [KM97].

We introduce the following sets of test functions. This very definition does not appear in the literature, but is quite close in spirit of the sets considered in metric viscosity solutions, where usually the test functions are defined as sums of various components with different roles of penalizations, as in [AF14, Eqs. (1.12) and (1.16)], [GŚ15b, Def. 2.2], or [CKT23a, Def. 3.13].

Definition 3.1.5 (Test functions). *Let $T > 0$. The sets of test functions are defined as*

$$\mathscr{T}_\pm := \left\{ \varphi : (0,T) \times \Omega \to \mathbb{R} \;\middle|\; \begin{array}{c} \varphi \text{ and } \partial_t \varphi \text{ are locally Lipschitz, and} \\ \text{for all } t \in (0,T),\ \pm\varphi(t,\cdot) \text{ is locally semiconcave} \end{array} \right\}. \tag{3.2}$$

The regularity of $\partial_t \varphi$ implies in particular that the directional derivative along time-space directions splits as the sum of the partial derivative in the time variable, and the directional derivative in the space variable.

As opposite to Chapter 2, the set $\mathscr{T}_+$ is made of locally semiconcave functions in the space variable, instead of semiconvex. This follows the curvature of the squared distance of the space. Apart from this, the definition of viscosity solutions is essentially unchanged.

Definition 3.1.6 (Viscosity solution of (3.1)). *An application $u : [0,T] \times \Omega \to \mathbb{R}$ is a*

- *viscosity subsolution of (3.1) if it is locally uniformly upper semicontinuous, and if for any $\varphi \in \mathscr{T}_+$ such that $u - \varphi$ reaches a maximum at $(t,x) \in [0,T) \times \Omega$, there holds*

$$-\partial_t \varphi(t,x) + H\big(x, D_x \varphi(t,x)\big) \leqslant 0, \tag{3.3}$$

- *viscosity supersolution of (3.1) if it is locally uniformly lower semicontinuous, and if for any $\varphi \in \mathscr{T}_-$ such that $u - \varphi$ reaches a minimum at $(t,x) \in [0,T) \times \Omega$, there holds*

$$-\partial_t \varphi(t,x) + H\big(x, D_x \varphi(t,x)\big) \geqslant 0, \tag{3.4}$$

- *viscosity solution of (3.1) if it is both a subsolution and a supersolution, and satisfies $u(T,\cdot) = \mathfrak{J}$.*

Remark 3.1.7. *This definition has two specificities: first, the choice of test functions, which is designed to define $H(x, D_x\varphi)$ unambiguously. Secondly, the regularity of semisolutions is stronger than upper semicontinuity/lower semicontinuity. This allows us to treat the case of a non locally compact space Ω in combination with an Ekeland-type principle, and is taken from a similar assumption in [FGŚ17] at the terminal time T.*

Remark 3.1.8. *The condition $u(T,\cdot) = \mathfrak{J}$ could be distributed in $u(T,\cdot) \leqslant \mathfrak{J}$ for subsolutions, and $u(T,\cdot) \geqslant \mathfrak{J}$ for supersolutions. The present formulation simplifies further statements. Similarly, the semisolutions could take values in $\mathbb{R} \cup \{-\infty\}$ for the subsolution, and $\mathbb{R} \cup \{+\infty\}$ for the supersolution.*

Remark 3.1.9 (Elements of comparison with metric viscosity solutions)**.** *In the case of Eikonal equations, one can define viscosity solutions through growth conditions, either on the metric slope as in [AF14], or along paths as in [GHN15]. The precise definitions are quite convoluted since adapted to the use of Ekeland-type principles, and we only compare ideas. In [AF14], test functions are taken as compositions of distance functions, on which the metric slope can be computed by maximum on the values of the directional derivatives. Consequently, we expect the corresponding definition to be morally equivalent to Definition 3.1.6. In [GHN15], the equation* $|Du| = f(x)$ *is understood as (up to regularity, boundaries and* ε*'s)*

- *for all curve* $\gamma \in AC(\mathbb{R};\Omega)$ *and test function* $\psi \in \mathcal{C}^1(\mathbb{R};\mathbb{R})$ *touching* $u \circ \gamma$ *from above at* $t = 0$*, there holds* $|\psi'(0)| \leqslant f(\gamma_0)$*;*
- *for all* x*, there exists* $\gamma \in AC(\mathbb{R};\Omega)$ *with* $\gamma_0 = x$ *such that* $|\psi'(t)| \geqslant f(\gamma_t)$ *for all* $\psi \in \mathcal{C}^1$ *touching* $u \circ \gamma$ *from below at some* t*.*

Define $H : \mathbb{T} \to \mathbb{R}$ *as* $H(x,p) := \sup_{v\in \mathrm{T}_x\Omega, |v|=1} -p(v) - f(x)$*. Let* u *satisfy the first item of the above. For any semiconcave test function* φ *touching* u *from above at* x*, the value of* $H(x, D_x\varphi)$ *can be approximated by* $-D_x\varphi(v_n) - f(x)$ *for a maximizing sequence* $(v_n)_n$ *of velocities of geodesics. By semiconcavity*[1]*,* $-D_x\varphi(v_n) = \lim_{t\searrow 0} -D_{\gamma^n_t}\varphi((\gamma_n)^+_t)$*, so applying the metric viscosity condition on the curves* γ^n *at suitable times* t_n *converging fast enough to 0, we recover that* u *is a subsolution in the sense of Definition 3.1.6. This indicates that the pathwise metric definition is stronger than the one presented in this manuscript. Conversely, if* u *satisfies (3.3) and* γ *is a given AC curve, one would like to extend any* ψ *in a suitable semiconcave test function. This seems to us more intricate.*

3.1.2 A preparatory lemma

Lemma 3.1.10 (Adaptation of Ekeland-Borwein-Preiss-Zhu)*. For convenience, denote* $Y := ([0,T]\times\Omega)^2$*. Let* $\Phi : Y \to \mathbb{R}$ *be upper semicontinuous and upper bounded,* $z^0 \in Y$ *be fixed such that* $A := \sup\Phi - \Phi(z^0) < \infty$*,* $C \geqslant 0$*, and assume that there exists* $R > 0$ *such that*

$$\left\{z \in Y \;\middle|\; \Phi(z) \geqslant \Phi(z^0)\right\} \subset \overline{\mathscr{B}}(z^0, R). \tag{3.5}$$

There exist $\omega_{T,R,A} : \mathbb{N} \to \mathbb{R}^+$ *going to* 0 *when* n *goes to* ∞*, and for each* n*, a point* $\xi_n = \big((t_n,x_n),(s_n,y_n)\big) \in \overline{\mathscr{B}}(z^0,R)$ *and a perturbation* $p_n : Y \to \mathbb{R}$*, such that*

$$\Phi - p_n \text{ reaches a global strict maximum at } \xi_n, \tag{3.6}$$

$$(t,x) \mapsto p_n((t,x),(s_n,y_n)) \in \mathscr{T}_+, \qquad (s,y) \mapsto p_n((t_n,x_n),(s,y)) \in \mathscr{T}_-, \tag{3.7}$$

$$\sum_{r\in\{s,t\}} \left|\partial_r p_n(\xi_n)\right| + C \sum_{z\in\{x,y\}} (1 + d(z_n,o))\|D_z p_n(\xi_n)\|_{z_n} \leqslant \omega_{T,R,A}(n), \tag{3.8}$$

$$\sup\Phi - \Phi(\xi_n) \leqslant \omega_{T,R,A}(n). \tag{3.9}$$

Proof. Denote $d_Y^2(\xi,\xi') := |t-t'|^2 + d^2(x,x') + |s-s'|^2 + d^2(y,y')$ the distance on Y. The metric space (Y, d_Y) is complete, and Φ satisfies all the assumptions of the Ekeland-Borwein-Preiss-Zhu theorem [BZ05, Theorem 2.5.2]. We consider the gauge-type function d_Y^2, and the choice of ponderation

$$\alpha_{n,m} := \frac{1}{(n+1)2^{m+1}}, \qquad \text{so that} \qquad \alpha_{n,0} = \frac{1}{2(n+1)} \quad \text{and} \quad \sum_{m\in\mathbb{N}} \alpha_{n,m} = \frac{1}{n+1}.$$

Applying Ekeland-Borwein-Preiss-Zhu, we get the existence of $z_n \in Y$ and $(z_{n,m})_{m\in\mathbb{N}} \subset Y$ such that

$$\begin{cases} d_Y^2(z^0,z_n) \leqslant \dfrac{A}{\alpha_{n,0}}, \qquad d_Y^2\left(z_{n,m},z_n\right) \leqslant \dfrac{A}{2^m\alpha_{n,0}} \qquad \forall m\in\mathbb{N}, & (3.10a)\\ \Phi(z_n) \geqslant \Phi(z^0) + \displaystyle\sum_{m\in\mathbb{N}} \alpha_{n,m} d_Y^2(z^0,z_{n,m}), & (3.10b)\\ \Phi(z_n) - \displaystyle\sum_{m\in\mathbb{N}} \alpha_{n,m} d_Y^2(z_n,z_{n,m}) > \Phi(\cdot) - \sum_{m\in\mathbb{N}} \alpha_{n,m} d_Y^2(\cdot,z_{n,m}) \quad \forall z \neq z_n. & (3.10c) \end{cases}$$

[1]The function $\kappa_n := t \mapsto -\varphi(\gamma^n_t)$ is semiconvex from $[0,1]$ to $\mathbb{R}$, so convex up to a quadratic term. By upper semicontinuity of the directional derivative of convex functions [Roc70, Theorem 24.5], $\kappa'_n(0) \geqslant \limsup_{t\searrow 0} \kappa'_n(t)$. The reverse inequality holds by semiconvexity, so $-D_x\varphi(v_n) = \kappa'_n(0) = \lim_{t\searrow 0}\kappa'_n(t) = \lim_{t\searrow 0} -D_{\gamma^n_t}\varphi((\gamma_n)^+_t)$.

Define $p_n : z \mapsto \sum_{m\in\mathbb{N}} \alpha_{n,m} d_Y^2(z, z_{n,m}) \geqslant 0$. Then using (3.10a),

$$p_n(z) = \sum_{m\in\mathbb{N}} \alpha_{n,m} d_Y^2(z, z_{n,m}) \leqslant 2 \sum_{m\in\mathbb{N}} \alpha_{n,m} \left(d_Y^2(z, z_n) + d_Y^2(z_n, z_{n,m})\right) \leqslant 2\frac{d_Y^2(z, z_n)}{n+1} + \frac{A}{(n+1)\alpha_0} \sum_{m=0} 4^{-m} < \infty.$$

Hence the map p_n is well-defined from Y to $\mathbb{R}^+$. By (3.10c), $\Phi - p_n$ reaches a global strict maximum in z_n.

We turn to Points (3.7) and (3.8). The application $p_n((\cdot, x), (s, y))$ is of the form $c + \sum_{m\in\mathbb{N}} 2^{-m-1} \frac{|\cdot - t_{n,m}|^2}{n+1}$, over a bounded interval, so uniformly convergent. Moreover, its derivative is Lipschitz in $[0, T]$ with constant $2/(n+1)$, and

$$\left|\partial_t p_n((t, x_n), (s_n, y_n))\right| \leqslant \frac{1}{n+1} \sum_{m\in\mathbb{N}} 2^{-m} \left|t - t_{n,m}\right| \leqslant \frac{2T}{n+1} \qquad \forall t \in [0, T]. \tag{3.11}$$

As $(t, x) \mapsto p_n((t, x), (s_n, y_n))$ writes as a sum of time and measure contributions, its derivative with respect to t is Lipschitz in the whole domain. Moreover, by (3.10b), $z_n \in \left\{\Phi \geqslant \Phi(z^0)\right\}$. As $x_{n,m} \to_m x_n$, the sequence $(x_{n,m})_m$ stays in a bounded set of Ω, and the partial function $p_n((t, \cdot), (s, y))$ is uniformly convergent for each n. Using the semiconcavity of $d^2(\cdot, x_{n,m})$, there holds for any unit-speed geodesic $\gamma \subset \Omega$ and $h \in [0, 1]$ that

$$\begin{aligned} p_n\left((t_n, \gamma(h)), (s_n, y_n)\right) &\geqslant \sum_{m\in\mathbb{N}} \alpha_{n,m} \left[(1-h) d^2(\gamma(0), x_{n,m}) + h d^2(\gamma(1), x_{n,m}) - h(1-h) d^2(\gamma(0), \gamma(1))\right] + \text{cte} \\ &= (1-h) p_n\left((t_n, \gamma(0)), (s_n, y_n)\right) + h p_n\left((t_n, \gamma(1)), (s_n, y_n)\right) - \frac{h(1-h)}{n+1} d^2(\gamma(0), \gamma(1)). \end{aligned}$$

Thus $p_n((t_n, \cdot), (s_n, y_n))$ is locally semiconcave. To prove that $(t, x) \mapsto p_n((t, x), (s_n, y_n))$ belongs to $\mathcal{T}_+$, there only stays to show the local Lipschitzianity in the space variable. For any $S > 0$ and $x, z \in \overline{\mathcal{B}}(o, S)$, one has

$$\frac{\left|p_n((t_n, x), (s_n, y_n)) - p_n((t_n, z), (s_n, y_n))\right|}{d(x, z)} \leqslant \sum_{m\in\mathbb{N}} \alpha_{n,m} \left(d(x, x_{n,m}) + d(z, x_{n,m})\right) \leqslant \frac{S + \sqrt{2(n+1)A}}{n+1} \sum_{m\in\mathbb{N}} \frac{1 + 2^{-m/2}}{2^m}.$$

Here we used (3.10a). This proves that $p_n(\cdot, (s_n, y_n)) \in \mathcal{T}_+$. Let $R > 0$ be given by the assumption (3.5) such that $d_Y\left(z_n, z^0\right) \leqslant R$ independantly of n. By the above, there holds

$$C\left(1 + d(x_n, o)\right) \|D_x p_n(z_n)\|_{x_n} \leqslant C(1+R) \frac{R + \sqrt{2(n+1)A}}{n+1} \sum_{m\in\mathbb{N}} \frac{1 + 2^{-m/2}}{2^m}. \tag{3.12}$$

Summing (3.11) and (3.12), we obtain a bound $\omega^{(0)}_{T,R,A}$ that decreases in $n^{-1/2}$. The reasoning over $(s, y) \mapsto p_n((t_n, x_n), (s, z))$ is symmetric.

Finally, notice that the supremum of Φ over Y is the same as the supremum of Φ over $\overline{\mathcal{B}}(z^0, R)$. In consequence, (3.10c) gives

$$\begin{aligned} \sup \Phi - \Phi(z_n) &= \sup_{z\in\overline{\mathcal{B}}(z^0,R)} \Phi(z) \leqslant \sup_{z\in\overline{\mathcal{B}}(z^0,R)} \sum_{m\in\mathbb{N}} \alpha_{n,m} \left[d_Y^2(z, z_{n,m}) - d_Y^2(z_n, z_{n,m})\right] \\ &\leqslant \sup_{z\in\overline{\mathcal{B}}(z^0,R)} \sum_{m\in\mathbb{N}} \alpha_{n,m} \left[d_Y(z, z_{n,m}) + d_Y(z_n, z_{n,m})\right] d_Y(z, z_n) \\ &\leqslant \sum_{m\in\mathbb{N}} \alpha_{n,m} \left[2R + 2 d_Y(z_n, z_{n,m})\right] 2R \leqslant \sum_{m\in\mathbb{N}} \frac{1}{2^{m+1}(n+1)} \left[2R + 2\sqrt{\frac{(n+1)A}{2^{m+1}}}\right] 2R. \end{aligned}$$

Hence Point (3.9) by choosing $\omega^{(1)}_{R,A} : n \mapsto \frac{4R^2}{n+1} + \frac{2R\sqrt{A}}{\sqrt{n+1}(2\sqrt{2}-1)}$, and $\omega_{T,R,A} := \max(\omega^{(0)}_{T,R,A}, \omega^{(1)}_{R,A})$. □

3.1.3 The comparison principle

We follow the vague program given in Algorithm 1 p. 17.

Viscosity subsolutions satisfy two sets of order relations. First, $v(T, x) \leqslant \mathfrak{J}(x)$ on the parabolic boundary. Secondly, on each $(t, x) \in [0, T) \times \Omega$, if

$$v \leqslant_{(t,x)} \varphi, \quad \text{in the sense that } v - \varphi \text{ reaches a maximum at } (t, x),$$

then the inequality (3.3) holds. The comparison principle states that a subsolution v and a supersolution w that are ordered at the boundary, are ordered on the whole domain, in the sense that $v(t,x) \leqslant w(t,x)$ for all $(t,x) \in [0,T] \times \Omega$.

The strategy is to assume that $v(t,x) > w(t,x)$ at some point, and to construct sequences $(t_n, x_n)_{n\in\mathbb{N}}$ and $(s_n, y_n)_{n\in\mathbb{N}}$, jointly with test functions $(\varphi_n)_{n\in\mathbb{N}} \subset \mathcal{T}_+$ and $(\psi_n)_{n\in\mathbb{N}} \subset \mathcal{T}_-$, such that $v \leqslant_{(t_n,x_n)} \varphi_n$ and $\psi_n \leqslant_{(s_n,y_n)} w$. The definition of semisolutions then provides a first inequality between the Hamiltonians. On the other hand, the sequences are constructed so that the opposite inequality holds strictly for n large enough, a contradiction.

The construction of φ_n and ψ_n uses the Kružkov method of doubling of variables [CIL92]. One considers

$$\Phi_{\iota,\varepsilon}((t,x),(s,y)) = v(t,x) - w(s,y) - G_\varepsilon((t,x),(s,y)) - \Pi_\iota((t,x),(s,y)) - \ell_\alpha((t,x),(s,y)).$$

Morally speaking, G_ε should be chosen so that when (t,x) is sufficiently close to (s,y), one has approximately

$$-\partial_t G_\varepsilon + H\big(x, D_x G_\varepsilon((t,x),(s,y))\big) \geqslant -\partial_s(-G_\varepsilon) + H\big(y, D_y(-G_\varepsilon)((t,x),(s,y))\big). \tag{3.13}$$

Classically, $G_\varepsilon((t,x),(s,y)) = \frac{1}{\varepsilon}\left(|t-s|^2 + d^2(x,y)\right)$. If $\mathrm{T}\Omega$ splits into $\Omega \times V$, (3.13) is granted as soon as H is Lipschitz with respect to x, the velocity variable being fixed. Otherwise, a form of (3.13) can be *assumed* as a regularity condition on H; see the discussion of [FK09, §1.2.2] on this topic. In the case of discontinuities, Imbert and Monneau [IM17] constructed G_ε by a fine examination of the level sets of the Hamiltonian. Here we stick to the usual choice.

Π_ι is a penalization term allowing to obtain maxima of $\Phi_{\iota,\varepsilon}$, or near-maxima from an Ekeland principle. This provides the sequence $((t_n,x_n),(s_n,y_n))$, and φ_n, ψ_n as the partial functions of $\Phi_{\iota,\varepsilon}$ when one variable is fixed. One has to control the derivatives of Π_ι appearing in the viscosity inequalities, either by a correct sign, or using the (near-)optimality to force the terms involving Π_ι to vanish. The "trick" in the proof below combines both.

The function ℓ_α adds a small parameter in (3.13) in order to make it strict, and eventually contradict the viscosity inequalities. In the parabolic case, it is simple to construct; in more general cases, it can be quite intricate. The existence of some ℓ_α, or a change of variable, or a general procedure to pass from the inequality $v(t,x) > w(t,x)$ to a strict inequality on the Hamiltonians, is the increasing monotonicity that discriminates the Hamiltonians for which uniqueness holds.

This introduction justifies the following assumptions on the Hamiltonian.

Assumption [A3.1.11] (Structure of the Hamiltonian). Let $o \in \Omega$ be fixed. There exists a constant $C \geqslant 0$ such that

$$\big|H(x,p) - H(x,q)\big| \leqslant C\,(1 + d(x,o))\,\|p-q\|_x \qquad \forall x \in \Omega,\ p,q \in \mathbb{T}_x. \tag{3.14}$$

Moreover, for any $r \geqslant 0$, there exists a constant $C_r \geqslant 0$ such that

$$H\big(y, -\lambda D_y d^2(x,\cdot)\big) - H\big(x, \lambda D_x d^2(\cdot,y)\big) \leqslant C_r\, d(x,y)\,\big(\lambda d(x,y) + 1\big) \qquad \forall x,y \in \overline{\mathscr{B}_\Omega}(o,r) \text{ and } \lambda \geqslant 0. \tag{3.15}$$

Theorem 3.1.12 (Comparison principle). *Assume that H satisfies **[A3.1.11]**. Let $v, w : [0,T] \times \Omega \to \mathbb{R}$ be locally bounded, and respectively subsolution and supersolution of (3.1) in the sense of Definition 3.1.6. Then*

$$\sup_{(t,x)\in[0,T]\times\Omega} v(t,x) - w(t,x) \leqslant \sup_{x\in\Omega} v(T,x) - w(T,x).$$

Proof. Assume without loss of generality that $\sup_{x\in\Omega} v(T,x) - w(T,x) < \infty$. Since the Hamiltonian does not depend on $u(x)$, we can add a constant to w in order that $\sup_{x\in\Omega} v(T,x) - w(T,x) \leqslant 0$, and still get a supersolution. Assume by contradiction that there exists $(t^0,x^0) \in [0,T] \times \Omega$ such that $v(t^0,x^0) > w(t^0,x^0)$.

Construction of ℓ_α and penalizations. By assumption, one can find some $\alpha > 0$ small enough as to let

$$\Gamma := v(t^0,x^0) - w(t^0,x^0) - \ell_\alpha(t^0) > 0,$$

where $\ell_\alpha(t) = \alpha(T-t)$. Since v and w are upper bounded on balls, one can construct an increasing growth function $g \in \mathcal{C}^1(\mathbb{R}^+;\mathbb{R}^+)$ such that

$$\sup_{(t,x)\in[0,T]\times\mathscr{B}(o,R)} \max(v(t,x), -w(t,x)) \leqslant g(R^2),$$

and $\lim_{R\to\infty} g(R^2) = \infty$. However, using g directly as a penalization of the space variable produces a term that is not controlled in the estimates in the end of the proof. We follow [FGŚ17] in using the time variable to compensate for the space one, by building a smooth function $h = h(r,z) : [0,T]\times\Omega \to \mathbb{R}$ satisfying

$$\partial_r h(r,z) + C(1+d(o,z))\|D_z h(r,z)\|_z \leqslant 0 \qquad \forall (r,z) \in [0,T]\times\Omega. \tag{3.16}$$

The exploding derivative in space is controlled by the (no less exploding) derivative in time. The map

$$h(r,z) := g^2\left((1+d^2(o,z))e^{4C(T-r)}\right)$$

satisfies (3.16) (using the coarse estimate $r + r^2 \leqslant 2(1+r^2)$ for all $r \geqslant 0$), and for any $\iota > 0$, the function $v(t,x) - \iota h(t,x)$ goes to $-\infty$ when $d(o,x) \to \infty$. The same holds for $-w(s,y) - \iota h(s,y)$. Moreover, both h and $\partial_r h$ are locally Lipschitz, and $z \mapsto h(r,z)$ is locally semiconcave by composition (see Remark 3.1.2), so that h can be used to construct test functions.

Construction of near-maxima and separating test functions. Consider $\Phi_{\iota,\varepsilon} : ([0,T]\times\Omega)^2 \to \mathbb{R}$ the function

$$\Phi_{\iota,\varepsilon}((t,x),(s,y)) := v(t,x) - w(s,y) - \ell_\alpha(t) - \frac{(t-s)^2 + d^2(x,y)}{2\varepsilon} - \iota\left(h(t,x) + h(s,y)\right). \tag{3.17}$$

For each $\iota, \varepsilon > 0$, the function $\Phi_{\iota,\varepsilon}$ is upper semicontinuous. Moreover, the set $\Phi_{\iota,\varepsilon} \geqslant \Phi_{\iota,\varepsilon}((T,o),(T,o))$ is contained in a ball of radius $R = R_\iota$, on which $\Phi_{\iota,\varepsilon}$ is bounded from above by a constant $A = A_\iota$ independently of ε. Applying Lemma 3.1.10, there exist $\omega_\iota : \mathbb{N} \to \mathbb{R}^+$ going to 0 when n goes to ∞, sequences of points $X_{\iota\varepsilon n} := \left((t_{\iota\varepsilon n}, x_{\iota\varepsilon n}),(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})\right)_{n\in\mathbb{N}} \subset [0,T]\times\overline{\mathscr{B}}(o,R_\iota)$ and penalizations $(p_{\iota\varepsilon n})_{n\in\mathbb{N}} \subset \mathcal{C}\left(([0,T]\times\Omega)^2;\mathbb{R}\right)$, such that

$$\Phi_{\iota,\varepsilon} - p_{\iota\varepsilon n} \text{ reaches a maximum at } ((t_{\iota\varepsilon n}, x_{\iota\varepsilon n}),(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})), \tag{3.18}$$

$$(t,x) \mapsto p_{\iota\varepsilon n}((t,x),(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})) \in \mathscr{T}_+, \qquad (s,y) \mapsto p_{\iota\varepsilon n}((t_{\iota\varepsilon n}, x_{\iota\varepsilon n}),(s,y)) \in \mathscr{T}_-, \tag{3.19}$$

$$\sup \Phi_{\iota,\varepsilon} - \Phi_{\iota,\varepsilon}\left((t_{\iota\varepsilon n}, x_{\iota\varepsilon n}),(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})\right) \leqslant \omega_\iota(n), \tag{3.20}$$

$$\sum_{\binom{r}{z}=\binom{t}{x},\binom{s}{y}} \left|\partial_r p_{\iota\varepsilon n}(X_{\iota\varepsilon n})\right| + C(1+d(z_{\iota\varepsilon n}, o))\|D_z p_{\iota\varepsilon n}(X_{\iota\varepsilon n})\|_{z_{\iota\varepsilon n}} \leqslant \omega_\iota(n). \tag{3.21}$$

Define the separating functions $\varphi_{\iota\varepsilon n}, \psi_{\iota\varepsilon n} : [0,T]\times\Omega \to \mathbb{R}$ by

$$\varphi_{\iota\varepsilon n}(t,x) := w(s_{\iota\varepsilon n}, y_{\iota\varepsilon n}) + \iota\left(h(t,x) + h(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})\right) + \frac{(t-s_{\iota\varepsilon n})^2 + d^2(x, y_{\iota\varepsilon n})}{2\varepsilon} + p_{\iota\varepsilon n}((t,x),(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})) + \ell_\alpha(t),$$

$$\psi_{\iota\varepsilon n}(s,y) := v(t_{\iota\varepsilon n}, x_{\iota\varepsilon n}) - \iota\left(h(t_{\iota\varepsilon n}, x_{\iota\varepsilon n}) + h(s,y)\right) - \frac{(t_{\iota\varepsilon n}-s)^2 + d^2(x_{\iota\varepsilon n}, y)}{2\varepsilon} - p_{\iota\varepsilon n}((t_{\iota\varepsilon n}, x_{\iota\varepsilon n}),(s,y)).$$

Both are regular enough so that $\varphi_{\iota\varepsilon n} \in \mathscr{T}_+$ and $\psi_{\iota\varepsilon n} \in \mathscr{T}_-$. Moreover, by (3.18), $v - \varphi_{\iota\varepsilon n}$ reaches a maximum at $(t_{\iota\varepsilon n}, x_{\iota\varepsilon n})$, and $w - \psi_{\iota\varepsilon n}$ a minimum at $(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})$. This will allow us to apply the definitions of semisolutions, provided that neither $t_{\iota\varepsilon n}$ nor $s_{\iota\varepsilon n}$ is equal to T.

Near-maximum points are getting closer to each other. The inequality (3.20) implies that

$$\begin{aligned}\sup \Phi_{\iota,\varepsilon} + \frac{(t_{\iota\varepsilon n}-s_{\iota\varepsilon n})^2 + d^2(x_{\iota\varepsilon n}, y_{\iota\varepsilon n})}{4\varepsilon} &\leqslant \Phi_{\iota,\varepsilon}\left((t_{\iota\varepsilon n}, x_{\iota\varepsilon n}),(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})\right) + \omega_\iota(n) + \frac{(t_{\iota\varepsilon n}-s_{\iota\varepsilon n})^2 + d^2(x_{\iota\varepsilon n}, y_{\iota\varepsilon n})}{4\varepsilon}\\ &\leqslant \sup \Phi_{\iota,\frac{\varepsilon}{2}} + \omega_\iota(n).\end{aligned}$$

For each fixed ι, the function $\varepsilon \to \sup \Phi_{\iota,\varepsilon}$ decreases when ε goes to 0, and is bounded below uniformly in ε, hence converges. Consequently,

$$\frac{(t_{\iota\varepsilon n}-s_{\iota\varepsilon n})^2 + d^2(x_{\iota\varepsilon n}, y_{\iota\varepsilon n})}{2\varepsilon} \leqslant 2\left(\sup \Phi_{\iota,\frac{\varepsilon}{2}} + \omega_\iota(n) - \sup \Phi_{\iota,\varepsilon}\right) \underset{\varepsilon\to 0, n\to\infty}{\longrightarrow} 0. \tag{3.22}$$

Fix now ι small enough so that $\iota\big(g^2(d^2(o,x^0)) + g^2(d^2(o,y^0))\big) \leqslant \Gamma/3$. In particular, testing with $\big((t^0,x^0),(t^0,x^0)\big)$, there holds $0 < 2\Gamma/3 \leqslant \sup\Phi_{\iota\varepsilon}$ for all $\varepsilon > 0$.

Near-maximum points do not lie on the boundary. Assume first that there exists a sequence $(\varepsilon_m, n_m)_{m\in\mathbb{N}}$ with $\varepsilon_m \searrow_m 0$ and $n_m \to_m \infty$ along which at least one of $t_{\iota\varepsilon_m n_m}$ or $s_{\iota\varepsilon_m n_m}$ is equal to T. Then, by (3.22), the other gets closer to T as $m \to \infty$. Since ι is fixed, the sequences $x_{\iota\varepsilon n}, y_{\iota\varepsilon n}$ stay in a ball of radius R_ι, so that

$$
\begin{aligned}
\frac{2}{3}\Gamma \leqslant \limsup_{m\to\infty} \sup\Phi_{\iota,\varepsilon_m} &\leqslant \limsup_{m\to\infty} \Phi_{\iota,\varepsilon_m}\big((t_{\iota\varepsilon_m n_m}, x_{\iota\varepsilon_m n_m}),(s_{\iota\varepsilon_m n_m}, y_{\iota\varepsilon_m n_m})\big) + \omega_\iota(n_m)\\
&\leqslant \lim_{r\searrow 0}\sup\Big\{v(t,x) - w(s,y) - \ell_\alpha(t)\ \Big|\ x,y \in \overline{\mathscr{B}}(o,R_\iota),\ t,s \in [T-r,T]\Big\} + \limsup_{m\to\infty}\omega_\iota(n_m)\\
&\leqslant \sup\Big\{v(T,x) - w(T,x)\ \Big|\ x \in \overline{\mathscr{B}}(o,R_\iota)\Big\} \leqslant 0.
\end{aligned}
$$

Here, we used the assumption of locally uniform upper semicontinuity of v and $-w$ on the decreasing sequence of closed bounded sets $[T-r,T]\times\overline{\mathscr{B}}(o,R_\iota)$, and the assumption that v is inferior to w on the parabolic boundary. This is absurd, and such a sequence does not exist.

A first inequality from the definition of semisolutions. Consider ε and n such that both $t_{\iota\varepsilon n}$ and $s_{\iota\varepsilon n}$ are strictly inferior to T. Then $v - \varphi_{\iota\varepsilon n}$ and $\psi_{\iota\varepsilon n} - w$ reach a maximum respectively at $(t_{\iota\varepsilon n}, x_{\iota\varepsilon n})$ and $(s_{\iota\varepsilon n}, y_{\iota\varepsilon n}) \in [0,T)\times\Omega$, so by the viscosity inequalities of Definition 3.1.6,

$$
-\partial_t\varphi_{\iota\varepsilon n}(t_{\iota\varepsilon n}, x_{\iota\varepsilon n}) + H\big(x_{\iota\varepsilon n}, D_{x_{\iota\varepsilon n}}\varphi_{\iota\varepsilon n}(t_{\iota\varepsilon n}, x_{\iota\varepsilon n})\big) \leqslant 0 \leqslant -\partial_s\psi_{\iota\varepsilon n}(s_{\iota\varepsilon n}, y_{\iota\varepsilon n}) + H\big(y_{\iota\varepsilon n}, D_{y_{\iota\varepsilon n}}\psi_{\iota\varepsilon n}(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})\big).
$$

A second inequality from the regularity of the Hamiltonian. On the one hand,

$$
\begin{aligned}
&\partial_t\varphi_{\iota\varepsilon n}(t_{\iota\varepsilon n}, x_{\iota\varepsilon n}) - \partial_s\psi_{\iota\varepsilon n}(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})\\
&\qquad = \frac{t_{\iota\varepsilon n} - s_{\iota\varepsilon n}}{\varepsilon} - \alpha + \frac{s_{\iota\varepsilon n} - t_{\iota\varepsilon n}}{\varepsilon} + \big(\partial_t p_{\iota\varepsilon n} + \iota\partial_t h + \partial_s p_{\iota\varepsilon n} + \iota\partial_s h\big)\big((t_{\iota\varepsilon n}, x_{\iota\varepsilon n}),(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})\big)\\
&\qquad \leqslant -\alpha + \sum_{\binom{r}{z} = \binom{t}{x},\binom{s}{y}} \Big[\big|\partial_r p_{\iota\varepsilon n}\big((t_{\iota\varepsilon n}, x_{\iota\varepsilon n}),(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})\big)\big| + \iota\partial_r h(r_{\iota\varepsilon n}, z_{\iota\varepsilon n})\Big].
\end{aligned}
$$

On the other hand, using both parts of the assumption **[A3.1.11]** on the Hamiltonian,

$$
\begin{aligned}
&H\big(y_{\iota\varepsilon n}, D_{y_{\iota\varepsilon n}}\psi_{\iota\varepsilon n}(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})\big) - H\big(x_{\iota\varepsilon n}, D_{x_{\iota\varepsilon n}}\varphi_{\iota\varepsilon n}(t_{\iota\varepsilon n}, x_{\iota\varepsilon n})\big)\\
&\leqslant H\left(y_{\iota\varepsilon n}, -D_{y_{\iota\varepsilon n}}\frac{d^2(x_{\iota\varepsilon n},\cdot)}{2\varepsilon}\right) - H\left(x_{\iota\varepsilon n}, D_{x_{\iota\varepsilon n}}\frac{d^2(\cdot, y_{\iota\varepsilon n})}{2\varepsilon}\right)\\
&\quad + C\sum_{\binom{r}{z} = \binom{t}{x},\binom{s}{y}} (1 + d(z_{\iota\varepsilon n}, o))\big(\|D_z p_{\iota\varepsilon n}\|_{z_{\iota\varepsilon n}} + \iota\|D_z h(r_{\iota\varepsilon n}, z_{\iota\varepsilon n})\|_{z_{\iota\varepsilon n}}\big)\\
&\leqslant C_{R_\iota} d(x_{\iota\varepsilon n}, y_{\iota\varepsilon n})\left(\frac{d(x_{\iota\varepsilon n}, y_{\iota\varepsilon n})}{2\varepsilon} + 1\right) + C\sum_{\binom{r}{z} = \binom{t}{x},\binom{s}{y}} (1 + d(z_{\iota\varepsilon n}, o))\big(\|D_z p_{\iota\varepsilon n}\|_{z_{\iota\varepsilon n}} + \iota\|D_z h(r_{\iota\varepsilon n}, z_{\iota\varepsilon n})\|_{z_{\iota\varepsilon n}}\big).
\end{aligned}
$$

Gathering the two estimates, and using the controls on the derivatives of $p_{\iota\varepsilon n}$ provided by (3.21), there holds

$$
\begin{aligned}
&\big[-\partial_s\psi_{\iota\varepsilon n}(s_{\iota\varepsilon n}, y_{\iota\varepsilon n}) + H\big(y_{\iota\varepsilon n}, D_{y_{\iota\varepsilon n}}\psi_{\iota\varepsilon n}(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})\big)\big] - \big[-\partial_t\varphi_{\iota\varepsilon n}(t_{\iota\varepsilon n}, x_{\iota\varepsilon n}) + H\big(x_{\iota\varepsilon n}, D_{x_{\iota\varepsilon n}}\varphi_{\iota\varepsilon n}(t_{\iota\varepsilon n}, x_{\iota\varepsilon n})\big)\big] + \alpha\\
&\leqslant C_{R_\iota} d(x_{\iota\varepsilon n}, y_{\iota\varepsilon n})\left(\frac{d(x_{\iota\varepsilon n}, y_{\iota\varepsilon n})}{2\varepsilon} + 1\right) + \omega_\iota(n) + \iota\sum_{\binom{r}{z} = \binom{t}{x},\binom{s}{y}} \big[\partial_r h(r_{\iota\varepsilon n}, z_{\iota\varepsilon n}) + C(1 + d(z_{\iota\varepsilon n}, o))\|D_z h(r_{\iota\varepsilon n}, z_{\iota\varepsilon n})\|_{z_{\iota\varepsilon n}}\big].
\end{aligned}
$$

Here, the last summand is nonpositive by the construction of the function h in (3.16). As ω_ι goes to 0 when $n \to \infty$, and using (3.22) to control the term including $d(x_{\iota\varepsilon n}, y_{\iota\varepsilon n})$, it is possible to choose ε small enough and n large enough so that $C_{R_\iota} d(x_{\iota\varepsilon n}, y_{\iota\varepsilon n})\left(\frac{d(x_{\iota\varepsilon n}, y_{\iota\varepsilon n})}{2\varepsilon} + 1\right) + \omega_\iota(n) \leqslant \alpha/2$. For this choice, one gets

$$
-\partial_s\psi_{\iota\varepsilon n}(s_{\iota\varepsilon n}, y_{\iota\varepsilon n}) + H\big(y_{\iota\varepsilon n}, D_{y_{\iota\varepsilon n}}\psi_{\iota\varepsilon n}(s_{\iota\varepsilon n}, y_{\iota\varepsilon n})\big) \leqslant -\partial_t\varphi_{\iota\varepsilon n}(t_{\iota\varepsilon n}, x_{\iota\varepsilon n}) + H\big(x_{\iota\varepsilon n}, D_{x_{\iota\varepsilon n}}\varphi_{\iota\varepsilon n}(t_{\iota\varepsilon n}, x_{\iota\varepsilon n})\big) - \frac{\alpha}{2}.
$$

By the previous steps, ε can be further decreased and n increased as to let $t_{\iota\varepsilon n}, s_{\iota\varepsilon n} < T$, in which case both inequalities conflict. This is absurd, and $v(t,x) \leqslant w(t,x)$ for all $(t,x) \in [0,T]\times\Omega$. □

The construction of the function h is inspired from [FGŚ17], in which it must satisfy second-order conditions as well. By tweaking the constants, it can also replace ℓ_α in providing the small parameter yielding the contradiction. As an example, consider the Eikonal equation with Hamiltonian $H : \mathbb{T} \to \mathbb{R}$ defined as

$$H(x,p) := \sup_{v \in \mathrm{T}_x\Omega, |v|_x \leqslant 1} -p(v) \qquad \forall (x,p) \in \mathbb{T}.$$

On the one hand, there holds for any $x \in \Omega$ and $p, q \in \mathbb{T}_x$ that

$$H(x,p) - H(x,q) \leqslant \sup_{v \in \mathrm{T}_x\Omega, |v|_x \leqslant 1} -p(v) + q(v) \leqslant \|p-q\|_x.$$

On the other hand, let $x \neq y$. For any reparametrized geodesic γ issued from y, the differential $D_y d^2(x,\cdot)(\gamma_0^+)$ can be computed by taking the directional derivative, and one has

$$D_y d^2(x,\cdot)(\gamma_0^+) = \lim_{h \searrow 0} \frac{d^2(x,\gamma_h) - d^2(x,y)}{h} \leqslant \limsup_{h \searrow 0} (d(x,\gamma_h) + d(x,y)) \frac{d(\gamma_h, y)}{h} \leqslant 2d(x,y)|\gamma_0^+|_y.$$

By continuity, $D_y d^2(x,\cdot)(w) \leqslant 2d(x,y)|w|_y$ for any $w \in \mathrm{T}_y\Omega$. Moreover, if γ is a geodesic between x and y, there holds $D_x d^2(\cdot,y)(\gamma) = -2d(x,y)|\gamma_0^+|_x$. Hence, for any $x,y \in \Omega$ and $a \geqslant 0$,

$$\begin{aligned} H(y, -aD_y d^2(x,\cdot)) - H(x, aD_x d^2(\cdot,y)) &= \sup_{w \in \mathrm{T}_y\Omega, |w|_y \leqslant 1} aD_y d^2(x,\cdot)(w) - \sup_{v \in \mathrm{T}_x\Omega, |v|_x \leqslant 1} -aD_x d^2(\cdot,y)(v) \\ &\leqslant 2ad(y,x) + \inf_{v \in \mathrm{T}_x\Omega, |v|_x \leqslant 1} aD_x d^2(\cdot,y)(v) \leqslant 2ad(y,x) - 2ad(x,y) = 0. \end{aligned}$$

Therefore the Hamiltonian of the Eikonal equation satisfies the assumptions of Theorem 3.1.12. To get other examples, we restrict to the Wasserstein space, where we may define optimal control problems.

3.2 Further results in the Wasserstein space

In this section, we consider the Wasserstein space $\Omega = \mathscr{P}_2(\mathbb{R}^d)$, endowed with the quadratic Wasserstein distance $d_{\mathcal{W}}(\cdot,\cdot)$. This is a CBB(0) space in the sense of Alexandrov, since the squared Wasserstein distance is 2-concave (see [AKP23, Corollary 8.25]). For the convenience of the reader, we recall the following notations.

$\Gamma(\mu,\nu) \subset \mathscr{P}((\mathbb{R}^d)^2)$	transport plans α such that $\pi_x \# \alpha = \mu$ and $\pi_y \# \alpha = \nu$	$\mu,\nu \in \mathscr{P}(\mathbb{R}^d)$
$\mathscr{P}_p(\mathbb{R}^d) \subset \mathscr{P}(\mathbb{R}^d)$	Borel probability measures μ with $\int_{x \in \mathbb{R}^d} \lvert x\rvert^p d\mu < \infty$	$\mu \in \mathscr{P}(\mathbb{R}^d)$, $p \in (0,\infty)$
$d_{\mathcal{W},p} : (\mathscr{P}_p(\mathbb{R}^d))^2 \to \mathbb{R}^+$	$p-$Wasserstein distance $\inf_{\alpha \in \Gamma(\mu,\nu)} \left(\int_{x,y \in \mathbb{R}^d} \lvert x-y\rvert^p d\alpha \right)^{\frac{1}{p}}$	$\mu,\nu \in \mathscr{P}_p(\mathbb{R}^d)$
$\mathrm{T}^n \mathbb{R}^d$	set of $(x, v_1, \cdots, v_n)$ with $x \in \mathbb{R}^d$, $v_i \in \mathrm{T}_x\mathbb{R}^d$ for all i	
$\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)$	measure fields on $\mathrm{T}\mathbb{R}^d$ such that $\pi_x \# \xi = \mu$	$\mu \in \mathscr{P}_2(\mathbb{R}^d)$
$\exp_\mu(h \cdot \xi) \in \mathscr{P}_2(\mathbb{R}^d)$	exponential $(\pi_x + s\pi_v)\#\xi$	$\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, $h \in \mathbb{R}$
$\exp_\mu^{-1}(\nu) \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$	partial inverse $\left\{ \xi \;\middle\vert\; \exp_\mu(\xi) = \nu,\ \lVert\xi\rVert_\mu = d_{\mathcal{W}}(\mu,\nu) \right\}$	$\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$

$\Gamma_\mu(\xi_0,\xi_1) \subset \mathscr{P}_2(\mathrm{T}^2\mathbb{R}^d)$	plans $\alpha = \alpha(dx, dv_0, dv_1)$ such that $(\pi_x, \pi_{v_i})\#\alpha = \xi_i$	$\xi_0,\xi_1 \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$
$W_\mu : (\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu)^2 \to \mathbb{R}^+$	cone distance $\inf_{\alpha \in \Gamma_\mu(\xi,\zeta)} \left(\int_{(x,v,w) \in \mathrm{T}^2\mathbb{R}^d} \lvert v-w\rvert^2 d\alpha \right)^{\frac{1}{2}}$	$\xi,\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$
$\langle\cdot,\cdot\rangle_\mu^\pm : (\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu)^2 \to \mathbb{R}^+$	applications $\pm \sup_{\alpha \in \Gamma_\mu(\xi,\zeta)} \pm \int_{(x,v,w) \in \mathrm{T}^2\mathbb{R}^d} \langle v,w \rangle \, d\alpha$	$\xi,\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$
$\mathbf{Tan}_\mu \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$	geometric tangent space $\overline{\mathbb{R}^+ \cdot \exp_\mu^{-1}(\mathscr{P}_2(\mathbb{R}^d))}^{W_\mu}$	$\mu \in \mathscr{P}_2(\mathbb{R}^d)$
$\mathrm{Tan}_\mu \subset L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$	regular tangent space $\overline{\nabla\mathcal{C}_c^\infty(\mathbb{R}^d;\mathbb{R})}^{L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)}$	$\mu \in \mathscr{P}_2(\mathbb{R}^d)$
$\pi_T^\mu : \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \to \mathbf{Tan}_\mu$	metric projection, which satisfies $\pi_T^\mu f \# \mu \in \mathrm{Tan}_\mu$	$\mu \in \mathscr{P}_2(\mathbb{R}^d)$

A subscript o on a set of plans denotes optimality, as in $\Gamma_o(\mu,\nu)$, $\Gamma_{\mu,o}(\xi,\zeta)$, and so on.

3.2.1 On viscosity solutions in $\mathscr{P}_2(\mathbb{R}^d)$

This section gathers some comments about L-differentiability and semidifferentials in $\mathscr{P}_2(\mathbb{R}^d)$, which are both used very frequently in the literature.

On the links with L-differentiable functions, we do not have a full answer. In the course of investigation, we reformulated L-differentiability on CBB(0) spaces that write as quotients of Hilbert spaces by isometries, with aim to distinguish the implications of the definition itself from the additional structure of the Wasserstein space. Precisely, in the construction by quotient, the fact that φ is L-differentiable at x does not trivially imply that there exists $p \in \mathrm{T}_x\mathbb{R}^d$ such that $D_x\varphi(q) = \langle p, q\rangle_x$ for all $q \in \mathrm{T}_x\mathbb{R}^d$. This section leaves a lot of room for discussion, beginning with: which CBB(0) spaces can be written under this form?

On semidifferentials, we provide an overview of the definitions in use in the literature, with the same dichotomy appearing between semidifferentials that allow the splitting of mass, and those that do not. In finite dimension, one can represent the elements of the semidifferential as gradients of test functions, and conversely. In the Wasserstein space, it it desirable to relax the definition to allow *approximation* of an element of the semidifferential by gradients, instead of an exact representation, since the test function constructed with the exact element can be quite irregular. We give such a construction in the case of the geometric semidifferentials.

3.2.1.1 Lift

The following definitions and computations are freely extracted from the theory developed in the Wasserstein space [Lio06; Car13; CD18b; Car+19]. They were written as a way to understand the case of $\mathscr{P}_2$ through a simpler setting, without measures. We provide them in case it might interest the reader; obviously, all ideas are from the cited references, and all mistakes are ours.

We consider the same construction as in Section 1.1.2, with E a Hilbert space, G a group of linear bijective isometries on E, Ω the set of equivalence classes for the relation $f^0 \sim f^1$ if there exists $g \in G$ such that $f^1 = g(f^0)$. Denote by $[f_0]$ the equivalence class of $f^0 \in E$. Let

$$d(x,y) := \inf_{f^0\in x, f^1\in y} \|f^0 - f^1\|_E.$$

We assume that each class is closed in E, and that the infimum is always attained. By Lemmata 1.1.12 and 1.1.13, (Ω, d) is a complete geodesic CBB space, and its geodesics are images of the geodesics of E in the following sense: $\gamma : [0,1] \to \Omega$ is a geodesic if and only if for any $f^0 \in \gamma_0 \subset E$, there exists $f^1 \in \gamma_1$ such that $\gamma_t = [(1-t)f^0 + tf^1]$ and $d(\gamma_t, \gamma_s) = |t-s|\,\|f^0 - f^1\|_E$ for all $s,t \in [0,1]$. Building a geometric tangent cone as in Chapter 1, we get many directionally differentiable maps. The L-differentiable ones would be the following.

Definition 3.2.1 (L-differentiable functions). *A Lipschitz function $\varphi : \Omega \to \mathbb{R}$ is said to be L-differentiable at $x \in \Omega$ if its lift*

$$\Phi : E \to \mathbb{R}, \qquad \Phi(f) := \varphi([f])$$

is Fréchet-differentiable in E at some point $f \in x$.

By definition of $d(\cdot,\cdot)$, the lift Φ is Lipschitz with the same constant as φ. It also admits the following properties, inspired from [Car13, Theorem 6.5] and [CD18a, Proposition 5.25].

Lemma 3.2.2 (Structure of the differential). *Let φ be Lipschitz and L−differentiable at x. Then*

- *Φ is Fréchet-differentiable at all point of the equivalence class x, and $D\Phi(f) = D\Phi(g(f)) \circ g$ for all $g \in G$,*
- *φ is directionally differentiable at x.*

Proof. Let $f^0 \in x$. For any $f^1 \in x$, there exists $g \in G$ such that $f^0 = g(f^1)$. In particular, $\Phi(f^1) = \Phi(f^0)$, and since g is linear, $\Phi(f^1 + h) = \Phi(f^0 + g(h))$ for all $h \in E$. As $\|g(h)\|_E = \|h\|_E$ goes to 0 when h does, one has

$$\left|\Phi(f^1 + h) - \Phi(f^1) - [D\Phi(f^0) \circ g](h)\right| = \left|\Phi(f^0 + g(h)) - \Phi(f^0) - D\Phi(f^0)(g(h))\right| \underset{h \to 0_E}{\longrightarrow} 0.$$

Hence Φ is Fréchet-differentiable at f^1 with differential $D\Phi(g(f^1)) \circ g$.

Since φ is Lipschitz, it is enough to show that it is directionally differentiable along geodesics, then extend to $\mathrm{T}_x\Omega$ by positive homogeneity and continuity. Let $f^0 \in x$. By Lemma 1.1.13, for any geodesic γ issued from x, there exist $f^1 \in \gamma_1$ such that $(1-t)f^0 + tf^1 \in \gamma_t$ and $d(\gamma_s, \gamma_t) = |t-s| \, \|f^0 - f^1\|_E$. Hence

$$\frac{\varphi(\gamma_t) - \varphi(x)}{t} = \frac{\Phi(f^0 + t(f^1 - f^0)) - \Phi(f^0)}{t} = D\Phi(f^0) \cdot (f^1 - f^0) + O(t),$$

and φ is directionally differentiable along geodesics. □

Assume now that for all $f^0 \in E$ and $h \in E$, the curve $s \mapsto [f^0 + sh]$ admits a velocity in $\mathrm{T}_x\Omega$ if and only if $s \mapsto [f^0 - sh]$ does. This is the case in the Wasserstein space, for instance. Then L-differentiable functions can admit a metric gradient only along directions of $\mathrm{T}_x\Omega$ that admit an opposite.

Lemma 3.2.3 (The metric gradient can only be a regular direction). *Let φ be L-differentiable. If there exists $p \in \mathrm{T}_x\Omega$ such that $D_x\varphi(v) = \langle p, v\rangle_x$ for all $v \in \mathrm{T}_x\Omega$, then p admits an opposite in $\mathrm{T}_x\Omega$, in the sense that there exists $q \in \mathrm{T}_x\Omega$ with the same norm as p and forming an angle π with it.*

Proof. Let $p \in \mathrm{T}_x\Omega$ be the metric gradient of φ, and $(p_n)_{n\in\mathbb{N}} \subset \mathrm{T}'_x\Omega$ be a Cauchy sequence with $d_x(p, p_n) \to_n 0$. W.l.o.g. assume that $p \neq 0_x$. Let $f^0 \in x$. Using Lemma 1.1.13, for any n, there exists $h^n \in E$ such that $s \mapsto [f^0 + sh^n]$ admits p_n as a velocity. Consider $q_n \in \mathrm{T}_x\Omega$ the velocity of $s \mapsto [f^0 - sh^n]$, that we assumed to exist. Then $|q_n|_x \leqslant |h^n|_E = |p_n|_x$, and

$$\langle p, p_n\rangle_x = D_x\varphi(p_n) = D\Phi(f^0)(h^n) = -D\Phi(f^0)(-h^n) = -D_x\varphi(q_n) = -\langle p, q_n\rangle_x.$$

Denote $\widehat{pq} := \frac{\langle p,q\rangle_x}{|p|_x|q|_x}$ the angle between $p, q \in \mathrm{T}_x\Omega$. Then $\langle p, p_n\rangle_x = |p|_x|q_n|_x\left(-\cos(\widehat{pq_n})\right)$. Estimating first $-\cos \leqslant 1$, we deduce that $|q_n|_x \geqslant |p_n|_x \cos(\widehat{pp_n}) \to_n |p|_x$. Plugging this back in the inequality, we obtain that $\widehat{pq_n} \to_n \pi$. As Ω is CBB, [AKP23, p. 8.1] implies that

$$\widehat{pq_n} + \widehat{q_nq_m} + \widehat{q_mp} \leqslant 2\pi,$$

so that $\lim_{n\to\infty} \sup_{m\geqslant n} d_x(q_n, q_m) = 0$, and the sequence $(q_n)_n$ is Cauchy in the complete space $\mathrm{T}_x\Omega$. Its limit q has the desired properties. □

This parallels the fact that in $\Omega = \mathscr{P}_2(\mathbb{R}^d)$, an application can admit a Wasserstein gradient only if the latter is induced by a map (see Remark 5.1.7). L-differentiable functions in $\mathscr{P}_2(\mathbb{R}^d)$ can admit a non-vanishing gradient everywhere, since all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ admit a subset of their tangent cone in which all directions have an opposite. This is not satisfied by the cone, or the surface of a cube; the apex, or the corners, have tangent cones in which there is no pair (p, q) forming an angle of π.

3.2.1.2 Semidifferentials in the Wasserstein space

There are various ways of defining sub and superdifferentials at $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. A general definition would read as follows: $\xi \in A_\mu \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ belongs to the subdifferential of a function $u : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ if

$$\liminf_{\nu\to\mu} \frac{u(\nu) - u(\mu) - \operatorname{opt}_{\eta\in G(\mu,\nu)} \operatorname{opt}_{\alpha\in\Gamma_\mu(\xi,\eta)} \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v, w\rangle \, d\alpha}{d_{\mathcal{W}}(\mu,\nu)} \geqslant 0 \qquad \forall \nu \in \mathscr{P}_2(\mathbb{R}^d).$$

Here $G(\mu, \nu)$ is a subset of the measure fields $\eta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ such that $\exp_\mu(\eta) = \nu$, and opt may be inf or sup. Then one has to choose between the following options:

– restrict A_μ to be a subset of $L^2_\mu(\mathbb{R}^d; \mathrm{T}\mathbb{R}^d)\#\mu$, or consider measure fields that may split mass,

– restrict A_μ to be a subset of $\mathbf{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$, or admit solenoidal components,
– restrict $G(\mu,\nu)$ to be the set of velocities of geodesics between μ and ν, or admit all measure fields $\eta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ such that $\exp_\mu(\eta) = \nu$,
– choose opt to be inf or sup.

This comes in addition of the variety of definitions in infinite dimension. Table 3.1 pictures some choices from the literature, sometimes up to small variations for the sake of classification.

		Subset of $\mathbf{Tan}_\mu$		Subset of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$	
		weak	strong	weak	strong
Map	inf			[AGS05, Def. 10.1.1, (10.1.3)$^{*\cap}$] [GT19, Def. 3.1 (i)$^{\cap}$]	
	sup	[GNT08, Def. 3.1] [GŚ14, Def. 6.2] [HK15, Def. 5.1] [JMQ23, Def. 3.1$^{\delta}$] [DS24, Def. 2.5*]	[MQ18, Def. 3.2$^{\delta}$] [JMQ20, Def. 3$^{\delta}$]	[AG08, Def. 3.2$^{\cap}$] [GŚ15a, Def. 2.5$^{\cap}$] [GT19, Def. 3.1 (ii)$^{\cap}$]	[AGS05, Def. 10.1.1, (10.1.4)$^{*\cap}$] [CQ08, Def. 1$^{\delta}$] [Cav+18, Def. 12$^{\delta}$] [BF22a, Def. 2.18$^{\delta}$]
Plan	inf	[Gig08, Def. 5.14]		[AGS05, Def. 10.3.1, (10.3.6)$^{\cap}$]	
	sup	[AF14, Def. 4.7]			[AGS05, Def. 10.3.1, (10.3.7)$^{\cap}$] [Ber24, Def. 2.2]

Table 3.1: Definitions of semidifferentials in the Wasserstein space.

The term weak *corresponds to* $G(\mu,\nu) = \exp_\mu^{-1}(\nu)$, *and* strong *to* $G(\mu,\nu)$ *taken as all measure fields sending* μ *to* ν. *A star* means that the definition is restricted to transport-regular measures, on which* inf *and* sup *coincide. A superscript* δ *denotes* δ*−semidifferentials for strict viscosity solutions, and* $\cap$ *indicates that the definition is written in* $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, *but the authors either mention or work in* $\mathbf{Tan}_\mu$.

We adopt the heavy but precise nomenclature of [weak/strong]−[inf/sup]−[map/plan]−[$\mathbf{Tan}_\mu$/$\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$] $\partial_\mu u$ for the corresponding semidifferential. If a property applies to both options, we replace it by a star: the $*$−inf−map−$\mathbf{Tan}_\mu$ semidifferential refers to the union of the weak−[$\cdots$] and strong−[$\cdots$] semidifferential.

Lemma 3.2.4 (Horizontal convexity). *Each set* $* -$ *sup* $- * - *$ $\partial_\mu u$ *is horizontally convex.*

Proof. Let $\xi^0, \xi^1 \in \partial_\mu u$ for the $* -$ sup $- * - *$ definition, which means that

$$\liminf_{\nu\to\mu} \frac{u(\nu) - u(\mu) - \sup_{\eta\in G(\mu,\nu)} \langle \eta, \xi\rangle_\mu^+}{d_{\mathcal{W}}(\mu,\nu)} \geqslant 0 \qquad \forall \nu \in \mathscr{P}_2(\mathbb{R}^d).$$

Let $\beta \in \Gamma_\mu(\xi^0,\xi^1)$, that is, $\beta = \beta(dx,dv,dw) \in \mathscr{P}_2(\mathrm{T}^2\mathbb{R}^d)_\mu$ and $(\pi_x,\pi_v)\#\beta = \xi^0$, $(\pi_x,\pi_w)\#\beta = \xi^1$. Let $t \in [0,1]$, and consider $\xi^t := (\pi_x, (1-t)\pi_v + t\pi_w)\#\beta$. If $\xi^0,\xi^1 \in \mathbf{Tan}_\mu$, then $\xi^t \in \mathbf{Tan}_\mu$ by [Gig08, Prop. 4.25]; if $\xi^i = (id, f^i)\#\mu$ for $i \in \{0,1\}$, then $\Gamma_\mu(\xi^0,\xi^1) = \{(id, f^0, f^1)\#\mu\}$, and ξ^t is itself induced by the map $(1-t)f^0 + tf^1$. Using the horizontal convexity of $\langle \eta, \cdot\rangle_\mu^+$ given by [Gig08, Prop. 4.27] for any $\eta \in G(\mu,\nu)$,

$$\liminf_{\nu\to\mu} \frac{u(\nu) - u(\mu) - \sup_{\eta\in G(\mu,\nu)} \langle \eta, \xi^t\rangle_\mu^+}{d_{\mathcal{W}}(\mu,\nu)} \geqslant \liminf_{\nu\to\mu} \frac{u(\nu) - u(\mu) - \sup_{\eta\in G(\mu,\nu)} (1-t)\langle \eta, \xi^0\rangle_\mu^+ + t\langle \eta, \xi^1\rangle_\mu^+}{d_{\mathcal{W}}(\mu,\nu)} \geqslant 0.$$

Hence $\xi^t \in \partial_\mu u$. □

One may wonder if $* -$ inf $- * - *$ $\partial_\mu u$ is vertically convex. However, neither $\mathbf{Tan}_\mu$ nor $(id, L^2_\mu)\#\mu$ are vertically convex, so that only $* -$ inf − plan − $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ $\partial_\mu u$ may be. Even in this case, the infimum over $\eta \in G(\mu,\nu)$ does not allow to conclude, and the argument can be carried only for weak − sup − plan − $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ $\partial_\mu u$ if μ is transport-regular.

The different definitions satisfy trivial inclusions since one option is more restrictive than the other. For instance, all other things being equal, the strong semidifferentials are included in the weak ones, the sup in the inf, etc. One nontrivial inclusion is proved in [GT19, Theorem 3.6], where it is showed that $*-\inf-\text{map}-*\ \partial_\mu u$ coincides with $*-\sup-\text{map}-*\ \partial_\mu u$. The argument is specific to map-induced measure fields. An illustration of the differences between definitions is provided in Section 5.1.3 on the squared distance.

Viscosity solutions with semidifferentials. Consider the formal first-order equation

$$H(\mu,\nabla_\mu u)=0 \qquad \mu\in\mathcal{O}\subset\mathscr{P}_2(\mathbb{R}^d), \tag{3.23}$$

supplemented with boundary conditions. Besides regularity of semisolutions, boundary conditions and specificities, the viscosity notions read as follows: a function $u:\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}$ is a viscosity

- subsolution of (3.23) if for all $\mu\in\mathcal{O}$ and $\xi\in\partial^+_\mu\xi$, there holds $H(\mu,\xi)\leqslant 0$,
- supersolution of (3.23) if for all $\mu\in\mathcal{O}$ and $\xi\in\partial^-_\mu\xi$, there holds $H(\mu,\xi)\geqslant 0$.

The notions of viscosity solutions associated to distinct semidifferential may coincide. Here we anticipate Lemma 5.1.20, proved with tools of Chapter 5 that do not depend on the present chapter. For instance, if the Hamiltonian H only depends on scalar products between ξ and maps, then by (5.8), there is no difference between imposing conditions on a $*-\sup-*-*$ semidifferential, or its restriction to map-induced elements. On the other hand, consider $u:\mu\mapsto d^2_{\mathcal{W}}(\mu,\nu)$. Then by (5.10), the weak−sup−map−$\mathbf{Tan}_\nu$ semidifferential is given by the singleton $\{0_\nu\}$, but the weak−sup−map−$\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\nu$ semidifferential contains all barycenters of the orthogonal of $\mathbf{Tan}_\nu$. In particular, u satisfies the viscosity inequalities at ν for the equation $\frac{1}{2}\|\nabla_\mu u\|^2_\mu=0$ with respect to the former semidifferential, but not with respect to the latter.

Comparison with test functions. In the classical theory, the definitions via test functions and semidifferentials coincide. This is seen from the fact that if φ touches u from above at x, then $\nabla_x\varphi$ belongs to the superdifferential of u at x. Conversely, given v in the superdifferential, one constructs a test function of the form $y\mapsto\langle v,y-x\rangle+m(|y-x|)$, with m a $\mathcal{C}^1$ function satisfying $m'(0)=0$ [BC97, Lemma 1.7].

In $\mathscr{P}_2(\mathbb{R}^d)$, one can reasonably obtain an approximate version of this construction, by first considering v_ε close to v with better regularity, then constructing a function with differential $\langle v_\varepsilon,\cdot\rangle$. For instance, if $\xi\in\mathbf{Tan}_\mu$ is a element of the superdifferential of a function u, one can consider $\xi_\varepsilon\in\mathbf{Tan}_\mu$ as close to $-\xi$ as desired with respect to $W_\mu(\cdot,\cdot)$, and such that $(\pi_x,\pi_x+s\pi_v)\#\xi_\varepsilon$ is optimal between its marginals for some $s=s_\varepsilon>0$. The function $\varphi_\varepsilon:\nu\mapsto\frac{1}{s}d^2_{\mathcal{W}}\left(\nu,\exp_\mu(\frac{s}{2}\cdot\xi_\varepsilon)\right)$ is semiconcave, directionally differentiable, and with differential

$$D_\mu\varphi_\varepsilon(\zeta)=\frac{2}{s}\langle\zeta,-\frac{s}{2}\xi_\varepsilon\rangle^-_\mu\sim\langle\zeta,\xi\rangle^-_\mu \qquad \forall\zeta\in\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu.$$

The above expression follows by the combination of Theorem 1.1.41 p. 14 and [AGS05, Lemma 7.2.1] to give uniqueness of the optimal transport plan. So, in the case of a definition using ε−semidifferentials in the weak−sup−plan−$\mathbf{Tan}_\mu$ sense, it is possible to retrieve the equivalence with test functions as defined in Definition 3.1.5.

It is also possible to consider applications with a prescribed differential, instead of approximations, for instance as follows.

Lemma 3.2.5 (Directional differentiability). *Let $\zeta\in\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ and $\varphi:\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}$ be one of*

$$\varphi_{+\pm}(\nu):=\sup_{\eta\in\exp_\mu^{-1}(\nu)}\langle\eta,\zeta\rangle^\pm_\mu, \qquad \varphi_{-\pm}(\nu):=\inf_{\eta\in\exp_\mu^{-1}(\nu)}\langle\eta,\zeta\rangle^\pm_\mu.$$

Then φ is directionally differentiable at μ along any $\xi\in\mathbb{R}^+\cdot\exp_\mu^{-1}(\mathscr{P}_2(\mathbb{R}^d))$, and

$$D_\mu\varphi_{++}(\xi)=D_\mu\varphi_{-+}(\xi)=\langle\xi,\zeta\rangle^+_\mu, \qquad D_\mu\varphi_{+-}(\xi)=D_\mu\varphi_{--}(\xi)=\langle\xi,\zeta\rangle^-_\mu. \tag{3.24}$$

The maps in (3.24) extend by continuity to Lipschitz maps over the tangent cone $\mathbf{Tan}_\mu$, so $\varphi_{\pm\pm}$ is differentiable in the sense of the metric structure on $\mathscr{P}_2(\mathbb{R}^d)$ inherited from the theory of CBB spaces. If moreover μ is such that $\xi\in\mathbf{Tan}_\mu$ implies $\lim_{h\searrow 0}\sup_{\zeta\in\frac{1}{h}\cdot\exp_\mu^{-1}(\exp_\mu(h\cdot\xi))}W_\mu(\zeta,\xi)=0$, then (3.24) holds for all $\xi\in\mathbf{Tan}_\mu$.

Proof. Consider $\xi = \lambda \cdot \eta$ for $\lambda > 0$ and $\eta \in \exp_\mu^{-1}(\mathscr{P}_2(\mathbb{R}^d))$. By [AGS05, Lemma 7.2.1], the set $\frac{1}{s} \cdot \exp_\mu^{-1}(\exp_\mu(h \cdot \xi))$ reduces to the singleton $\{\xi\}$ when $0 < s < \lambda^{-1}$. Consequently, for such s,

$$\frac{\varphi(\exp_\mu(s \cdot \xi)) - \varphi(\mu)}{s} = \frac{\langle s \cdot \xi, \zeta \rangle_\mu^\pm - 0}{s} = \langle \xi, \zeta \rangle_\mu^\pm. \tag{3.25}$$

This shows the first part of the statement. Assume now that μ is as in the statement. For each $\xi, \overline{\xi} \in \mathbf{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$, the Lipschitz-continuity of the metric scalar products (see [Gig08, Proposition 4.21]) yields

$$\begin{aligned}
\varphi(\exp_\mu(s \cdot \xi)) - \varphi(\exp_\mu(s \cdot \overline{\xi})) &\leqslant \sup_{\eta \in \exp_\mu^{-1}(\exp_\mu(s \cdot \xi))} \inf_{\overline{\eta} \in \exp_\mu^{-1}(\exp_\mu(s \cdot \overline{\xi}))} \langle \eta, \zeta \rangle_\mu^\pm - \langle \overline{\eta}, \zeta \rangle_\mu^\pm \\
&\leqslant \|\zeta\|_\mu \sup_{\eta \in \frac{1}{s} \cdot \exp_\mu^{-1}(\exp_\mu(s \cdot \xi))} \inf_{\overline{\eta} \in \frac{1}{s} \cdot \exp_\mu^{-1}(\exp_\mu(s \cdot \overline{\xi}))} s W_\mu(\eta, \overline{\eta}) \\
&\leqslant s \|\zeta\|_\mu \left(O_\xi(s) + O_{\overline{\xi}}(s) + W_\mu(\xi, \overline{\xi}) \right).
\end{aligned}$$

As the roles of ξ and $\overline{\xi}$ are symmetric, this gives

$$\frac{\left| \varphi(\exp_\mu(s \cdot \xi)) - \varphi(\exp_\mu(s \cdot \overline{\xi})) \right|}{s} \leqslant \|\zeta\|_\mu \left(O_\xi(s) + O_{\overline{\xi}}(s) + W_\mu(\xi, \overline{\xi}) \right). \tag{3.26}$$

Fix $\xi \in \mathbf{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$. Evaluating the above with $\overline{\xi} = 0_\mu$ (for which $O_{\overline{\xi}} \equiv 0$) gives that the quotients $\frac{\varphi(\exp_\mu(s \cdot \xi)) - \varphi(\mu)}{s}$ are bounded uniformly in $s \in (0,1]$. Passing to a vanishing subsequence $(s_n)_{n \in \mathbb{N}}$, one may assume that

$$\lim_{n \to \infty} \frac{\varphi(\exp_\mu(s_n \cdot \xi)) - \varphi(\mu)}{s_n} =: A \in \mathbb{R}$$

exists. For any $\overline{\xi} \in \mathbb{R}^+ \cdot \exp_\mu^{-1}\left(\mathscr{P}_2(\mathbb{R}^d)\right)$, taking the limit in $n \to \infty$ in (3.26) yields $|A - D_\mu \varphi(\overline{\xi})| = |A - \langle \overline{\xi}, \zeta \rangle_\mu^\pm| \leqslant \|\zeta\|_\mu W(\xi, \overline{\xi})$. Recalling that $\mathbf{Tan}_\mu = \overline{\mathbb{R}^+ \cdot \exp_\mu^{-1}\left(\mathscr{P}_2(\mathbb{R}^d)\right)}^{W_\mu}$ by definition, and using the continuity of the metric scalar product, we deduce that $A = \langle \xi, \zeta \rangle_\mu^\pm$. This holds independently of the chosen subsequence $(s_n)_{n \in \mathbb{N}}$, so that the directional derivative of φ exists and is given by (3.24) for any $\xi \in \mathbf{Tan}_\mu$. □

However, this is not satisfactory, since the applications in Lemma 3.2.5 may be discontinuous. Indeed, consider $\varphi = \varphi_{++}$, the dimension $d = 2$ and

$$A = (1,0), \quad B = (0,-1), \quad C = (-1,0), \quad D = (0,1), \qquad \mu := \frac{\delta_A + \delta_C}{2}, \qquad \sigma := \frac{\delta_B + \delta_D}{2}.$$

The set $\exp_\mu^{-1}(\sigma)$ contains the two distinct velocities

$$\xi = \frac{\delta_{(A,D-A)} + \delta_{(D,B-C)}}{2}, \qquad \zeta := \frac{\delta_{(A,B-A)} + \delta_{(C,D-C)}}{2}.$$

Let us approximate $\varphi(\sigma) = \varphi(\exp_\mu(\xi)) = \varphi(\exp_\mu(\zeta))$ along two different paths. Using [AGS05, Lemma 7.2.1], the geodesic between μ and $\exp_\mu(s \cdot \xi)$ is unique whenever $s < 1$, that is,

$$\exp_\mu^{-1}\left(\exp_\mu(s \cdot \xi)\right) = \{s \cdot \xi\}, \qquad \exp_\mu^{-1}\left(\exp_\mu(s \cdot \zeta)\right) = \{s \cdot \zeta\}.$$

Then

$$\lim_{s \to 1} \varphi(\exp_\mu(s \cdot \zeta)) = \lim_{s \to 1} \langle s\zeta, \xi \rangle_\mu^+ = 0, \qquad \lim_{s \to 1} \varphi(\exp_\mu(s \cdot \xi)) = \lim_{s \to 1} \langle s\xi, \xi \rangle_\mu^+ = 1,$$

and φ is discontinuous at σ. For this reason, embedding $\varphi_{\pm\pm}$ directly into sets of test functions would force to consider very large sets $\mathscr{T}_\pm$.

To conclude, it seems to us that viscosity solutions in $\mathscr{P}_2(\mathbb{R}^d)$ split into two branches: a "regular" branch, with smooth test functions and only barycentric elements, and a "metric" branch, with general semidifferentials allowing to recover the metric slope. The definitions are not equivalent, but one expects that for regular Hamiltonians, the sets of semisolutions coincide, as for Hamiltonians depending only on scalar products between ξ and some map. We do not know the exact condition for this equivalence.

3.2.2 Applicability of the comparison principle

From now on, we consider viscosity solutions as defined in Definition 3.1.6, with test functions defined in Definition 3.1.5 using the CBB structure.

In order to make the theory applicable to optimal control problems, we have to take into account that the continuity equation has *a priori* no link with the metric structure. More precisely, a probability vector field $\mu \mapsto F[\mu] \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ may be Lipschitz with respect to μ even if $\mu \mapsto \pi_T^\mu F[\mu]$ is discontinuous. Hence the well-posedness results for $\partial_t \mu_t + \operatorname{div}(F[\mu_t]\#\mu_t) = 0$ may apply to F, but not to its projection on $\mathbf{Tan}_\mu$. The Lipschitz-continuity is understood here in the sense of the application $W_{\mu,\nu} : \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \times \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\nu \to \mathbb{R}^+$, that we defined in Definition 1.1.39 by

$$W_{\mu,\nu}^2(\xi,\zeta) := \inf\left\{ \int_{(x,v),(y,w)\in \mathrm{T}\mathbb{R}^d} |v-w|^2 \,\middle|\, \alpha \in \Gamma(\xi,\zeta) \text{ and } (\pi_x,\pi_y)\#\alpha \in \Gamma_o(\mu,\nu) \right\}.$$

For instance, consider $f : x \mapsto Rx$ a rotation in $\mathbb{R}^2$. The probability vector field $F[\mu] := (id, f)\#\mu$ satisfies the strong condition

$$W_{\mu,\nu}\big(F[\mu], F[\nu]\big) \leqslant d_{\mathcal{W}}(\mu,\nu),$$

but the projection $\pi_T^\mu F[\mu]$ is discontinuous along a sequence $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathbb{R}^d)$ of purely atomic measures with a finite number of atoms, on which $\pi_T^{\mu_n} F[\mu_n] = F[\mu_n]$, converging towards the Hausdorff measure on the unit circle, on which $\pi_T^\mu F[\mu] = 0_\mu$. The following result gives sufficient regularity conditions on a dynamic with values in $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ in order to satisfy the assumptions of the comparison principle Theorem 3.1.12.

Proposition 3.2.6 (Hamiltonians supporting a comparison principle). *Let A, B be sets. Let $F : \mathscr{P}_2(\mathbb{R}^d) \times A \times B \to \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)$ be a probability vector field such that*

- *$F[\mu,a,b] \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $(a,b) \in A \times B$,*
- *F grows at most linearly: there exists $C^F \geqslant 0$ such that $\sup_{(a,b)\in A\times B} \|F[\mu,a,b]\|_\mu \leqslant C^F \left(1 + d_{\mathcal{W}}(\delta_0,\mu)\right)$ for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$,*
- *F is locally Lipschitz-continuous in the strong sense: for any $r > 0$, there exists $C_r^F \geqslant 0$ such that for all $\mu,\nu \in \overline{\mathscr{B}}(\delta_0,r)$, $W_{\mu,\nu}\left(F[\mu,a,b], F[\nu,a,b]\right) \leqslant C_r^F d_{\mathcal{W}}(\mu,\nu)$.*

Let $\ell : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ be such that

- *$\sup_{(a,b)\in A\times B} \left|\ell(\mu,a,b)\right| < \infty$ for each $\mu \in \mathscr{P}_2(\mathbb{R}^d)$,*
- *ℓ is locally Lipschitz-continuous with respect to $d_{\mathcal{W}}$: for any $r > 0$, there exists $C_r^\ell \geqslant 0$ such that for all $\mu,\nu \in \overline{\mathscr{B}}(\delta_0,r)$, $\left|\ell(\mu,a,b) - \ell(\nu,a,b)\right| \leqslant C_r^\ell d_{\mathcal{W}}(\mu,\nu)$.*

The Hamiltonian $H : \mathbb{T} \to \mathbb{R}$ defined by

$$H(\mu,p) := \sup_{a\in A} \inf_{b\in B} -p\left(\pi_T^\mu F[\mu,a,b]\right) - \ell(\mu,a,b)$$

satisfies the assumptions of the comparison principle given in Theorem 3.1.12.

The proof uses the estimate of Lemma 1.1.42, which states that for all measures $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and measure fields $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ and $\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\nu$,

$$D_\mu d_{\mathcal{W}}^2(\cdot,\nu)(\xi) + D_\nu d_{\mathcal{W}}^2(\mu,\cdot)(\zeta) \leqslant 2 d_{\mathcal{W}}(\mu,\nu) W_{\mu,\nu}(\xi,\zeta). \tag{3.27}$$

Proof. For each $(\mu,p) \in \mathbb{T}$, the value of the Hamiltonian is finite. Using that $\operatorname{opt}_{x\in X} f(x) - \operatorname{opt}_{x\in X} g(x) \leqslant \sup_{x\in X} f(x) - g(x)$ for $\operatorname{opt} \in \{\inf,\sup\}$, one gets

$$H(\mu,p) - H(\mu,q) \leqslant \sup_{a\in A}\sup_{b\in B}\left[-p\left(\pi_T^\mu F[\mu,a,b]\right) + q\left(\pi_T^\mu F[\mu,a,b]\right)\right] \leqslant \|p-q\|_\mu \sup_{(a,b)\in A\times B} \|\pi_T^\mu F[\mu,a,b]\|_\mu.$$

The projection on the tangent cone can only decrease the norm, so that the assumption $\|F[\mu,a,b]\|_\mu \leq C^F\left(1+d_{\mathcal{W}}(\delta_0,\mu)\right)$ proves the first part of the claim. Secondly, for any $r>0$, $\mu,\nu\in\overline{\mathscr{B}}(\delta_0,r)$, and $\lambda\geq 0$, the same argument yields

$$\begin{aligned}&H\left(\nu,-\lambda D_\nu d_{\mathcal{W}}^2(\mu,\cdot)\right)-H\left(\mu,\lambda D_\mu d_{\mathcal{W}}^2(\cdot,\nu)\right)\\ &\leq \sup_{(a,b)\in A\times B}\lambda\left(D_\nu d_{\mathcal{W}}^2(\mu,\cdot)\left(\pi_T^\nu F[\nu,a,b]\right)+D_\mu d_{\mathcal{W}}^2(\cdot,\nu)\left(\pi_T^\mu F[\mu,a,b]\right)\right)+\ell(\mu,a,b)-\ell(\nu,a,b).\end{aligned}$$

On the one hand, $\ell(\mu,a,b)-\ell(\nu,a,b)\leq C_r^\ell d_{\mathcal{W}}(\mu,\nu)$. On the other hand, the directional derivative of the squared Wasserstein distance along $\pi_T^\mu\xi$ is equal to that along ξ. By (3.27), we get

$$\lambda\left(D_\nu d_{\mathcal{W}}^2(\mu,\cdot)\left(F[\nu,a,b]\right)+D_\mu d_{\mathcal{W}}^2(\cdot,\nu)\left(F[\mu,a,b]\right)\right)\leq 2\lambda d_{\mathcal{W}}(\mu,\nu)W_{\mu,\nu}\left(F[\nu,a,b],F[\mu,a,b]\right)\leq 2\lambda C_r^F d_{\mathcal{W}}^2(\mu,\nu).$$

Combining both estimates, we conclude to the validity of **[A3.1.11]** with $C_r:=\max(C_r^F,C_r^\ell)$. □

We point that locally uniform upper semicontinuity in $\left(\mathscr{P}_2(\mathbb{R}^d),d_{\mathcal{W}}\right)$ is not comparable to a continuity in the narrow topology, in the sense that neither implies the other. Indeed, the squared Wasserstein distance is locally uniformly continuous, hence luusc, but is not narrowly upper semicontinuous. On the other hand, consider $u:=\mathbb{I}_G$, where $G\subset\mathscr{P}_2(\mathbb{R}^2)$ is the narrowly closed set

$$G:=\{\delta_{(0,0)}\}\cup\left\{(1-\lambda^2)\delta_{(0,0)}+\lambda^2\delta_{(1/\lambda,0)}\;\middle|\;0<\lambda\leq 1\right\}.$$

Consider the decreasing family of sets

$$B_n:=B_n^0\cup B^1,\qquad \text{where}\qquad \begin{cases}B_n^0:=\left\{(1-\lambda^2)\delta_{(0,0)}+\lambda^2\delta_{(1/\lambda,0)}\;\middle|\;0<\lambda\leq e^{-n}\right\},\\ B^1:=\left\{(1-\lambda)\left[(1-\lambda^2)\delta_{(0,0)}+\lambda^2\delta_{(1/\lambda,0)}\right]+\lambda\delta_{(0,1)}\;\middle|\;0<\lambda\leq 1\right\}.\end{cases}$$

Each B_n is contained in the unit ball centred in $\delta_{(0,0)}$. Moreover, B_n is closed with respect to the Wasserstein distance. Indeed, a Cauchy sequence $(\mu_m)_m$ of B_n^0 is identified by the corresponding sequence of parameters $(\lambda_m)_{m\in\mathbb{N}}$. If $\lambda_m\to_m 0$, then the sequence $(\mu_m)_m$ converges to its narrow limit point $\delta_{(0,0)}$, which is absurd since $d_{\mathcal{W}}(\mu_m,\delta_{(0,0)})=1$ for all m. Hence $(\lambda_m)_m$ is bounded away from 0, and admits a limit point, that is easily proved to be a parameter of the limit. The same argument proves the closedness of B^1. The intersection of the sets B_n is reduced to B^1, and there holds

$$\sup_{\mu\in B_n}\inf_{\nu\in B_1}d_{\mathcal{W}}(\mu,\nu)\leq\sup_{\lambda\leq e^{-n}}d_{\mathcal{W}}\left(\mu_\lambda,(1-\lambda)\mu_\lambda+\lambda\delta_{(0,1)}\right)\leq\sup_{\lambda\leq e^{-n}}\sqrt{(1-\lambda)\times 0+\lambda d_{\mathcal{W}}^2\left(\mu_\lambda,\delta_{(0,1)}\right)}\leq 2e^{-2n}.$$

Here $\mu_\lambda:=(1-\lambda^2)\delta_{(0,0)}+\lambda^2\delta_{(1/\lambda,0)}$, and we used the vertical convexity of the squared Wasserstein distance and the triangular inequality. As $\sup_{\mu\in B_n}u(\mu)=1$ for all n, but $\sup_{\mu\in B_1}u(\mu)=0$, we conclude that u is not locally uniformly upper semicontinuous.

3.2.3 HJB characterization of the value function

In this section, we apply the theory of Section 3.1.3 to characterize the value function of a control problem as the unique viscosity solution of a certain Hamilton-Jacobi-Bellman equation.

3.2.3.1 Continuity equations

In the Wasserstein space, the role of ODEs is played by continuity equations. This is an analogy, but also a generalization in the mathematical sense, since ODEs are particular cases of continuity equations when the initial point is a Dirac mass. To introduce the definition, denote by $X\subset\mathcal{C}(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ the set of locally Lipschitz vector fields with linear growth, endowed with the topology of uniform convergence on compact sets.

Definition 3.2.7 (Continuity equation). *Let $I\subset\mathbb{R}$ be a nonempty open interval, and $f\in L^1(\mathscr{P}_2(\mathbb{R}^d)\times I;X)$. A curve of measure $(\mu_s)_{s\in I}\in AC(I;\mathscr{P}_2(\mathbb{R}^d))$ is a solution to the continuity equation*

$$\partial_s\mu_s+\operatorname{div}\left(f[\mu(s),s]\#\mu_s\right)=0\qquad s\in I$$

if it is so in the sense of distributions, i.e. if for any $\varphi \in C_c^\infty\left(I \times \mathbb{R}^d; \mathbb{R}\right)$, there holds

$$\int_{(s,x)\in I\times\mathbb{R}^d} \partial_s\varphi(s,x) + \langle \nabla_x\varphi(s,x), f[\mu_s,s](x)\rangle \, d\mu_s(x)ds = 0. \tag{3.28}$$

The gradient term in (3.28) is the $L^2_{\mu_s}$ scalar product between the element $(id, \nabla\varphi)\#\mu_s \in \mathrm{Tan}_{\mu_s}$ and the vector field $f[\mu_s,s]$. It it therefore equal to the scalar product between $(id,\nabla\varphi)\#\mu_s$ and the projection $\pi_T^{\mu_s} f[\mu_s,s]$ on the regular tangent cone to μ_s. This shows that a solution of the continuity equation automatically satisfies (3.28) where $f[\mu_s,s]$ is replaced by its projection; however, it is useful to work with general vector fields instead of elements of the various tangent cones, since the application $\mu \mapsto \pi_T^\mu$ is quite irregular.

To go further in the geometric extensions, let us point that an extension of continuity equations to general measure fields has been proposed in [Pic19]. These Measure Differential Equations (MDEs) allow for the mass to split, and some existence results are available by considering limits of finite volume schemes [Cam+21]. The general picture that is emerging is that whenever the measure field is regular, for instance when the barycenter is Lipschitz as a function of space, then the solution of the MDE coincides with the solution of the continuity equation driven by the barycenter field. In [CSS23a; CSS23b], the authors clarify the situation for dissipative probability vector fields, by giving the precise regularity conditions on the measure field that allow for the convergence of the explicit and/or implicit Euler scheme directly at the level of MDEs. It is noticeable that the said conditions are weaker if one considers the equation directly in the Wasserstein space, instead of lifting it to a Hilbert space and use the theory of dissipative equations there.

We are interested in controlled continuity equations, where the dynamic f can be oriented by a parameter $u \in U$. More precisely, let U be a compact metric space of controls. The dynamic is taken as follows.

Assumption [A3.2.8] (Structure of the dynamic). Assume that $f : \mathscr{P}_2(\mathbb{R}^d) \times U \to X$ is valued in the set X of locally Lipschitz vector field with at most linear growth, and that

– there exists a constant $\mathrm{Lip}(f) \geq 0$ such that for all $u \in U$, $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ and $x, y \in \mathbb{R}^d$,

$$\left|f[\mu,u](x) - f[\nu,u](y)\right| \leq \mathrm{Lip}(f)\left(|x-y| + d_{\mathcal{W}}(\mu,\nu)\right).$$

– There exists a modulus of continuity $m_{f,U} : \mathbb{R}^+ \to \mathbb{R}^+$ such that for all $u, v \in U$, $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$,

$$\left|f[\mu,u](x) - f[\mu,v](x)\right| \leq (1+|x|)m_{f,U}(d_U(u,v)).$$

– There exists $|f|_{0,\infty} \geq 0$ such that for all $u \in U$ and $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, there holds $\left|f[\mu,u](0)\right| \leq |f|_{0,\infty}$.

We quote from Theorems 2.18, Proposition 2.22 and Theorem 4.2 of [BF24] the following well-posedness and representation result.

Proposition 3.2.9 (Existence, uniqueness and representation). *Assume that f satisfies **[A3.2.8]**. For each $\nu \in \mathscr{P}_2(\mathbb{R}^d)$, $t \in [0,T)$ and $u \in L^1(t,T;U)$, there exists a unique solution $\left(\mu_s^{t,\nu,u}\right)_{s\in[t,T]} \in AC\left([t,T]; \mathscr{P}_2(\mathbb{R}^d)\right)$ of the continuity equation*

$$\partial_s\mu_s + \mathrm{div}\left(f[\mu_s,u(s)]\#\mu_s\right) = 0 \qquad s \in (t,T), \qquad\qquad \mu_t = \nu. \tag{3.29}$$

Moreover, this solution is given by $\mu_s^{t,\nu,u} = \Phi_s^t\#\nu$, where

$$\frac{d}{ds}\Phi_s^t(x) = f[\mu_s^{t,\nu,u}, u(s)](\Phi_s^t(x)), \qquad \Phi_t^t(x) = x.$$

The representation by pushforward allows to obtain various estimates from iterated applications of the Grönwall lemma. The only idea that one has to have is to take an optimal transport plan between the initial points, and "push it" along the trajectories; the remaining computations are classical.

Lemma 3.2.10 (Grönwall estimates). *Assume [A3.2.8]. Let $(\mu_s^{t,\nu,u})_{s\in[t,T]}$ denote the solution of (3.29) issued from $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ at time $t \in [0,T]$ and driven by the control $u \in L^1(t,T;U)$. Let $0 \leqslant t \leqslant \overline{s} \leqslant s \leqslant T$, $\nu, \overline{\nu} \in \mathscr{P}_2(\mathbb{R}^d)$ and $u, \overline{u} \in L^1(t,T;U)$. Then*

$$d_{\mathcal{W}}\left(\mu_s^{t,\nu,u}, \mu_{\overline{s}}^{t,\overline{\nu},\overline{u}}\right) \leqslant (s-\overline{s})\left(Lip(f)(\overline{s}-t)e^{(s-t)Lip(f)} + e^{(s-\overline{s})Lip(f)}\right)\left(Lip(f)d_{\mathcal{W}}(\delta_0,\nu) + |f|_{0,\infty}\right)$$
$$+ e^{Lip(f)(\overline{s}-t)\left(1+e^{Lip(f)(\overline{s}-t)}\right)}\left(d_{\mathcal{W}}(\nu,\overline{\nu}) + E_{t,\overline{s},\overline{\nu},u,\overline{u}}\right),$$

where

$$E_{t,\overline{s},\overline{\nu},u,\overline{u}} := \left(1 + (\overline{s}-t)\left(|f|_{0,\infty} + Lip(f)d_{\mathcal{W}}(\delta_0,\overline{\nu})\right)e^{(\overline{s}-t)Lip(f)}\right)\int_{r=t}^{\overline{s}} m_{f,U}\left(d_U(u(r),\overline{u}(r))\right)dr.$$

As particular cases, we record that

$$d_{\mathcal{W}}(\mu_{t+h}^{t,\nu,u},\nu) \leqslant h\left(\mathrm{Lip}(f)d_{\mathcal{W}}(\delta_0,\nu) + |f|_{0,\infty}\right)e^{h\mathrm{Lip}(f)},$$
$$d_{\mathcal{W}}\left(\mu_T^{t,\nu,u}, \mu_T^{t,\overline{\nu},u}\right) \leqslant e^{\mathrm{Lip}(f)(T-t)\left(1+e^{\mathrm{Lip}(f)(T-t)}\right)}d_{\mathcal{W}}(\nu,\overline{\nu}).$$

Proof. Denote $\left(\Phi_\tau^{t,\nu,u}\right)_{\tau\in[t,s]}$ and $\left(\Phi_\tau^{t,\overline{\nu},\overline{u}}\right)_{\tau\in[t,\overline{s}]}$ the flows of the ODEs

$$\dot{y}_\tau = f[\mu_\tau^{t,\nu,u}, u(\tau)](y_\tau), \qquad \dot{\overline{y}}_\tau = f[\mu_\tau^{t,\overline{\nu},\overline{u}}, \overline{u}(\tau)](\overline{y}_\tau).$$

On the one hand, for $t \leqslant r \leqslant \tau \leqslant s$,

$$\left|\Phi_\tau^{t,\nu,u}(x) - \Phi_r^{t,\nu,u}(x)\right| \leqslant \int_{\theta=r}^{\tau}\left|f\left[u(\theta),\mu_\theta^{t,\nu,u}\right]\left(\Phi_\theta^{t,\nu,u}(x)\right)\right|d\theta$$
$$\leqslant \int_{\theta=r}^{\tau}\left[\mathrm{Lip}(f)\left(|\Phi_\theta^{t,\nu,u}(x) - \Phi_r^{t,\nu,u}(x)| + \left|\Phi_r^{t,\nu,u}(x)\right|\right) + |f|_{0,\infty}\right]d\theta,$$

so that a Grönwall lemma yields

$$\left|\Phi_\tau^{t,\nu,u}(x) - \Phi_r^{t,\nu,u}(x)\right| \leqslant (\tau-r)\left(\mathrm{Lip}(f)\left|\Phi_r^{t,\nu,u}(x)\right| + |f|_{0,\infty}\right)e^{(\tau-r)\mathrm{Lip}(f)}. \tag{3.30}$$

In particular, choosing $\tau = t$,

$$\left|\Phi_\tau^{t,\nu,u}(x) - \Phi_r^{t,\nu,u}(x)\right| \leqslant (\tau-r)\left(\mathrm{Lip}(f)\left|\Phi_r^{t,\nu,u}(x) - x\right| + \mathrm{Lip}(f)|x| + |f|_{0,\infty}\right)e^{(\tau-r)\mathrm{Lip}(f)}$$
$$\leqslant (\tau-r)\left(\mathrm{Lip}(f)(r-t)e^{(r-t)\mathrm{Lip}(f)} + 1\right)\left(\mathrm{Lip}(f)|x| + |f|_{0,\infty}\right)e^{(\tau-r)\mathrm{Lip}(f)}.$$

Taking the square of each side and integrating with respect to ν, we get by the pushforward representation

$$\begin{aligned} d_{\mathcal{W}}(\mu_\tau^{t,\nu,u}, \mu_r^{t,\nu,u}) &\leqslant \sqrt{\int_{x\in\mathbb{R}^d}\left|\Phi_\tau^{t,\nu,u}(x) - \Phi_r^{t,\nu,u}(x)\right|^2 d\nu(x)} \\ &\leqslant (\tau-r)\left(\mathrm{Lip}(f)(r-t)e^{(r-t)\mathrm{Lip}(f)} + 1\right)\left(\mathrm{Lip}(f)d_{\mathcal{W}}(\delta_0,\nu) + |f|_{0,\infty}\right)e^{(\tau-r)\mathrm{Lip}(f)}. \end{aligned} \tag{3.31}$$

On the other hand, for $\tau \in [t,\overline{s}]$,

$$\left|\Phi_\tau^{t,\nu,u}(x) - \Phi_\tau^{t,\overline{\nu},\overline{u}}(y)\right| \leqslant |x-y| + \mathrm{Lip}(f)\int_{r=t}^{\tau}\left|\Phi_r^{t,\nu,u}(x) - \Phi_r^{t,\overline{\nu},\overline{u}}(y)\right| + d_{\mathcal{W}}\left(\mu_r^{t,\nu,u}, \mu_r^{t,\overline{\nu},\overline{u}}\right)dr$$
$$+ \left(1 + (\tau-t)\left(|f|_{0,\infty} + \mathrm{Lip}(f)|y|\right)e^{(\tau-t)\mathrm{Lip}(f)}\right)\int_{r=t}^{\tau} m_{f,U}\left(d(u(r),\overline{u}(r))\right)dr.$$

Applying a second Grönwall lemma,

$$\left|\Phi_\tau^{t,\nu,u}(x) - \Phi_\tau^{t,\overline{\nu},\overline{u}}(y)\right| \leqslant \left(|x-y| + \mathrm{Lip}(f)\int_t^{\tau} d_{\mathcal{W}}\left(\mu_r^{t,\nu,u}, \mu_r^{t,\overline{\nu},\overline{u}}\right)dr + E_{t,\tau,y,u,\overline{u}}\right)e^{\mathrm{Lip}(f)(\tau-t)},$$

where $E_{t,\tau,y,u,\overline{u}} := \left(1+(\tau-t)\left(\left|f\right|_{0,\infty}+\mathrm{Lip}(f)\left|y\right|\right)e^{(\tau-t)\mathrm{Lip}(f)}\right)\int_{r=t}^{\tau} m_{f,U}\left(d(u(r),\overline{u}(r))\right)dr$. Now, let $\eta\in\Gamma_o(\nu,\overline{\nu})$. The "pushforwarded" plan $\eta_{\overline{s}} := \left(\Phi_{\overline{s}}^{t,\nu,u},\Phi_{\overline{s}}^{t,\overline{\nu},\overline{u}}\right)\#\eta$ belongs to $\Gamma(\mu_{\overline{s}}^{t,\nu,u},\mu_{\overline{s}}^{t,\overline{\nu},\overline{u}})$, so that

$$\begin{aligned} d_{\mathcal{W}}(\mu_{\overline{s}}^{t,\nu,u},\mu_{\overline{s}}^{t,\overline{\nu},\overline{u}}) &\leqslant \sqrt{\int_{(\mathbb{R}^d)^2}\left|\Phi_{\overline{s}}^{t,\nu,u}(x)-\Phi_{\overline{s}}^{t,\overline{\nu},\overline{u}}(y)\right|^2 d\eta(x,y)} \\ &\leqslant e^{\mathrm{Lip}(f)(\overline{s}-t)}\sqrt{\int_{(\mathbb{R}^d)^2}\left(\left|x-y\right|+\int_{r=t}^{\overline{s}}\mathrm{Lip}(f)d_{\mathcal{W}}\left(\mu_r^{t,\nu,u},\mu_r^{t,\overline{\nu},\overline{u}}\right)dr+E_{t,\overline{s},y,u,\overline{u}}\right)^2 d\eta(x,y)} \\ &\leqslant e^{\mathrm{Lip}(f)(\overline{s}-t)}\left(d_{\mathcal{W}}(\nu,\overline{\nu})+\int_{r=t}^{\overline{s}}\mathrm{Lip}(f)d_{\mathcal{W}}\left(\mu_r^{t,\nu,u},\mu_r^{t,\overline{\nu},\overline{u}}\right)dr+\sqrt{\int_{y\in\mathbb{R}^d}E^2_{t,\overline{s},y,u,\overline{u}}d\overline{\nu}(y)}\right). \end{aligned}$$

Denote $E^2_{t,\overline{s},\overline{\nu},u,\overline{u}} := \int_{y\in\mathbb{R}^d}E^2_{t,\overline{s},y,u,\overline{u}}d\overline{\nu}(y)$. As

$$\begin{aligned} E^2_{t,\overline{s},\overline{\nu},u,\overline{u}} &= \int_{\mathbb{R}^d}\left(1+(\overline{s}-t)\left(\left|f\right|_{0,\infty}+\mathrm{Lip}(f)\left|y\right|\right)e^{(\overline{s}-t)\mathrm{Lip}(f)}\right)^2\left(\int_t^{\tau} m_{f,U}\left(d(u(r),\overline{u}(r))\right)dr\right)^2 d\overline{\nu}(y) \\ &\leqslant \left(1+(\overline{s}-t)\left(\left|f\right|_{0,\infty}+\mathrm{Lip}(f)d_{\mathcal{W}}\left(\delta_0,\overline{\nu}\right)\right)e^{(\overline{s}-t)\mathrm{Lip}(f)}\right)^2\left(\int_{r=t}^{\tau} m_{f,U}\left(d(u(r),\overline{u}(r))\right)dr\right)^2 < \infty, \end{aligned}$$

we are ready to apply our third Grönwall lemma to get

$$d_{\mathcal{W}}(\mu_{\overline{s}}^{t,\nu,u},\mu_{\overline{s}}^{t,\overline{\nu},\overline{u}}) \leqslant e^{\mathrm{Lip}(f)(\overline{s}-t)}\left(d_{\mathcal{W}}(\nu,\overline{\nu})+E_{t,\overline{s},\overline{\nu},u,\overline{u}}\right)e^{\mathrm{Lip}(f)(\overline{s}-t)e^{\mathrm{Lip}(f)(\overline{s}-t)}}. \tag{3.32}$$

Combining (3.31) and (3.32), we get the desired result. □

We readily deduce the following. For any $\nu\in\mathscr{P}_2(\mathbb{R}^d)$ and $0\leqslant t\leqslant T$, define the reachable set

$$\mathcal{R}_T^{t,\nu} := \left\{\mu_T^{t,\nu,u} \mid u\in L^1(t,T;U)\right\}. \tag{3.33}$$

Corollary 3.2.11 (Lipschitz-continuity of the reachable sets). *Assume* ***[A3.2.8]****. There exists a constant Lip*$(\mathcal{R})$ *depending only on* f *and* T *such that for all* $t\in[0,T]$ *and* $\nu,\overline{\nu}\in\mathscr{P}_2(\mathbb{R}^d)$,

$$\max\left(\sup_{\mu\in\mathcal{R}_T^{t,\nu}}\inf_{\overline{\mu}\in\mathcal{R}_T^{t,\overline{\nu}}}d_{\mathcal{W}}(\mu,\overline{\mu}),\ \sup_{\overline{\mu}\in\mathcal{R}_T^{t,\overline{\nu}}}\inf_{\mu\in\mathcal{R}_T^{t,\nu}}d_{\mathcal{W}}(\mu,\overline{\mu})\right)\leqslant Lip(\mathcal{R})d_{\mathcal{W}}(\nu,\overline{\nu}).$$

Proof. By Lemma 3.2.10, two solutions $s\mapsto\mu_s^{t,\nu,u}$ and $s\mapsto\mu_s^{t,\overline{\nu},u}$ driven by the same $u\in L^1(t,T;U)$ satisfy

$$d_{\mathcal{W}}\left(\mu_T^{t,\nu,u},\mu_T^{t,\overline{\nu},u}\right)\leqslant\exp\left(\mathrm{Lip}(f)(T-t)\left(e^{\mathrm{Lip}(f)(T-t)}+1\right)\right)d_{\mathcal{W}}(\nu,\overline{\nu}).$$

The claim follows by approximating each $\mu=\mu_T^{t,\nu,u}\in\mathcal{R}_T^{t,\nu}$ by $\mu_T^{t,\overline{\nu},u}$, with $\mathrm{Lip}(\mathcal{R}) := e^{\mathrm{Lip}(f)T\left(e^{\mathrm{Lip}(f)T}+1\right)}$. □

Corollary 3.2.12 (Smooth case). *Assume* ***[A3.2.8]****. Let* $0\leqslant t\leqslant T$ *and* $u\in L^1(t,T;U)$ *be a constant control* $u(s)\equiv\overline{u}\in U$. *Then the unique solution* $(\mu_s^{t,\nu,u})_{s\in[t,T]}$ *issued from* $\nu\in\mathscr{P}_2(\mathbb{R}^d)$ *satisfies*

$$\lim_{h\searrow 0}\frac{d_{\mathcal{W}}\left(\mu_{t+h}^{t,\nu,u},\exp_\nu(h\cdot f[\nu,\overline{u}]\#\nu)\right)}{h}=0.$$

Proof. Let $s\mapsto\Phi_s^t$ be the flow of the ODE $\frac{d}{ds}y_s=f[\mu_s^{t,\nu,u},\overline{u}](y_s)$, such that $\mu_s^{t,\nu,u}=\Phi_s^t\#\nu$. There holds

$$\begin{aligned} d^2_{\mathcal{W}}\left(\mu_{t+h}^{t,\nu,u},\exp_\nu(h\cdot f[\nu,\overline{u}]\#\nu)\right) &\leqslant \int_{x\in\mathbb{R}^d}\left|\Phi_{t+h}^t(x)-(x+hf[\nu,\overline{u}](x))\right|^2 d\nu \\ &\leqslant \int_{x\in\mathbb{R}^d}h\int_{s=t}^{t+h}\left|f[\mu_s^{t,\nu,u},\overline{u}](\Phi_s^t(x))-f[\nu,\overline{u}](x)\right|^2 ds d\nu \\ &\leqslant \int_{x\in\mathbb{R}^d}h\int_{s=t}^{t+h}\mathrm{Lip}(f)^2\left(d_{\mathcal{W}}(\mu_s^{t,\nu,u},\nu)+\left|\Phi_s^t(x)-x\right|\right)^2 ds d\nu. \end{aligned}$$

Using the estimates of Lemma 3.2.10, we get to

$$\limsup_{h\searrow 0}\frac{d_{\mathcal{W}}\left(\mu_{t+h}^{t,\nu,u},\exp_\nu(h\cdot f[\nu,\overline{u}]\#\nu)\right)}{h}\leqslant\lim_{h\searrow 0}2h\mathrm{Lip}(f)\left(\mathrm{Lip}(f)d_{\mathcal{W}}(\nu,\delta_0)+\left|f\right|_{0,\infty}\right)e^{\mathrm{Lip}(f)h}=0.$$

Hence the result. □

3.2.3.2 The Mayer problem

Consider the following Mayer problem:

$$\begin{aligned}&\text{Minimize} \quad \mathfrak{J}(\mu_T^{0,\nu,u}) \quad \text{on } u(\cdot)\in L^1(0,T;U),\\ &\text{where } \partial_s\mu_s + \operatorname{div}(f[\mu_s,u(s)]\#\mu_s)=0, \text{ and } \mu_0^{0,\nu,u}=\nu.\end{aligned}$$

We assume at first that the terminal cost is quite regular. An extension is provided in Section 3.2.4.4.

Assumption [A3.2.13] (Structure of the terminal cost). The terminal cost $\mathfrak{J}:\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}$ is locally uniformly continuous, i.e. for each $R>0$, there exists a modulus of continuity $m_{\mathfrak{J},R}:\mathbb{R}^+\to\mathbb{R}^+$ such that

$$\left|\mathfrak{J}(\mu)-\mathfrak{J}(\nu)\right| \leqslant m_{\mathfrak{J},R}(d_{\mathcal{W}}(\mu,\nu)) \qquad \forall \mu,\nu\in\overline{\mathscr{B}}(\delta_0,R).$$

Introduce the value function $V:[0,T]\times\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}$ of the Mayer control problem as

$$V(t,\nu) := \inf_{u\in L^1(t,T;U)} \mathfrak{J}\left(\mu_T^{t,\nu,u}\right).$$

In this deterministic setting, we retrieve the classical Dynamic Programming Principle (DPP): for each $0<h\leqslant T-t$,

$$V(t,\nu) = \inf_{u\in L^1(t,t+h;U)} V(t+h,\mu_{t+h}^{t,\nu,u}). \tag{3.34}$$

Lemma 3.2.14 (Local uniform continuity of the value function). *Assume **[A3.2.8]** and **[A3.2.13]**. The value function V is locally uniformly continuous in time and space, i.e. for all $R>0$, there exists a modulus $m_{V,R}:\mathbb{R}^+\to\mathbb{R}^+$ such that*

$$\left|V(s,\nu)-V(t,\mu)\right| \leqslant m_{V,R}\left(|t-s|+d_{\mathcal{W}}(\mu,\nu)\right) \qquad \forall t,s\in[0,T] \text{ and } \mu,\nu\in\overline{\mathscr{B}}(\delta_0,R).$$

Proof. Let $R>0$, and denote $R_T := R + T\exp \mathrm{Lip}(f)T\left(\mathrm{Lip}(f)R+\left|f\right|_{0,\infty}\right)$ a radius large enough so that $\mathcal{R}_T^{0,\nu}\subset\overline{\mathscr{B}}(\delta_0,R_T)$ for all $\nu\in\overline{\mathscr{B}}(\delta_0,R)$. Let $m_{\mathfrak{J},R_T}$ be a local modulus of continuity of $\mathfrak{J}$ in the ball $\overline{\mathscr{B}}(\delta_0,R_T)$. According to the $\mathrm{Lip}(\mathcal{R})$–Lipschitz continuity of the reachable sets given by Corollary 3.2.11, we have for all $t\in[0,T]$ and $\nu,\overline{\nu}\in\overline{\mathscr{B}}(\delta_0,R)$ that

$$V(t,\nu)-V(t,\overline{\nu}) \leqslant \sup_{\overline{\mu}\in\mathcal{R}_T^{t,\overline{\nu}}}\inf_{\mu\in\mathcal{R}_T^{t,\nu}} \mathfrak{J}(\mu)-\mathfrak{J}(\overline{\mu}) \leqslant \sup_{\overline{\mu}\in\mathcal{R}_T^{t,\overline{\nu}}}\inf_{\mu\in\mathcal{R}_T^{t,\nu}} m_{\mathfrak{J},R_T}\left(d_{\mathcal{W}}(\mu,\overline{\mu})\right) \leqslant m_{\mathfrak{J},R_T}\left(\mathrm{Lip}(\mathcal{R})d_{\mathcal{W}}(\nu,\overline{\nu})\right).$$

On the other hand, let $0\leqslant t\leqslant s\leqslant T$ and $\nu\in\overline{\mathscr{B}}(\delta_0,R)$. The DPP (3.34) and the estimates of Lemma 3.2.10 give

$$\begin{aligned}V(t,\nu)-V(s,\nu) &= \inf_{\mu\in\mathcal{R}_s^{t,\nu}} V(s,\mu)-V(s,\nu) \leqslant \inf_{\mu\in\mathcal{R}_s^{t,\nu}} m_{\mathfrak{J},R_T}\left(\mathrm{Lip}(\mathcal{R})d_{\mathcal{W}}(\mu,\nu)\right)\\ &\leqslant m_{\mathfrak{J},R_T}\left(\mathrm{Lip}(\mathcal{R})\,|s-t|\,e^{\mathrm{Lip}(f)T}\left(\mathrm{Lip}(f)R+|f|_{0,\infty}\right)\right),\end{aligned}$$

and in the same way,

$$\begin{aligned}V(s,\nu)-V(t,\nu) &= \sup_{\mu\in\mathcal{R}_s^{t,\nu}} V(s,\mu)-V(s,\nu) \leqslant \sup_{\mu\in\mathcal{R}_s^{t,\nu}} m_{\mathfrak{J},R_T}\left(\mathrm{Lip}(\mathcal{R})d_{\mathcal{W}}(\mu,\nu)\right)\\ &\leqslant m_{\mathfrak{J},R_T}\left(\mathrm{Lip}(\mathcal{R})\,|s-t|\,e^{\mathrm{Lip}(f)T}\left(\mathrm{Lip}(f)R+|f|_{0,\infty}\right)\right).\end{aligned}$$

Hence V is locally uniformly continuous with a modulus depending on $\mathfrak{J}$, f and T. □

3.2.3.3 Characterization as the solution of a HJ PDE

Introduce now the control Hamiltonian

$$H:\mathbb{T}\to\mathbb{R}, \qquad H(\mu,p) := \sup_{u\in U} -p(\pi_T^{\mu}f[\mu,u]\#\mu). \tag{3.35}$$

Recall that the metric cotangent bundle $\mathbb{T}$ is the set of all pairs (μ,p), where $\mu\in\mathscr{P}_2(\mathbb{R}^d)$, and $p:\mathbf{Tan}_{\mu}\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}$ is a positively homogeneous and Lipschitz application.

Remark 3.2.15 (Projection)**.** *It may happen that $f[\mu,u]\#\mu$ does not belong to the tangent cone $\mathbf{Tan}_\mu$, hence the projection π_T^μ on the tangent cone in the definition. This has in fact very few consequences: by [Gig08, Corollary 4.37], there holds $W_\mu(\pi_T^\mu\xi,\pi_T^\mu\zeta) \leqslant W_\mu(\xi,\zeta)$ for any two measure fields $\xi,\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, so that all the estimates on $f[\mu,\cdot](\cdot)$ imply the same estimates on the projections. One should however be careful when taking directional derivatives along f or its projection. In our case, we can use results that are specific to measure fields induced by maps; if one were to consider the general case of a control problem over solutions of MDEs, then test functions should be taken such that $D_\mu\varphi(\xi) = D_\mu\varphi(\pi_T^\mu\xi)$ for all ξ.*

We state the next result for a dynamic with convex values, and postpone the general case to Section 3.2.4.1.

Theorem 3.2.16 (Characterization of the value function). *Assume **[A3.2.8]**, **[A3.2.13]** and that each set $f[\mu,U]$ is convex as a subset of X. The value function is the unique viscosity solution in the sense of Definition 3.1.6 of the Hamilton-Jacobi-Bellman equation*

$$\begin{cases} -\partial_t u(t,\mu) + H\big(\mu, D_\mu u(t,\mu)\big) = 0 & (t,\mu) \in [0,T) \times \mathscr{P}_2(\mathbb{R}^d), \\ u(T,\mu) = \mathfrak{J}(\mu) & \mu \in \mathscr{P}_2(\mathbb{R}^d). \end{cases} \tag{3.36}$$

The assumption of convex images of the dynamic allows to get the following classical result, whose proof is recalled for completeness.

Lemma 3.2.17 (Right linear approximation). *Assume **[A3.2.8]** and that each set $f[\mu,U] \subset X$ is convex. Let $\overline{s} > 0$, $(\mu_s)_{s\in[0,\overline{s}]}$ be the solution of (3.29) for some control $u \in L^1(0,\overline{s};U)$. Then there exists $b \in f[\mu_0,U]$ and a sequence $(s_n) \searrow 0$ such that*

$$\lim_{n\to\infty} \frac{d_{\mathcal{W}}\left(\mu_{s_n}, \exp_{\mu_0}(s_n \cdot b\#\mu)\right)}{s_n} = 0.$$

Proof. Let $(s_n)_n \searrow 0$, and define $b_n : \mathbb{R}^d \mapsto T\mathbb{R}^d$ by

$$b_n(x) := \frac{1}{s_n}\int_{s=0}^{s_n} f[\mu_0, u(s)](x)ds.$$

Since $f[\mu_0,U]$ is convex, $b_n \in X$, and there exists $u_n \in U$ such that $b_n = f[\mu_0,u_n]$. By assumption, U is compact, and along a (non relabeled) subsequence, u_n converges to some $u \in U$. Let $b := f[\mu_0,u]$. The regularity of f implies that $|b_n(x) - b(x)| \leqslant (1+|x|)m_{f,U}(d_U(u,u_n))$. By Proposition 3.2.9, $\mu_s = \Phi_s^{0,u}\#\mu_0$, where $\Phi_s^{0,u}$ is the flow of the underlying ODE $\frac{d}{ds}y_s = f[\mu_s,u(s)](y_s)$. Along the sequence $(s_n)_n$, we have that

$$\begin{aligned} \left|\Phi_{s_n}^{0,u}(x) - (x + s_n b(x))\right| &= \left|\int_{s=0}^{s_n} f[\mu_s,u(s)]\left(\Phi_s^{0,u}(x)\right)ds - s_n b(x)\right| \\ &\leqslant \int_{s=0}^{s_n} \left|f[\mu_s,u(s)]\left(\Phi_s^{0,u}(x)\right) - f[\mu_0,u(s)](x)\right| ds + s_n|b_n(x) - b(x)| \\ &\leqslant \mathrm{Lip}(f)\int_{s=0}^{s_n} d_{\mathcal{W}}(\mu_s,\mu_0) + \left|\Phi_s^{0,u}(x) - x\right| ds + s_n(1+|x|)m_{f,U}(d_U(u,u_n)). \end{aligned}$$

Using Lemma 3.2.10 and integrating the squares against μ_0, we get that $d_{\mathcal{W}}(\mu_{s_n}, \exp_{\mu_0}(s_n\cdot b\#\mu_0)) = o(s_n)$. □

We will also need the following technical result, proved in Corollary 5.3.9.

Lemma 3.2.18. *Let $b \in L_\mu^2(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$. Then $d_{\mathcal{W}}\left(\exp_\mu(h\cdot b\#\mu), \exp_\mu(h\cdot\pi_T^\mu b\#\mu)\right) = o(h)$.*

We can now turn to the proof of Theorem 3.2.16.

Proof. Let us show that it is a viscosity solution of (3.36). By Lemma 3.2.14, V is locally uniformly continuous, hence simultaneously luusc and lulsc. By definition, $V(T,\cdot) = \mathfrak{J}$, so that we only have to verify the viscosity inequalities.

Subsolution inequality. Let $\varphi \in \mathcal{T}_+$ such that $V - \varphi$ reaches a maximum at $(t,\mu) \in [0,T) \times \mathcal{P}_2(\mathbb{R}^d)$, and $\mathrm{Lip}(\varphi)$ a local Lipschitz constant in a neighbourhood of (t,μ) containing all trajectories issued from μ up to time T. By Corollary 3.2.12, each constant control $u \in L^1(t,T;U)$ given by $u(s) \equiv u \in U$ generates a smooth solution $(\mu_s^{t,\nu,u})_{s\in[t,T]}$, in the sense that

$$\lim_{h\searrow 0} \frac{d_{\mathcal{W}}\left(\mu_{t+h}^{t,\nu,u}, \exp_\nu(h\cdot f[\nu,u]\#\nu)\right)}{h} = 0.$$

Moreover, by Lemma 3.2.18, there holds

$$\lim_{h\searrow 0} \frac{d_{\mathcal{W}}\left(\exp_\nu(h\cdot f[\nu,u]\#\nu), \exp_\nu(h\cdot \pi_T^\nu f[\nu,u]\#\nu)\right)}{h} = 0.$$

Consequently, we may pass from the inequality on $h \mapsto \mu_{t+h}^{t,\nu,u}$ given by the dynamic programming principle (3.34), to the directional derivative along the exponential of the projection, i.e.

$$\begin{aligned}\varphi(t,\nu) = V(t,\nu) \leqslant V(t+h,\mu_{t+h}^{t,\nu,u}) \leqslant \varphi(t+h,\mu_{t+h}^{t,\nu,u}) \quad &\Rightarrow \quad \frac{\varphi(t,\nu) - \varphi(t+h,\mu_{t+h}^{t,\nu,u})}{h} \leqslant 0 \\ &\Rightarrow \quad \frac{\varphi(t,\nu) - \varphi(t+h,\exp_\mu(h\cdot\pi_T^\nu f[\nu,u]\#\nu))}{h} \leqslant O(h).\end{aligned}$$

Taking the limit in $h \searrow 0$ and using the Lipschitz-continuity of $\partial_s\varphi$ to split the directional derivative into its time and measure components, we obtain

$$-\partial_t\varphi(t,\nu) - D_\nu\varphi(t,\nu)(\pi_T^\nu f[\nu,u]\#\nu) \leqslant 0.$$

Taking the supremum over $u \in U$, we recover the subsolution inequality (3.3), so that V is a subsolution.

Supersolution inequality. Let $\psi \in \mathcal{T}_-$ such that $V - \psi$ reaches a minimum in $(t,\nu) \in [0,T) \times \mathcal{P}_2(\mathbb{R}^d)$. Since under **[A3.2.8]**, the set of solutions issued from (t,μ) is compact in the topology of uniform convergence (see [BF24, Theorem 4.5]), we may find $u \in L^1(t,T;U)$ such that $V(t,\nu) = V(t+h,\mu_{t+h}^{t,\nu,u})$ for all $h \in [0,T-t]$. Let $\mathrm{Lip}(\psi)$ be a local Lipschitz constant of ψ as above. Applying Lemma 3.2.17, there exist $(h_n)_n \searrow 0$ and $b \in f[\nu,U]$ such that $d_{\mathcal{W}}\left(\mu_{h_n}^{t,\nu,u}, \exp_\nu(h_n\cdot b\#\nu)\right) = o(h_n)$. Applying Lemma 3.2.18 again, the same approximation holds for $\pi_T^\nu b\#\nu$ instead of $b\#\nu$. Then

$$\begin{aligned}\psi(t,\nu) = V(t,\nu) = V(t+h_n,\mu_{t+h_n}^{t,\nu,u}) \geqslant \psi(t+h_n,\mu_{t+h_n}^{t,\nu,u}) \quad &\Rightarrow \quad \frac{\psi(t,\nu) - \psi(t+h_n,\mu_{t+h_n}^{t,\nu,u})}{h_n} \geqslant 0 \\ &\Rightarrow \quad \frac{\psi(t,\nu) - \psi(t+h_n,\exp_\mu(h_n\cdot\pi_T^\nu b\#\nu))}{h_n} \geqslant O(h_n).\end{aligned}$$

Taking the limit in $n \to \infty$, using the regularity of ψ and taking the supremum over $u \in U$, we obtain the supersolution inequality.

Uniqueness. To conclude, assume that there exists another viscosity solution $W : [0,T] \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ of (3.36). We check that the Hamiltonian defined in (3.35) falls under the assumptions of Proposition 3.2.6. Pick u_0 in the compact U. By the assumption of Lipschitz-continuity of f,

$$\begin{aligned}\sup_{u\in U} \|f[\mu,u]\#\mu\|_\mu &= \sup_{u\in U} \sqrt{\int_{x\in\mathbb{R}^d} \left|f[\mu,u](x)\right|^2 d\mu} \\ &\leqslant \sup_{u\in U} \sqrt{\int_{x\in\mathbb{R}^d} \left|f[\delta_0,u](x)\right|^2 d\mu} + \sqrt{\int_{x\in\mathbb{R}^d} \left|f[\delta_0,u](x) - f[\mu,u](x)\right|^2 d\mu} \\ &\leqslant \sup_{u\in U} \sqrt{\int_{x\in\mathbb{R}^d} \left(\mathrm{Lip}(f)|x| + |f|_{0,\infty}\right)^2 d\mu} + \mathrm{Lip}(f)\, d_{\mathcal{W}}(\delta_0,\mu) \leqslant 2\mathrm{Lip}(f)\, d_{\mathcal{W}}(\mu,\delta_0) + |f|_{0,\infty}.\end{aligned}$$

This proves the assumption of linear growth. On the other hand, let $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and $\eta \in \Gamma_o(\mu,\nu)$. Then

$$W^2_{\mu,\nu}(f[\mu,u]\#\mu, f[\nu,u]\#\nu) \leqslant \int_{(x,y)\in\mathbb{R}^{d^2}} \left|f[\mu,u](x) - f[\nu,u](y)\right|^2 d\eta \leqslant \int_{(x,y)\in\mathbb{R}^{d^2}} \mathrm{Lip}(f)^2 \left(d_{\mathcal{W}}(\mu,\nu) + |x-y|^2\right) d\eta,$$

from which we deduce that the probability vector field is Lipschitz in the strong sense with constant $2\mathrm{Lip}(f)$. Hence Proposition 3.2.6 holds, and the comparison principle follows. Applying Theorem 3.1.12 to the couples (V,W) and (W,V), we have $V \geqslant W$ and $W \geqslant V$ on $[0,T]\times\mathscr{P}_2(\mathbb{R}^d)$, so that $V = W$. □

3.2.4 Extensions

3.2.4.1 The case of non-convex dynamics

So far, we assumed the dynamics $f : \mathscr{P}_2(\mathbb{R}^d)\times U \to X$ to have convex values, in the sense that each set $f[\mu,U]\subset X$ is convex. If such is not the case, one can still formulate the control problem over the solutions of the controlled continuity equation (3.29), and compute the value function V. However, without convexity, the set of solutions of (3.29) may not be closed, and the infimum may not be attained. This is due to the possible chattering of the control; if one chooses $u^n(\cdot)\in L^1(0,T;U)$ so that $f[\cdot,u^n]$ oscillates between two given values with a frequency increasing with n, the limiting curves will be driven by the mean of the said values.

Characterization of V by a relaxed problem. This has the following consequences on the PDE characterization of the value function. Under **[A3.2.8]**, the results of Bonnet-Weill and Frankowska [BF24] imply that the closure in $\mathrm{AC}([t,T];\mathscr{P}_2(\mathbb{R}^d))$ of the set of solutions of the controlled continuity equation (3.29) is given by the set of solutions of the relaxed controlled continuity equation, where the latter is parametrized by controls $\omega\in\mathscr{P}_2(U)$ as

$$f^{\text{relaxed}}[\omega,\mu] = \int_{u\in U} f[\mu,u]\,d\omega(u)\in X.$$

If (U,d_U) is a compact metric space, then so is $\left(\mathscr{P}_2(U), d_{\mathcal{W},U}\right)$, where $d_{\mathcal{W},U}$ is the Wasserstein distance induced on $\mathscr{P}_2(U)$ by d_U [Vil09, Remark 6.19]. The images of the relaxed dynamic are convex by definition, and all the results in Section 3.2.3 apply to the value function V^{relaxed} computed as the infimum over the trajectories of the relaxed system. If the terminal cost $\mathfrak{J}$ is continuous, then $V = V^{\text{relaxed}}$, and the former is characterized by the Hamilton-Jacobi-Bellman equation associated to the relaxed problem, i.e. with Hamiltonian

$$H^{\text{relaxed}}(\mu,p) := \sup_{\omega\in\mathscr{P}_2(U)} -p\left(\pi^{\mu}_{T} f^{\text{relaxed}}[\omega,\mu]\#\mu\right).$$

Link with the original equation. Let us comment on the relation between H^{relaxed} and the original Hamiltonian $H(\mu,p) = \sup_{u\in U} -p(\pi^{\mu}_{T} f[\mu,u]\#\mu)$. If we were in a Hilbert space with $\mathcal{C}^1$ test functions, both would coincide, since the projection π^{μ}_{T} would reduce to the identity, and p would be linear. The supremum of $-p(b)$ over $b\in\overline{\mathrm{conv}}\, f[x,U]$ is reached at extremal points of $\overline{\mathrm{conv}}\, f[x,U]$, which are contained in $f[x,U]$, and $H(x,p) = H^{\text{relaxed}}(x,p)$.

However, in our setting, p may not be linear, and the projection does not reduce to the identity. For the second point, we are still able to say something owing to the fact that the relaxed dynamics is induced by a map; in this case, the projection coincides with the linear projection on Tan_{μ} in $L^2_{\mu}(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$, and for any $\omega\in\mathscr{P}_2(U)$,

$$\pi^{\mu}_{T} f^{\text{relaxed}}[\omega,\mu]\#\mu = \int_{u\in U}\pi^{\mu}_{T} f[\mu,u]\#\mu\, d\omega(u).$$

Consequently, if the variable p in the Hamiltonian H^{relaxed} is concave, then the supremum is attained for ω a Dirac mass, and $H(\mu,p) = H^{\text{relaxed}}(\mu,p)$. This shows that at best, the subsolutions of the HJB equation with original or relaxed Hamiltonian coincide. Without restricting the set of test functions, there is no reason to think that the sets of supersolutions coincide, and we conclude here our comments.

3.2.4.2 The case of a Bolza problem

Let $\ell : \mathscr{P}_2(\mathbb{R}^d) \times U \to \mathbb{R}$ be a running cost, and consider the Bolza problem

$$\begin{aligned}&\text{Minimize} \quad \int_{s=0}^{T} \ell(\mu_s^{0,\nu,u}, u(s))ds + \mathfrak{J}(\mu_T^{0,\nu,u}) \quad \text{on } u(\cdot) \in L^1(0,T;U),\\ &\text{where } \partial_s\mu_s + \operatorname{div}(f[\mu_s, u(s)]\#\mu_s) = 0, \text{ and } \mu_0^{0,\nu,u} = \nu.\end{aligned}$$

We suppose that $\ell : \mathscr{P}_2(\mathbb{R}^d) \times U \to \mathbb{R}$ is locally Lipschitz with respect to μ and continuous with respect to u. In addition, we suppose that the sets $\ell(\mu, U)$ are convex in $\mathbb{R}$: otherwise, one should consider a relaxed Hamiltonian, as in Section 3.2.4.1. Consider the value function

$$V(t,\nu) = \inf_{u(\cdot)\in L^1(t,T;U)} \int_{s=t}^{T} \ell(\mu_s^{t,\nu,u}, u(s))ds + \mathfrak{J}(\mu_T^{t,\nu,u}).$$

One expects V to be the unique viscosity solution of the Hamilton-Jacobi-Bellman equation with Hamiltonian

$$H(\mu, p) := \sup_{u\in U} -p\left(\pi_T^{\mu} f[\mu, u]\#\mu\right) - \ell(\mu, u).$$

By Proposition 3.2.6, uniqueness indeed holds for the HJB equation understood in the viscosity sense. Regarding the satisfaction of the HJB equation by the value function, one should pay attention to limits. In the subsolution part of Theorem 3.2.16, the inequality

$$\frac{V(t+h, \mu_{t+h}^{t,\nu,u}) - V(t,\mu)}{h} \geqslant -\frac{1}{h}\int_{s=t}^{t+h} \ell(\mu_{t+h}^{t,\nu,u}, \overline{u})ds \geqslant -\ell(\nu, \overline{u}) - O(h)$$

follows from the regularity of ℓ, and the fact that the control is taken constant. In the supersolution part, one considers instead

$$\frac{V(t+h, \mu_{t+h}^{t,\nu,u}) - V(t,\mu)}{h} = -\frac{1}{h}\int_{s=t}^{t+h} \ell(\mu_s^{t,\nu,u}, u(s))ds = -\frac{1}{h}\int_{s=t}^{t+h} \ell(\nu, u(s))ds + O(h).$$

Here, the integral term rewrites as $\int_{v\in U} \ell(\nu, v)d\omega_h(v)$, where $\omega_h := \frac{1}{h}\int_{s=t}^{t+h} \delta_{u(s)}ds$ is a probability measure on U. Since $\mathscr{P}_2(U)$ is compact in the topology induced by the Wasserstein distance, one can extract a converging subsequence towards some measure $\varpi \in \mathscr{P}_2(U)$, and obtain a limit point given by $\int_{v\in U} \ell(\nu, v)d\varpi(v)$. Since the sets $\ell(\nu, U)$ are convex in $\mathbb{R}$, we conclude that V is a viscosity supersolution.

3.2.4.3 The case of continuity with respect to the inductive narrow topology

Up to now, we considered the Wasserstein topology on $\mathscr{P}_2(\mathbb{R}^d)$, with respect to which balls are not compact. If one is interested in viscosity solutions that are continuous with respect to a weaker topology, it is possible to argue differently. We consider for instance the inductive narrow topology $\tau := \tau_2$ defined in Definition 1.1.27p. 10: a sequence $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathbb{R}^d)$ converges towards $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ with respect to τ, which we denote $\mu_n \xrightarrow{\tau}_n \mu$, if it converges narrowly and $\sup_{n\in\mathbb{N}} d_{\mathcal{W}}(\mu_n, \delta_0) < \infty$. *In this section, we refer to the topology τ as the weak topology on $\mathscr{P}_2(\mathbb{R}^d)$.*

We first adapt the definition of viscosity solutions and the comparison principle. We then focus on a Mayer control problem with a weakly continuous dynamic, and show that the associated value function is $\tau-$continuous. This is stronger than continuity with respect to the Wasserstein distance: $\mu \mapsto d_{\mathcal{W}}(\mu, \nu)$ is obviously $d_{\mathcal{W}}-$continuous, but only $\tau-$lower semicontinuous. In consequence, the value function will satisfy the HJB equation, in the viscosity sense that we now detail. The sets of test functions are taken as before (in Definition 3.1.5).

Definition 3.2.19 (Viscosity solution of (3.1)). *An application $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ is a*

- *viscosity subsolution of (3.1) if it is $\tau-$upper semicontinuous, and if for any $\varphi \in \mathscr{T}_+$ such that $u - \varphi$ reaches a maximum at $(t,x) \in [0,T) \times \mathbb{R}^d$, there holds*

$$-\partial_t\varphi(t,x) + H\left(x, D_x\varphi(t,x)\right) \leqslant 0,$$

– *viscosity supersolution of (3.1) if it is τ−lower semicontinuous, and if for any $\varphi \in \mathcal{T}_-$ such that $u-\varphi$ reaches a minimum at $(t,x) \in [0,T) \times \mathbb{R}^d$, there holds*

$$-\partial_t \varphi(t,x) + H\big(x, D_x\varphi(t,x)\big) \geqslant 0,$$

– *viscosity solution of (3.1) if it is both a subsolution and a supersolution, and satisfies $u(T,\mu) = \mathfrak{J}(\mu)$.*

Proposition 3.2.20 (Comparison principle in the case of the inductive weak topology τ). *Assume that H satisfies **[A3.1.11]**. Let $v, w : [0,T] \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ be locally bounded, and respectively subsolution and supersolution of the HJ equation (3.1) in the sense of Definition 3.2.19. Then* $\sup_{[0,T]\times\mathscr{P}_2(\mathbb{R}^d)} v - w \leqslant \sup_{\{T\}\times\mathscr{P}_2(\mathbb{R}^d)} v - w$.

Proof. Instead of reproducing the proof of Theorem 3.1.12 p. 65, we point at the modifications. One can consider the same definition of the doubling function Φ in (3.17). As the Wasserstein distance is τ−lower semicontinuous, Φ is itself τ−upper semicontinuous. Moreover, its upper level sets are bounded in the Wasserstein distance, so compact with respect to τ by Proposition 1.1.28. In consequence, one can directly take maximizing points $(t_{\iota\varepsilon}, \mu_{\iota\varepsilon}), (s_{\iota\varepsilon}, \nu_{\iota\varepsilon})$, instead of applying the Ekeland-type Lemma 3.1.10. This corresponds to perturbations $p_{\iota\varepsilon n} \equiv 0$, and accordingly, $\omega_\iota \equiv 0$. The rest of the argument is unchanged: for a sufficiently small but fixed $\iota > 0$, maximizers are getting close to each other, do not lie on the parabolic boundary when ε goes to 0, so the viscosity inequalities can be applied, but are contradicted for small $\varepsilon > 0$ by the term coming from ℓ_α. □

We now apply this Hamilton-Jacobi theory to the case of a Mayer control problem where the dynamic is weakly continuous, in addition to the assumptions ensuring well-posedness of the controlled trajectories. Precisely, we consider the control problem

$$\begin{aligned}&\text{Minimize} \quad \mathfrak{J}(\mu_T^{0,\nu,u}) \quad \text{on } u(\cdot) \in L^1(0,T;U),\\ &\text{where } \partial_s\mu_s + \operatorname{div}(f[\mu_s, u(s)]\#\mu_s) = 0, \text{ and } \mu_0^{0,\nu,u} = \nu.\end{aligned}$$

We assume the dynamic f to satisfy the following.

Assumption [A3.2.21] (Structure of the dynamic). The dynamic $f : \mathbb{R}^d \times U \to X$ satisfies **[A3.2.8]**, has convex images in the sense that each set $f[\mu, U] \subset X$ is convex, and τ−upper semicontinuous in the set-valued sense: for any $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $\mathcal{O} \subset X$ an open neighbourhood of $f[\mu, U]$, there exists an open neighbourhood $\mathcal{O}' \in \tau$ of μ such that $f[\nu, U] \subset \mathcal{O}$ for any $\nu \in \mathcal{O}'$.

Under this assumption, we can prove a convergence theorem for trajectories of the controlled continuity equation, akin to [BF24, Thm. 4.5] but with a dependence on the initial condition taken with respect to the inductive topology τ. For convenience, let us bring up the following small lemma.

Lemma 3.2.22 (Sufficient subset of $(L^1(I;X))^*$). *Let I be a nontrivial compact interval, and $A \subset L^1(I;X)$ be a subset such that for almost each $s \in I$, the set A_s is convex and relatively compact. Then $f \in A$ if and only if*

$$\int_{s\in I} r(s) q(f_s)\, ds \leqslant \int_{s\in I} \sup_{a\in A_s} r(s) q(a)\, ds \qquad \forall r \in L^\infty(I;\mathbb{R}^+) \text{ and } q \in X^*. \tag{3.37}$$

Proof. If $f \in A$, the property (3.37) is direct. On the other hand, assume that $f \notin A$, and let us construct r and q such that (3.37) fails. As X is separable, the weak−∗ topology over X^* is metrizable and renders $\overline{\mathscr{B}}_{X^*}(0,1)$ compact. Let $(q_n)_{n\in\mathbb{N}}$ be a countable weak−∗ dense set in $\overline{\mathscr{B}}_{X^*}(0,1)$, and define $s \mapsto \sigma_{A_s}(x) := \sup_{m\in\mathbb{N}} \inf_{a\in A_s} q_m(x-a)$. The function $s \mapsto \sigma_s(f_s)$ is bounded by $|f_s|_{\mathrm{ucc}} + \sup_{a\in A_s} |a|_{\mathrm{ucc}}$, and measurable by the marginal map theorem [AF90, Theorem 8.2.11]. For each $s \in I$ such that A_s is convex and relatively compact, there holds $x \in A_s$ if and only if $\sigma_{A_s}(x) \leqslant 0$; indeed, if $x \notin A_s$, Hahn-Banach provides a strictly separating hyperplane $p \in \overline{\mathscr{B}}_{X^*}(0,1)$, that may be approximated uniformly over the compact $A_s \cup \{x\}$ by some q_m with an error as small as desired, thus proving $\sigma_{A_s}(x) > 0$. Now, if $f \notin A$, there exists $\rho \in \mathcal{C}(I;\mathbb{R}^+)$ such that

$$0 < \int_{s\in I} \rho(s)\sigma_{A_s}(f_s)\, ds = \int_{s\in I} \sup_{m\in\mathbb{N}} \inf_{a\in A_s} \rho(s) q_m(f_s - a)\, ds.$$

Hence there exists $m \in \mathbb{N}$ and $I_m \subset I$ of positive measure such that $\rho(s)q_m(f_s) - \sup_{a\in A_s} \rho(s) q_m(a) > 0$ for a.e. $s \in I_m$, and the couple $r = \mathbb{1}_{I_m}\rho$ and $q = q_m$ satisfies the desired property. □

We can now investigate the convergence of trajectories when the initial condition converges with respect to the topology τ. Note that there is no hope to get convergence for the Wasserstein topology: taking $f \equiv 0$ trivially satisfies **[A3.2.21]**, and the (stationary) trajectories converge with respect to τ only. However, owing to Proposition 1.1.30, the convergence with respect to τ implies that with respect to any $d_{\mathcal{W},p}$ for $p \in [1,2)$, and we may work for such a p.

Lemma 3.2.23 (Convergence of trajectories). *Assume **[A3.2.21]** holds, and let $(t,\nu) \in [0,T) \times \mathscr{P}_2(\mathbb{R}^d)$. Let $(\nu^n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathbb{R}^d)$ be such that $\nu^n \xrightarrow{\tau}_n \nu$. Let $(\mu^n_s)_{s\in[t,T]}$ be the solution of (3.29) provided by Proposition 3.2.9 associated with the control $u^n \in L^1(t,T;U)$. Then, for any $p \in [1,2)$, there is a control $u \in L^1(t,T;U)$ such that*

$$\liminf_{n\to\infty} \sup_{s\in[t,T]} d_{\mathcal{W},p}\left(\mu^n_s, \mu^{t,\nu,u}_s\right) = 0. \tag{3.38}$$

By the Grönwall estimates of Lemma 3.2.10, one can bound the second moment of μ^n_s uniformly in n and s, and deduce that $\mu^n_s \xrightarrow{\tau}_n \mu^{t,\nu,u}_s$ for any s; however, (3.38) gives a useful quantitative control over $[t,T]$, shall it be in a weaker topology.

Proof. The proof is divided in the following steps: first obtain a candidate to the solution by Arzelà-Ascoli, then extract a weakly convergent subsequence of the dynamics, and conclude by showing that the limit solves the original equation. The first extraction cannot be done in $(\mathscr{P}_2(\mathbb{R}^d), d_{\mathcal{W}})$ directly, since the latter does not have compact balls. For this reason, we extract with respect to the topology of a p−Wasserstein distance for some $p \in [1,2)$. The weak continuity of the dynamic allows to gain sufficient compactness in the second step.

Limiting curve. Fix $p \in [1,2)$. Since $\nu^n \xrightarrow{\tau}_n \nu$, one has $\sup_{n\in\mathbb{N}} d_{\mathcal{W}}(\nu_n, \delta_0) \leq R$ for some $R < \infty$. From the Grönwall estimates of Lemma 3.2.10, we deduce that the family $(\mu^n)_n$ is equiLipschitz with respect to the distance $d_{\mathcal{W}}$, and contained in a $d_{\mathcal{W}}$−ball of radius $R_T \geq 0$ independent of n.

Consequently, the sections $\left(\mu^n_s\right)_{n\in\mathbb{N}}$ are relatively compact in $(\mathscr{P}_p(\mathbb{R}^d), d_{\mathcal{W},p})$, and by Hölder's inequality, $(\mu^n)_n$ is also equiLipschitz with respect to $d_{\mathcal{W},p}$. By the Arzelà-Ascoli theorem, there exists $(\mu_s)_s \in \mathrm{AC}([t,T];\mathscr{P}_p(\mathbb{R}^d))$ and a (non relabeled) subsequence $(\mu^n)_{n\in\mathbb{N}}$ so that

$$\lim_{n\to\infty} \sup_{s\in[t,T]} d_{\mathcal{W},p}(\mu^n_s, \mu_s) = 0 \quad \text{and} \quad \sup_{s\in[t,T]} d_{\mathcal{W}}(\mu_s, \delta_0) \leq R_T. \tag{3.39}$$

Extraction of a dynamic. Denote $b^n \in L^1(t,T;X)$ the element defined by $b^n_s := f[\mu^n_s, u^n(s)]$. Let us work the whole way up to the weak compactness that we need, inspiring ourselves from [BF24, Thm. 2.1].

By the weak continuity assumed in **[A3.2.21]**, $f[\mu_s, U]$ is convex and compact in $(X, |\cdot|_{\mathrm{ucc}})$. It is therefore weakly compact in the Banach space X by James' Theorem [Jam64, Theorem 5]. The set-valued map $\mu \mapsto f[\mu, U]$ is upper semicontinuous from $(\mathscr{P}_2(\mathbb{R}^d), \tau)$ to $(X, |\cdot|_{\mathrm{ucc}})$: since each weakly open set of X is also open, it is also upper semicontinuous into X equipped with its weak topology. We may then apply [Ber59, Theorem 3 p.116] to get that the union A of the images $f[\mu, U]$ when μ ranges in the τ−compact $\overline{\mathscr{B}}(\delta_0, R_T)$ is again weakly compact. By Diestel's theorem [DU77, Proposition 7], $L^1(0,T;A)$ is then relatively weakly compact. As it is closed and convex, it is additionally weakly closed, hence weakly compact in $L^1(0,T;X)$. In conclusion, there is a (non relabeled) subsequence of $(b^n)_n$ that converges weakly to some $b \in L^1(t,T;X)$.

Let us show that $b_s \in f[\mu_s, U]$ for a.e. $s \in [t,T]$. Consider $r \in L^\infty(t,T;\mathbb{R}^+)$ and $q \in X^*$. For any $n \in \mathbb{N}$, there holds

$$\int_{s=t}^T r(s) q(b^n_s) ds \leq \int_{s=t}^T \sup_{\beta\in f[\mu^n_s, U]} r(s) q(\beta) ds. \tag{3.40}$$

Since the linear form defined by $\psi \mapsto \int_{s=t}^T r(s) q(\psi_s) ds$ is continuous with respect to $\psi \in L^2(t,T;X)$, we can pass to the limit in the right hand-side by weak convergence of b^n towards b in $L^1(t,T;X)$. Using upper semi-continuity of $\mu \mapsto f[\mu, U]$ from $(\mathscr{P}_2(\mathbb{R}^d), \tau)$ to X, jointly with the fact that $\mu^s_n \xrightarrow{\tau}_n \mu_s$ for all s, there holds $\limsup_{n\to\infty} \sup_{\beta\in f[\mu^n_s, U]} q(\beta) \leq \sup_{\beta\in f[\mu_s, U]} q(\beta)$ for any s. Taking the limit sup in $n \to \infty$ in (3.40),

$$\int_{s=t}^T r(s) q(b_s) ds \leq \int_{s=t}^T \sup_{\beta\in f[\mu_s, U]} r(s) q(\beta) ds.$$

By Lemma 3.2.22, this implies $b_s \in f[\mu_s, U]$ for almost every $s \in [t, T]$.

The limit curve solves the equation. Let $\varphi \in \mathcal{C}_c^\infty((t,T)\times\mathbb{R}^d;\mathbb{R})$ arbitrary, but fixed, and let $M \in \mathbb{N}$ be the radius of a ball containing supp $\varphi(s,\cdot)$ for all $s \in (t,T)$. Since $(\mu_s^n)_{s\in[t,T]}$ solves (3.29), for any $n \in \mathbb{N}$ we have

$$0 = \int_{s=t}^T \int_{x\in\mathbb{R}^d} \left[\partial_s\varphi + \langle \nabla_x\varphi, b_s^n\rangle\right] d\mu_s^n ds = \int_{s=t}^T \int_{x\in\mathbb{R}^d} \left[\partial_s\varphi + \langle \nabla_x\varphi, b_s\rangle\right] d\mu_s ds - (I_1^n + I_2^n + I_3^n),$$

where the three error terms are respectively given by

$$I_1^n := \int_{s=t}^T \left[\int_{x\in\mathbb{R}^d} \partial_s\varphi d\mu_s - \int_{x\in\mathbb{R}^d} \partial_s\varphi d\mu_s^n\right] ds, \quad I_2^n := \int_{s=t}^T \int_{x\in\mathbb{R}^d} \langle \nabla_x\varphi, b_s - b_s^n\rangle\, d\mu_s ds$$

and

$$I_3^n := \int_{s=t}^T \left[\int_{x\in\mathbb{R}^d} \langle \nabla_x\varphi, b_s^n\rangle\, d\mu_s - \int_{x\in\mathbb{R}^d} \langle \nabla_x\varphi, b_s^n\rangle\, d\mu_s^n\right] ds.$$

The term I_1^n is bounded in absolute value by $\mathrm{Lip}(\partial_s\varphi)(T-t)\sup_{s\in[t,T]} d_{\mathcal{W},p}(\mu_s^n,\mu_s)$, which goes to 0 when n goes to ∞ by construction of $(\mu_s)_s$. Secondly, the real-valued linear operator

$$L^1(t,T;X) \ni \beta \mapsto \int_{s=t}^T \int_{\mathbb{R}^d} \langle \nabla_x\varphi, \beta_s\rangle\, d\mu_s ds \leqslant \int_{s=t}^T \int_{x\in\overline{\mathscr{B}}(0,M)} \left|\nabla_x\varphi\right| \|\beta_s\| d\mu_s ds \leqslant 2^M \|\nabla_x\varphi\|_\infty \int_{s=t}^T \left|\beta_s\right|_{\mathrm{ucc}} ds$$

is bounded, thus an element of $[L^1(t,T;X)]^*$. Therefore, $I_2^n \to 0$ when $n \to \infty$.

We turn to I_3^n. For any $n \in \mathbb{N}$, let $\eta^n \in L^1(t,T;\mathscr{P}_2((\mathbb{R}^d)^2))$ be a curve of optimal transport plans for $d_{\mathcal{W},p}$ between μ_s and μ_s^n, i.e. such that $\eta_s^n \in \Gamma_{o,p}(\mu_s,\mu_s^n)$ for a.e. $s \in [t,T]$. This can be constructed by using the compact-valuedness and upper semicontinuity of the application that to a pair of measures, associates the set of optimal plans between them, first by taking a measurable selection, then by proving integrability using the marginals. Then

$$I_3^n = \int_{s=t}^T \int_{(x,y)\in(\mathbb{R}^d)^2} \left[\langle \nabla_x\varphi(s,x), b_s^n(x)\rangle - \langle \nabla_x\varphi(s,y), b_s^n(y)\rangle\right] d\eta_s^n(x,y) ds.$$

Recall that all the trajectories μ^n and μ are contained in a $d_{\mathcal{W}}-$ball of radius R_T. Since φ is supported in a ball of radius M, and the dynamic f is assumed Lipschitz in the measure and space variable, uniformly in the control variable, with a uniform bound at a given point in space, the vectors $b_s^n(x)$ and $b_s^n(y)$ are bounded from above by a constant $C_{\varphi,R_T,M,f} \geqslant 0$ for $x,y \in \overline{\mathscr{B}}(0,M)$. Hence

$$I_3^n \leqslant \int_{s=t}^T \left[\mathrm{Lip}(\nabla_x\varphi) C_{\varphi,R_T,M,f} + \|\nabla_x\varphi\|_\infty \mathrm{Lip}(f)\right] |x-y| d\eta_s^n(x,y) ds \leqslant \widetilde{C}_{\varphi,R_T,M,f} \sup_{s\in[t,T]} d_{\mathcal{W},p}(\mu_s^n,\mu_s).$$

Sending n to ∞, we conclude that $\int_{s=t}^T \int_{x\in\mathbb{R}^d} \left[\partial_s\varphi + \langle \nabla_x\varphi, b_s\rangle\right] d\mu_s ds = 0$ for all $\varphi \in \mathcal{C}_c^\infty((t,T)\times\mathbb{R}^d;\mathbb{R})$, and the limit point solves the continuity inclusion. □

Lemma 3.2.24. *Assume **[A3.2.21]**. Consider* $(t,\nu) \in [0,T) \times \mathscr{P}_2(\mathbb{R}^d)$ *and a control* $u \in L^1(t,T;U)$. *Let* $(\mu_s^{t,\nu,u})_{s\in[t,T]}$ *be the unique solution of* (3.29) *provided by Proposition 3.2.9. Then, for any sequence* $(\nu^n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathbb{R}^d)$ *such that* $\nu^n \xrightarrow{\tau}_n \nu$ *and any* $p \in [1,2)$, *we have* $d_{\mathcal{W},p}\left(\mu_T^{t,\nu_n,u}, \mu_T^{t,\nu,u}\right) \to 0$ *as* $n \to \infty$.

Proof. The proof is a refinement of the argument used in Lemma 3.2.23, and we just point at the modifications. By the definition of the topology τ and the Grönwall estimates of Lemma 3.2.10, the curves $(\mu^{t,\nu_n,u})_n$ are all contained in a ball of radius $B \geqslant 0$ independent of n. Following the proof of Lemma 3.2.23, the sequence $\left((\mu_s^{t,\nu_n,u})_{s\in[t,T]}\right)_{n\in\mathbb{N}}$ admits a limit point $(\mu_s)_{s\in[t,T]}$ in $\mathrm{AC}\left([t,T];(\mathscr{P}_2(\mathbb{R}^d), d_{\mathcal{W},p})\right)$. Let $q \in L^\infty(t,T;X^*)$, and consider

$$\int_{s=t}^T q(f[\mu_s^n, u(s)]) ds.$$

The integrand converges pointwise towards $q(f[\mu_s,u(s)])$ by weak convergence, and is bounded by

$$\|q_s\|_{X^*} \left|f[\mu_s^n,u(s)]\right|_{\mathrm{ucc}} \leqslant \|q_s\|_{X^*} \left(d_{\mathcal{W}}(\mu_s^n,\delta_0) + \left|f[\delta_0,u(s)]\right|_{\mathrm{ucc}}\right) \leqslant \|q_s\|_{X^*} \left(B + \left|f[\delta_0,u(s)]\right|_{\mathrm{ucc}}\right) \in L^1(t,T;\mathbb{R})$$

uniformly in n. By Lebesgue' dominated convergence, it converges towards $\int_{s=t}^{T} q(f[\mu_s, u(s)])ds$. Hence the control $u(\cdot) \in L^1(t,T;U)$ provides a parametrisation of the weak limit point $b \in L^1(t,T;X)$ by $b_s = f[\mu_s, u(s)]$. The same argument as in Lemma 3.2.23 shows that $(\mu_s)_{s\in[t,T]}$ solves the continuity equation with dynamics b, so that $\mu_s = \mu_s^{t,\nu,u}$ by definition. □

The following Lemma is trivial in the case where the dynamic f is Lipschitz (as in our case), but could be quite intricate if f was only dissipative, and false if f was anti-dissipative.

Lemma 3.2.25 (Backward extension). *Let $\nu \in \mathscr{P}_2(\mathbb{R}^d)$. For any $\delta > 0$ and control $u \in L^1(0,\delta;U)$, there exists $\omega \in \mathscr{P}_2(\mathbb{R}^d)$ such that the unique solution $(\mu_s^{0,\omega,u})_{s\in[0,\delta]}$ of the controlled continuity equation (3.29) satisfies $\mu_\delta^{t,\omega,u} = \nu$.*

Proof. The dynamic $\widetilde{f} : \mathscr{P}_2(\mathbb{R}^d) \times U \to X$ given by $\widetilde{f}[\mu,u] = -f[\mu,u]$ satisfies **[A3.2.21]**. Let $\widetilde{u} \in L^1(0,\delta;U)$ be given by $\widetilde{u}(s) := u(\delta - s)$, and consider the trajectory $(\widetilde{\mu}_s^{0,\nu,\widetilde{u}})_{s\in[0,\delta]}$ solution of the continuity equation with dynamic $\widetilde{f}$ driven by the control $\widetilde{u}$. Let $\mu_s := \widetilde{\mu}_{\delta-s}^{0,\nu,\widetilde{u}}$. Then $(\mu_s)_{s\in[0,\delta]} \in \mathrm{AC}([0,\delta];\mathscr{P}_2(\mathbb{R}^d))$ satisfies $\mu_\delta = \nu$, and for any $\varphi \in \mathcal{C}_c^\infty((0,\delta)\times\mathbb{R}^d;\mathbb{R})$,

$$\begin{aligned}
&\int_{s\in[0,\delta]}\int_{x\in\mathbb{R}^d} \left[\partial_s\varphi(s,x) + \langle\nabla_x\varphi(s,x), f[\mu_s,u(s)]\rangle\right] d\mu_s ds \\
&\qquad = \int_{s\in[0,\delta]}\int_{x\in\mathbb{R}^d} \left[\partial_s\varphi(s,x) + \langle\nabla_x\varphi(s,x), f[\mu_s,u(s)]\rangle\right] d\widetilde{\mu}_{\delta-s}^{0,\nu,\widetilde{u}} ds \\
&\qquad = \int_{s\in[0,\delta]}\int_{x\in\mathbb{R}^d} \left[\partial_s\varphi(\delta-s,x) + \langle\nabla_x\varphi(\delta-s,x), f[\mu_{\delta-s},u(\delta-s)]\rangle\right] d\widetilde{\mu}_s^{0,\nu,\widetilde{u}} ds \\
&\qquad = \int_{s\in[0,\delta]}\int_{x\in\mathbb{R}^d} \left[\partial_s\varphi(\delta-s,x) + \langle\nabla_x\varphi(\delta-s,x), f[\widetilde{\mu}_s^{0,\nu,\widetilde{u}},\widetilde{u}(s)]\rangle\right] d\widetilde{\mu}_s^{0,\nu,\widetilde{u}} ds.
\end{aligned}$$

Introduce $\psi : (s,x) \mapsto \varphi(\delta - s, x)$. Then $\psi \in \mathcal{C}_c^\infty((0,\delta)\times\mathbb{R}^d;\mathbb{R})$, and recalling that $\widetilde{f} = -f$, the above rewrites

$$\begin{aligned}
&\int_{s\in[0,\delta]}\int_{x\in\mathbb{R}^d} \left[\partial_s\varphi(s,x) + \langle\nabla_x\varphi(s,x), f[\mu_s,u(s)]\rangle\right] d\mu_s ds \\
&= \int_{s\in[0,\delta]}\int_{x\in\mathbb{R}^d} \left[-\partial_s\psi(s,x) - \langle\nabla_x\psi(s,x), \widetilde{f}[\widetilde{\mu}_s^{0,\nu,\widetilde{u}},\widetilde{u}(s)]\rangle\right] d\widetilde{\mu}_s^{0,\nu,\widetilde{u}} ds = 0.
\end{aligned}$$

Hence $\mu_s = \mu_s^{0,\omega,u}$ for $\omega := \widetilde{\mu}_\delta^{0,\nu,\widetilde{u}}$. □

We can now prove τ−lower semicontinuity of the value function.

Proposition 3.2.26 (Regularity of the value function). *Assume that the dynamic f satisfies **[A3.2.21]**, and that $\mathfrak{J} : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}\cup\{\infty\}$ is lower bounded and not identically ∞, and τ−lower semicontinuous. Let $V : [0,T]\times\mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}\cup\{\pm\infty\}$ be the value function of the Mayer problem with terminal cost $\mathfrak{J}$. Then each $V(t,\cdot)$ is not identically ∞, V is lower bounded and τ-lower semicontinuous. Moreover, if $\mathfrak{J}$ is bounded and τ-continuous, then so is V.*

Proof. Lower boundedness of V follows from that of $\mathfrak{J}$. Let $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ such that $\mathfrak{J}(\nu) < \infty$. From Lemma 3.2.25, for any $u \in L^1(0,T;U)$, there is $\tilde{\nu} \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu_T^{0,\tilde{\nu},u} = \nu$. Hence $V(t, \mu_t^{0,\tilde{\nu},u}) \leqslant \mathfrak{J}(\nu) < \infty$.

τ−lower semicontinuity. From Lemma 1.1.29 p. 10, lower semicontinuity and sequential lower semicontinuity coincide in the topology τ. Let $(t_n,\nu_n)_n \subset [0,T]\times\mathscr{P}_2(\mathbb{R}^d)$ be such that $t_n \to t \in [0,T]$ and $\nu_n \xrightarrow{\tau}_n \nu \in \mathscr{P}_2(\mathbb{R}^d)$. For each $n\in\mathbb{N}$, there is $u_n \in L^1(t_n,T;U)$ such that $V(t_n,\nu_n) \geqslant \mathfrak{J}(\mu_T^{t_n,\nu_n,u_n}) - e^{-n}$.

We claim that there is $\tilde{\nu}_n \in \mathscr{P}_2(\mathbb{R}^d)$ such that $\tilde{\nu}_n \xrightarrow{\tau}_n \nu \in \mathscr{P}_2(\mathbb{R}^d)$, and $\tilde{u}_n \in L^1(t,T;U)$ such that $\mu_T^{t,\tilde{\nu}_n,\tilde{u}_n} = \mu_T^{t_n,\nu_n,u_n}$. Indeed, if $t_n < t$, set $\tilde{u}_n = u_n|_{[t,T]}$ a.e. and $\tilde{\nu}_n = \mu_t^{t_n,\nu_n,u_n}$. By the Grönwall estimates, $d_{\mathcal{W}}(\tilde{\nu}_n,\nu) \to 0$ as $n\to\infty$, so that $\tilde{\nu}_n \xrightarrow{\tau}_n \nu$ as well. On the other hand, if $t_n > t$, it is enough take any $\tilde{u}_n \in L^1(t,T;U)$ such that $\tilde{u}_n = u_n|_{[t,T]}$ a.e. and take $\tilde{\nu}_n \in \mathscr{P}_2(\mathbb{R}^d)$ given by Lemma 3.2.25. By the same reasoning, $\tilde{\nu}_n \xrightarrow{\tau}_n \nu$ as $n\to\infty$.

Therefore, from Lemma 3.2.23, possibly along a subsequence, there is a measurable control $u \in L^1(t,T;U)$ such that $d_{\mathcal{W},1}\left(\mu_T^{t_n,\nu_n,u_n}, \mu_T^{t_n,\nu,u}\right) \to_n 0$. As moreover, all measures $\mu_T^{t_n,\nu_n,u_n}$ stay in a bounded ball around ν by the definition of $\nu_n \xrightarrow{\tau}_n \nu$ and the Grönwall estimates of Lemma 3.2.10, there holds $\mu_T^{t_n,\nu_n,u_n} \xrightarrow{\tau}_n \mu_T^{t,\nu,u}$. Then, by lower semicontinuity of $\mathfrak{J}$ in τ, we get

$$\liminf_{n\to\infty} V(t_n,\nu_n) \geqslant \liminf_{n\to\infty} \mathfrak{J}(\mu_T^{t_n,\nu_n,u_n}) - e^{-n} \geqslant \mathfrak{J}(\mu_T^{t,\nu,u}) \geqslant V(t,\nu). \tag{3.41}$$

$\tau-$**continuity.** Assume now that $\mathfrak{J}$ is bounded and τ-continuous. Then V shares the same bound by definition. To prove that V is $\tau-$upper semicontinuous, fix $(t,\nu) \in [0,T] \times \mathscr{P}_2(\mathbb{R}^d)$ and take a sequence $(t_n,\nu_n)_n \subset [0,T]\times\mathscr{P}_2(\mathbb{R}^d)$ such that $t_n \to t$ and $\nu_n \xrightarrow{\tau}_n \nu$. Up to extraction of a subsequence, we may assume that $\limsup_{n\to\infty} V(t_n,\nu_n) = \lim_{n\to\infty} V(t_n,\nu_n)$. For each $\varepsilon > 0$, pick $u_\varepsilon \in L^1(t,T;U)$ such that $V(t,\nu) \geqslant \mathfrak{J}(\mu_T^{t,\nu,u_\varepsilon}) - \varepsilon$. Either by restriction if $t_n \geqslant t$, or by extension by a constant if $t_n < t$, one can construct $u_\varepsilon^n \in L^1(t,T;U)$ that coincides with u^ε on $[t,T] \cap [t_n,T]$. Then, combining the Grönwall estimates of Lemma 3.2.10 to control the perturbation around t, and Lemma 3.2.24, there holds

$$d_{\mathcal{W},p}\left(\mu_T^{t_n,\nu_n,u_\varepsilon^n}, \mu_T^{t,\nu,u_\varepsilon}\right) \underset{n\to\infty}{\longrightarrow} 0.$$

By Lemma 3.2.10, the second moment of $\mu_T^{t_n,\nu_n,u_\varepsilon^n}$ is bounded by a constant depending only on T and $d_{\mathcal{W}}(\nu_n,\delta_0)$, with $d_{\mathcal{W}}(\nu_n,\delta_0)$ bounded uniformly in n by definition of τ, the trajectories issued from ν_n all have bounded moments, and we get that $\mu_T^{t,\tilde{\nu}_n,u_\varepsilon^n} \xrightarrow{\tau}_n \mu_T^{t,\nu,u_\varepsilon}$. Then, since $\mathfrak{J}$ is τ-continuous,

$$\lim_{n\to\infty} V(t_n,\nu_n) \leqslant \limsup_{t\to\infty} \mathfrak{J}(\mu_T^{t_n,\nu_n,u_\varepsilon^n}) = \mathfrak{J}(\mu_T^{t,\nu,u_\varepsilon}) \leqslant V(t,\nu) + \varepsilon. \tag{3.42}$$

Letting $\varepsilon \searrow 0$, we conclude that V is $\tau-$continuous. □

By Theorem 3.2.16, the value function V satisfies the viscosity inequalities. Hence, in the case of a $\tau-$continuous and bounded terminal cost, Proposition 3.2.26 is enough to prove that V is a viscosity solution of the HJB equation with control Hamiltonian (3.35), in the sense of Definition 3.2.19.

3.2.4.4 The case of a lower semicontinuous terminal cost

Consider again the Mayer problem

$$\begin{aligned} &\text{Minimize} \quad \mathfrak{J}(\mu_T^{0,\nu,u}) \quad \text{on } u(\cdot) \in L^1(0,T;U), \\ &\text{where } \partial_s\mu_s + \operatorname{div}(f[\mu_s,u(s)]\#\mu_s) = 0, \text{ and } \mu_0^{0,\nu,u} = \nu. \end{aligned}$$

One can force the terminal points $\mu_T^{t,\nu,u}$ of optimal trajectories to land in a given set $K \subset \mathscr{P}_2(\mathbb{R}^d)$ by setting $\mathfrak{J}(\mu) = \infty$ outside of K. In this case, the value function $V(t,\nu) = \inf_{u\in L^1(t,T;U)} \mathfrak{J}(\mu_T^{t,\nu,u})$ may take the value ∞. If the dynamic f satisfies **[A3.2.8]** and has convex images, V is still a supersolution of the HJB equation with Hamiltonian $H(\mu,p) = \sup_{u\in U} -p(\pi_T^\mu f[\mu,u]\#\mu)$ by the argument of Theorem 3.2.16, but has no chance to be a viscosity subsolution. However, we may still prove that it is the smallest supersolution in the pointwise sense. The argument proceeds by truncation and regularization, relying on classical inf-convolution. Since it is slightly more delicate with the topology τ, we provide it in this case, and trust the reader to trim out the specificities of τ to obtain the proof in $(\mathscr{P}_2(\mathbb{R}^d), d_{\mathcal{W}})$.

Lemma 3.2.27. *Let $\mathfrak{J} : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{\infty\}$ be lower bounded and $\tau-$lower semicontinuous. Then for each $N \in \mathbb{N}$, there is a nondecreasing sequence of bounded continuous maps $\mathfrak{J}_n : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ that converge pointwise towards $\mathfrak{J}$ over $\overline{\mathscr{B}}(\delta_0,N)$.*

Proof. Denote $\mathbb{I}_N : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{\infty\}$ the characteristic function of $\overline{\mathscr{B}}(\delta_0,N)$, i.e. $\mathbb{I}_N(\nu) = 0$ if $d_{\mathcal{W}}(\delta_0,\nu) \leqslant N$, and $\mathbb{I}_N(\nu) = \infty$ otherwise. Since $\overline{\mathscr{B}}(\delta_0,N)$ is $\tau-$compact, $\mathbb{I}_N$ is $\tau-$lower semicontinuous. Moreover, the function $\nu \mapsto \mathfrak{J}(\nu) + \mathbb{I}_N(\nu)$ is narrowly lower semicontinuous. Indeed, this is due to the fact that the topology τ coincides with the narrow topology on $\overline{\mathscr{B}}(\delta_0,N)$. Let $\mathrm{d}\mathbb{I} : \mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^+$ be a distance inducing the topology of narrow convergence over $\mathscr{P}_2(\mathbb{R}^d)$ (e.g. [AGS05, Section 5.1]), and

$$\mathfrak{J}_n(\mu) := \min\left(n, \inf_{\nu\in\mathscr{P}_2(\mathbb{R}^d)} (\mathfrak{J} + \mathbb{I}_N)(\nu) + n\mathrm{d}\mathbb{I}(\mu,\nu)\right).$$

We directly have $\mathfrak{J}_n(\mu) \leqslant \min\big(n, \mathfrak{J}(\mu)\big) \leqslant \mathfrak{J}(\mu)$ for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. As $\mathfrak{J}_n$ is lower bounded, using that $\inf(g_1) - \inf(g_2) \leqslant \sup(g_1 - g_2)$, for each $\mu_0, \mu_1 \in \mathscr{P}_2(\mathbb{R}^d)$ we get

$$\mathfrak{J}_n(\mu_0) - \mathfrak{J}_n(\mu_1) \leqslant \max\left(0, \sup_{\nu \in \mathscr{P}_2(\mathbb{R}^d)} n\big(\mathrm{d\!I}(\mu_0,\nu) - \mathrm{d\!I}(\mu_1,\nu)\big)\right) \leqslant n\mathrm{d\!I}(\mu_0,\mu_1).$$

By symmetry, $\mathfrak{J}_n$ is n-Lipschitz w.r.t. $\mathrm{d\!I}$, thus τ-continuous. It is moreover valued in $[\min(0, \inf(\mathfrak{J})), n]$. To prove pointwise convergence, let $\mu \in \overline{\mathscr{B}}(\delta_0, N)$ be fixed. Assume by contradiction that there exists $M < \mathfrak{J}(\mu)$ such that $\mathfrak{J}_n(\mu) \leqslant M$ for all n. Let $\varepsilon := \min(1, \mathfrak{J}(\mu) - M) > 0$. Since $\mathfrak{J} + \mathbb{I}_N$ is narrowly lower semicontinuous, there exists $r > 0$ such that $\mathrm{d\!I}(\mu,\nu) < r$ implies $(\mathfrak{J} + \mathbb{I}_N)(\nu) \geqslant M + \varepsilon/2$. In particular, $\inf_{\mathrm{d\!I}(\mu,\nu)<r} (\mathfrak{J} + \mathbb{I}_N)(\nu) + n\mathrm{d\!I}(\mu,\nu) \geq M + \varepsilon/2$. From the definition of $\mathfrak{J}_n$ we have

$$\mathfrak{J}_n(\mu) = \min\left(n, \inf_{\mathrm{d\!I}(\mu,\nu)<r} (\mathfrak{J} + \mathbb{I}_N)(\nu) + n\mathrm{d\!I}(\mu,\nu), \inf_{\mathrm{d\!I}(\mu,\nu)\geqslant r} (\mathfrak{J} + \mathbb{I}_N)(\nu) + n\mathrm{d\!I}(\mu,\nu)\right),$$

and so

$$\mathfrak{J}_n(\mu) \geq \min\left(n, M + \frac{\varepsilon}{2}, \inf_{\mathrm{d\!I}(\mu,\nu)\geqslant r} (\mathfrak{J} + \mathbb{I}_N)(\nu) + nr\right) \geq \min\left(n, M + \frac{\varepsilon}{2}, \inf \mathfrak{J} + nr\right).$$

Taking n large enough so that $nr \geqslant M + \frac{\varepsilon}{2} - \inf(\mathfrak{J})$ and $n \geqslant M + \varepsilon/2$, we get $\mathfrak{J}_n(\mu) \geqslant M + \frac{\varepsilon}{2}$, which is absurd. □

This regularization allows us to implement the strategy of [LS84, Prop. 1.1].

Theorem 3.2.28 (Minimal supersolution in the unbounded case). *Assume **[A3.2.8]**, and that $\mathfrak{J} : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{\infty\}$ is τ−lower semicontinuous, lower bounded and not identically equal to ∞. Then, for any supersolution $v : [0,T] \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{\infty\}$ of the HJB equation (3.1) in the sense of Definition 3.2.19 such that $v(T,\cdot)$ is not identically ∞, there holds*

$$v(t,\nu) \geqslant V(t,\nu) \qquad \forall (t,\nu) \in [0,T] \times \mathscr{P}_2(\mathbb{R}^d). \tag{3.43}$$

Consequently, the value function V is the smallest viscosity supersolution of (3.1) in the sense of Definition 3.1.6.

Proof. Let $(t_0,\nu_0) \in [0,T] \times \mathscr{P}_2(\mathbb{R}^d)$. By the Grönwall estimates of Lemma 3.2.10, there is $N > 0$ such that $\mu_T^{t_0,\nu_0,u} \in \overline{\mathscr{B}}(\delta_0, N)$ for any control $u \in L^1(t_0,T;U)$. If $v(t_0,\nu_0) = \infty$, the inequality (3.43) is trivially satisfied. Assume now that $v(t_0,\nu_0) < \infty$, and let $(\mathfrak{J}_n)_n$ be given by Lemma 3.2.27. By Theorem 3.2.16, the HJB equation

$$-\partial_t v_n(t,\mu) + H\big(\mu, D_\mu v_n(t,\mu)\big) = 0, \qquad v_n(T,\mu) = \mathfrak{J}_n(\mu) \tag{3.44}$$

admits a unique solution given by

$$V_n(t,\nu) := \inf\{\mathfrak{J}_n(\mu_T^{t,\nu,u}) \mid u \in L^1(t,T;U)\} \qquad \forall (t,\nu) \in [0,T] \times \mathscr{P}_2(\mathbb{R}^d).$$

Note that the map v is a supersolution of each regularized problem (3.44). Let $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$ such that $v(T,\sigma) < \infty$: since $v(T,\sigma) \geqslant \mathfrak{J}(\sigma) \geqslant \mathfrak{J}_n(\sigma) = V_n(T,\sigma)$, we have $-\infty < V_n(T,\sigma) - v(T,\sigma) \leqslant 0$. In consequence, we can apply the comparison principle of Theorem 3.1.12 (see Proposition 3.2.6 for the applicability in our case), and deduce that $v(t,\nu) \geqslant V_n(t,\nu)$ for any $(t,\nu) \in \mathscr{P}_2(\mathbb{R}^d)$.

By the same comparison principle, the solutions V_n are ordered in the sense that $V_{n+1}(t,\nu) \geqslant V_n(t,\nu)$ for all n. Moreover, $\mathfrak{J}_n \leqslant \mathfrak{J}$ implies that the subsolutions V_n are smaller than the supersolution V. Hence the sequence $(V_n(t_0,\nu_0))_n$ is nondecreasing and upper bounded by $v(t_0,\nu_0) < \infty$, thus converges. For each $n \in \mathbb{N}$, let $u^n \in L^1(t,T;U)$ be a measurable control such that $V_n(t_0,\nu_0) \geqslant \mathfrak{J}_n\left(\mu_T^{t_0,\nu_0,u^n}\right) - e^{-n}$. Using Lemma 3.2.23, some (non relabeled) subsequence converges in τ to $\mu_T^{t_0,\nu_0,u}$ for some $u \in L^1(t,T;U)$. Using the monotonicity of the family $(\mathfrak{J}_n)_n$ and the continuity in τ of each $\mathfrak{J}_m$ for a fixed m,

$$\lim_{n\to\infty} V_n(t_0,\nu_0) \geqslant \liminf_{n\to\infty} \mathfrak{J}_n\left(\mu_T^{t_0,\nu_0,u^n}\right) - e^{-n} \geqslant \liminf_{n\to\infty, n\geqslant m} \mathfrak{J}_m\left(\mu_T^{t_0,\nu_0,u^n}\right) - e^{-n} = \mathfrak{J}_m\left(\mu_T^{t_0,\nu_0,u}\right).$$

Since $\mu_T^{t_0,\nu_0,u^n} \in \overline{\mathscr{B}}(\delta_0, N)$, the conclusion follows from taking the limit in $m \to \infty$ to obtain

$$v(t_0,\nu_0) \geqslant \lim_{n\to\infty} V_n(t,\nu) \geqslant \mathfrak{J}\left(\mu_T^{t_0,\nu_0,u}\right) \geqslant V(t_0,\nu_0).$$

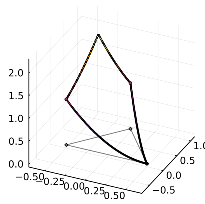
Since $(t_0,\nu_0) \in [0,T] \times \mathscr{P}_2(\mathbb{R}^d)$ is arbitrary, we conclude. □

Chapter 4

Directional differentiability of the Wasserstein distance over a network

This small chapter is entirely devoted to the existence and explicit expression of the directional derivative of the squared Wasserstein distance over a network.

Unlike in Chapter 2, we consider a class of one-dimensional networks Ω that may admit loops. Ω is equipped with the shortest path distance $d(\cdot,\cdot)$, and all edges are linearly parametrized for simplicity. This results in a metric space that is locally CAT(0), in that all points admit a neighbourhood that is isometric to a CAT(0) space − taking a sufficiently small ball around x so that $\overline{\mathscr{B}}(x,r)\setminus\{x\}$ does not contain any junction. However, as soon as the network contains a loop, it is not globally CAT(0).



We are interested into the space $\mathscr{P}_2(\Omega)$ of probability measures with finite second moment on Ω. Our main result is Theorem 4.3.2 below, that reads as follows.

Theorem. *Let (Ω, d) be an admissible network, ξ a probability measure on the unit-speed geodesics of Ω, and σ a measure on Ω, both with bounded second moments. Denote by e_h the evaluation of geodesics ar time $h \in [0,1]$. Then $h \mapsto d^2_{\mathcal{W}}(e_h\#\xi, \sigma)$ admits a limit at $h = 0$.*

The result includes the expression of the limit. This generalises the case where the underlying space Ω is Euclidean, treated in [AGS05; Gig08]. The case of manifolds is given in [Gig11]. With respect to these proofs, we argue slightly differently, not relying on the existence of the limit but directly estimating the limit inf and the limit sup.

If Ω is now a network, two difficulties appear: the junction points, at which the derivative of the underlying squared distance is always discontinuous, and the cut locus of $d(\cdot, z)$, which is precisely the remaining set of discontinuities. Most of the chapter is devoted to the behaviour of the cut locus with respect to z, and estimates of the measure that it can receive from a monotone plan (Lemma 4.2.7). The estimates imply that the discontinuities cannot play any role at the limit, and we can conclude.

Contents of Chapter 4

4.1 A direct proof in $\mathbb{R}^d$

As an introduction, let us consider an application of the form

$$u(\nu) := \inf_{\alpha\in\Gamma(\mu,\nu)} \int_{(x,y)\in(\mathbb{R}^d)^2} c(x,y)\,d\alpha(x,y), \tag{4.1}$$

where $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ is fixed. The infimum in (4.1) is reached as soon as $c(\cdot,\cdot)$ is lsc and coercive [Vil09, Th. 4.1].

Proposition 4.1.1 (Directional derivative). *Assume that $c \in \mathcal{C}^1((\mathbb{R}^d)^2;\mathbb{R}^+)$ grows at infinity, has a globally Lipschitz first partial derivative $\partial_x c(\cdot,\cdot)$, and admits a modulus $m : \mathbb{R}^+ \to \mathbb{R}^+$ and $p : \mathrm{T}\mathbb{R}^d \times \mathbb{R}^d$ quadratically growing such that*

$$c(x+hv,z) - c(x,z) \geqslant h\left[D_x c(x,z)(v) - p((x,v),z)m(h)\right] \qquad \textit{for all } h>0,\ (x,v)\in \mathrm{T}\mathbb{R}^d \textit{ and } z\in\mathbb{R}^d.$$

Then, for any $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, any measure field $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ and any $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$, there holds

$$\lim_{h\searrow 0} \frac{u(\exp_\mu(h\cdot\xi)) - u(\mu)}{h} = \min_{\substack{\alpha\in\Gamma(\xi,\sigma)\\ (\pi_x,\pi_z)\#\alpha\in\Gamma_o(\mu,\sigma)}} \int_{((x,v),z)\in\mathrm{T}\mathbb{R}^d\times\mathbb{R}^d} D_x c(x,z)(v)\,d\alpha((x,v),z). \tag{4.2}$$

Proof. By elementary estimates, there exists $k = k_c$ such that $|c(x+hv,z) - c(x,z)| \leqslant hk\left(|(x,v)|^2 + |z|^2 + 1\right)$ for all $h>0$ and $((x,v),z)\in\mathrm{T}\mathbb{R}^d\times\mathbb{R}^d$. Sending h to 0, we get that the right hand-side of (4.2) is finite.

First inequality. Let $\alpha \in \Gamma(\xi,\sigma)$ be such that $(\pi_x,\pi_z)\#\alpha$ realizes the inf in (4.1). Then, for any $h>0$,

$$\frac{u(\exp_\mu(h\cdot\xi)) - u(\mu)}{h} \leqslant \int_{((x,v),z)\in\mathrm{T}\mathbb{R}^d\times\mathbb{R}^d} \frac{c(x+hv,z) - c(x,z)}{h}\,d\alpha((x,v),z). \tag{4.3}$$

Owing to the estimate on $|c(x+hv,z) - c(x,z)|$, we may pass to the limit sup in $h\searrow 0$ and obtain a first inequality in (4.2).

Second inequality. Let $(h_n)_{n\in\mathbb{N}} \subset (0,1]$ be a vanishing sequence such that

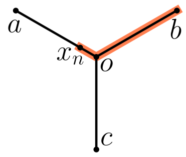
$$\liminf_{h\searrow 0} \frac{u(\exp_\mu(h\cdot\xi)) - u(\mu)}{h} = \lim_{n\to\infty} \frac{u(\exp_\mu(h_n\cdot\xi)) - u(\mu)}{h_n}. \tag{4.4}$$

For each n, let $\alpha_n \in \Gamma(\xi,\sigma)$ such that $(\pi_x + h_n\pi_v, \pi_z)\#\alpha_n$ is optimal for the cost $c(\cdot,\cdot)$. Since $\Gamma(\xi,\sigma)$ is compact with respect to $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d\times\mathbb{R}^d}(\cdot,\cdot)$ by Lemma 1.1.25 p. 9, some subsequence of $(\alpha_n)_{n\in\mathbb{N}}$ admits a limit α^* for this topology, and still satisfies (4.4). By classical arguments, one shows that $(\pi_x,\pi_z)\#\alpha^*$ is optimal for $c(\cdot,\cdot)$. Using the assumed regularity of $c(\cdot,\cdot)$,

$$\begin{aligned}\frac{u(\exp_\mu(h_n\cdot\xi)) - u(\mu)}{h_n} &\geqslant \int_{((x,v),z)} \frac{c(x+h_nv,z) - c(x,z)}{h_n}\,d\alpha_n \geqslant \int_{((x,v),z)} D_x c(x,z)(v) - p((x,v),z)m(h_n)\,d\alpha_n \\ &\underset{n\to\infty}{\longrightarrow} \int_{((x,v),z)\in\mathrm{T}\mathbb{R}^d\times\mathbb{R}^d} D_x c(x,z)(v)\,d\alpha^*.\end{aligned}$$

This provides the other inequality in (4.2), and we conclude. □

Let us comment on the proof to highlight the difficulties in the case of networks. The first inequality only needs a Lipschitz estimate on $c(\cdot,\cdot)$, and may be generalized without major efforts. The second inequality relies on three steps: a compactness argument, that generalizes quite well, a kind of stability of optimality for a varying cost, which also passes in networks, and a uniform lower bound. This third point is the problematic one in networks, since in the case $c(x,z) = d^2(x,z)$, the application $((x,v),z) \mapsto D_x c(x,z)(v)$ is not lower semicontinuous.



For instance, in the tripod network $[oa] \cup [ob] \cup [oc]$ with branches of length one glued at the junction o, consider the points $x_n = (1 - e^{-n})o \oplus e^{-n} a$. For any $n \in \mathbb{N}$, there holds $D_x d^2(x_n, x)(\overrightarrow{x_n b}) = -2d(x_n, c)d(x_n, b)$, but $D_x d^2(o, z)(\overrightarrow{ob}) = 2d(o, c)$. This shows that there is no lower semicontinuity to hope for in the last equations, and we cannot directly pass to the limit. However, the discontinuities happen only at certain points; the strategy in the sequel is to identify these points and control their influence. This is not absolutely trivial, since μ and σ are arbitrary in $\mathscr{P}_2(\Omega)$.

4.2 One-dimensional networks

We are interested in the following class of networks.

Definition 4.2.1 (Admissible network). *A one-dimensional network is a complete proper geodesic metric space (Ω, d) in which each point $x \in \Omega$ admits a neighbourhood that is isometric to a finite number of intervals glued at a common endpoint identified with x. The points at which at least three such intervals are glued are called junctions, and the set of junctions is denoted $\mathcal{J}$.*

Since Ω is locally CAT(0), we consider the same notations for the tangent cones $\mathrm{T}_x\Omega$, tangent bundle $\mathrm{T}\Omega$, and geodesics $[xy]$ whenever the latter are unique. However, a network may have branching geodesics, and the exponential is defined only for a small time that depends on $(x, v) \in \mathrm{T}\Omega$. To avoid unnecessary complexities, we consider the set $\mathscr{G} \subset \mathrm{AC}([0,1];\Omega)$ of geodesics as a surrogate of $\mathrm{T}\mathbb{R}^d$, and curves of measures defined by superposition measures $\xi \in \mathscr{P}_2(\mathscr{G})$. More precisely, we use the following notations.

$\mathscr{G} \subset \mathrm{AC}([0,1];\Omega)$	unit-speed geodesics, with $d_\infty(\gamma, \gamma') := \sup_{s \in [0,1]} d(\gamma_s, \gamma' s)$	$\gamma, \gamma' \in \mathscr{G}$
$e_h : \mathscr{G} \to \Omega$	evaluation map $e_h(\gamma) = \gamma_h$ at time $h \in [0,1]$	$\gamma \in \mathscr{G}$
$\exp_\mu(h \cdot \xi) \in \mathscr{P}_2(\Omega)$	"exponential" $e_h \# \xi$, with $\mu := e_0 \# \xi$	$\xi \in \mathscr{P}_2(\mathscr{G})$
$\Gamma_{o,h}(\xi, \sigma) \subset \Gamma(\xi, \sigma)$	plans $\alpha = \alpha(d\gamma, dz)$ such that $(e_h(\pi_\gamma), \pi_z)\#\alpha$ is optimal	$\xi \in \mathscr{P}_2(\mathscr{G})$, $\sigma \in \mathscr{P}_2(\Omega)$
$\exp_x^{-1}(z) \subset \mathrm{T}_x\Omega$	velocities γ_0^+ of geodesics $\gamma \in \mathscr{G}$ sending x to z	$x, z \in \Omega$

Definition 4.2.2 (Cut locus). *Let (Ω, d) be an admissible network in the sense of Definition 4.2.1. For each $z \in \Omega$, denote $\mathfrak{B}_z \subset \Omega$ the set of branching points, given by*

$$\mathfrak{B}_z := \left\{ x \in \Omega \;\middle|\; \textit{there exists several elements in } \exp_x^{-1}(z) \right\}.$$

Since all geodesics from x to z share the same length, hence the same norm of their initial velocities, the condition $x \in \mathfrak{B}_z$ really means that geodesics issued from x and going to z take different *directions* in the tangent cone. If (Ω, d) is CAT(0), then $\mathfrak{B}_z = \emptyset$ for all $z \in \Omega$.

4.2.1 Properties of the cut locus

We provide the argument of some elementary properties of the application $z \mapsto \mathfrak{B}_z$.

Lemma 4.2.3 (Upper semicontinuity). *Let $(x_n, z_n)_{n \in \mathbb{N}} \subset \Omega^2$ converge to (x, z) with $x_n \in \mathfrak{B}_{z_n}$ for all n. Then $x \in \mathfrak{B}_z$.*

Proof. Let $r > 0$ be small enough such that $\mathscr{B}(x, r)$ is isometric to a finite collection of segments glued at x. Up to extraction, we may consider that all points x_n belong to the same branch $]xa[$ for some point $a \in \Omega$. The space of directions at x_n is reduced to $\{\uparrow_{x_n}^a, \uparrow_{x_n}^x\}$, so that for each n, there exist two geodesics $\gamma^n, \overline{\gamma}^n$ from x_n to z_n with $a \in \gamma^n([0,1])$ and $x \in \gamma^n([0,1])$. Up to further extraction, we may consider that all geodesics $\overline{\gamma}^n$ pass through the same branch $[xb]$ for some $b \in \Omega$. The sequences $(\gamma^n)_{n \in \mathbb{N}}$ and $(\overline{\gamma}^n)_{n \in \mathbb{N}}$ are both equiLipschitz with constant $\sup_n d(x_n, z_n)$, and have equibounded sections in the locally compact space Ω, hence admit limit points γ^∞ and $\overline{\gamma}^\infty$ by Arzelà-Ascoli. These limit points are themselves geodesics sending x to z, and by construction, $a \in \gamma^\infty([0,1])$, and $b \in \overline{\gamma}^\infty([0,1])$. This shows that $x \in \mathfrak{B}_z$. □

Lemma 4.2.4 (Isolated points). *Assume (Ω, d) is an admissible network in the sense of Definition 4.2.1. For any $z \in \Omega$, the set $\mathfrak{B}_z$ is made of isolated points (each $x \in \mathfrak{B}_z$ is the only element of $\mathfrak{B}_z$ in a sufficiently small ball around x).*

Proof. Let $x \in \Omega \setminus \mathcal{J}$, and denote e the connected component of $\Omega \setminus \mathcal{J}$ containing x. By definition, each point of e admits exactly two directions in its tangent cone, so that e is isometric to a (possibly infinite) interval of $\mathbb{R}$. If $e \sim \mathbb{R}$, then the whole network is isometric to $\mathbb{R}$ (since it is geodesic, thus connected), and $\mathfrak{B}_z = \emptyset$ for all z. If $e \sim (0, \infty)$, it cannot be a subset of a closed loop, and $e \cap \mathfrak{B}_z = \emptyset$ for any $z \in \Omega$. Consider now a constant-speed parametrization $e : [0,1] \to \Omega$, with possibly $e(0) = e(1)$. Let $z \in \Omega$, and assume that there exist $t_1 \leqslant t_2 \in]0,1[$ such that $e_{t_i} \in \mathfrak{B}_z$ for $i \in \{1,2\}$.

Case 1: $z = e_{t_3}$ with $t_3 \in]t_1, t_2[$. Since the space of directions of e_{t_i} is reduced to $\{\uparrow_{e_{t_i}}^{e_0}, \uparrow_{e_{t_i}}^{e_1}\}$, the definition of $\mathfrak{B}_z$ implies that there exist two distinct geodesics $\gamma_i, \overline{\gamma}_i$ linking e_{t_i} to z with $\gamma_i = e_{[t_i, t_3]}$ and $e_0, e_1 \in \overline{\gamma}_i([0,1])$. Hence

$$d(e_{t_1}, z) = d(e_{t_1}, e_0) + d(e_0, e_1) + d(e_1, z) = d(e_{t_1}, e_0) + d(e_0, e_1) + d(e_1, e_{t_2}) + d(e_{t_2}, z),$$
$$d(e_{t_2}, z) = d(e_{t_2}, e_1) + d(e_1, e_0) + d(e_0, z) = d(e_{t_2}, e_1) + d(e_1, e_0) + d(e_0, e_{t_1}) + d(e_{t_1}, z).$$

Summing both and simplifying, we obtain $0 = d(e_{t_1}, e_0) + d(e_0, e_1) + d(e_1, e_{t_2})$, so that $t_1 = 0$ and $t_2 = 1$, which is absurd.

Case 2: $z \neq e_{[t_1, t_2]}$. Since the space of directions of e_{t_i} is reduced to $\{\uparrow_{e_{t_i}}^{e_0}, \uparrow_{e_{t_i}}^{e_1}\}$, the definition of $\mathfrak{B}_z$ implies that there exist two distinct geodesics $\gamma_i, \overline{\gamma}_i$ linking e_{t_i} to z with $e_{t_2} \in \gamma_1([0,1])$ and $e_{t_1} \in \gamma_2([0,1])$. Hence

$$d(e_{t_1}, z) = d(e_{t_1}, e_{t_2}) + d(e_{t_2}, z) = d(e_{t_1}, e_{t_2}) + d(e_{t_2}, e_{t_1}) + d(e_{t_1}, z),$$

and $t_1 = t_2$. This shows that each point located in the interior of an edge is isolated. On the other hand, if $y \in \mathfrak{B}_z$ is a junction point, it is at positive distance of the other junctions, and the common end point of a finite number of edges: consequently, if these edges were to contain a point of $\mathfrak{B}_z$, it would still stay at positive distance of y. □

Lemma 4.2.5 (Lipschitz-continuity). *Assume that (Ω, d) is an admissible network in the sense of Definition 4.2.1. The map $\mathfrak{B} : \Omega \rightrightarrows \Omega$ is Lipschitz-continuous with constant 1 in the Hausdorff sense, i.e.*

$$\max\left(\sup_{x \in \mathfrak{B}_z} \inf_{x' \in \mathfrak{B}_{z'}} d(x, x'), \sup_{x' \in \mathfrak{B}_{z'}} \inf_{x \in \mathfrak{B}_z} d(x, x') \right) \leqslant d(z, z') \qquad \forall z, z' \in \Omega.$$

Proof. It is enough to show that for any $z, z' \in \Omega$ and $x \in \mathfrak{B}_z$, there exists $x' \in \mathfrak{B}_{z'}$ such that $d(x, x') \leqslant d(z, z')$. We first prove the result for z' sufficiently close to z, then proceed by bootstrap. By definition, there exist several geodesics linking x to z, with distinct initial velocities in $\mathrm{T}_x\Omega$. Let $r > 0$ be sufficiently small so that any point $z' \in \mathscr{B}(z, r)$ is either on a geodesic linking x to z, or such that any geodesic linking x to z' goes by z. In the second case, one has $\exp_x^{-1}(z) = \exp_x^{-1}(z')$, and we might take $x' = x$. Consider now a point z' belonging to at least one geodesic from x to z. If z' belongs to the graph of two geodesics with distinct initial velocities, then by restriction, there exist two geodesics from x to z' with distinct initial velocities, and $x' := x$ belongs to $\mathfrak{B}_{z'}$. In the last case, all geodesics from x to z passing by z' share the same initial velocity. Let $\gamma : [0,1] \to \Omega$ be one of them, with $\gamma_0 = x$, $\gamma_1 = z$ and $\gamma_s = z'$ for some $s \in (0,1)$. Let $\gamma' : [0,1] \to \Omega$ be another geodesic from x to z with $\gamma_0^+ \neq (\gamma')_0^+$. We claim that there is no $\tau \in (0, s]$ such that $\gamma_\tau = \gamma'_\tau$. Indeed, if it were the case, the concatenation

$$t \mapsto \begin{cases} \gamma'_t & t \in [0, \tau] \\ \gamma_t & t \in]\tau, 1] \end{cases}$$

would furnish a geodesic linking x to z, passing by z' and with initial velocity distinct from γ_0^+. Let $t \in [s, 1]$ be the smallest time for which $\gamma_t = \gamma'_t$, and let $y = \gamma'_{t_y}$ be the unique point of $\gamma'([0,t])$ such that $d(x, y) = d(z', \gamma_t)$. Then

$$d(z', y) \leqslant d(z', x) + d(x, y) = d(z', x) + d(z', \gamma_t) = d(x, \gamma_t) = d(x, y) + d(y, \gamma_t) = d(z', \gamma_t) + d(\gamma_t, y).$$

This means that the concatenations $\gamma([0,s])\cup\gamma'([0,t_y])$ and $\gamma'([t_y,t])\cup\gamma([s,t])$ have the same length. If these curves are geodesics, then, noticing that they have distinct velocities at y, we deduce that $x' := y \in \mathfrak{B}_{z'}$. Otherwise, there exists a shorter path γ'' from y to z'. By continuity, there must exist a point y' in $\gamma'([0,t_y])$ admitting two geodesics from y' to z', one passing through $\gamma([0,s])$, and one through γ''. Since $d(x,y') \leqslant d(x,y) \leqslant d(z',\gamma_t) \leqslant d(z',z)$, we might take $x' := y'$.

This shows that for any z and $x \in \mathfrak{B}_z$, there exists an open set containing z in which all z' admit a point $x' \in \mathfrak{B}_{z'}$ at distance inferior to $d(z,z')$ of x. Consider $z, z' \in \Omega$ arbitrarily far from each other, and let $\xi : [0,1] \to \Omega$ be a geodesic between z and z'. Let $I \subset [0,1]$ be the largest interval containing 0 on which the following property is valid: for any $s \in I$, there exists $x_s \in \mathfrak{B}_{\xi_s}$ such that $d(x,x_s) \leqslant d(z,\xi_s)$. We show that I is both closed and open in $[0,1]$. If $s \in I$, then the previous step guarantees that there exists $r > 0$ such that $]s-r, s+r[\cap[0,1] \subset I$. On the other hand, in our case, I is closed if and only if $s := \max(I) \in I$. Let $s_n \nearrow_n s$, and for each n, pick $x_n \in \mathfrak{B}_{\xi_{s_n}}$ such that $d(x,x_n) \leqslant d(z,\xi_{s_n})$. By local compactness of Ω, the family $(x_n)_{n\in\mathbb{N}}$ admits a limit point x_∞. By Lemma 4.2.3, $x_\infty \in \mathfrak{B}_{\xi_s}$, and satisfies $d(x,x_\infty) = \lim_{n\to\infty} d(x,x_n) \leqslant \lim_{n\to\infty} d(z,\xi_{s_n}) = d(z,\xi_s)$. So I is closed, hence $I = [0,1]$, and we conclude. □

4.2.2 Preparatory steps for the measure case

Junctions and points in the cut locus are the obstructions to the continuity of the directional derivative of the underlying squared distance.

Lemma 4.2.6 (Differential of the underlying squared distance). *Let (Ω,d) be an admissible network in the sense of Definition 4.2.1. For any $z \in \Omega$, the application $d^2(\cdot,z)$ is differentiable everywhere along geodesics. Moreover, assume that for some geodesic $\gamma \in \mathscr{G}$ and $h > 0$, the segment $\gamma(]0,h[)$ does not intersect $\mathcal{J}$ nor $\mathfrak{B}_z$. Then*

$$\left|\frac{d^2(\gamma_s,z)-d^2(\gamma_0,z)}{s} - \frac{d^2(\gamma_{s'},z)-d^2(\gamma_0,z)}{s'}\right| = |s-s'||\gamma_0^+|_{\gamma_0}^2 \qquad \forall 0 < s,s' \leqslant h.$$

Proof. Let $z \in \Omega$, and $\gamma \in \mathscr{G}$ be a geodesic. By assumption, there exists a neighbourhood of γ_0 which is isometric to a CAT(0) network with unique junction identified with γ_0. If $\gamma_0^+ \in \exp_{\gamma_0}^{-1}(z)$, then γ coincides on a small interval with a geodesic linking γ_0 to z, and the directional derivative is given by $-2d(\gamma_0,z)|\gamma_0^+|_{\gamma_0}$. Otherwise, using the contraposition of Lemma 4.2.3, there must be a small interval $[0,\varepsilon]$ such that the only geodesic linking γ_s to z passes by γ_0 for all $s \in [0,\varepsilon]$. Therefore $d(\gamma_s,z) = d(\gamma_s,\gamma_0) + d(\gamma_0,z)$, and the directional derivative equals $2d(\gamma_0,z)|\gamma_0^+|_{\gamma_0}$.

Assume now that for some $h > 0$, the segment $\gamma(]0,h[)$ does not contain any junction point nor points of $\mathfrak{B}_z$. If $z = \gamma_\tau$ for some $\tau \in [0,h]$, then – by restriction of geodesics – the segment $\gamma([\tau,s])$ furnishes a geodesic linking γ_τ to γ_s for any $s \in [0,h]$, and $d(\gamma_s,z) = |\tau - s||\gamma_0^+|_{\gamma_0}$. Consequently, for $0 < s,s' \leqslant h$,

$$\left|\frac{d^2(\gamma_s,z)-d^2(\gamma_0,z)}{s} - \frac{d^2(\gamma_{s'},z)-d^2(\gamma_0,z)}{s'}\right| = \left|\frac{|s-\tau|^2-\tau^2}{s} - \frac{|s'-\tau|^2-\tau^2}{s'}\right||\gamma_0^+|_{\gamma_0}^2 = |s-s'||\gamma_0^+|_{\gamma_0}^2.$$

Consider now that $z \notin \gamma([0,h])$. Since $\gamma(]0,h[)$ does not contain any junction, the set of directions of any of its points y is reduced to $\{\uparrow_y^{\gamma_0}, \uparrow_y^{\gamma_h}\}$. Assume that for $0 < s < s' < h$, the geodesics linking γ_s and $\gamma_{s'}$ to z choose different directions. Then either $\uparrow_{\gamma_s}^z = \uparrow_{\gamma_s}^{\gamma_h}$ and $\uparrow_{\gamma_{s'}}^z = \uparrow_{\gamma_{s'}}^{\gamma_0}$, in which case

$$d(\gamma_s,z) = d(\gamma_s,\gamma_{s'}) + d(\gamma_{s'},z) = d(\gamma_s,\gamma_{s'}) + d(\sigma_{s'},\gamma_s) + d(\gamma_s,z),$$

so that $s = s'$. This is absurd. If, on the other hand, $\uparrow_{\gamma_s}^z = \uparrow_{\gamma_s}^{\gamma_0}$ and $\uparrow_{\gamma_{s'}}^z = \uparrow_{\gamma_{s'}}^{\gamma_h}$, the geodesic linking γ_s to z passes through γ_0, and that linking $\gamma_{s'}$ to z passes through γ_h. In other words, the continuous function

$$\tau \in [s,s'] \mapsto \left[d(\gamma_\tau,\gamma_s) + d(\gamma_s,z)\right] - \left[d(\gamma_\tau,\gamma_{s'}) + d(\gamma_{s'},z)\right]$$

is negative for $\tau = s$ and positive for $\tau = s'$, thus vanishes at some $\tau \in]s,s'[$, which must belong to $\mathfrak{B}_z$. This is impossible by assumption. Hence, for all points $s \in]0,h[$, the geodesics linking γ_s to z all pass by the same endpoint. If all geodesics pass through γ_0, then $d(\gamma_s,z) = d(\gamma_0,z) + d(\gamma_s,\gamma_0) = d(\gamma_0,z) + s|\gamma_0^+|_{\gamma_0}$, and for all $0 < s,s' \leqslant h$,

$$\left|\frac{d^2(\gamma_s,z)-d^2(\gamma_0,z)}{s} - \frac{d^2(\gamma_{s'},z)-d^2(\gamma_0,z)}{s'}\right| = \left|\frac{s^2|\gamma_0^+|_{\gamma_0}^2 + 2s|\gamma_0^+|_{\gamma_0}d(\gamma_0,z)}{s} - \frac{(s')^2|\gamma_0^+|_{\gamma_0}^2 + 2s'|\gamma_0^+|_{\gamma_0}d(\gamma_0,z)}{s'}\right|$$
$$= |s-s'||\gamma_0^+|_{\gamma_0}^2.$$

The case where all geodesics pass through γ_h is symmetric. □

The following result is the key step to isolate the discontinuities of the derivative of the squared distance. The statement is quite intricate, but sharp; we provide some illustrations of each point after the proof.

Lemma 4.2.7 (Monotone subset of the cut locus). *Let (Ω, d) be an admissible network in the sense of Definition 4.2.1, and let $S \subset \Omega^2$ be a closed set such that*

4.2.7.a) for each couple $(x,z) \in S$, there holds $x \in \mathfrak{B}_z$,

4.2.7.b) S is monotone, i.e. for any $(x,z),(x',z') \in S$, there holds $d^2(x,z) + d^2(x',z') \leqslant d^2(x,z') + d^2(x',z)$.

Then for any compact $K \subset \Omega^2$, the set $\pi_x(S \cap K) := \{x \in \Omega \mid (x,z) \in S \cap K\}$ is made of isolated points.

Proof. Assume by contradiction that there exists $(x_n, z_n) \subset S \cap K$ such that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence of points that are two-by-two distinct, and converge towards $x \in \Omega$. Up to extraction, we may assume that $(z_n)_{n\in\mathbb{N}}$ converges to some $z \in \Omega$, and since S is closed, the pair (x,z) belongs to S. If there were n_0 large enough such that $z = z_n$ for all $n \geqslant n_0$, then all x_n for $n \geqslant n_0$ would belong to the set $\mathfrak{B}_z$, which is made of isolated points by Lemma 4.2.4. This is absurd, and up to further extraction, we may assume that $d(z_n, z) > 0$ for all n.

Let $r > 0$ be sufficiently small so that $\mathscr{B}(z,r)$ identifies with a finite collection of segments glued at their common endpoint z. Up to further extraction, we may assume that all points z_n belong to the same branch, identified with $[za]$ for some $a \in \Omega$.

If there exists a subsequence $(n_k)_{k\in\mathbb{N}}$ such that *all* geodesics linking x_{n_k} to z_{n_k} go through a, then by restriction, each x_{n_k} belongs to $\mathfrak{B}_a$. As the latter set is made of isolated points by Lemma 4.2.4, this is only possible if the sequence $(x_{n_k})_{n_k}$ is eventually stationary, which contradicts the assumption that $x_n \neq x$. By the same reasoning, there cannot exist a subsequence $(n_k)_{k\in\mathbb{N}}$ such that *all* geodesics linking x_{n_k} to z_{n_k} go through z, and up to a shift of indexes, we may consider that for all n, there exist two geodesics γ^n and $\overline{\gamma}^n$ linking x_n to z_n, with $a \in \gamma^n_{[0,1]}$ and $z \in \overline{\gamma}^n_{[0,1]}$. Consequently,

$$d(z_n, x_n) = d(z_n, z) + d(z, x_n) \qquad \forall n \in \mathbb{N}.$$

(The identity $d(z_n, x_n) = d(z_n, a) + d(a, x_n)$ is also valid but not used.) By Arzelà-Ascoli, there exists a limit point $\gamma^\infty \in \mathrm{AC}([0,1];\Omega)$ of the family $(\gamma^n)_{n\in\mathbb{N}}$, which is a geodesic linking x to z and passing through a. As each z_n belongs to the segment $[z,a]$, we get that

$$d(z,x) = d(z,z_n) + d(z_n,x).$$

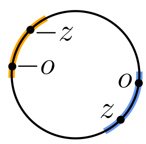
By the monotonicity of S, there holds for all $n \in \mathbb{N}$ that

$$\begin{aligned} d^2(x_n,z_n) + d^2(x,z) \leqslant d^2(x_n,z) + d^2(x,z_n) &= (d(z_n,x_n) - d(z_n,z))^2 + (d(z,x) - d(z,z_n))^2 \\ &= d^2(z_n,x_n) - 2d(z_n,x_n)d(z_n,z) + d^2(z_n,z) + d^2(z,x) - 2d(z,x)d(z,z_n) + d^2(z,z_n). \end{aligned}$$

This implies

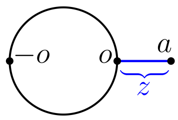
$$0 \leqslant -2d(z_n,z)(d(z_n,x_n) + d(z,x)) + 2d^2(z_n,z) \leqslant -2d(z_n,z)d(z,x) + 2d^2(z_n,z) = 2d(z_n,z)\left(d(z_n,z) - d(z,x)\right).$$

For n large enough, the quantity $d(z_n,z) - d(z,x)$ becomes negative, and we get an absurdity. □



The second point in Lemma 4.2.7 is not dispensable. For instance, consider $\Omega = \mathbb{S}^1$ with the shortest path distance. For any $z \in \Omega$, the only point of $\mathfrak{B}_z$ is the antipodal point $-z$. Hence we may construct a closed set S by the union of the points $(z,-z)$ for z varying in a geodesic. In this case, $\pi_x(S)$ is not made of isolated points. However, such a set will not be monotone; the antipodal point $-z$ is characterized as the *furthest* point of $\mathbb{S}^1$, so that any choice $z \neq z'$ will satisfy $d^2(z,-z) + d^2(z',-z') > d^2(z,-z') + d^2(z',-z)$.

Similarly, it is necessary to consider only the first coordinates of S. For instance,



consider the pan network made of the gluing of $\mathbb{S}^1$ to a segment $[oa]$. Let $S := \bigcup_{z\in[oa]} \mathfrak{B}_z \times \{z\}$. As all points $z \in [oa]$ share the same set $\mathfrak{B}_z$, reduced of the singleton $\{-o\}$. The set S is monotone (by commutativity of the addition), closed, satisfies $\pi_x(S) = \{-o\}$, but S is not itself made of isolated points.

4.3 Directional differentiability of the squared Wasserstein distance

The underlying squared distance $d^2(\cdot, z)$ is directionally differentiable along any reparametrized geodesic γ, and by Lemma 4.2.6, the differential quotients are Lipschitz if $\gamma(]0,h[)$ does not cross any junction nor intersect the cut locus of z. The junctions do not depend on z, and may be treated quite easily. However, the cut locus does move with z. For any $s > 0$, consider the problematic set

$$Z_s := \left\{(\gamma, z) \in X \,\middle|\, \gamma(]0,s[) \text{ intersects } \mathfrak{B}_z\right\}.$$

The following result allows us to control the measure of this set.

Lemma 4.3.1 (The bad set disappears). *Let $\xi \in \mathscr{P}_2(\mathscr{G})$ and $\sigma \in \mathscr{P}_2(\Omega)$. For each $h > 0$, let $\alpha_h \in \Gamma(\xi,\sigma)$ be such that $(e_h(\pi_\gamma), \pi_z)\#\alpha_h$ is an optimal transport plan between its marginals. Then for each compact $K \subset \mathscr{G} \times \Omega$,*

$$\limsup_{s\searrow 0} \limsup_{h\searrow 0} \alpha_h (Z_s \cap K) = 0.$$

Proof. Assume this not to hold. First pick a vanishing sequence $(s_n)_{n\in\mathbb{N}}$ such that $\limsup_{h\searrow 0} \alpha_h(Z_{s_n} \cap K) > 0$, and then, for each n, some h_n sufficiently small so that $\alpha_n := \alpha_{h_n} \in \Gamma_{o,h_n}(\xi,\sigma)$ satisfies $\alpha_n(Z_{s_n} \cap K) \geqslant \iota$ for some $\iota > 0$ that is independent of n. Let $C \geqslant 0$ be a bound on $\mathrm{Lip}(\gamma)$ for $(\gamma, z) \in K$.

Taking submeasures. Define

$$\beta_n := \frac{\alpha_n(\cdot \cap Z_{s_n} \cap K)}{\alpha_n(Z_{s_n} \cap K)}. \tag{4.5}$$

By restriction of optimality, each $(e_{h_n}(\pi_\gamma), \pi_z)\#\beta_n$ is optimal between its marginals, and supported in K. Since K is compact, we might extract a subsequence $(\beta_{n_k})_{k\in\mathbb{N}}$ converging with respect to $d_{\mathcal{W},\mathscr{G}\times\Omega}(\cdot,\cdot)$ towards some $\beta^* \in \mathscr{P}_2(\mathscr{G}\times\Omega)$. By stability of optimality [Vil09, Theorem 5.20], $(e_0(\pi_\gamma), \pi_z)\#\beta^*$ is itself optimal between its marginals.

The limit puts mass on geodesics issued from a finite set. Since the family $\left(\overline{Z_{s_n} \cap K}\right)_n$ is nonincreasing with n in the sense of inclusion, for any fixed $n \in \mathbb{N}$, there holds

$$\operatorname{supp} \beta_m \subset \overline{Z_{s_m} \cap K} \subset \overline{Z_{s_n} \cap K} \qquad \forall m \geqslant n \qquad \Longrightarrow \qquad \operatorname{supp} \beta^* \subset \overline{Z_{s_n} \cap K}.$$

Taking the intersection over $n \in \mathbb{N}$, we obtain

$$\operatorname{supp} \beta^* \subset \bigcap_{n\in\mathbb{N}} \overline{Z_{s_n} \cap K}.$$

Consider the set $S := \left\{(\gamma_0, z) \subset \Omega^2 \,\middle|\, (\gamma, z) \in \operatorname{supp} \beta^*\right\}$. Then S is closed as the image of the compact $\operatorname{supp} \beta^*$ by the continuous map $(\gamma, z) \mapsto (\gamma_0, z)$. Moreover,

- S is monotone. Indeed, $S \subset \operatorname{supp}(e_0, \pi_z)\#\beta^*$, which is cyclically monotone since optimal.
- For each couple $(x, z) \in S$, there holds $x \in \mathfrak{B}_z$. Indeed, let $(x, z) \in S$. By construction, there exists $(x^n, z^n) = (\gamma_0^n, z^n)$ such that $(x^n, z^n) \to_n (x, z)$ and $(\gamma^n, z^n) \in \overline{Z_{s_n} \cap K}$. By definition, for all $n \in \mathbb{N}$, the segment $\gamma^n([0, s_n])$ contains an element $y^n \in \mathfrak{B}_{z^n}$. Using that $\mathrm{Lip}(\gamma^n) \leqslant C$ over K uniformly in n, one has $d(x, y^n) \leqslant d(x, x^n) + s_n C \to_n 0$. By Lemma 4.2.3, we deduce that $x \in \mathfrak{B}_z$.

By Lemma 4.2.7, $\pi_x(S)$ is a finite set. Denoting $(x_i)_{i=1}^N$ its elements, we may write $\operatorname{supp} \beta^* = \bigsqcup_{i\in[\![1,N]\!]} B_i$ for some two-by-two disjoint sets $B_i \subset K$, each satisfying $\gamma_0 = x_i \in \mathfrak{B}_z$ for all $(\gamma, z) \in B_i$. Let $\jmath := \beta^*(B_1) > 0$, and consider $\beta_1^* := \beta^*(\cdot \cap B_1)$ a measure of mass $\jmath$. Tracking the mass transported by optimal transport plans, we may define a family $\beta_{n,1}$ of submeasures of β_n, each of mass $\jmath$, converging with respect to $d_{\mathcal{W},\mathcal{G}\times\Omega}(\cdot,\cdot)$ towards β_1^*.

The set $\pi_z(B_1)$ receives mass before the limit. Since supp $\beta_1^* = B_1$, one has $\pi_z\#\beta_1^*(\pi_z(B_1)) = \jmath$. As π_z is continuous, the family $(\pi_z\#\beta_{n,1})_{n\in\mathbb{N}}$ converges towards $\pi_z\#\beta_1^*$ with respect to $d_{\mathcal{W},\Omega}(\cdot,\cdot)$. We claim that $\liminf_{n\to\infty}\pi_z\#\beta_{n,1}(\pi_z(B_1)) = \jmath$. This cannot be deduced directly since $\pi_z(B_1)$ is not open; the additional argument is that all $\pi_z\#\beta_{n,1}$ are submeasures of σ.

Assume by contradiction that $\liminf_{n\to\infty}\pi_z\#\beta_{n,1}(\pi_z(B_1)) \leqslant \jmath-\rho$ for some $\rho>0$. Let $(\mathcal{O}_\varepsilon)_{\varepsilon>0}\subset\Omega$ be a decreasing family of open sets such that $\pi_z(B_1) = \bigcap_{\varepsilon>0}\mathcal{O}_\varepsilon$. For each $\varepsilon>0$, one has

$$\jmath = \pi_z\#\beta_1^*(\pi_z(B_1)) \leqslant \pi_z\#\beta_1^*(\mathcal{O}_\varepsilon) \leqslant \liminf_{n\to\infty}\pi_z\#\beta_{n,1}(\mathcal{O}_\varepsilon) = \liminf_{n\to\infty}\pi_z\#\beta_{n,1}(\mathcal{O}_\varepsilon\setminus\pi_z(B_1)) + \pi_z\#\beta_{n,1}(\pi_z(B_1)). \tag{4.6}$$

Recalling the construction of β_n in (4.5), and the assumption that $\alpha_n(Z_{s_n}\cap K)\geqslant\iota$, one has

$$\pi_z\#\beta_{n,1}(\mathcal{O}_\varepsilon\setminus\pi_z(B_1)) \leqslant \frac{1}{\iota}\pi_z\#\alpha_n(\mathcal{O}_\varepsilon\setminus\pi_z(B_1)) = \frac{1}{\iota}\sigma(\mathcal{O}_\varepsilon\setminus\pi_z(B_1))$$

independently of n. Consequently, (4.6) becomes

$$\jmath = \pi_z\#\beta_1^*(\pi_z(B_1)) \leqslant \frac{1}{\iota}\sigma(\mathcal{O}_\varepsilon\setminus\pi_z(B_1)) + \liminf_{n\to\infty}\pi_z\#\beta_{n,1}(\pi_z(B_1)) \leqslant \frac{1}{\iota}\sigma(\mathcal{O}_\varepsilon\setminus\pi_z(B_1)) + \jmath - \rho$$

for all $\varepsilon>0$. The family $(\mathcal{O}_\varepsilon\setminus\pi_z(B))_{\varepsilon>0}$ is decreasing with empty intersection, so $\sigma(\mathcal{O}_\varepsilon\setminus\pi_z(B)) \xrightarrow[\varepsilon\searrow 0]{} 0$, and this is absurd.

The measures $\beta_{n,1}$ put mass on geodesics passing through $\overline{x}$ at some positive time. For each $z\in\pi_z(B_1)$, the set $\mathfrak{B}_z$ contains $\overline{x}$ and is made of isolated points by Lemma 4.2.4. Consequently, the function $\kappa(z) := \inf_{y\in\mathfrak{B}_z\setminus\{\overline{x}\}} d(\overline{x},y)$ is positive for all $z\in\pi_z(B_1)$, and continuous since $z\mapsto\mathfrak{B}_z$ is 1-Lipschitz in the Hausdorff distance by Lemma 4.2.5. Let $\overline{\kappa}>0$ be a lower bound of $\kappa(\cdot)$ on the relatively compact set $\pi_z(B_1)\subset\pi_z(K)$. Consider

$$A_{s_n} := \left\{(\gamma,z)\subset Z_{s_n}\cap K \;\middle|\; \gamma_0\in\mathscr{B}(\overline{x},\overline{\kappa}/3),\ z\in\pi_z(B_1)\right\}.$$

We claim that for n large enough, $\overline{x}\in\gamma(]0,s_n[)$ for any $(\gamma,z)\in A_{s_n}$. Indeed, let n_0 be such that $s_nC\leqslant\overline{\kappa}/3$ for all $n\geqslant n_0$. By definition of $\overline{\kappa}$, for any $y\in\mathfrak{B}_z\setminus\{\overline{x}\}$, one has

$$\inf_{s\in[0,s_n]} d(y,\gamma_s) \geqslant \inf_{s\in[0,s_n]} d(y,\overline{x}) - d(\gamma_s,\gamma_0) - d(\gamma_0,\overline{x}) \geqslant \overline{\kappa} - 2\overline{\kappa}/3 = \overline{\kappa}/3 > 0.$$

So $\gamma(]0,s_n[)$ cannot intersect any other point of $\mathfrak{B}_z$ than $\overline{x}$. As this intersection is nonempty since $(\gamma,z)\in Z_{s_n}$, we deduce that $\overline{x}\in\gamma(]0,s_n[)$ for all $(\gamma,z)\in A_{s_n}$.

By definition, $\beta_{n,1}$ puts mass only on subsets of $Z_{s_n}\cap K$, and $\pi_z\#\beta_{n,1}(\pi_z(B_1))\to_n \jmath$ from the previous step. Moreover, since $(\gamma,z)\mapsto\gamma_0$ is 1-Lipschitz, the measures $e_0\circ\pi_\gamma\#\beta_{n,1}$ converge with respect to $d_{\mathcal{W},\Omega}$ towards $e_0\circ\pi_\gamma\#\beta_1^* = \jmath\delta_{\overline{x}}$. So there exists $n_1\geqslant n_0$ large enough so that $\beta_{n,1}(A_{s_n})\geqslant \jmath/2$ for all $n\geqslant n_0$. Then in particular,

$$\begin{aligned}\frac{\jmath}{2} &\leqslant \beta_{n,1}\left\{(\gamma,z)\in K \;\middle|\; \overline{x}\in\gamma(]0,s_n[)\right\} \leqslant \beta_{n,1}\left\{(\gamma,z)\in\mathcal{G}\times\Omega \;\middle|\; 0<d(\overline{x},\gamma_0)<s_nC\right\}\\ &\leqslant \frac{1}{\iota}[e_0\circ\pi_\gamma\#\alpha_n]\left\{y\in\Omega \;\middle|\; 0<d(\overline{x},y)<s_nC\right\} = \frac{1}{\iota}\mu\big(\mathscr{B}(\overline{x},s_nC)\setminus\{\overline{x}\}\big).\end{aligned}$$

The family $\big(\mathscr{B}(\overline{x},s_nC)\setminus\{\overline{x}\}\big)_n$ is nondecreasing with empty intersection, so for n large enough, we get a contradiction. □

We can now conclude with the main result of this chapter. Recall that $\mathcal{G}$ stands for the set of geodesics of Ω, each being an element of $\mathrm{AC}([0,1];\Omega)$; that e_h is the evaluation map at time h, and that $\Gamma_{o,h}(\xi,\sigma)$ stands for the subset of transport plans $\alpha=\alpha(d\gamma,dz)\in\Gamma(\xi,\sigma)$ such that $(e_h\circ\pi_\gamma,\pi_z)\#\alpha$ is an optimal transport plan between its marginals.

Theorem 4.3.2 (Directional differentiability of the squared Wasserstein distance). *Let (Ω,d) be an admissible network in the sense of Definition 4.2.1. Let $\xi \in \mathscr{P}_2(\mathscr{G})$ and $\sigma \in \mathscr{P}_2(\Omega)$. The squared Wasserstein distance $d^2_{\mathcal{W}}(\cdot,\sigma)$ is directionally differentiable at $\mu := e_0 \circ \pi_\gamma \# \xi$ along the curve $h \mapsto \exp_\mu(h\cdot\xi) := e_h \circ \pi_\gamma \# \xi$, and there holds*

$$\lim_{h\searrow 0} \frac{d^2_{\mathcal{W}}(\exp_\mu(h\cdot\xi),\sigma) - d^2_{\mathcal{W}}(\mu,\sigma)}{h} = \min_{\alpha\in\Gamma_{o,0}(\xi,\sigma)} \int_{(\gamma,z)\in\mathscr{G}\times\Omega} \lim_{h\searrow 0} \frac{d^2(\gamma_h,z) - d^2(\gamma_0,z)}{h} d\alpha(\gamma,z). \tag{4.7}$$

Proof. The first part of the proof is the same as in Proposition 4.1.1. We repeat it to spare the reader the adaptation of notations and technicalities. By coarse elementary estimates, there exists $k \geqslant 0$ such that for all $h > 0$, $\gamma \in \mathscr{G}$ and $z \in \Omega$,

$$\frac{\left|d^2(\gamma_h,z) - d^2(\gamma_0,z)\right|}{h} \leqslant \left(d(\gamma_h,z) + d(\gamma_0,z)\right)\left|\gamma_0^+\right|_{\gamma_0} \leqslant k\left(d^2_\infty(\gamma,o_{\mathscr{G}}) + d^2(o,z)\right) =: \varphi(\gamma,z). \tag{4.8}$$

Hence the quotients in the left-hand side of (4.7) are bounded uniformly in h. We treat separately the limit sup and inf.

Limit sup. For each $\varepsilon > 0$, let $\alpha^\varepsilon \in \Gamma_{o,0}(\xi,\sigma)$ be $\varepsilon-$optimal for the infimum in the right hand-side of (4.7). Since $(e_h \circ \pi_\gamma, \pi_z)\#\alpha^\varepsilon \in \Gamma(\exp_\mu(h\cdot\xi),\sigma)$, there holds

$$\frac{d^2_{\mathcal{W}}(\exp_\mu(h\cdot\xi),\sigma) - d^2_{\mathcal{W}}(\mu,\sigma)}{h} \leqslant \int_{(\gamma,z)\in\mathscr{G}\times\Omega} \frac{d^2(\gamma_h,z) - d^2(\gamma_0,z)}{h} d\alpha^\varepsilon(\gamma,z).$$

By Lemma 4.2.6, the integrand converges pointwise towards the directional derivative of the underlying squared distance, and is uniformly bounded by the quadratic estimate of (4.8). By Lebesgue' dominated convergence,

$$\begin{aligned}
\limsup_{h\searrow 0} \frac{d^2_{\mathcal{W}}(\exp_\mu(h\cdot\xi),\sigma) - d^2_{\mathcal{W}}(\mu,\sigma)}{h} &\leqslant \int_{(\gamma,z)\in\mathscr{G}\times\Omega} \frac{d}{dh}_{|h=0} d^2(\gamma(\cdot),z)\, d\alpha^\varepsilon \\
&\leqslant \inf_{\alpha\in\Gamma_{o,0}(\xi,\sigma)} \int_{(\gamma,z)\in\mathscr{G}\times\Omega} \frac{d}{dh}_{|h=0} d^2(\gamma(\cdot),z)\, d\alpha + \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we obtain a first inequality.

Limit inf. Let $(h_n)_{n\in\mathbb{N}} \subset]0,1]$ be a vanishing sequence such that $h_n \searrow 0$ and

$$\liminf_{h\searrow 0} \frac{d^2_{\mathcal{W}}(\exp_\mu(h\cdot\xi),\sigma) - d^2_{\mathcal{W}}(\mu,\sigma)}{h} = \lim_{n\searrow 0} \frac{d^2_{\mathcal{W}}(\exp_\mu(h_n\cdot\xi),\sigma) - d^2_{\mathcal{W}}(\mu,\sigma)}{h_n}. \tag{4.9}$$

For each n, let $\alpha_n \in \Gamma_{o,h_n}(\xi,\sigma) \subset \Gamma(\xi,\sigma)$. By the compactness of $\Gamma(\xi,\sigma)$ (see Lemma 1.1.25 p. 9), up to a non-relabeled subsequence that still retains (4.9), the sequence α_n converges with respect to $d_{\mathcal{W},\mathscr{G}\times\Omega}$ towards some $\alpha^* \in \Gamma(\xi,\sigma)$. Using respectively that the convergence with respect to $d_{\mathcal{W},\mathscr{G}\times\Omega}$ is equivalent to the convergence of the integrals against quadratically growing maps, the uniform estimate (4.8), and any plan $\alpha^0 \in \Gamma_{o,0}(\xi,\sigma)$ to furnish a transport plan $(e_{h_n}(\pi_\gamma),\pi_z)\#\alpha^0$ between $\exp_\mu(h_n\cdot\xi)$ and σ, we get

$$\begin{aligned}
\int_{(\gamma,z)\in\mathscr{G}\times\Omega} d^2(\gamma_0,z)\, d\alpha^* &= \lim_{n\to\infty} \int_{(\gamma,z)\in\mathscr{G}\times\Omega} d^2(\gamma_0,z)\, d\alpha_n \leqslant \lim_{n\to\infty} \int_{(\gamma,z)\in\mathscr{G}\times\Omega} \left[d^2(\gamma_{h_n},z) + h_n\varphi(\gamma,z)\right] d\alpha_n \\
&= \lim_{n\to\infty} d^2_{\mathcal{W}}(\exp_\mu(h_n\cdot\xi),\sigma) + h_n \int_{(\gamma,z)\in\mathscr{G}\times\Omega} \varphi(\gamma,z)\, d\alpha_n \\
&\leqslant \lim_{n\to\infty} d^2_{\mathcal{W}}(\mu,\sigma) + h_n \int_{(\gamma,z)\in\mathscr{G}\times\Omega} \varphi(\gamma,z)\, d(\alpha_n + \alpha^0) \\
&\leqslant d^2_{\mathcal{W}}(\mu,\sigma) + \lim_{n\to\infty} 2h_n \left[\int_{\gamma\in\mathscr{G}} d^2_\infty(\gamma,o_{\mathscr{G}})\, d\xi + d^2_{\mathcal{W}}(\sigma,\delta_o)\right] = d^2_{\mathcal{W}}(\mu,\sigma).
\end{aligned}$$

Hence $\alpha^* \in \Gamma_{o,0}(\xi,\sigma)$. The remaining part of the proof is new with respect to Proposition 4.1.1. Denote again

$$X_h := \{(\gamma,z)\in\mathscr{G}\times\Omega \mid \gamma(]0,h[) \text{ has empty intersection with } \mathcal{J}\cap\mathfrak{B}_z\},$$
$$Y_h := \mathscr{G}_h\times\Omega, \quad \text{with} \quad \mathscr{G}_h := \{\gamma\in\mathscr{G} \mid \gamma(]0,h[) \text{ intersects } \mathcal{J}\},$$
$$Z_h := \{(\gamma,z)\in\mathscr{G}\times\Omega \mid \gamma(]0,h[) \text{ intersects } \mathfrak{B}_z\}$$

Then $X_h\cup Y_h\cup Z_h = \mathscr{G}\times\Omega$, and X_h is the complement of $Y_h\cup Z_h$. The set Y_h writes as the countable union of the closed sets $\{\gamma\in\mathscr{G} \mid \gamma([e^{-m},h-e^{-m}])\cap\mathcal{J}\neq\emptyset\}\times\Omega$ for $m\in\mathbb{N}$, and Z_h as the countable union of the sets $\{(\gamma,z) \mid \gamma([e^{-m},h-e^{-m}])\cap\mathfrak{B}_z\neq\emptyset\}$. As $z\mapsto\mathfrak{B}_z$ is continuous in the set-valued sense by Lemma 4.2.5, these sets are closed, and all X_h, Y_h and Z_h are measurable. If $h\leqslant h'$, then $X_h\supset X_{h'}$, and $\bigcup_{h>0}X_h = \mathscr{G}\times\Omega$. By Lemma 4.2.6, there exists a continuous and quadratically growing map $m:\mathscr{G}\times\Omega\to\mathbb{R}^+$ such that for any $h>0$, $(\gamma,z)\in X_h$ and $0<h_n\leqslant h$, there holds

$$\left|\frac{d^2(\gamma_h,z)-d^2(\gamma_0,z)}{h} - \frac{d^2(\gamma_{h_n},z)-d^2(\gamma_0,z)}{h_n}\right| \leqslant |h-h_n|\, m(\gamma,z).$$

Fix $h>0$, and let $n\in\mathbb{N}$ be large enough so that $0<h_n\leqslant h$. Estimating $d^2_{\mathcal{W}}(\mu,\sigma)$ from above with the transport plan $(e_0\circ\pi_\gamma,\pi_z)\#\alpha_n$, we get that

$$\begin{aligned}\frac{d^2_{\mathcal{W}}(\exp_\mu(h_n\cdot\xi),\sigma)-d^2_{\mathcal{W}}(\mu,\sigma)}{h_n} &\geqslant \int_{(\gamma,z)\in X_h\sqcup X_h^c} \frac{d^2(\gamma_{h_n},z)-d^2(\gamma_0,z)}{h_n}\,d\alpha_n(\gamma,z)\\ &\geqslant \int_{(\gamma,z)\in X_h}\left[\frac{d^2(\gamma_h,z)-d^2(\gamma_0,z)}{h} - |h_n-h|\,m(\gamma,z)\right]d\alpha_n - \int_{X_h^c}\varphi(\gamma,z)\,d\alpha_n\\ &\geqslant \int_{(\gamma,z)\in\mathscr{G}\times\Omega}\frac{d^2(\gamma_h,z)-d^2(\gamma_0,z)}{h}\,d\alpha_n - |h_n-h|\int m\,d\alpha_n - 2\int_{X_h^c}\varphi(\gamma,z)\,d\alpha_n.\end{aligned}$$

The application $(\gamma,z)\mapsto\frac{d^2(\gamma_h,z)-d^2(\gamma_0,z)}{h}$ is continuous and quadratically growing, so we may pass to the limit in $n\to\infty$ in the two first terms. This yields

$$\liminf_{n\to\infty}\frac{d^2_{\mathcal{W}}(\exp_\mu(h_n\cdot\xi),\sigma)-d^2_{\mathcal{W}}(\mu,\sigma)}{h_n} \geqslant \int_{\mathscr{G}\times\Omega}\frac{d^2(\gamma_h,z)-d^2(\gamma_0,z)}{h}\,d\alpha^* - h\int m\,d\alpha^* - 2\limsup_{n\to\infty}\int_{X_h^c}\varphi(\gamma,z)\,d\alpha_n.$$

On the one hand, the quotients $\frac{d^2(\gamma_h,z)-d^2(\gamma_0,z)}{h}$ converge pointwise to the directional derivative of the underlying squared distance, and are uniformly bounded by a quadratic map. We may then apply Lebesgue' dominated convergence to the first term. On the other hand, the set $\{\alpha_n\}_{n\in\mathbb{N}}$ is relatively compact with respect to $d_{\mathcal{W},\mathscr{G}\times\Omega}(\cdot,\cdot)$, which is equivalent to the fact that for all $\varepsilon>0$, there exists a compact $K\subset\mathscr{G}\times\Omega$ satisfying $\sup_{n\in\mathbb{N}}\int_{(\gamma,z)\in K^c}\varphi(\gamma,z)\,d\alpha_n\leqslant\varepsilon$. Denote $\|\varphi_{|K}\|_\infty$ a constant bounding φ over K. Recalling that $X_h^c = Y_h\cup Z_h$,with $Y_h=\mathscr{G}_h\times\Omega$, we have

$$\limsup_{h\searrow 0}\limsup_{n\to\infty}\int_{(\gamma,z)\in X_h^c}\varphi(\gamma,z)\,d\alpha_n \leqslant \varepsilon + \|\varphi_{|K}\|_\infty\limsup_{h\searrow 0}\xi(\mathscr{G}_h) + \|\varphi_{|K}\|_\infty\limsup_{h\searrow 0}\limsup_{n\to\infty}\alpha_n(Z_h\cap K).$$

The family $(\mathscr{G}_h)_{h>0}$ has nonempty intersection, thus $\limsup_{h\searrow 0}\xi(\mathscr{G}_h)=0$. By Lemma 4.3.1, the last summand also vanishes. Combining all estimates, we get

$$\liminf_{n\to\infty}\frac{d^2_{\mathcal{W}}(\exp_\mu(h_n\cdot\xi),\sigma)-d^2_{\mathcal{W}}(\mu,\sigma)}{h_n} \geqslant \int_{(\gamma,z)\in\mathscr{G}\times\Omega}\lim_{h\searrow 0}\frac{d^2(\gamma_h,z)-d^2(\gamma_0,z)}{h}\,d\alpha^* - 2\varepsilon.$$

As $\varepsilon>0$ is arbitrary and $\alpha^*\in\Gamma_{o,0}(\xi,\sigma)$, the limit inf and the limit sup coincide, and (4.7) holds. □

Chapter 5

Geometry of the Wasserstein space

The first part of this chapter collects some algebraic results in $\mathscr{P}_2(\mathbb{R}^d)$. The aim is to study orthogonal decompositions, with two particular cases of interest: the barycentric/centred decomposition, and the tangent/solenoidal one. A large part of the chapter is devoted to the possibility to classify tangent and solenoidal measure fields based on directional derivatives of the squared distance. The last topic is the decomposition of any measure μ as a sum $\mu = \mu_0 + \cdots + \mu_d$ of submeasures that have centred tangent and solenoidal of dimension summing to d. The reader may be helped by thinking that the support of μ_k is "of dimension k", with Dirac masses having a support of dimension 0. This is only a visual analogy, without proper foundations so far, but may simplify the reading of the last sections.

The content of this chapter is partially derived from [Aus25].

Contents of Chapter 5

In all the sequel, unless explicitly stated otherwise, μ is an element of $\mathscr{P}_2(\mathbb{R}^d)$. The topics of this chapter are tightly linked to that of [Gig08, Chap. 4], to which we constantly refer. If a proposition is not explicitly cited as coming from [Gig08], it should be read as new.

5.1 Properties of metric scalar products

Here we focus on the maps $\langle\cdot,\cdot\rangle_\mu^\pm : \left(\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu\right)^2 \to \mathbb{R}$ defined in Definition 1.1.40. We recall some of their properties and extend them in preparation of the following sections. For $\xi,\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ disintegrated as $\xi = \xi_x \otimes \mu$ and $\zeta = \zeta_x \otimes \mu$, [Gig08, Proposition 4.2] yields the equivalent expressions

$$\langle\xi,\zeta\rangle_\mu^+ = \sup_{\alpha\in\Gamma_\mu(\xi,\zeta)} \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v,w\rangle\, d\alpha = \int_{x\in\mathbb{R}^d} \langle\xi_x,\zeta_x\rangle_{\delta_x}^+\, d\mu = \frac{1}{2}\left[\|\xi\|_\mu^2 + \|\zeta\|_\mu^2 - W_\mu^2(\xi,\zeta)\right], \tag{5.1a}$$

$$\langle\xi,\zeta\rangle_\mu^- = \inf_{\alpha\in\Gamma_\mu(\xi,\zeta)} \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v,w\rangle\, d\alpha = \int_{x\in\mathbb{R}^d} \langle\xi_x,\zeta_x\rangle_{\delta_x}^-\, d\mu = -\langle-\xi,\zeta\rangle_\mu^+ = -\langle\xi,-\zeta\rangle_\mu^+. \tag{5.1b}$$

Both $\langle\cdot,\cdot\rangle_\mu^\pm$ are symmetric and positively homogeneous with respect to both variables. By [Gig08, Prop. 4.21],

$$\left|\langle\xi,\zeta\rangle_\mu^+ - \langle\bar{\xi},\zeta\rangle_\mu^+\right| \leqslant W_\mu(\xi,\bar{\xi})\|\zeta\|_\mu \qquad \forall \xi,\bar{\xi},\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu. \tag{5.2}$$

The same estimate holds for $\langle\cdot,\cdot\rangle_\mu^-$. Since W_μ is lower semicontinuous with respect to the narrow and τ topologies on $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, the applications $\xi \mapsto \pm\langle\xi,\zeta\rangle_\mu^\pm$ are upper semicontinuous in these topologies.

5.1.1 Horizontal and vertical interpolation

We start by the convexity properties of $\langle\cdot,\cdot\rangle_\mu^\pm$ along both types of interpolation in $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$.

Definition 5.1.1 (Horizontal interpolation). *For $\xi_0,\xi_1 \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ and $t\in[0,1]$, let $\beta\in\Gamma_\mu(\xi_0,\xi_1)$ and*

$$\vec{\xi}_t^\beta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu, \qquad \vec{\xi}_t^\beta = \left(\pi_x, (1-t)\pi_{v_0} + t\pi_{v_1}\right)\#\beta.$$

In the sequel, *geodesic interpolation* (or *displacement interpolation*) refers to horizontal interpolation *along optimal transport plans only.* A set can be geodesically convex even if it is not horizontally convex.

Remark 5.1.2 (Geodesic and horizontal convex hulls)**.** *Consider the dimension $d = 1$. Let $\mu = \delta_0$ and $\xi = \frac{1}{2}\delta_{(0,-1)} + \frac{1}{2}\delta_{(0,1)}$. The only geodesic in $\left(\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu, W_\mu\right)$ linking ξ to itself is induced by $(\pi_x,\pi_v,\pi_v)\#\xi$. However, there is more than one transport plan in $\Gamma_\mu(\xi,\xi)$, and in fact*

$$\Gamma_\mu(\xi,\xi) = \left\{\alpha\frac{\delta_{(0,-1,-1)}+\delta_{(0,1,1)}}{2} + (1-\alpha)\frac{\delta_{(0,-1,1)}+\delta_{(0,1,-1)}}{2} \;\middle|\; \alpha\in[0,1]\right\}.$$

Hence the closed horizontally convex hull of the singleton $\{\xi\}$ is not reduced to ξ, and contains all the elements $(\pi_x,(1-t)\pi_v + t\pi_w)\#\eta$ for $\eta\in\Gamma_\mu(\xi,\xi)$.

Definition 5.1.3 (Vertical interpolation). *For $\xi_0,\xi_1 \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ and $t\in[0,1]$, let*

$$\xi_t^\uparrow \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu, \qquad \xi_t^\uparrow = (1-t)\xi^0 + t\xi^1.$$

The sum is understood here in the Banach sense: for any measurable $A\subset\mathrm{T}\mathbb{R}^d$, $\xi_t^\uparrow(A) = (1-t)\xi^0(A) + t\xi^1(A)$.

A simple but most useful property of vertical interpolation is that is preserves the support, in the sense that $\operatorname{supp}\xi_t^\uparrow = \operatorname{supp}\xi_0 \cup \operatorname{supp}\xi_1$ for all $t\in(0,1)$. Indeed, if $(x,v)\in\operatorname{supp}\xi_t^\uparrow$, then there exists a ball $B\subset\mathrm{T}\mathbb{R}^d$ of positive radius centred in (x,v) such that $(1-t)\xi_0(B) + t\xi_1(B) > 0$, which happens if and only if $\xi_0(B) > 0$ or $\xi_1(B) > 0$. Consequently, any vertical combination of measure fields whose supports are contained in a closed subset $A\subset\mathrm{T}\mathbb{R}^d$ stays supported in A.

The metric scalar products have mixed convexity properties along these two types of curves. The horizontal case has been treated in [Gig08].

Lemma 5.1.4 (Convexity properties). *For any $\xi_0, \xi_1, \zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ and $\beta \in \Gamma_\mu(\xi,\zeta)$, we have that*

$$t \mapsto \langle \vec{\xi}^{\beta}_t, \zeta\rangle^+_\mu \text{ and } t \mapsto \langle \xi^{\uparrow}_t, \zeta\rangle^-_\mu \text{ are convex,} \qquad t \mapsto \langle \vec{\xi}^{\beta}_t, \zeta\rangle^-_\mu \text{ and } t \mapsto \langle \xi^{\uparrow}_t, \zeta\rangle^+_\mu \text{ are concave.}$$

Proof. The fact that $t \mapsto \langle \vec{\xi}^{\beta}_t, \zeta\rangle^+_\mu$ is convex is deduced from [Gig08, Proposition 4.27]. The horizontal concavity of $\langle \cdot, \zeta\rangle^-_\mu$ follows from the relation $\langle \xi, \zeta\rangle^-_\mu = -\langle \xi, -\zeta\rangle^+_\mu$. For the same reason, it is enough to prove that $\langle \cdot, \zeta\rangle^+_\mu$ is vertically concave to conclude. To this aim, we use the expression of $\langle \cdot, \cdot\rangle^+_\mu$ by cone distance. First notice that

$$\|\xi^{\uparrow}_t\|^2_\mu = (1-t)\int_{(x,v)\in\mathrm{T}\mathbb{R}^d} |v|^2\, d\xi_0(x,v) + t\int_{(x,v)\in\mathrm{T}\mathbb{R}^d} |v|^2\, d\xi_1(x,v) = (1-t)\|\xi^0\|^2_\mu + t\|\xi^1\|^2_\mu.$$

On the other hand, let $\alpha_i \in \Gamma_{\mu,o}(\xi_i,\zeta)$ for $i \in \{1,2\}$. Then $\alpha^{\uparrow}_t \in \Gamma_\mu(\xi^{\uparrow}_t,\zeta)$, and

$$W^2_\mu(\xi^{\uparrow}_t,\zeta) \leqslant \int_{(x,v,w)\in T^2\mathbb{R}^d} |v-w|^2\, d\alpha^{\uparrow}_t(x,v,w) = (1-t)\, W^2_\mu(\xi_0,\zeta) + t W^2_\mu(\xi_1,\zeta).$$

Hence

$$\langle \xi^{\uparrow}_t, \zeta\rangle^+_\mu = \frac{1}{2}\left[\|\xi^{\uparrow}_t\|^2_\mu + \|\zeta\|^2_\mu - W^2_\mu(\xi^{\uparrow}_t,\zeta)\right] \geqslant (1-t)\langle \xi^0,\zeta\rangle^+_\mu + t\langle \xi^1,\zeta\rangle^+_\mu,$$

and we conclude. □

If $\xi = (id, f)\#\mu$, then $\Gamma_\mu(\xi,\zeta)$ is reduced to one element for each $\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, and

$$\langle \xi,\zeta\rangle^+_\mu = \langle \xi,\zeta\rangle^-_\mu = \int_{x\in\mathbb{R}^d} \langle f(x), \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\zeta)(x)\rangle\, d\mu(x)$$

for all $\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. Conversely, the cases where $\langle \cdot,\cdot\rangle^\pm_\mu$ coincide characterize map-induced elements.

Lemma 5.1.5 (Characterization of elements induced by maps). *Let $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. Then the following propositions are equivalent:*

(a) there exists $f \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ such that $\xi = f\#\mu$,

(b) $\langle \zeta,\xi\rangle^-_\mu = \langle \zeta,\xi\rangle^+_\mu$ for all $\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$,

(c) $\langle \xi,\xi\rangle^-_\mu = \langle \xi,\xi\rangle^+_\mu$.

Proof. Assume (a). Then for each $\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, the set $\Gamma_\mu(\xi,\zeta)$ is reduced to the unique element

$$\alpha = \int_{x\in\mathbb{R}^d} \zeta_x \otimes \delta_{f(x)}\, d\mu(x), \qquad \text{where } \zeta = \zeta_x \otimes \mu.$$

Then the inf and sup in (5.1) coincide, and (b) holds. As (b) trivially implies (c), there only stays to show that (c) implies (a). Assume that $\langle \xi,\xi\rangle^-_\mu = \langle \xi,\xi\rangle^+_\mu$. As

$$\begin{aligned}\langle \xi,\xi\rangle^-_\mu &= \inf_{\alpha\in\Gamma_\mu(\xi,\xi)} \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v,w\rangle\, d\alpha(x,v,w)\\ &\leqslant \int_{x\in\mathbb{R}^d}\int_{(v,w)\in\mathrm{T}^2_x\mathbb{R}^d} \langle v,w\rangle\, d[\xi_x\otimes\xi_x](v,w)\,d\mu(x) = \int_{x\in\mathbb{R}^d} \left|\mathrm{Bary}_{\mathrm{T}_x\mathbb{R}^d}(\xi_x)\right|^2 d\mu(x),\end{aligned}$$

where $\mathrm{Bary}_{\mathrm{T}_x\mathbb{R}^d}(\xi_x) := \int_{v\in\mathrm{T}_x\mathbb{R}^d} v\, d\xi_x(v)$ is defined for μ−almost every $x \in \mathbb{R}^d$, there holds

$$0 = \langle \xi,\xi\rangle^+ - \langle \xi,\xi\rangle^-_\mu = \int_{x\in\mathbb{R}^d}\left(\int_{v\in\mathrm{T}_x\mathbb{R}^d} |v|^2\, d\xi_x(v) - \left|\int_{v\in\mathrm{T}_x\mathbb{R}^d} v\, d\xi_x(v)\right|^2\right) d\mu(x).$$

Since $|\cdot|^2$ is strictly convex, $v = \mathrm{Bary}_{\mathrm{T}_x\mathbb{R}^d}(\xi_x)$ for ξ−almost all (x,v), and (a) holds with $f := \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi)$. □

In general, there is no linearity of $\langle\cdot,\cdot\rangle_\mu^\pm$ to hope for, as shown by the following example. Consider $d=1$, $\mu=\delta_0$, and the tangent elements $\xi_0=\frac{3}{4}\delta_{(0,-1)}+\frac{1}{4}\delta_{(0,1)}$, $\xi_1=\frac{1}{4}\delta_{(0,-1)}+\frac{3}{4}\delta_{(0,1)}$ and $\zeta=\frac{1}{2}\delta_{(0,-1)}+\frac{1}{2}\delta_{(0,1)}$. Let $\beta:=\frac{3}{4}\delta_{(0,-1,1)}+\frac{1}{4}\delta_{(0,1,-1)}$ be a transport plan in $\Gamma(\xi_0,\xi_1)$ that corresponds to a symmetry with respect to 0, i.e. $\vec{\xi}_t^\beta=(2t-1)\cdot\xi_0$. Then

$$\langle\xi_t^\uparrow,\zeta\rangle_\mu^+=\frac{1}{2}+\frac{1}{2}\min(t,1-t),\qquad \langle\vec{\xi}_t^\beta,\zeta\rangle_\mu^+=\frac{1}{2}-\frac{1}{2}\min(t,1-t),\qquad (1-t)\langle\xi_0,\zeta\rangle_\mu^++t\langle\xi_1,\zeta\rangle_\mu^+\equiv\frac{1}{2}.$$

It turns out that $\langle\cdot,\zeta\rangle_\mu^\pm$ is horizontally linear only in the map-induced case.

Lemma 5.1.6 (Linearity). *Let $\zeta\in\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. The applications $\langle\cdot,\zeta\rangle_\mu^\pm:\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu\to\mathbb{R}$ are horizontally linear if and only if ζ is induced by a map.*

Proof. By Lemma 5.1.5, if ζ is of the form $f\#\mu$ for some $f\in L^2(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$, then the applications $\langle\cdot,\zeta\rangle_\mu^+$ and $\langle\cdot,\zeta\rangle_\mu^-$ coincide on $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. By Lemma 5.1.4, they are both concave and convex along horizontal interpolating curves, thus linear. Assume now that $\langle\cdot,\zeta\rangle_\mu^+$ is horizontally linear. Let $\beta:=(\pi_x,\pi_v,-\pi_v)\#\zeta\in\Gamma_\mu(\zeta,-\zeta)$, so that $\vec{\zeta}_{1/2}^\beta=\left(\pi_x,\frac{\pi_v+\pi_w}{2}\right)\#\beta=0_\mu$. Then

$$\langle\zeta,\zeta\rangle_\mu^+-\langle\zeta,\zeta\rangle_\mu^-=\langle\zeta,\zeta\rangle_\mu^++\langle\zeta,-\zeta\rangle_\mu^+=2\langle\zeta,\vec{\zeta}_{1/2}^\beta\rangle_\mu^+=0.$$

By Lemma 5.1.5, ζ is induced by a map. □

Remark 5.1.7 (Wasserstein gradient)**.** *As an application, we might give a simpler justification of the fact that there is no Wasserstein gradient that would not be induced by a map, already pointed at in Lemma 3.2.3. A Wasserstein gradient $\nabla_W u(\mu)$ of a map $u:\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}$, if it exists, is an element $\xi\in$* **Tan**$_\mu$ *that satisfies*

$$\begin{aligned}&\textit{a subdifferential condition:} && u(\nu)-u(\mu)\geqslant\sup\nolimits_{\eta\in\exp_\mu^{-1}(\nu)}\langle\xi,\eta\rangle_\mu^++o(d_{\mathcal{W}}(\mu,\nu)) && \forall\nu\in\mathscr{P}_2(\mathbb{R}^d),\\ &\textit{a superdifferential condition:} && u(\nu)-u(\mu)\leqslant\inf\nolimits_{\eta\in\exp_\mu^{-1}(\nu)}\langle\xi,\eta\rangle_\mu^-+o(d_{\mathcal{W}}(\mu,\nu)) && \forall\nu\in\mathscr{P}_2(\mathbb{R}^d).\end{aligned}$$

By definition of **Tan**$_\mu$*, for any $\varepsilon>0$, there exists ξ^ε optimal on $[0,\alpha]$ for some $\alpha>0$, and $W_\mu(\xi,\xi^\varepsilon)\leqslant\varepsilon$. For any $s\in[0,\alpha)$, the plan $(\pi_x,\pi_x+s\pi_v)\#\xi^\varepsilon$ is the unique optimal plan between μ and $\nu_s:=\exp_\mu(s\cdot\xi^\varepsilon)$ by [AGS05, Lemma 7.2.1], so that*

$$\langle\xi,s\cdot\xi^\varepsilon\rangle_\mu^-+o(d_{\mathcal{W}}(\mu,\nu_s))\geqslant u(\nu_s)-u(\mu)\geqslant\langle\xi,s\cdot\xi^\varepsilon\rangle_\mu^++o(d_{\mathcal{W}}(\mu,\nu_s)).$$

Dividing by $s>0$ and sending $s\searrow 0$, we get that $\langle\xi,\xi^\varepsilon\rangle_\mu^-\geqslant\langle\xi,\xi^\varepsilon\rangle_\mu^+$, which, by the local Lipschitz-continuity of $\langle\cdot,\cdot\rangle_\mu^\pm$, implies $\langle\xi,\xi\rangle_\mu^-\geqslant\langle\xi,\xi\rangle_\mu^+$. As the opposite inequality always holds, both coincide, and by Lemma 5.1.5, ξ is induced by a map. The same results holds if inf and sup are inverted in the definition of $\nabla_W u$, but we cannot conclude if $\langle\cdot,\cdot\rangle_\mu^\pm$ are exchanged.

5.1.2 Closed horizontally convex subsets of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$

We are interested into horizontal convexity in $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, with respect to the plans in $\Gamma_\mu(\xi,\zeta)$. One could consider horizontally convex sets in $\mathscr{P}_2(\mathbb{R}^k)$ instead, with any transport plan between measures; this is done in [Gig08, Chap. 5]. Using $\mathbb{R}^{2d}\sim\mathrm{T}\mathbb{R}^d$, one could hope to get all results by simple application of this study. However, the transport plans in $\Gamma_\mu(\xi,\zeta)$ along which horizontal interpolation is considered form a strict subset of $\Gamma(\xi,\zeta)$. It is true that the cone distance W_μ identifies with a Monge-Kantorovich distance between measures on $\mathbb{R}^{2d}$, for the cost

$$c((x,v),(y,w)):=|v-w|^2+\mathbb{I}_0(x-y),$$

where $\mathbb{I}_0(z)=0$ if $z=0$, and ∞ otherwise. The set $\Gamma_\mu(\xi,\zeta)$ identifies with the subset of plans $\alpha\in\Gamma(\xi,\zeta)$ for which $\int c\,d\alpha<\infty$. This difference forbids us to apply the results in $\mathscr{P}_2(\mathbb{R}^{2d})$, and not only by a matter of definitions; some results that are valid in $\mathscr{P}_2(\mathbb{R}^{2d})$ are plainly false for $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. For instance, $d_{\mathcal{W}}-$closed horizontally convex subsets of $\mathscr{P}_2(\mathbb{R}^{2d})$ are $\tau-$closed [Gig08, Theorem 5.8]. (The topology τ is defined in Definition 1.1.27; the reader may think "narrowly closed in $d_{\mathcal{W}}-$balls"). This is false in general in $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, and we consider one of our main results the fact that the metric orthogonal of $\mathbf{Tan}_\mu$ is $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}-$closed (at least in dimension one, see Theorem 5.2.24 below). As a rule of thumb, one can check if a property depends on the continuity of the cost; if it does, we kindly ask the reader to be very careful when trying to prove it in $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$.

Definition 5.1.8 ($W_\mu-$closed horizontally convex hull). *Let $A \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. Its horizontally convex hull* $\operatorname{conv} A$ *is the smallest set $C \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ containing A that is horizontally convex in the sense that for all $\xi^0, \xi^1 \in C$, $\beta \in \Gamma_\mu(\xi,\zeta)$ and $t \in [0,1]$, there holds $\vec{\xi}_t^\beta \in C$. This set exists, since any intersection of horizontally convex sets is again so. We define in the same way the $W_\mu-$closed horizontally convex hull of A, denoted $\overrightarrow{\operatorname{conv}} A$.*

In the sequel, we drop the prefix $W_\mu-$ and say that a subset of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ is closed if it is $W_\mu-$closed.

Remark 5.1.9 (Balls)**.** *The first example of closed horizontally convex subsets of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ is given by $\overline{\mathscr{B}}(0_\mu, R)$ for any R. Indeed, if $\xi_0, \xi_1 \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ satisfy $\|\xi_i\|_\mu \leqslant R$, then for any $\beta \in \Gamma_\mu(\xi_0,\xi_1)$ and $t \in [0,1]$,*

$$\|\vec{\xi}_t^\beta\|_\mu = \sqrt{\int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} |(1-t)v + tw|^2 \, d\beta(x,v,w)} \leqslant (1-t)\|\xi_0\|_\mu + t\|\xi_1\|_\mu \leqslant R.$$

This implies that if A is bounded with respect to W_μ, so is $\overrightarrow{\operatorname{conv}} A$ with the same bound.

Remark 5.1.10 (Iterative formula)**.** *Denote*

$$C_0 := C, \qquad C_{n+1} := \left\{ \vec{\xi}_t^\beta = (\pi_x, (1-t)\pi_v + t\pi_w)\#\beta \;\middle|\; \xi_0, \xi_1 \in C_n,\ \beta \in \Gamma_\mu(\xi_0,\xi_1),\ t \in [0,1] \right\}.$$

By induction, the increasing family $(C_n)_{n\in\mathbb{N}}$ is contained into $\operatorname{conv} C$*, and consequently $D := \bigcup_{n\in\mathbb{N}} C_n \subset$* $\operatorname{conv} C$*. Moreover, D is horizontally convex: indeed, let $\xi_0,\xi_1 \in D$ and $\beta \in \Gamma_\mu(\xi_0,\xi_1)$. By definition, there exists $n,m \in \mathbb{N}$ such that $\xi_0 \in C_n$ and $\xi_1 \in C_m$. Then, by construction, $\vec{\xi}_t^\beta \in C_{\max(n,m)+1}$ for any $t \in [0,1]$. This shows that* $\operatorname{conv} C \subset D$*, and equality holds.*

The following result is proved in [Gig08, Proposition 4.30] in the case of $\mathbf{Tan}_\mu$. The proof in the case of a generic closed and horizontally convex set is exactly the same if one substitutes C for $\mathbf{Tan}_\mu$.

Proposition 5.1.11 (Well-defined projection [Gig08, Proposition 4.30]). *Let $C \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ be closed and horizontally convex. Then, for any $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, there exists a unique $\pi_C^\mu \xi \in C$ minimizing $W_\mu(\xi,\cdot)$ over C.*

The projection is induced by a map in a certain sense. This is proved in [Gig08, Proposition 4.32] in the case of $\mathbf{Tan}_\mu$, and we do not claim originality: for completeness, we provide the argument in the general case.

Proposition 5.1.12 (The projection is induced by a map). *Let $C \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ be closed and horizontally convex. Let $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, and $\pi_C^\mu\xi$ be its projection on C. There exists a unique optimal transport plan $\gamma \in \Gamma_{\mu,o}(\xi,\pi_C^\mu\xi)$, and an application $T \in L^2_\xi(\mathrm{T}\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ such that $\pi_x(T(x,v)) = x$ and $\gamma = (\pi_x,\pi_v,\pi_v(T(\pi_x,\pi_v)))\#\xi$.*

Proof. Pick any $\alpha \in \Gamma_{\mu,o}(\xi,\pi_C^\mu\xi)$, and disintegrate it in $\alpha = \alpha_{(x,v)} \otimes \xi$. Construct a transport plan $\beta \in \Gamma_\mu(\pi_C^\mu\xi,\pi_C^\mu\xi)$ by $\beta(dx,du,dv) = \int_{(z,w)\in\mathrm{T}\mathbb{R}^d} \delta_z(dx) \otimes \alpha_{(z,w)}(du) \otimes \alpha_{(z,w)}(dv) d\xi(z,w)$, in that for all $\varphi \in C_b(\mathrm{T}^2\mathbb{R}^d;\mathbb{R})$,

$$\int_{(x,u,v)\in\mathrm{T}^2\mathbb{R}^d} \varphi(x,u,v)\, d\beta(x,u,v) = \int_{(z,w)\in\mathrm{T}\mathbb{R}^d} \int_{(u,v)\in\mathrm{T}_x^2\mathbb{R}^d} \varphi(z,u,v)\, d\left[\alpha_{(z,w)} \otimes \alpha_{(z,w)}\right](u,v) d\xi(z,w).$$

Then $\left(\pi_x, \frac{\pi_v+\pi_w}{2}\right)\#\beta \in C$ by horizontal convexity, and

$$\begin{aligned} W_\mu^2\left(\xi, \left(\pi_x, \frac{\pi_v+\pi_w}{2}\right)\#\beta\right) &\leqslant \int_{(z,w)\in\mathrm{T}\mathbb{R}^d} \int_{(u,v)\in\mathrm{T}_x^2\mathbb{R}^d} \left|w - \frac{u+v}{2}\right|^2 d\left[\alpha_{(z,w)} \otimes \alpha_{(z,w)}\right](u,v) d\xi(z,w) \\ &\leqslant \frac{1}{2} W_\mu^2\left(\xi,\pi_C^\mu\xi\right) + \frac{1}{2} W_\mu^2\left(\xi,\pi_C^\mu\xi\right) - \frac{1}{4} \int_{(z,w)} \int_{(u,v)} |u-v|^2 \, d\left[\alpha_{(z,w)} \otimes \alpha_{(z,w)}\right](u,v) d\xi. \end{aligned}$$

Since $\pi_C^\mu \xi$ is the projection of ξ on C, the last integral vanishes, which happens only if $\alpha_{(z,w)}$ is concentrated on a single point for $\xi-$almost every (z,w). This shows that every optimal plan is induced by a map in the sense given to it in the statement. If there were two distinct optimal plans γ^0 and γ^1, the (vertical) convex combination $\frac{1}{2}(\gamma^0+\gamma^1)$ would still be optimal but not induced by a map, which is absurd; hence the optimal plan is unique. □

A similar argument shows the following.

Lemma 5.1.13 (The barycenter belongs to the closed convex hull). *Let $C \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ be closed and horizontally convex. Then, for any $\xi \in C$, the element $(id, Bary_{\mathrm{T}\mathbb{R}^d}(\xi))\#\mu$ also belongs to C.*

Proof. It is enough to prove that $(id, \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi))\#\mu$ belongs to $\overrightarrow{\mathrm{conv}}\{\xi\} \subset \overrightarrow{\mathrm{conv}}\, C$. Disintegrate ξ in $\xi_x \otimes \mu$, and consider the transport plan $\beta := (\xi_x \otimes \xi_x) \otimes \mu$. Then $\zeta := (\pi_x, \frac{1}{2}\pi_v + \frac{1}{2}\pi_w)\#\beta$ belongs to $\overrightarrow{\mathrm{conv}}\{\xi\}$. Denote $b := \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi)$. As $(id, b)\#\mu$ is induced by a map, there is only one element in $\Gamma_\mu(\zeta, (id, b)\#\mu)$, and

$$W_\mu^2(\zeta, (id,b)\#\mu) = \int_{(x,v)\in\mathrm{T}\mathbb{R}^d} |v-b(x)|^2\, d\zeta = \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \left|\frac{v+w}{2} - b(x)\right|^2 d\beta$$
$$= \frac{1}{4}\int_{(x,v)\in\mathrm{T}\mathbb{R}^d} |v-b(x)|^2\, d\xi + \frac{1}{4}\int_{(x,w)\in\mathrm{T}\mathbb{R}^d} |w-b(x)|^2\, d\xi + \frac{1}{2}\int_{\mathrm{T}^2\mathbb{R}^d} \langle v-b(x), w-b(x)\rangle\, d\beta.$$

By the same argument, the first two terms are equal to $\frac{1}{4}W_\mu^2(\xi, (id,b)\#\mu)$. As

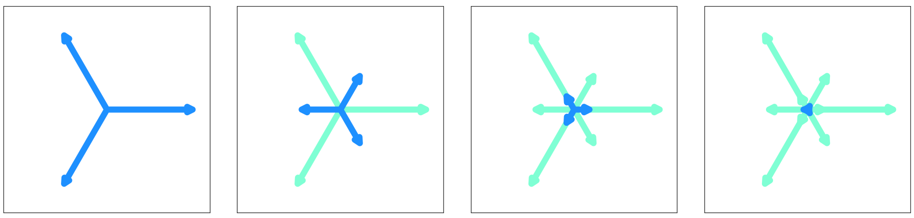
$$\int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v-b(x), w-b(x)\rangle\, d\beta = \int_{(x,v)\in\mathrm{T}^2\mathbb{R}^d} \langle v-b(x), \int_{w\in\mathrm{T}_x\mathbb{R}^d} w d\xi_x(w) - b(x)\rangle\, d\xi(x,v) = 0,$$

we obtain $W_\mu^2(\zeta, b\#\mu) \leqslant \frac{1}{2}W_\mu^2(\xi, b\#\mu)$. Since $\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\zeta) = \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi) = b$, we may employ the same argument with ζ in place of ξ to find by induction that

$$\inf_{\gamma\in\overrightarrow{\mathrm{conv}}\{\xi\}} W_\mu^2(\gamma, b\#\mu) \leqslant \inf_{n\in\mathbb{N}} \frac{1}{2^n} W_\mu^2(\xi, b\#\mu) = 0.$$

As $\overrightarrow{\mathrm{conv}}\{\xi\}$ is $W_\mu-$closed, it contains $b\#\mu$. □

It may happen that the barycenter is reached at the limit in n only, as for instance with $\mu = \delta_0$ and $\xi := \frac{1}{3}\sum_{i=0}^2 \delta_{(0,v_i)}$ for $v_i = (\cos(2\pi i/3), \sin(2\pi i/3))$.



The element of minimal norm of the closed convex hull of a set C is often of particular interest. In the case where C is in addition *vertically convex*, i.e. the vertical convex combination $(1-\lambda)\xi_0 + \lambda\xi_1$ belongs to C whenever $\xi_0, \xi_1 \in C$ and $\lambda \in [0,1]$, we can represent this element as a barycenter of an element of C.

Lemma 5.1.14 (The smallest element as a barycenter). *Let $C \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ be vertically convex and compact in the topology induced by $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}(\cdot,\cdot)$. The metric projection of 0_μ on $\overrightarrow{\mathrm{conv}}\, C$ is induced by a map $b \in L_\mu^2(\mathbb{R}^d; \mathrm{T}\mathbb{R}^d)$ owing to Proposition 5.1.12; additionally, there exists $\xi \in C$ such that $b = Bary_{\mathrm{T}\mathbb{R}^d}(\xi)$.*

Proof. Let $(C_n)_{n\in\mathbb{N}}$ be the sequence given by Remark 5.1.10, with $C = C_0 \subset C_1 \subset \cdots \subset \overline{\bigcup_n C_n} = \overrightarrow{\mathrm{conv}}\, C$. By construction, there exists $(\xi_n)_{n\in\mathbb{N}} \subset \overrightarrow{\mathrm{conv}}\, C$ such that $W_\mu(\xi_n, b\#\mu) \to_n 0$ and $\xi_n \in C_n$. By Lemma 5.1.13, the element $\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi_n)\#\mu$ belongs to $\overrightarrow{\mathrm{conv}}\, C$ for any n, and as $\|b\#\mu\|_\mu \leqslant \|\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi_n)\#\mu\|_\mu \leqslant \|\xi_n\|_\mu \to \|b\#\mu\|_\mu$, the sequence $(\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi_n))_{n\in\mathbb{N}}$ converges towards the unique minimizer $b\#\mu$ of $\|\cdot\|_\mu$ on $\overrightarrow{\mathrm{conv}}\, C$. For each n, ξ_n writes as the result of finitely many horizontally convex combinations of plans, so it can be written as

$$\xi^n = \left(\pi_x, \sum_{i=1}^{2^n} \lambda_i v_i\right)\#\gamma^n$$

for some $\gamma^n \in \mathscr{P}_2(\mathrm{T}^{2^n}\mathbb{R}^d)_\mu$ with 2^n marginals in C, and coefficients $(\lambda_i)_{i\in[\![1,2^n]\!]} \subset [0,1]$ summing to 1. Let $\zeta^n := \sum_{i=1}^{2^n} \lambda_i (\pi_x, \pi_{v_i})\#\gamma^n$ in the vertical sense. Since C is vertically convex, $\zeta^n \in C$ for all n. For any $\varphi \in \mathcal{C}(\mathrm{T}\mathbb{R}^d;\mathbb{R})$ such that any $\varphi(x,\cdot)$ is a linear application with operator norm satisfying $\int_{x\in\mathbb{R}^d} \|\varphi(x,\cdot)\|^2 d\mu < \infty$, one has

$$\int_{(x,v)\in\mathrm{T}\mathbb{R}^d} \varphi(x,v)d\zeta^n = \sum_{i=1}^{2^n} \lambda_i \int_{(x,v_1,\cdots,v_{2^n})\in\mathrm{T}^{2^n}\mathbb{R}^d} \varphi(x,v_i)d\gamma^n = \int_{(x,v_1,\cdots,v_{2^n})\in\mathrm{T}^{2^n}\mathbb{R}^d} \varphi\left(x, \sum_{i=1}^{2^n} \lambda_i v_i\right) d\gamma^n$$
$$= \int_{(x,v)\in\mathrm{T}\mathbb{R}^d} \varphi(x,v)d\left[\left(\pi_x, \sum_{i=1}^{2^n} \lambda_i v_i\right)\#\gamma^n\right] = \int_{(x,v)\in\mathrm{T}\mathbb{R}^d} \varphi(x,v)d\xi^n.$$

So $\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi_n) = \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\zeta_n)$. As C is compact in the topology induced by $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}$, the sequence (ζ_n) admits a limit point $\zeta^* \in C$. Passing to the limit yields $b = \lim_{n\to\infty} \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi_n) = \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\zeta^*)$. □

In Hilbert spaces, the projection p_v of some point v over a closed convex set C is characterized by

$$\langle u - p_v, p_v - v\rangle \geqslant 0 \qquad \forall u \in C.$$

This generalizes into a necessary condition in the CBB(0) space $(\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu, W_\mu)$.

Lemma 5.1.15 (Necessary condition). *Let $C \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ be nonempty, closed and horizontally convex. Let $\xi \notin C$ and $\pi^\mu_C\xi \in C$ be its projection on C. Then for any $\eta \in C$, and any plan $\alpha = \alpha(dx,du,dv,dp_v) \in \Gamma_\mu(\eta,\xi,\pi^\mu_C\xi)$ such that $(\pi_x,\pi_v,\pi_{p_v})\#\alpha \in \Gamma_{\mu,o}(\xi,\pi^\mu_C\xi)$, there holds*

$$\int_{(x,u,v,p_v)\in\mathrm{T}^3\mathbb{R}^d} \langle u - p_v, p_v - v\rangle \, d\alpha \geqslant 0. \tag{5.3}$$

Proof. By horizontal convexity, the curve $\eta_h := h \mapsto \left(\pi_x, (1-h)\pi_{p_v} + h\pi_u\right)\#\alpha$ lies in C. Hence

$$W^2_\mu\left(\pi^\mu_C\xi,\xi\right) \leqslant W^2_\mu\left(\eta_h,\xi\right) \leqslant \int_{(x,u,v,p_v)\in\mathrm{T}^3\mathbb{R}^d} \left|((1-h)p_v + hu) - v\right|^2 d\alpha$$
$$= \int_{(x,u,v,p_v)\in\mathrm{T}^3\mathbb{R}^d} h^2\left|u-p_v\right|^2 + 2h\langle u-p_v, p_v - v\rangle + \left|p_v - v\right|^2 d\alpha$$
$$= h^2 \|(\pi_x, \pi_u - \pi_{p_v})\#\alpha\|^2_\mu + 2h \int_{(x,u,v,w)\in\mathrm{T}^3\mathbb{R}^d} \langle u - p_v, p_v - v\rangle \, d\alpha + W^2_\mu\left(\pi^\mu_C\xi,\xi\right).$$

Dividing by h and letting $h \searrow 0$, we obtain the desired result. □

The analogy with Hilbert space also provides intuition for the following corollary. Denoting again p_v the projection of v on C, one has for all $u \in C$ that

$$|p_v - v|^2 = \langle p_v - v, p_v - v\rangle = \langle p_v - u, p_v - v\rangle + \langle u - v, p_v - v\rangle \leqslant \langle u, p_v - v\rangle - \langle v, p_v - v\rangle.$$

Corollary 5.1.16 (Control of the distance to the projection). *Consider the assumptions of Lemma 5.1.15. By Proposition 5.1.12, $\Gamma_\mu(\xi,\pi^\mu_C\xi)$ is reduced to a single element $\gamma = \gamma(dx,dv,dp_v)$. This allows to define unambiguously a "difference" $\omega := (\pi_x, \pi_{p_v} - \pi_v)\#\gamma$ between $\pi^\mu_C\xi$ and ξ. Then*

$$W^2_\mu(\pi^\mu_C\xi,\xi) \leqslant \langle \eta,\omega\rangle^-_\mu - \langle \xi,\omega\rangle^-_\mu \qquad \forall \eta \in C.$$

Proof. Let $\eta \in C$, and $\beta = \beta(dx,du,dw) \in \Gamma_\mu(\eta,\omega)$ realize $\langle\eta,\omega\rangle^-_\mu$. By Lemma 1.1.38 p. 13, there exists $\alpha = \alpha(dx,du,dv,dp_v) \in \Gamma_\mu(\eta,\omega)$ such that $(\pi_x,\pi_u,\pi_{p_v} - \pi_v)\#\alpha = \beta$. Using this plan,

$$\langle\eta,\omega\rangle^-_\mu = \int_{(x,u,w)\in\mathrm{T}^2\mathbb{R}^d} \langle u, w\rangle \, d\beta = \int_{(x,u,v,p_v)\in\mathrm{T}^3\mathbb{R}^d} \langle u, p_v - v\rangle \, d\alpha$$
$$= \int_{(x,u,v,p_v)\in\mathrm{T}^3\mathbb{R}^d} \langle u - p_v, p_v - v\rangle \, d\alpha + \int_{(x,u,v,p_v)\in\mathrm{T}^3\mathbb{R}^d} \langle p_v - v, p_v - v\rangle \, d\alpha + \int_{(x,u,v,p_v)\in\mathrm{T}^3\mathbb{R}^d} \langle v, p_v - v\rangle \, d\alpha.$$

Here the first term is greater than 0 by Lemma 5.1.15, the second is equal to $\|\omega\|_\mu^2 = W_\mu^2(\pi_C^\mu\xi,\xi)$, and the third is the integral of a transport plan between $\xi = (\pi_x,\pi_v)\#\alpha$ and $\omega = (\pi_x,\pi_{p_v}-\pi_v)\#\alpha$, so by definition larger than $\langle\xi,\omega\rangle_\mu^-$. Hence

$$\langle\eta,\omega\rangle_\mu^- \geqslant W_\mu^2(\pi_C^\mu\xi,\xi)+\langle\xi,\omega\rangle_\mu^-,$$

and we conclude. □

Lemma 5.1.17 (Attainment on the vertices). *Let $\varphi:\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu\to\mathbb{R}$ be horizontally concave. Then, for any $C\subset\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$,*

$$\inf_C\varphi = \inf_{\operatorname{conv}C}\varphi.$$

Proof. Since $C\subset\operatorname{conv}C$, there holds $\inf_C\varphi\geqslant\inf_{\operatorname{conv}C}\varphi$. On the other hand, let $(C_n)_{n\in\mathbb{N}}$ be the family of sets given by Remark 5.1.10, with $C = C_0\subset C_1\subset\cdots\subset\bigcup_n C_n = \operatorname{conv}C$. We have $\inf_{C_0}\varphi = \inf_C\varphi$. Assume by induction that $\inf_{C_n}\varphi\geqslant\inf_C\varphi$ for some $n\in\mathbb{N}$. Then for any $\xi_0,\xi_1\in C_n$ and $\beta\in\Gamma_\mu(\xi_0,\xi_1)$,

$$\varphi\left(\vec{\xi}_t^\beta\right)\geqslant(1-t)\varphi(\xi_0)+t\varphi(\xi_1)\geqslant\inf_{C_n}\varphi\geqslant\inf_C\varphi\qquad\forall t\in[0,1].$$

Hence $\inf_{C_{n+1}}\varphi\geqslant\inf_{C_n}\varphi\geqslant\inf_C\varphi$, thus $\inf_{\operatorname{conv}C}\varphi = \inf_{n\in\mathbb{N}}\inf_{C_n}\varphi\geqslant\inf_C\varphi$, and equality holds. □

As a first application, we may formulate a simple minimax result.

Lemma 5.1.18 (Minimax lemma). *Let $A,B\subset\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ be two nonempty, horizontally convex and bounded sets, with A relatively compact with respect to W_μ. Then*

$$\sup_{\alpha\in A}\inf_{\beta\in B}\langle\alpha,\beta\rangle_\mu^+ = \inf_{\beta\in B}\sup_{\alpha\in A}\langle\alpha,\beta\rangle_\mu^+ = \inf_{\beta\in B}\sup_{\alpha\in A}\langle\alpha,\beta\rangle_\mu^- = \sup_{\alpha\in A}\inf_{\beta\in B}\langle\alpha,\beta\rangle_\mu^-. \tag{5.4}$$

Proof. Since A,B are bounded, all terms are finite. By Lemma 5.1.13, for each $\xi\in A$, the element $\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi)\#\mu$ belongs to $\overline{A}$. As $\langle\cdot,\cdot\rangle_\mu^\pm$ is continuous with respect to W_μ, there holds

$$\sup_{\alpha\in A}\inf_{\beta\in B}\langle\alpha,\beta\rangle_\mu^\pm = \sup_{\alpha\in\overline{A}}\inf_{\beta\in\overline{B}}\langle\alpha,\beta\rangle_\mu^\pm\geqslant\sup_{a\in\overline{\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(A)}}\inf_{\beta\in B}\langle(id,a)\#\mu,\beta\rangle_\mu^\pm.$$

However, for any $\beta\in B$, there holds $\langle a\#\mu,\beta\rangle_\mu^\pm = \langle a\#\mu,(id,\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(\beta))\#\mu\rangle_\mu^\pm = \langle a,\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(\beta)\rangle_{L^2_\mu}$. Applying the same reasoning with the infimum over B in place of the supremum over A, we get the two inequalities

$$\sup_{\alpha\in A}\inf_{\beta\in B}\langle\alpha,\beta\rangle_\mu^\pm\geqslant\sup_{a\in\overline{\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(A)}}\inf_{b\in\overline{\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(B)}}\langle a,b\rangle_{L^2_\mu},\qquad\inf_{\beta\in B}\sup_{\alpha\in A}\langle\alpha,\beta\rangle_\mu^\pm\leqslant\inf_{b\in\overline{\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(B)}}\sup_{a\in\overline{\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(A)}}\langle a,b\rangle_{L^2_\mu}.$$

Since A is compact with respect to W_μ, and $\xi\mapsto\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi)$ is continuous from $(\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu,W_\mu)$ to L^2_μ, the set $\overline{\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(A)}$ is compact in L^2_μ. Applying the Ky Fan minimax theorem [Yan+08, Theorem 3], there holds

$$\sup_{a\in\overline{\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(A)}}\inf_{b\in\overline{\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(B)}}\langle a,b\rangle_{L^2_\mu} = \inf_{b\in\overline{\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(B)}}\sup_{a\in\overline{\operatorname{Bary}_{\mathrm{T}\mathbb{R}^d}(A)}}\langle a,b\rangle_{L^2_\mu}. \tag{5.5}$$

As $\sup_{\alpha\in A}\inf_{\beta\in B}\langle\alpha,\beta\rangle_\mu^\pm\leqslant\inf_{\beta\in B}\sup_{\alpha\in A}\langle\alpha,\beta\rangle_\mu^\pm$, and the terms in (5.5) do not depend on $\pm$, we conclude that (5.4) holds. □

5.1.3 Application: superdifferential of the squared distance

This section follows the notations of Section 3.2.1.2, in which the various notions of semidifferentials in the Wasserstein space are discussed. Let $\sigma\in\mathscr{P}_2(\mathbb{R}^d)$ be fixed. The fact that the elements of $\exp_\mu^{-1}(\sigma)$ belong to the superdifferential of $u:\nu\mapsto d^2_{\mathcal{W}}(\nu,\sigma)$ appears in various places [AF14; GT19], but to our knowledge, there is no complete description of this set. To be precise, we are interested into the set of $\xi\in\mathbf{Tan}_\mu$ satisfying

$$\limsup_{\nu\to\mu}\frac{u(\nu)-u(\mu)-\inf_{\eta\in\exp_\mu^{-1}(\nu)}\langle\xi,\eta\rangle_\mu^-}{d_{\mathcal{W}}(\mu,\nu)}\leqslant 0. \tag{5.6}$$

In the terminology of Section 3.2.1.2, this is the weak–sup–plan–$\mathbf{Tan}_\mu$ superdifferential of u, that we simply denote ∂_μ^+u.

Theorem 5.1.19 (Superdifferential of the squared Wasserstein distance). *For any $\mu,\sigma \in \mathscr{P}_2(\mathbb{R}^d)$, the superdifferential of $\nu \mapsto d^2_{\mathcal{W}}(\nu,\sigma)$ at μ is given by*

$$\partial^+_\mu u = \overrightarrow{\operatorname{conv}}\left\{-2\cdot \exp^{-1}_\mu(\sigma)\right\} = \overrightarrow{\operatorname{conv}}\left\{(\pi_x, -2(\pi_y - \pi_x))\#\alpha \mid \alpha \in \Gamma_o(\mu,\sigma)\right\}$$

Proof. Let $C := \overrightarrow{\operatorname{conv}}\{-2\cdot\exp^{-1}_\mu(\sigma)\}$. Since $\mathbf{Tan}_\mu$ is stable by rescaling and horizontal convex combinations by [Gig08, Propositions 4.25 and 4.29], the set C is a subset of $\mathbf{Tan}_\mu$. We proceed by double inclusion.

First inclusion. Let $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and $\eta \in \exp^{-1}_\mu(\nu)$. By the semiconcavity of $d^2_{\mathcal{W}}(\cdot,\sigma)$, there holds

$$d^2_{\mathcal{W}}(\exp_\mu(h\cdot\eta),\sigma) \geqslant (1-h)d^2_{\mathcal{W}}(\mu,\sigma) + hd^2_{\mathcal{W}}(\nu,\sigma) - h(1-h)d^2_{\mathcal{W}}(\mu,\nu).$$

Rearranging the terms, and using the expression of the directional derivative in Theorem 1.1.41,

$$\begin{aligned}\inf_{\zeta\in -2\cdot\exp^{-1}_\mu(\sigma)} \langle\zeta,\eta\rangle^-_\mu &= \lim_{h\searrow 0} \frac{d^2_{\mathcal{W}}(\exp_\mu(h\cdot\eta),\sigma) - d^2_{\mathcal{W}}(\mu,\sigma)}{h}\\ &\geqslant \lim_{h\searrow 0} d^2_{\mathcal{W}}(\nu,\sigma) - d^2_{\mathcal{W}}(\mu,\sigma) - (1-h)d^2_{\mathcal{W}}(\mu,\nu) = d^2_{\mathcal{W}}(\nu,\sigma) - d^2_{\mathcal{W}}(\mu,\sigma) - d^2_{\mathcal{W}}(\mu,\nu).\end{aligned}$$

By Lemma 5.1.4, $\zeta \mapsto \langle\zeta,\eta\rangle^-_\mu$ is horizontally concave, and Lemma 5.1.17 yields that $\inf_{\zeta\in-2\cdot\exp^{-1}_\mu(\sigma)}\langle\zeta,\eta\rangle^-_\mu = \inf_{\zeta\in C}\langle\zeta,\eta\rangle^-_\mu$. Minimizing over $\eta\in\exp^{-1}_\mu(\nu)$, there holds for any $\zeta\in C$ that

$$d^2_{\mathcal{W}}(\nu,\sigma) - d^2_{\mathcal{W}}(\mu,\sigma) \leqslant \inf_{\eta\in\exp^{-1}_\mu(\nu)} \langle\zeta,\eta\rangle^-_\mu + d^2_{\mathcal{W}}(\mu,\nu)$$

Consequently, $C \subset \partial^+_\mu d^2_{\mathcal{W}}(\cdot,\sigma)$.

Second inclusion. We argue by contradiction. Let $\xi\in\mathbf{Tan}_\mu$ such that $\xi\notin C$. Applying Propositions 5.1.11 and 5.1.12, there exists a unique metric projection $\pi^\mu_C\xi$ of ξ over C, for which the set $\Gamma_{\mu,o}(\xi,\pi^\mu_C\xi)$ is reduced to a unique element $\gamma = \gamma(dx,dv,dp_v)$. Since $\pi^\mu_C\xi\in\mathbf{Tan}_\mu$ by definition of C, the measure field $\omega := (\pi_x,\pi_{p_v}-\pi_v)\#\gamma \in \pi^\mu_C\xi\oplus(-\xi)$ also belongs to $\mathbf{Tan}_\mu$ by Proposition 5.2.8. By Corollary 5.1.16, ω satisfies

$$0 < W^2_\mu(\pi^\mu_C\xi,\xi) \leqslant \inf_{\zeta\in C}\langle\zeta,\omega\rangle^-_\mu - \langle\xi,\omega\rangle^-_\mu.$$

Let $\varpi\in\mathbf{Tan}_\mu$ be optimal on $[0,\lambda]$ for some $\lambda>0$, and such that $W_\mu(\omega,\varpi) \leqslant \frac{W^2_\mu(\pi^\mu_C\xi,\xi)}{2(2d_{\mathcal{W}}(\mu,\nu)+\|\xi\|_\mu)}$. We may assume that $\|\varpi\|_\mu>0$. By Remark 5.1.9, any $\zeta'\in C$ satisfies $\|\zeta'\|_\mu \leqslant 2d_{\mathcal{W}}(\mu,\nu)$, and using the Lipschitz estimate (5.2) on $\langle\cdot,\cdot\rangle^\pm_\mu$,

$$\inf_{\zeta\in C}\langle\zeta,\omega\rangle^-_\mu - \langle\xi,\omega\rangle^-_\mu \leqslant \inf_{\zeta\in C}\langle\zeta,\varpi\rangle^-_\mu - \langle\xi,\varpi\rangle^-_\mu + W_\mu(\varpi,\omega)\left(\sup_{\zeta'\in C}\|\zeta'\|_\mu + \|\xi\|_\mu\right) \leqslant \inf_{\zeta\in C}\langle\zeta,\varpi\rangle^-_\mu - \langle\xi,\varpi\rangle^-_\mu + \frac{W^2_\mu(\pi^\mu_C\xi,\xi)}{2}.$$

On the one hand,

$$\inf_{\zeta\in C}\langle\zeta,\varpi\rangle^-_\mu = \inf_{\zeta\in\overrightarrow{\operatorname{conv}}\{-2\cdot\exp^{-1}_\mu(\sigma)\}}\langle\zeta,\varpi\rangle^-_\mu = \inf_{\zeta\in-2\cdot\exp^{-1}_\mu(\sigma)}\langle\zeta,\varpi\rangle^-_\mu = \lim_{h\searrow 0}\frac{d^2_{\mathcal{W}}(\exp_\mu(h\cdot\varpi),\sigma) - d^2_{\mathcal{W}}(\mu,\sigma)}{h}.$$

On the other hand, for any $h\in(0,\lambda)$, the measure field $h\cdot\varpi$ is the unique element in $\exp^{-1}_\mu(\exp_\mu(h\cdot\varpi))$ by [AGS05, Lemma 7.2.1]. Hence, for such h,

$$\langle\xi,\varpi\rangle^-_\mu = \frac{1}{h}\langle\xi,h\cdot\varpi\rangle^-_\mu = \frac{1}{h}\inf_{\eta\in\exp^{-1}_\mu(\exp_\mu(h\cdot\varpi))}\langle\xi,\eta\rangle^-_\mu.$$

Since $d_{\mathcal{W}}(\mu,\exp_\mu(h\cdot\varpi)) = h\|\varpi\|_\mu$ for such h, this yields (taking $\nu=\exp_\mu(h\cdot\varpi)$)

$$\begin{aligned}0 < \frac{W^2_\mu(\pi^\mu_C\xi,\xi)}{2} &\leqslant \lim_{h\searrow 0}\frac{d^2_{\mathcal{W}}(\exp_\mu(h\cdot\varpi),\sigma) - d^2_{\mathcal{W}}(\mu,\sigma)}{h} - \frac{1}{h}\inf_{\eta\in\exp^{-1}_\mu(\exp_\mu(h\cdot\varpi))}\langle\xi,\eta\rangle^-_\mu\\ &\leqslant \limsup_{\nu\to\mu}\frac{u(\nu)-u(\mu)-\inf_{\eta\in\exp^{-1}_\mu(\nu)}\langle\xi,\eta\rangle^-_\mu}{d_{\mathcal{W}}(\mu,\nu)/\|\varpi\|_\mu}.\end{aligned}$$

Consequently, ξ does not belong to $\partial^+_\mu u$. □

The superdifferential that we computed is neither the only one, nor the most common in the literature. The following table provides the expression of some other variants, with the notations of Table 3.1. We do not know how to fill the remaining cells.

		Subset of $\mathbf{Tan}_\mu$		Subset of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$	
		weak	strong	weak	strong
Map	inf	$\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}\left(\partial_\mu^+ u\right)$		$\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}\left(\partial_\mu^+ u\right) \oplus \mathrm{Sol}_\mu$	
	sup	$\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}\left(\partial_\mu^+ u\right)$		$\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}\left(\partial_\mu^+ u\right) \oplus \mathrm{Sol}_\mu$	
Plan	inf				
	sup	$\partial_\mu^+ u := \overrightarrow{\mathrm{conv}}\left\{-2\cdot \exp_\mu^{-1}(\sigma)\right\}$		$\partial_\mu^+ u \oplus \mathbf{Sol}_\mu$	

This is a consequence of the following general inclusions. The first is proved by Gangbo and Tudorascu [GT19], and relies on the fact that for $f \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$, the difference between the supremum and the infimum of $\langle f, \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\eta)\rangle_{L^2_\mu}$ over $\eta \in \exp_\mu^{-1}(\nu)$ is itself of order $o(d_{\mathcal{W}}(\mu,\nu))$.

Lemma 5.1.20 (Inclusions). *There holds*

$$\begin{aligned}
-inf-map-\ \partial_\mu u &= *-sup-map-*\ \partial_\mu u, && (5.7)\\
-sup-map-\ \partial_\mu u &= Bary_{\mathrm{T}\mathbb{R}^d}\left(*-sup-plan-*\ \partial_\mu u\right), && (5.8)\\
weak-*-map-\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu\ \partial_\mu u &= weak-*-map-\mathbf{Tan}_\mu\ \partial_\mu u \oplus \mathrm{Sol}_\mu && (5.9)\\
weak-*-plan-\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu\ \partial_\mu u &= weak-*-plan-\mathbf{Tan}_\mu\ \partial_\mu u \oplus \mathbf{Sol}_\mu && (5.10)
\end{aligned}$$

In the above, $*$ should be replaced by the same value on both sides: for instance, (5.7) implies that weak−inf−map−$\mathbf{Tan}_\mu\ \partial_\mu u \subset$ weak−sup−map−$\mathbf{Tan}_\mu\ \partial_\mu u$, that strong−inf−map−$\mathbf{Tan}_\mu\ \partial_\mu u \subset$ strong−sup−map−$\mathbf{Tan}_\mu\ \partial_\mu u$, etc.

Proof. The equality (5.7) is given by [GT19, Theorem 3.6]. The inclusion $\subset$ in (5.8) is trivial. Conversely, let $\xi \in *-\mathrm{sup}-\mathrm{plan}-*\ \partial_\mu u$. Denote again $G(\mu,\nu)$ the set $\exp_\mu^{-1}(\nu)$ for weak definitions, or $(\pi_x, \pi_y - \pi_x)\#\Gamma(\mu,\nu)$ for strong definitions. Then for any $\eta \in G(\mu,\nu)$, there holds

$$\langle \eta, \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi)\rangle_\mu^+ = \int_{(x,v)\in\mathrm{T}\mathbb{R}^d} \langle v, \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi)(x)\rangle\, d\eta = \int_{(x,v,w)\in\mathrm{T}\mathbb{R}^d} \langle v, w\rangle\, d\alpha \leqslant \langle \eta,\xi\rangle_\mu^+$$

for $\alpha = (\eta_x \otimes \xi_x)\otimes \mu \in \Gamma_\mu(\eta,\xi)$ the pointwise product measure. Hence

$$\liminf_{\nu\to\mu} \frac{u(\nu)-u(\mu)-\sup_{\eta\in G(\mu,\nu)}\langle \eta, \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi)\rangle_\mu^+}{d_{\mathcal{W}}(\mu,\nu)} \geqslant \liminf_{\nu\to\mu} \frac{u(\nu)-u(\mu)-\sup_{\eta\in G(\mu,\nu)}\langle \eta,\xi\rangle_\mu^+}{d_{\mathcal{W}}(\mu,\nu)} \geqslant 0.$$

The argument for (5.9) and (5.10) is the same, with the difference that the former is restricted to map-induced fields. We write it in the case of (5.10). If ξ belongs to a weak−$*$−plan−$\mathbf{Tan}_\mu\ \partial_\mu u$ and $\zeta \in \mathbf{Sol}_\mu$, then for any plan $\alpha \in \Gamma_\mu(\xi,\zeta)$ and $\eta \in \mathbf{Tan}_\mu$, there holds $\langle \eta, (\pi_x, \pi_v + \pi_w)\#\alpha\rangle_\mu^\pm = \langle \eta,\xi\rangle_\mu^\pm$ by Remark 5.2.14. Hence $\xi \oplus \mathbf{Sol}_\mu \subset$ weak−$*$−plan−$\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu\ \partial_\mu u$. Conversely, if ξ belongs to the latter set, then $\langle \eta,\xi\rangle_\mu^\pm = \langle \eta, \pi_T^\mu \xi\rangle_\mu^\pm$ for any $\eta \in \mathbf{Tan}_\mu$, so that $\pi_T^\mu \eta \in \partial_\mu u$. □

5.1.4 Application: refined bounds on the sup-convolution of the squared distance

The following modification of the Wasserstein distance was used by Gallouët, Natale, and Todeschi [GNT22] to extrapolate geodesics beyond their maximal interval of definition, and Bertucci and Lions [BL24] as an instance of an L-differentiable test function.

Definition 5.1.21 (Sup-convolution). *Let $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ be fixed. For any $0<\delta<1$ and $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, let*

$$\Phi_\delta(\mu) := \sup_{\omega\in\mathscr{P}_2(\mathbb{R}^d)} d^2_{\mathcal{W}}(\omega,\nu) - \frac{1}{\delta} d^2_{\mathcal{W}}(\omega,\mu). \tag{5.11}$$

If one considers $\mu = \delta_x$ and $\nu = \delta_y$ for some points $x, y \in \mathbb{R}^d$, then for any $\omega \in \mathscr{P}_2(\mathbb{R}^d)$, the supremand in (5.11) reduces to

$$\int_{z\in\mathbb{R}^d} \left[|z-y|^2 - \frac{1}{\delta}|z-x|^2 \right] d\omega = -\left(\frac{1}{\delta}-1\right) \int_{z\in\mathbb{R}^d} \left| z - \frac{x/\delta - y}{1/\delta - 1} \right|^2 d\omega + \left(\frac{1}{\delta}-1\right) \left| \frac{x/\delta - y}{1/\delta - 1} \right|^2 + |y|^2 - \frac{1}{\delta}|x|^2.$$

The sup is attained in the Dirac mass $\omega := \delta_{x_\delta}$ for $x_\delta := \frac{x/\delta - y}{1/\delta - 1} = y + \frac{x-y}{1-\delta}$. This heuristic justifies the interpretation of the point of maximum as an extension of the geodesic going from ν to μ beyond the point μ. In [BL24], it is proved that

$$d_{\mathcal{W}}^2(\mu,\nu) \leqslant \Phi_\delta(\mu) \leqslant \frac{d_{\mathcal{W}}^2(\mu,\nu)}{1-\delta}, \tag{5.12}$$

that the supremum in (5.11) is reached at a unique point $\omega \in \mathscr{P}_2(\mathbb{R}^d)$ for which $\exp_\mu^{-1}(\omega)$ is reduced to a unique element $T\#\mu$ for some $T \in L^2_\mu(\mathbb{R}^d; \mathrm{T}\mathbb{R}^d)$, and that the function Φ_δ is L-differentiable, of class $\mathcal{C}^1$ in the Lions sense, with Wasserstein gradient given by $\nabla_w \Phi_\delta(\mu) = \frac{2}{\delta} T$. This provides a simple way to build regular functions, that can be used as test functions without changing the topology.

We are interested into the limit of the Wasserstein gradient $\nabla_w \Phi_\delta(\mu)$ when δ goes to 0. If there exists a unique optimal transport plan from μ to ν, that is induced by an application $S \in L^2_\mu(\mathbb{R}^d; \mathrm{T}\mathbb{R}^d)$, then $b := (-S)\#\mu$ is the Wasserstein gradient of $\frac{1}{2} d_{\mathcal{W}}^2(\cdot,\nu)$ at μ. In the general case, consider $\overrightarrow{\mathrm{conv}}\{-\exp_\mu^{-1}(\nu)\}$, where the closed horizontally convex hull is defined in Definition 5.1.8. By Propositions 5.1.11 and 5.1.12, there exists a unique projection of 0_μ on this set, which is induced by an application $b \in L^2_\mu(\mathbb{R}^d; \mathrm{T}\mathbb{R}^d)$. Moreover, $\exp_\mu^{-1}(\nu)$ is vertically convex since convex combinations of optimal transport plans are still optimal transport plans, and compact with respect to $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}$ by the compactness of transport plans (see Lemma 1.1.25 p. 9). By Lemma 5.1.14, b writes as $-\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi)$ for some $\xi \in \exp_\mu^{-1}(\nu) \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. We show that the Wasserstein gradient of Φ_δ converges towards $2b$ when δ goes to 0.

In this section, we systematically identify maps $f \in L^2_\mu(\mathbb{R}^d; \mathrm{T}\mathbb{R}^d)$ with their velocity component, and write $x + f(x)$ in place of $x + \pi_v(f(x))$.

Lemma 5.1.22 (Bounds). *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ and $\delta \in (0,1)$. Denote $\omega_\delta := \exp_\mu\left(\frac{\delta}{1-\delta} \cdot b\#\mu\right)$, and let ξ_δ be the measure field sending ω_δ to ν defined as $\xi_\delta := \left(\pi_x + \frac{\delta}{1-\delta} b(\pi_x), \pi_v - \frac{\delta}{1-\delta} b(\pi_x)\right)\#\xi$. Let T be the unique optimal transport map between μ and the point ω realising the supremum in (5.11). Then*

$$d_{\mathcal{W}}^2(\mu,\nu) + \frac{\delta}{1-\delta}\|b\|^2_{L^2_\mu} - \left(\|\xi_\delta\|^2_{\omega_\delta} - d_{\mathcal{W}}^2(\omega_\delta,\nu)\right) \leqslant \Phi_\delta(\mu) \leqslant d_{\mathcal{W}}^2(\mu,\nu) + \frac{1-\delta}{\delta}\left(\left\| \frac{\delta b}{1-\delta} \right\|^2_{L^2_\mu} - \left\| T - \frac{\delta b}{1-\delta} \right\|^2_{L^2_\mu} \right). \tag{5.13}$$

Moreover,

$$\left\| \nabla_w \Phi_\delta(\mu) - \frac{2}{1-\delta} b \right\|^2_{L^2_\mu} \leqslant \frac{4}{\delta(1-\delta)} \left(\|\xi_\delta\|^2_{\omega_\delta} - d_{\mathcal{W}}^2(\omega_\delta,\nu) \right) = O(\delta).$$

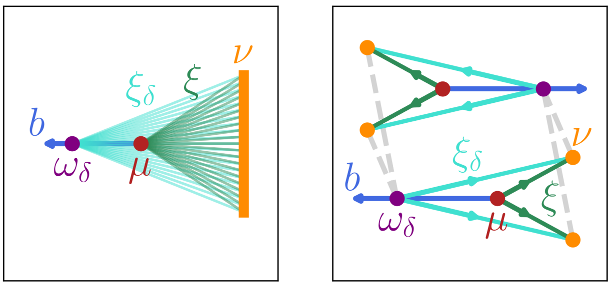


Figure 5.1: Notations.

Left: the measure field ξ is optimal between μ and ν. The vector field b is the opposite of the barycenter of ξ, and drags it along to produce ξ_δ. Right: ξ_δ is not optimal between ω_δ and ν, since the competitor in dashed line is better.

Since $\|b\|^2_{L^2_\mu} = \|\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi)\|^2_{L^2_\mu} \leqslant \|\xi\|^2_\mu = d^2_{\mathcal{W}}(\mu,\nu)$, the upper bound in (5.13) is finer than the upper bound in (5.12). It is less trivial that the lower bound is an improvement, although it becomes better than (5.12) as δ goes to 0.

The bounds provide information on pathological cases: the problems arise when ξ_δ is not optimal, i.e. when $\|\xi_\delta\|^2_{\omega_\delta} - d^2_{\mathcal{W}}(\omega_\delta,\nu) > 0$. In simple cases, as in both examples of Figure 5.1, there exists some threshold $\delta_0 > 0$ under which ξ_δ becomes optimal, and both bounds in (5.13) collapse to $\Phi_\delta(\mu) = d^2_{\mathcal{W}}(\mu,\nu) + \frac{\delta}{1-\delta}\|b\|^2_{L^2_\mu}$. Consequently, the map T is given by $\frac{\delta}{1-\delta}b$, and the optimal point ω is explicitly known. This is the case when μ is given by a finite sum of Dirac masses, for instance. This behaviour of being exact in some regular situations is pointed at in Proposition 3.1 of [BL24] for measures μ with smooth densities.

Proof. We begin by the upper bound. Recall that the unique point of maximum writes as $\omega = \exp_\mu(T\#\mu)$ for some $T \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$, and that $\xi \in \exp_\mu^{-1}(\nu)$ is an optimal velocity such that $b = -\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi)$. Then $(\pi_x + T(\pi_x), \pi_v - T(\pi_x))\#\xi$ is a measure field sending ω on ν, and

$$\begin{aligned}\Phi_\delta(\mu) &= d^2_{\mathcal{W}}(\omega,\nu) - \frac{1}{\delta}\|T\|^2_{L^2_\mu} \leqslant \int_{(x,v)\in\mathrm{T}\mathbb{R}^d} |v - T(x)|^2\, d\xi - \frac{1}{\delta}\|T\|^2_{L^2_\mu} \\ &= \int_{(x,v)\in\mathrm{T}\mathbb{R}^d} \left[|v|^2 - 2\langle T(x), v\rangle + |T(x)|^2\right] d\xi - \frac{1}{\delta}\|T\|^2_{L^2_\mu} = d^2_{\mathcal{W}}(\mu,\nu) + 2\langle T, b\rangle_{L^2_\mu} - \left(\frac{1}{\delta} - 1\right)\|T\|^2_{L^2_\mu} \\ &= d^2_{\mathcal{W}}(\mu,\nu) + \left(\frac{1}{\delta} - 1\right)\left(\left\|\frac{\delta}{1-\delta}b\right\|^2_{L^2_\mu} - \left\|T - \frac{\delta}{1-\delta}b\right\|^2_{L^2_\mu}\right).\end{aligned}$$

We turn to the lower bound. Consider ω_δ and ξ_δ as in the statement. Using that $b = -\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi)$,

$$\begin{aligned}\Phi_\delta(\mu) &\geqslant d^2_{\mathcal{W}}(\omega_\delta,\nu) - \frac{1}{\delta}d^2_{\mathcal{W}}(\omega_\delta,\mu) \geqslant \|\xi_\delta\|^2_{\omega_\delta} - \frac{1}{\delta}\left\|\frac{\delta}{1-\delta}b\right\|^2_{L^2_\mu} - \left(\|\xi_\delta\|^2_{\omega_\delta} - d^2_{\mathcal{W}}(\omega_\delta,\nu)\right) \\ &= \int_{(x,v)\in\mathrm{T}\mathbb{R}^d}\left|v - \frac{\delta}{1-\delta}b(x)\right|^2 d\xi - \frac{1}{\delta}\left\|\frac{\delta}{1-\delta}b\right\|^2_{L^2_\mu} - \left(\|\xi_\delta\|^2_{\omega_\delta} - d^2_{\mathcal{W}}(\omega_\delta,\nu)\right) \\ &= d^2_{\mathcal{W}}(\mu,\nu) - 2\left\langle \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi), \frac{\delta}{1-\delta}b\right\rangle_{L^2_\mu} - \left(\frac{1}{\delta} - 1\right)\left\|\frac{\delta}{1-\delta}b\right\|^2_{L^2_\mu} - \left(\|\xi_\delta\|^2_{\omega_\delta} - d^2_{\mathcal{W}}(\omega_\delta,\nu)\right) \\ &= d^2_{\mathcal{W}}(\mu,\nu) + \frac{2\delta}{1-\delta}\|b\|^2_{L^2_\mu} - \left(\frac{1}{\delta} - 1\right)\left\|\frac{\delta}{1-\delta}b\right\|^2_{L^2_\mu} - \left(\|\xi_\delta\|^2_{\omega_\delta} - d^2_{\mathcal{W}}(\omega_\delta,\nu)\right).\end{aligned}$$

This yields the lower bound in (5.13). Combining both inequalities and using that $\frac{2}{\delta}T = \nabla_w\Phi_\delta(\mu)$, we get

$$\left(\frac{1}{\delta} - 1\right)\left\|T - \frac{\delta}{1-\delta}b\right\|^2_{L^2_\mu} = \frac{1-\delta}{\delta}\frac{\delta^2}{4}\left\|\nabla_w\Phi_\delta(\mu) - \frac{2}{1-\delta}b\right\|^2_{L^2_\mu} \leqslant \|\xi_\delta\|^2_{\omega_\delta} - d^2_{\mathcal{W}}(\omega_\delta,\nu). \tag{5.14}$$

To bound the right hand-side, we use the directional differentiability of the squared Wasserstein distance. By Theorem 1.1.41, there exists function $m : \mathbb{R}^+ \to \mathbb{R}$ with $\lim_{r\to 0} m(r) = 0$ such that

$$d^2_{\mathcal{W}}(\omega_\delta,\nu) = d^2_{\mathcal{W}}\left(\exp_\mu\left(\frac{\delta}{1-\delta}\cdot b\#\mu\right),\nu\right) = d^2_{\mathcal{W}}(\mu,\nu) + D_\mu d^2_{\mathcal{W}}(\cdot,\nu)\left(\frac{\delta}{1-\delta}\cdot b\#\mu\right) + \left\|\frac{\delta b}{1-\delta}\right\|_{L^2_\mu} m\left(\left\|\frac{\delta b}{1-\delta}\right\|_{L^2_\mu}\right),$$

where using the horizontal concavity of $\langle\cdot,\cdot\rangle^-_\mu$ and Lemma 5.1.17,

$$D_\mu d^2_{\mathcal{W}}(\cdot,\nu)\left(\frac{\delta}{1-\delta}\cdot b\#\mu\right) = \frac{2\delta}{1-\delta}\inf_{\eta\in-\exp_\mu^{-1}(\nu)}\langle\eta, b\#\mu\rangle^-_\mu = \frac{2\delta}{1-\delta}\inf_{\eta\in\overrightarrow{\mathrm{conv}}\{-\exp_\mu^{-1}(\nu)\}}\langle\eta, b\#\mu\rangle^-_\mu.$$

Recall that $b\#\mu$ is the projection of 0_μ on the horizontally convex and W_μ−closed set $C := \overrightarrow{\mathrm{conv}}\{-\exp_\mu^{-1}(\nu)\}$. For any $\eta \in C$, the set of $\alpha = \alpha(dx,du,dv,dp_v) \in \Gamma_\mu(\eta, 0_\mu, b\#\mu)$ is reduced to the element $(\pi_x, \pi_v, 0, b(\pi_x))\#\eta$. Applying Lemma 5.1.15, we get that $\int_{(x,u,v,p_v)}\langle u - p_v, p_v - v\rangle\, d\alpha = \langle\eta, b\#\mu\rangle^-_\mu - \|b\|^2_{L^2_\mu} \geqslant 0$. So the infimum of $\langle \eta, b\#\mu\rangle_\mu^-$ over $\eta \in C$ is reached at $b\#\mu$, and $D_\mu d^2_{\mathcal{W}}(\cdot,\nu)(\frac{\delta}{1-\delta}\cdot b\#\mu) = \frac{2\delta}{1-\delta}\|b\|^2_{L^2_\mu}$. On the other hand, the explicit expression of ξ_δ yields that

$$\|\xi_\delta\|^2_{\omega_\delta} = \int_{(x,v)\in \mathrm{T}\mathbb{R}^d} \left| v - \frac{\delta}{1-\delta} b(x)\right|^2 d\xi = d^2_{\mathcal{W}}(\mu,\nu) + \frac{2\delta}{1-\delta}\|b\|^2_{L^2_\mu} + \frac{\delta^2}{(1-\delta)^2}\|b\|^2_{L^2_\mu}.$$

Plugging the two last lines in (5.14), we get

$$\begin{aligned}\left\|\nabla_w \Phi_\delta(\mu) - \frac{2}{1-\delta} b\right\|^2_{L^2_\mu} &\leqslant \frac{4}{\delta(1-\delta)}\left[\frac{\delta^2}{(1-\delta)^2}\|b\|^2_{L^2_\mu} - \left\|\frac{\delta}{1-\delta}b\right\|_{L^2_\mu} m\left(\left\|\frac{\delta}{1-\delta}b\right\|_{L^2_\mu}\right)\right] \\ &= \frac{4\delta\|b\|^2_{L^2_\mu}}{(1-\delta)^3} - \frac{\|b\|_{L^2_\mu}}{(1-\delta)^2} m\left(\left\|\frac{\delta}{1-\delta}b\right\|_{L^2_\mu}\right).\end{aligned}$$

This goes to 0 when δ does. The bound could be improved if one has *a priori* information on m. □

We give a few examples to conclude. Consider the dimension $d = 1$ and $\nu := \frac{\delta_{-1}+\delta_1}{2}$. Let $\mu_a := \frac{\delta_{-a}+\delta_a}{2}$ for $a \in [-1,1]$. The optimal transport plan between μ_a and ν is unique, and under the previous notations,

$$\xi = \frac{\delta_{(-|a|,-1+|a|)} + \delta_{(|a|,1-|a|)}}{2}, \quad b = \begin{cases}\delta_{(0,0)} & a = 0,\\ -\xi & a \neq 0,\end{cases} \quad \xi_\delta = \begin{cases}\xi & a = 0,\\ \frac{1}{2}\left[\delta_{\left(\frac{\delta-|a|}{1-\delta}, \frac{1}{1-\delta}(-1+|a|)\right)} + \delta_{\left(\frac{|a|-\delta}{1-\delta}, \frac{1}{1-\delta}(1-|a|)\right)}\right] & a \neq 0.\end{cases}$$

Here ξ_δ is not optimal for large δ, since when a is close to 0, the directions on which b puts mass are crossing each other. However, for a fixed, ξ_δ becomes optimal for δ sufficiently small. On this example, the value of Φ_δ can be computed, and

$$\Phi_\delta(\mu_a) = 1 - \frac{a^2}{\delta} \text{ if } |a| \leqslant \delta, \text{ and } \frac{(1-|a|)^2}{1-\delta} \text{ otherwise.}$$

This is to be compared with $d^2_{\mathcal{W}}(\mu_a,\nu) = (1-|a|)^2$. Figure 5.2 provides a visual comparison.

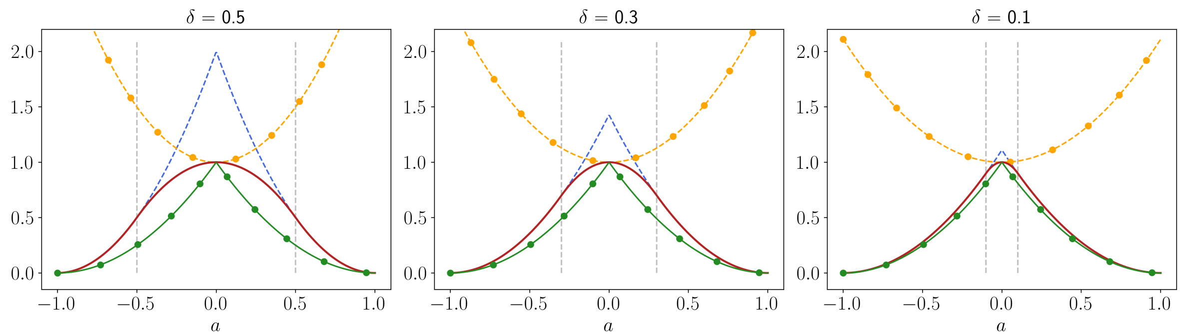


Figure 5.2: Explicit computation of Φ_δ.

The solid marked green line is the exact value of $d^2_{\mathcal{W}}(\mu_a,\nu)$ as a function of a. The solid red line is the approximation $\Phi_\delta(\mu_a)$ as a function of a, for values of δ ranging in $\{0.5, 0.3, 0.1\}$ from left to right. The dotted blue line is the upper bound $\frac{1}{1-\delta}d^2_{\mathcal{W}}(\mu_a,\nu)$ provided by (5.12), while the marked dotted yellow line is the upper bound $\frac{\delta}{1-\delta}\|b\#\mu\|^2_\mu$ provided by (5.13), without the (nonpositive) term in T. The vertical lines indicate $|a| = \delta$.

To get a pathological case, one can consider the dimension $d = 2$, and let μ be the 1-dimensional Hausdorff measure on a corner, i.e. the union of $[O,A] := [(0,0),(1,0)]$ and $[O,B] := [(0,0),(0,1)]$. Consider $\nu = (id + f)\#\mu$ for $f : \mathbb{R}^2 \to \mathbb{R}^2$ the vector field equal to $(0,-1)$ on $[O,A]$ and $[-1,0]$ on $[O,B]$. The measure field $(id, f)\#\mu$ is the unique optimal transport between μ and ν, but $-\lambda\cdot(id,f)\#\mu$ is not optimal for any $\lambda \in (0,1]$. Hence the upper and lower bounds in (5.13) do not coincide for $\delta > 0$.

5.2 Orthogonality and algebra in $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$

5.2.1 Orthogonal decompositions

We are interested in two orthogonal decompositions of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. The first one is quite trivial, but not as uninteresting as it seems. The second one is more involved but has a long mathematical history, for the good reason that it is fascinating. The intersections of both decompositions are the main focus of the remaining of the chapter.

5.2.1.1 Barycentric and centred components

Denote as follows the barycentric and centred measure fields.

$$L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)\#\mu := \left\{b\#\mu \;\middle|\; b\in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)\right\}, \qquad \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)^0_\mu := \left\{\xi\in\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \;\middle|\; \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi) = 0_{L^2_\mu}\right\}.$$

Definition 5.2.1 (Barycentric and centred components). *For any $\xi\in\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, denote $b_\xi := Bary_{\mathrm{T}\mathbb{R}^d}(\xi)\in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$, and $\xi^0\in\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)^0_\mu$ the centred field obtained by removing the barycenter of ξ, given by $\xi^0 := (\pi_x,\pi_v-\pi_v(b_\xi(\pi_x)))\#\xi$.*

Then ξ is the only element in $b_\xi\#\mu\oplus\xi^0$. Our interest comes from the following Pythagoras equality, which is an instance of the general fact that the integral of a function with mean 0 against a constant is 0.

Lemma 5.2.2 (Pythagoras for barycentric/centred decomposition). *Let $\xi,\zeta\in\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. Under the notations of Definition 5.2.1, there holds*

$$W^2_\mu(\xi,\zeta) = W^2_\mu(b_\xi\#\mu, b_\zeta\#\mu) + W^2_\mu\left(\xi^0,\zeta^0\right). \tag{5.15}$$

Consequently, $\langle\xi,\zeta\rangle^\pm_\mu = \langle b_\xi,b_\zeta\rangle_{L^2_\mu} + \langle\xi^0,\zeta^0\rangle^\pm_\mu$.

Proof. The mental picture is that an optimal transport plan between two measures moves the barycenter to the barycenter, and “the rest” to “the rest”. Up to boring the reader, we now detail the computations. First consider $\alpha^0\in\Gamma_\mu(\xi^0,\zeta^0)$, and define $\alpha := (\pi_x,\pi_v+b_\xi(\pi_x),\pi_w+b_\zeta(\pi_x))\#\alpha^0$. Then $\alpha\in\Gamma_\mu(\xi,\zeta)$, and

$$\begin{aligned}
W^2_\mu(\xi,\zeta) &\leqslant \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \left|v+b_\xi(x)-(w+b_\zeta(x))\right|^2 d\alpha^0\\
&= \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} |b_\xi(x)-b_\zeta(x)|^2 + 2\langle b_\xi(x)-b_\zeta(x), v-w\rangle + |v-w|^2\, d\alpha^0\\
&= W^2_\mu(b_\xi\#\mu,b_\zeta\#\mu) + 2\int_{x\in\mathbb{R}^d}\langle b_\xi(x)-b_\zeta(x), \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}\left(\xi^0\right)(x) - \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}\left(\zeta^0\right)(x)\rangle\, d\mu + W^2_\mu(\xi^0,\zeta^0).
\end{aligned}$$

The middle term vanishes since $\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}\left(\xi^0\right)(x) = \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}\left(\zeta^0\right)(x) = 0$ for $\mu-$a.e. $x\in\mathbb{R}^d$. On the other hand, consider $\alpha\in\Gamma_o(\xi,\zeta)$, and write it as $\alpha = b_\alpha\#\mu\otimes\alpha^0$, where the barycenter $b_\alpha\in L^2(\mathbb{R}^d;\mathrm{T}^2\mathbb{R}^d)$ is defined for $\mu-$a.e. $x\in\mathbb{R}^d$ by $b_\alpha(x) := \left(x,\int_{(v,w)\in\mathrm{T}^2_x\mathbb{R}^d}(v,w)\,d\alpha_x(v,w)\right)$, and $\alpha^0\in\mathscr{P}_2(\mathrm{T}^2\mathbb{R}^d)_\mu$ is given by $\alpha^0 = (\pi_x,\pi_v-\pi_v(b_\alpha(\pi_x)),\pi_w-\pi_w(b_\alpha(\pi_x)))\#\alpha$. Then $b_\alpha\#\mu\in\Gamma_\mu\left(b_\xi\#\mu,b_\zeta\#\mu\right)$; this was checked in the proof of Lemma 1.1.37 with the exact same object. On the other hand, for any $\varphi\in\mathcal{C}_b(\mathrm{T}\mathbb{R}^d;\mathbb{R})$,

$$\begin{aligned}
\int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d}\varphi(x,v)\,d\alpha^0 &= \int_{\mathrm{T}^2\mathbb{R}^d}\varphi\left(x,\pi_v-\pi_v(b_\alpha(x))\right)d\alpha = \int_{\mathrm{T}^2\mathbb{R}^d}\varphi\left(x,\pi_v-\int_{(v',w')\in\mathrm{T}^2_x\mathbb{R}^d} v'\,d\alpha_x(v',w')\right)d\alpha\\
&= \int_{(x,v)\in\mathrm{T}\mathbb{R}^d}\varphi\left(x,\pi_v-\int_{v'\in\mathrm{T}_x\mathbb{R}^d} v'\,d\xi_x(v')\right)d\xi = \int_{(x,v)\in\mathrm{T}\mathbb{R}^d}\varphi(x,v)\,d\xi^0(x,v).
\end{aligned}$$

Hence $(\pi_x,\pi_v)\#\alpha^0 = \xi^0$. Similarly, $(\pi_x,\pi_w)\#\alpha^0 = \zeta^0$, and

$$\begin{aligned}
W^2_\mu(\xi,\zeta) &= \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d}|v-w|^2\,d\alpha = \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d}|v+\pi_v(b_\alpha(x))-(w+\pi_w(b_\alpha(x)))|^2\,d\alpha^0\\
&= \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d}|v-w|^2\,d[b_\alpha\#\mu] + 0 + \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d}|v-w|^2\,d\alpha^0 \geqslant W^2_\mu(b_\xi\#\mu,b_\zeta\#\mu) + W^2_\mu\left(\xi^0,\zeta^0\right).
\end{aligned}$$

In the above, the middle term vanishes for the same reason as before. Hence the desired inequality. □

Proposition 5.2.3 (Barycentric/centred decomposition of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$). *Both sets $L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)\#\mu$ and $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)^0_\mu$ are W_μ−closed horizontally convex subsets of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. For each $b \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ and $\xi^0 \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)^0_\mu$, there holds*

$$\langle b\#\mu, \xi^0\rangle^+_\mu = \langle b\#\mu, \xi^0\rangle^-_\mu = 0.$$

Let $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, and write it as $\xi = b_\xi\#\mu \oplus \xi^0$ according to Definition 5.2.1. Then $b_\xi\#\mu$ and ξ^0 are respectively the metric projections of ξ on $L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)\#\mu$ and $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)^0_\mu$, in the sense that they minimize $W_\mu(\xi,\cdot)$ on these subsets.

Proof. The restriction of $W_\mu(\cdot,\cdot)$ to $L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)\#\mu$ coincides with the L^2_μ−norm between the inducing maps, with respect to which $L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ is closed. By Lemma 1.1.37, the barycenter is 1-Lipschitz with respect to W_μ, so that the limit of a Cauchy sequence $(\xi^0_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)^0_\mu$ is centred as well. We turn to horizontal convexity: if $\xi = f\#\mu$ and $\zeta = g\#\mu$, the set $\Gamma_\mu(\xi,\zeta)$ is reduced to $\beta := (\pi_x, \pi_v(f(\pi_x)), \pi_v(g(\pi_x)))\#\mu$, and for any $t \in [0,1]$, the plan $(\pi_x, (1-t)\pi_v + t\pi_w)\#\beta = ((1-t)f + tg)\#\beta$ is also induced by a map. Let now $\beta \in \Gamma_\mu(\xi^0,\zeta^0)$ and $t \in [0,1]$. For any $\varphi \in \mathcal{C}(\mathrm{T}\mathbb{R}^d;\mathbb{R})$ with quadratic growth that is linear in its second argument,

$$\begin{aligned}\int_{(x,v)} \varphi(x,v)\,d(\pi_x,(1-t)\pi_x + t\pi_w)\#\beta &= \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \varphi(x,(1-t)v + tw)\,d\beta\\ &= (1-t)\int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \varphi(x,v)\,d\beta + t\int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \varphi(x,w)\,d\beta = 0\end{aligned}$$

since $\xi^0 = (\pi_x,\pi_v)\#\beta$ and $\zeta^0 = (\pi_x,\pi_w)\#\beta$ are centred. Take now $b \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ and $\xi^0 \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)^0_\mu$. The unique element in $\Gamma_\mu(b\#\mu,\xi^0)$ is given by $(\pi_x, \pi_v(b(\pi_x)), \pi_v)\#\xi^0$, so that (abusing the notation by removing the π_v in the scalar product)

$$\langle b\#\mu, \xi^0\rangle^\pm_\mu = \int_{(x,v)\in\mathrm{T}\mathbb{R}^d} \langle b(x), v\rangle\, d\xi^0 = \int_{x\in\mathbb{R}^d} \langle b(x), \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}\left(\xi^0\right)(x)\rangle\, d\mu = 0.$$

The fact that b_ξ, ξ^0 are the metric projections of ξ on the barycentric and centred measure fields is a direct consequence of the Pythagoras equality of Lemma 5.2.2. □

5.2.1.2 Tangent and solenoidal components

Let us recall from Chapter 1 the definition of the tangent cone $\mathbf{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$. In the meantime, we define the divergence-free, or *solenoidal*, measure fields.

Definition 5.2.4 (Tangent and solenoidal measure fields). *A measure field $\xi \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ is* optimal on $[0,\alpha]$ *for $0 < \alpha \leqslant 1$ if $(\pi_x, \pi_x + \alpha\pi_v)\#\xi$ is an optimal transport plan between its marginals. The geometric tangent cone is defined as*

$$\mathbf{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d) := \overline{\left\{\,\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \;\middle|\; \xi \textit{ is optimal on } [0,\alpha] \textit{ for some } \alpha > 0\,\right\}}^{W_\mu} = \overline{\mathbb{R}^+\cdot \exp^{-1}_\mu\left(\mathscr{P}_2(\mathbb{R}^d)\right)}^{W_\mu}. \tag{5.16}$$

A measure field $\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ is solenoidal if

$$\langle\zeta,\eta\rangle^+_\mu = \langle\zeta,\eta\rangle^-_\mu = 0 \qquad \forall \eta \in \mathbf{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d). \tag{5.17}$$

The set of all solenoidal measure fields is denoted $\mathbf{Sol}_\mu\mathscr{P}_2(\mathbb{R}^d)$.

The notations $\mathbf{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$ and $\mathbf{Sol}_\mu\mathscr{P}_2(\mathbb{R}^d)$ are abbreviated $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$. The latter set is not empty, since it contains 0_μ. The rest of this subsection is devoted to algebraic properties of $\mathbf{Tan}_\mu$ (mostly quoting [Gig08]) and $\mathbf{Sol}_\mu$, fundamental in the sequel. The reader who longs for examples of solenoidal fields may prefer to read Theorem 5.2.12 below first.

Remark 5.2.5 (Terminology)**.** *In the literature of fluid dynamics, solenoidal fields are the orthogonal complement of gradient fields in L^2. Definition 5.2.4 generalizes the L^2_μ case, since $\langle f\#\mu, g\#\mu\rangle_\mu = \langle f, g\rangle_{L^2_\mu}$ for any $f, g \in L^2_\mu(\mathbb{R}^d; \mathrm{T}\mathbb{R}^d)$.*

Remark 5.2.6 (Interpretation with plans)**.** *Using the formula (5.1) for $\langle\cdot,\cdot\rangle^\pm_\mu$, (5.17) rewrites as*

$$0 = \inf_{\alpha\in\Gamma_\mu(\xi,\zeta)} \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v, w\rangle \, d\alpha = \sup_{\alpha\in\Gamma_\mu(\xi,\zeta)} \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v, w\rangle \, d\alpha = 0 \qquad \forall \xi \in \mathbf{Tan}_\mu.$$

Consequently, a measure field ζ is solenoidal if and only if for any $\xi \in \mathbf{Tan}_\mu$ and any transport plan $\alpha \in \Gamma_\mu(\xi,\zeta)$, the integral of $(x, v, w) \mapsto \langle v, w\rangle$ with respect to α vanishes.

The following consequence of Remark 5.2.6 is direct but fundamental.

Proposition 5.2.7 (All transport plans are optimal). *Let $\eta \in \mathbf{Tan}_\mu$ and $\zeta \in \mathbf{Sol}_\mu$. Then $W^2_\mu(\eta,\zeta) = \|\eta\|^2_\mu + \|\zeta\|^2_\mu$, and any transport plan $\alpha \in \Gamma_\mu(\eta,\zeta)$ realises the infimum in the definition of $W^2_\mu(\eta,\zeta)$.*

Proof. For any $\alpha \in \Gamma_\mu(\eta,\zeta)$, there holds $\int |v - w|^2 d\alpha = \|\eta\|^2_\mu + 0 + \|\zeta\|^2_\mu = W^2_\mu(\eta,\zeta)$, in which the middle term vanishes by Remark 5.2.6. Since this coincides with $\|\eta\|^2_\mu - 2\langle\eta,\zeta\rangle^+_\mu + \|\zeta\|^2_\mu = W^2_\mu(\eta,\zeta)$, the plan α is optimal. □

Let us gather Propositions 4.25, 4.29, 4.30 and 4.33 in [Gig08] as one statement.

Proposition 5.2.8 (Properties of the tangent space). *The set $\mathbf{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$*

- *is W_μ−closed and horizontally convex,*
- *is stable by scalar multiplication, i.e. $\lambda\cdot\xi \in \mathbf{Tan}_\mu$ whenever $\xi \in \mathbf{Tan}_\mu$ and $\lambda \in \mathbb{R}$,*
- *satisfies $\langle\xi,\gamma\rangle^+_\mu = \langle\pi^\mu_T\xi,\gamma\rangle^+_\mu$ for any $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ and $\gamma \in \mathbf{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$.*

Proposition 5.2.8 deserves some comment. The horizontal convexity of $\mathbf{Tan}_\mu$ is the fundamental property on which the rest is built. It is deduced from the horizontal convexity of optimal transport plans sharing the same marginal. In the case where $\eta_i = (id, \nabla\varphi_i)\#\mu$ for $i \in \{0,1\}$, and φ_i convex applications, the only convex combinations are given by $(id, (1-t)\nabla\varphi_0 + t\nabla\varphi_1)\#\mu$, which is itself induced by the gradient of the *convex* function $(1-t)\varphi_0 + t\varphi_1$. In the case of plans, one passes by c−cyclical monotonicity. The fact that any multiple of a tangent measure field by a positive scalar stays tangent is trivial; the fact that the same holds for negative scalars is, to our opinion, a deep result, particularly since it is not shared by all CBB spaces. The proof is purely algebraic, and actually goes way beyond simple scalar multiplication. The behaviour of the scalar product with respect to projections is expected, but precious.

Solenoidal measure fields share the same "algebraic" properties.

Proposition 5.2.9 (Properties of the solenoidal space). *The set $\mathbf{Sol}_\mu\mathscr{P}_2(\mathbb{R}^d)$*

- *is W_μ−closed and horizontally convex,*
- *is stable by scalar multiplication, i.e. $\lambda\cdot\xi \in \mathbf{Sol}_\mu$ whenever $\xi \in \mathbf{Sol}_\mu$ and $\lambda \in \mathbb{R}$.*

Proof. The closedness with respect to W_μ holds by continuity of $\langle\cdot,\cdot\rangle^\pm_\mu$ with respect to W_μ (see (5.2)). Let $\eta \in \mathbf{Tan}_\mu$, $\zeta_0, \zeta_1 \in \mathbf{Sol}_\mu$, and $\beta \in \Gamma_\mu(\zeta_0,\zeta_1)$. By Lemma 5.1.4, $\langle\cdot,\eta\rangle^+_\mu$ is horizontally convex and $\langle\cdot,\xi\rangle^-_\mu$ horizontally concave, so that for any $t \in [0,1]$,

$$0 = (1-t)\langle\zeta_0,\eta\rangle^-_\mu + t\langle\zeta_1,\eta\rangle^-_\mu \leqslant \langle\vec{\zeta}^\beta_t,\eta\rangle^-_\mu \leqslant \langle\vec{\zeta}^\beta_t,\eta\rangle^+_\mu \leqslant (1-t)\langle\zeta_0,\eta\rangle^+_\mu + t\langle\zeta_1,\eta\rangle^+_\mu = 0.$$

Hence $\vec{\zeta}^\beta_t$ is solenoidal. For any $\eta \in \mathbf{Tan}_\mu$, $\lambda \in \mathbb{R}$ and $\zeta \in \mathbf{Sol}_\mu$, there holds $\langle\lambda\cdot\xi,\eta\rangle^\pm_\mu = \lambda\langle\xi,\eta\rangle^{\mathrm{sign}(\lambda)}_\mu = 0$. Thus $\lambda\cdot\zeta \in \mathbf{Sol}_\mu$. □

We pursue with the first occurrence of a link between $\mathbf{Sol}_\mu$ and the variation of $d^2_{\mathcal{W}}(\cdot,\cdot)$.

Lemma 5.2.10 (Characterization of solenoidal measure fields). *The following propositions are equivalent:*

1. $\zeta \in \mathbf{Sol}_\mu$,
2. $\langle \zeta, \eta \rangle_\mu^+ = 0$ *for all* η *that is the velocity of a geodesic, that is,* $(\pi_x, \pi_x + \pi_v)\#\eta$ *is optimal between its marginals,*
3. $D_\mu d^2_{\mathcal{W}}(\cdot, \nu)(\zeta) = 0$ *for all* $\nu \in \mathscr{P}_2(\mathbb{R}^d)$,
4. $\pi_T^\mu \zeta = 0_\mu$.

Proof. The implication $1 \implies 2$ holds by definition. Assume that 2 holds. Using the formula (1.16) of the directional derivative of the squared Wasserstein distance,

$$D_\mu d^2_{\mathcal{W}}(\cdot, \nu)(\zeta) = \inf_{\eta \in \exp_\mu^{-1}(\nu)} -2 \langle \zeta, \eta \rangle_\mu^+ = 0.$$

Assume now that 3 holds for some $\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, and denote $\pi_T^\mu \zeta$ its metric projection on the tangent cone. By definition, there exists $(a_n, \eta_n)_{n \in \mathbb{N}} \subset \mathbb{R}^+ \times \mathbf{Tan}_\mu$ such that $\eta_n \in \exp_\mu^{-1}(\exp_\mu(\eta_n))$, and $W_\mu(\pi_T^\mu \zeta, a_n \cdot \eta_n) \to_n 0$. By [AGS05, Lemma 7.2.1], the unique optimal transport plan between μ and $\exp_\mu(1/2 \cdot \eta_n)$ is induced by $1/2 \cdot \eta_n$. Then

$$\langle \pi_T^\mu \zeta, \eta_n \rangle_\mu^+ = \langle \zeta, \eta_n \rangle_\mu^+ = -2 \langle \zeta, -\frac{1}{2} \cdot \eta_n \rangle_\mu^- = -D_\mu d^2_{\mathcal{W}}(\cdot, \exp_\mu(1/2 \cdot \eta_n))(\zeta) = 0.$$

Multiplying by $a_n \geqslant 0$ and using the continuity of $\langle \cdot, \cdot \rangle_\mu^+$, we get $0 = \langle \pi_T^\mu \xi, a_n \cdot \eta_n \rangle_\mu^+ \to_n \langle \pi_T^\mu \xi, \pi_T^\mu \xi \rangle_\mu = \|\pi_T^\mu \xi\|_\mu^2$. Thus 4 holds. If 4 holds, then for any $\eta \in \mathbf{Tan}_\mu$, one has $\langle \zeta, \eta \rangle_\mu^+ = \langle \pi_T^\mu \zeta, \eta \rangle_\mu^+ = 0$. Since $-\eta$ belongs to $\mathbf{Tan}_\mu$ as soon as η does by Proposition 5.2.8, there also holds $0 = \langle \zeta, -\eta \rangle_\mu^+ = -\langle \zeta, \eta \rangle_\mu^-$, and 1 holds. □

To conclude on this section, we provide a simple characterization of $\mathbf{Tan}_\mu$ by the relation with the differential of the squared distance. In essence, this tells that tangent fields are the ones that can faithfully represent this differential, at least at the limit. This is not used in the sequel, but gives a counterpart to Point 3 in Lemma 5.2.10.

Lemma 5.2.11 (Characterization of $\mathbf{Tan}_\mu$). *A measure field* $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \setminus \{0_\mu\}$ *belongs to* $\mathbf{Tan}_\mu$ *if and only if for some (thus all)* $R > 0$*, there exists* $(\mu_n)_{n \in \mathbb{N}} \subset \mathscr{P}_2(\mathbb{R}^d) \setminus \{\mu\}$ *such that*

$$\lim_{n \to \infty} \sup_{\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu, \|\zeta\|_\mu \leqslant R} \left| \langle \xi, \zeta \rangle_\mu^+ - \left(-\frac{D_\mu d^2_{\mathcal{W}}(\cdot, \mu_n)(\zeta)}{2 d_{\mathcal{W}}(\mu, \mu_n) / \|\xi\|_\mu} \right) \right| = 0. \tag{5.18}$$

Proof. Assume first that $\xi \in \mathbf{Tan}_\mu$, and let $(\xi_n)_n$ be a sequence of measure fields, each optimal on some nontrivial interval $[0, \tau_n] \subset [0,1]$, converging towards ξ with respect to W_μ. Let $\mu_n := \exp_\mu(\tau_n/2 \cdot \xi_n)$, which is distinct from μ for n large enough since $\|\xi\|_\mu > 0$. Then, the unique optimal transport plan between μ and μ_n is exactly $\frac{1}{2} \cdot \xi_n$ by [AGS05, Lemma 7.2.1], and

$$D_\mu d^2_{\mathcal{W}}(\cdot, \mu_n)(\zeta) = \langle -2 \cdot \frac{1}{2} \xi_n, \zeta \rangle_\mu^- = -\langle \xi_n, \zeta \rangle_\mu^+ \qquad \forall \zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu.$$

Let $R > 0$. As $2 d_{\mathcal{W}}(\mu, \mu_n) / \|\xi\|_\mu = 2 \|\frac{1}{2} \xi_n\|_\mu / \|\xi\|_\mu \xrightarrow[n \to \infty]{} 1$, using the estimate (5.2) on the scalar products, we get

$$\begin{aligned}
\sup_{\|\zeta\|_\mu \leqslant R} \left| \langle \xi, \zeta \rangle_\mu^+ - \left(-\frac{D_\mu d^2_{\mathcal{W}}(\cdot, \mu_n)(\zeta)}{2 d_{\mathcal{W}}(\mu, \mu_n) / \|\xi\|_\mu} \right) \right| &= \sup_{\|\zeta\|_\mu \leqslant R} \left| \langle \xi, \zeta \rangle_\mu^+ - \frac{\langle \xi_n, \zeta \rangle_\mu^+}{2 d_{\mathcal{W}}(\mu, \mu_n) / \|\xi\|_\mu} \right| \\
&\leqslant \sup_{\|\zeta\|_\mu \leqslant R} \left| \langle \xi, \zeta \rangle_\mu^+ - \langle \xi_n, \zeta \rangle_\mu^+ \right| + \left| \langle \xi_n, \zeta \rangle_\mu^+ \left(1 - \frac{1}{2 d_{\mathcal{W}}(\mu, \mu_n) / \|\xi\|_\mu} \right) \right| \\
&\leqslant R W_\mu(\xi, \xi_n) + \|\xi_n\|_\mu R \left| 1 - \frac{1}{2 d_{\mathcal{W}}(\mu, \mu_n) / \|\xi\|_\mu} \right| \xrightarrow[n \to \infty]{} 0.
\end{aligned}$$

Conversely, assume that (5.18) holds. Let $R \gg \|\xi\|_\mu$. In particular, there exists a vanishing sequence $(\varepsilon_n)_n$ such that

$$\sup_{\|\zeta\|_\mu \leqslant R} -\frac{D_\mu d^2_{\mathcal{W}}(\cdot,\mu_n)(\zeta)}{2 d_{\mathcal{W}}(\mu,\mu_n)/\|\xi\|_\mu} - \langle \xi,\zeta\rangle_\mu^+ \leqslant \varepsilon_n.$$

Recall that $D_\mu d^2_{\mathcal{W}}(\cdot,\mu_n)(\zeta) = \inf_{\eta\in\exp_\mu^{-1}(\mu_n)} -2\langle\eta,\zeta\rangle_\mu^+$ and $W_\mu^2(\xi,\zeta) = \|\xi\|_\mu^2 - 2\langle\xi,\zeta\rangle_\mu^+ + \|\zeta\|_\mu^2$. Then

$$\sup_{\|\zeta\|_\mu \leqslant R} \sup_{\eta\in\exp_\mu^{-1}(\mu_n)} \frac{\langle\eta,\zeta\rangle_\mu^+}{d_{\mathcal{W}}(\mu,\mu_n)/\|\xi\|_\mu} + \frac{1}{2}\left[W_\mu^2(\xi,\zeta) - \|\xi\|_\mu^2 - \|\zeta\|_\mu^2\right] \leqslant \varepsilon_n.$$

Restricting the supremum over $\zeta \in \frac{\|\xi\|_\mu}{d_{\mathcal{W}}(\mu,\mu_n)}\cdot\exp_\mu^{-1}(\mu_n)$, and picking $\eta = \frac{d_{\mathcal{W}}(\mu,\mu_n)}{\|\xi\|_\mu}\cdot\zeta$, we get

$$\begin{aligned}
\varepsilon_n &\geqslant \sup_{\zeta\in\frac{\|\xi\|_\mu}{d_{\mathcal{W}}(\mu,\mu_n)}\cdot\exp_\mu^{-1}(\mu_n)} \|\zeta\|_\mu^2 + \frac{1}{2}\left[W_\mu^2(\xi,\zeta) - \|\xi\|_\mu^2 - \|\zeta\|_\mu^2\right] \\
&= \|\xi\|_\mu^2 + \frac{1}{2}\left[\sup_{\zeta\in\frac{\|\xi\|_\mu}{d_{\mathcal{W}}(\mu,\mu_n)}\cdot\exp_\mu^{-1}(\mu_n)} W_\mu^2(\xi,\zeta) - \|\xi\|_\mu^2 - \|\xi\|_\mu^2\right] = \frac{1}{2}\sup_{\zeta\in\frac{\|\xi\|_\mu}{d_{\mathcal{W}}(\mu,\mu_n)}\cdot\exp_\mu^{-1}(\mu_n)} W_\mu^2(\xi,\zeta).
\end{aligned}$$

Hence the sequence of sets $\frac{\|\xi\|_\mu}{d_{\mathcal{W}}(\mu,\mu_n)}\cdot\exp_\mu^{-1}(\mu_n)$ converge towards $\{\xi\}$ in the Hausdorff distance induced by W_μ, which is more than enough to show that the latter measure field is tangent. □

5.2.2 Helmholtz-Hodge decomposition

Up to now, the solenoidal fields were defined as the elements that are orthogonal to any tangent measure field. The next result allows to construct them, and provides a decomposition of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ in "direct measure sum" of $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$.

The statement is meant to be symmetric with respect to $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$, even if the case of $\mathbf{Tan}_\mu$ was already treated. Precisely, Point 1 and (5.19) are proved in [Gig08], and the arguments of Point 2 and (5.20) follow the same vein. The new part is Point 3, which links the projections on $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$ by a sum formula that generalizes the Helmholtz-Hodge decomposition of vector fields. It also provides a way to compute the solenoidal component by removing the tangent component from ξ, that is, $\pi_S^\mu\xi = (\pi_x,\pi_v - \pi_v(T(\pi_x,\pi_v)))\#\xi$. Symmetrically, $\pi_T^\mu\xi = (\pi_x,\pi_v - \pi_v(S(\pi_x,\pi_v)))\#\xi$. We are interested into the partial Pythagoras identities of Point 4 since the full Pythagoras does not hold, as per Remark 5.2.15.

Theorem 5.2.12 (Helmholtz-Hodge decomposition). *Let $\xi\in\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. Then*

1. *there exists a unique projection of ξ on $\mathbf{Tan}_\mu$, denoted $\pi_T^\mu\xi$, and given by $T\#\xi$ for some $T\in L^2_\xi(\mathrm{T}\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ satisfying $\pi_x(T(x,v)) = x$ for $\xi-$a.e. $(x,v)\in\mathrm{T}\mathbb{R}^d$. Moreover $(\pi_x,\pi_v,\pi_v(T(\pi_x,\pi_v)))\#\xi$ is the unique element of $\Gamma_{\mu,o}(\xi,\pi_T^\mu\xi)$;*
2. *There exists a unique projection of ξ on $\mathbf{Sol}_\mu$, denoted $\pi_S^\mu\xi$, and given by $S\#\xi$ for some $S\in L^2_\xi(\mathrm{T}\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ satisfying $\pi_x(S(x,v)) = x$ for $\xi-$a.e. $(x,v)\in\mathrm{T}\mathbb{R}^d$. Moreover $(\pi_x,\pi_v,\pi_v(S(\pi_x,\pi_v)))\#\xi$ is the unique element of $\Gamma_{\mu,o}(\xi,\pi_S^\mu\xi)$;*
3. *The applications T,S satisfy $\pi_v(T(x,v)) + \pi_v(S(x,v)) = v$ for $\xi-$a.e. $(x,v)\in\mathrm{T}\mathbb{R}^d$;*
4. *The following identities hold:*

$$\langle\xi,\eta\rangle_\mu^\pm = \langle\pi_T^\mu\xi,\eta\rangle_\mu^\pm \quad \text{and} \quad W_\mu^2(\xi,\eta) = W_\mu^2(\xi,\pi_T^\mu\xi) + W_\mu^2(\pi_T^\mu\xi,\eta) \quad \forall\eta\in\mathbf{Tan}_\mu, \tag{5.19}$$

$$\langle\xi,\zeta\rangle_\mu^\pm = \langle\pi_S^\mu\xi,\zeta\rangle_\mu^\pm \quad \text{and} \quad W_\mu^2(\xi,\zeta) = W_\mu^2(\xi,\pi_S^\mu\xi) + W_\mu^2(\pi_S^\mu\xi,\zeta) \quad \forall\zeta\in\mathbf{Sol}_\mu, \tag{5.20}$$

and one has $W_\mu^2(\pi_T^\mu\xi,\pi_S^\mu\xi) = \|\pi_T^\mu\xi\|_\mu^2 + \|\pi_S^\mu\xi\|_\mu^2 = \|\xi\|_\mu^2$.

Proof. Since $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$ are both W_μ−closed and horizontally convex by Propositions 5.2.8 and 5.2.9, Points 1 and 2 follow from Propositions 5.1.11 and 5.1.12. To show Point 3, we consider the element $\omega := (\pi_x, \pi_v - \pi_v(T(\pi_x,\pi_v)))\#\xi$, prove that it is equal to $\pi^\mu_S\xi$, and that the plan $(\pi_x,\pi_v,\pi_v - \pi_v(T(\pi_x,\pi_v)))\#\xi$ is optimal between ξ and ω. Since the unique optimal transport plan between ξ and $\pi^\mu_S\xi$ is given by $(\pi_x,\pi_v,\pi_v(S(\pi_x,\pi_v)))\#\xi$, this will imply that $\pi_v - \pi_v(T(\pi_x,\pi_v)) = \pi_v(S(\pi_x,\pi_v))$ for ξ−a.e. $(x,v) \in \mathrm{T}\mathbb{R}^d$.

The plan ω is solenoidal. Let $\eta \in \mathbf{Tan}_\mu$ be a tangent measure field, and $\alpha \in \Gamma_\mu(\eta,\omega)$ an arbitrary transport plan. Denote γ the unique element of $\Gamma_{\mu,o}(\xi,\pi^\mu_T\xi)$. Using Lemma 1.1.38 p. 13, there exists $\beta = \beta(dx,du,dv,dw) \in \Gamma_\mu(\eta,\gamma) \subset \mathscr{P}_2(\mathrm{T}^3\mathbb{R}^d)$ such that $(\pi_x,\pi_u,\pi_v - \pi_w)\#\beta = \alpha$. Note that both $\eta = (\pi_x,\pi_u)\#\beta$ and $\pi^\mu_T\xi = (\pi_x,\pi_w)\#\beta$ belong to the tangent cone. The latter is stable by scalar multiplication and horizontal interpolation by Proposition 5.2.8, so $(\pi_x,\pi_w + h\pi_u)\#\beta \in \mathbf{Tan}_\mu$ for any $h \in [0,1]$. Then

$$W^2_\mu(\xi,\pi^\mu_T\xi) \leq W^2_\mu\left(\xi,(\pi_x,\pi_w + h\pi_u)\#\beta\right) \leq \int_{(x,u,v,w)\in\mathrm{T}^3\mathbb{R}^d} |v-(w-hu)|^2\, d\beta.$$

Developing and using the optimality of $\gamma = (\pi_x,\pi_v,\pi_w)\#\beta$ between ξ and $\pi^\mu_T\xi$, there holds

$$W^2_\mu(\xi,\pi^\mu_T\xi) \leq W^2_\mu(\xi,\pi^\mu_T\xi) - 2h\int_{(x,u,v,w)\in\mathrm{T}^3\mathbb{R}^d} \langle u, v-w\rangle\, d\beta + h^2\|\eta\|^2_\mu.$$

Dividing by $h > 0$ and sending h to 0,

$$0 \leq -\int_{(x,u,v,w)\in\mathrm{T}^3\mathbb{R}^d} \langle u, v-w\rangle\, d\beta = -\int_{(x,u,v)\in\mathrm{T}^2\mathbb{R}^d} \langle u, v\rangle\, d\alpha.$$

Taking the infimum over $\alpha \in \Gamma_\mu(\eta,\omega)$ and recalling that $\langle\cdot,\cdot\rangle^+_\mu$ writes as a supremum over such plans (5.1), we recover that $0 \leq -\langle\eta,\omega\rangle^+_\mu$. Repeating the argument with $-\eta$ instead of η, which belongs to $\mathbf{Tan}_\mu$ by Proposition 5.2.8, we get that $0 \leq -\langle-\eta,\omega\rangle^+_\mu = \langle\eta,\omega\rangle^-_\mu$. As $\langle\eta,\omega\rangle^-_\mu \leq \langle\eta,\omega\rangle^+_\mu \leq 0$, equality holds everywhere, and ω is solenoidal.

The plan ω minimizes the W_μ−distance. Let $\zeta \in \mathbf{Sol}_\mu$. Using the expression of ω, there holds

$$\langle\xi,\zeta\rangle^+_\mu = \sup_{\alpha\in\Gamma_\mu(\xi,\zeta)} \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v - \pi_v(T(x,v)), w\rangle\, d\alpha + \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle \pi_v(T(x,v)), w\rangle\, d\alpha.$$

As $(\pi_x,\pi_v(T(\pi_x,\pi_v)),\pi_w)\#\alpha \in \Gamma_\mu(\pi^\mu_T\xi,\zeta)$ is a transport plan between $\pi^\mu_T\xi \in \mathbf{Tan}_\mu$ and $\zeta \in \mathbf{Sol}_\mu$, the last summand vanishes by Remark 5.2.6. Then, using Lemma 1.1.38 p. 13 to get the parametrization formula $(\pi_x,\pi_v - \pi_v(T(\pi_x,\pi_v)),\pi_w)\#\Gamma_\mu(\xi,\zeta) = \Gamma_\mu(\omega,\zeta)$, there holds

$$\langle\xi,\zeta\rangle^+_\mu = \sup_{\alpha\in\Gamma_\mu(\xi,\zeta)} \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v - \pi_v(T(x,v)), w\rangle\, d\alpha = \sup_{\beta\in\Gamma_\mu(\omega,\zeta)} \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v, w\rangle\, d\beta = \langle\omega,\zeta\rangle^+_\mu. \tag{5.21}$$

Taking in particular $\zeta = \omega$ yields $\langle\xi,\omega\rangle^+_\mu = \|\omega\|^2_\mu$. Consequently, for any $\zeta \in \mathbf{Sol}_\mu$,

$$\begin{aligned} W^2_\mu(\xi,\zeta) &= \|\xi\|^2_\mu - 2\langle\xi,\zeta\rangle^+_\mu + \|\zeta\|^2_\mu = \|\xi\|^2_\mu - 2\langle\omega,\zeta\rangle^+_\mu + \|\zeta\|^2_\mu + \left[2\|\omega\|^2_\mu - 2\langle\xi,\omega\rangle^+_\mu\right] \\ &= \left[\|\xi\|^2_\mu - 2\langle\xi,\omega\rangle^+_\mu + \|\omega\|^2_\mu\right] + \left[\|\omega\|^2_\mu - 2\langle\omega,\zeta\rangle^+_\mu + \|\zeta\|^2_\mu\right] = W^2_\mu(\xi,\omega) + W^2_\mu(\omega,\zeta). \end{aligned} \tag{5.22}$$

This shows that ω realises the minimum of $W_\mu(\xi,\cdot)$ over $\zeta \in \mathbf{Sol}_\mu$, hence $\omega = \pi^\mu_S\xi$.

Optimal transport plan. On the one hand, using the equality $\langle\xi,\omega\rangle^+_\mu = \|\omega\|^2_\mu$ of the previous step,

$$W^2_\mu(\xi,\pi^\mu_S\xi) = \|\xi\|^2_\mu - 2\langle\xi,\pi^\mu_S\xi\rangle^+_\mu + \|\pi^\mu_S\xi\|^2_\mu = \|\xi\|^2_\mu - \|\pi^\mu_S\xi\|^2_\mu.$$

On the other hand, denoting $\beta := (\pi_x,\pi_v,\pi_v - \pi_v(T(\pi_x,\pi_v)))\#\xi$,

$$\int_{\mathrm{T}^2\mathbb{R}^d} |v-w|^2 d\beta = \int |v|^2 - 2\langle v-w, w\rangle - |w|^2 d\beta = \|\xi\|^2_\mu - 2\int_{\mathrm{T}\mathbb{R}^d} \langle\pi_v(T(x,v)), v - \pi_v(T(x,v))\rangle\, d\xi - \|\pi^\mu_S\xi\|^2_\mu,$$

where the middle term equals 0 since $(\pi_x, \pi_v(T(\pi_x,\pi_v)), \pi_v - \pi_v(T(\pi_x,\pi_v)))\#\xi$ is a transport plan between the tangent plan $\pi_T^\mu\xi$ and the solenoidal plan $\pi_S^\mu\xi$. Consequently, $\pi_v(S(x,v)) = v - \pi_v(T(x,v))$ for ξ−a.e. $(x,v) \in \mathrm{T}\mathbb{R}^d$, and Point 3 is proved.

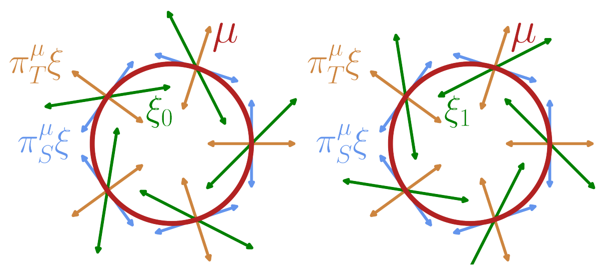
We turn to the identities of 4. The equality of metric scalar products in (5.19) is proved in [Gig08, Theorem 4.33]. In particular, $\langle \xi, \pi_T^\mu\xi\rangle_\mu^+ = \|\pi_T^\mu\xi\|_\mu^2$, and the partial Pythagoras identity follows by the same reasoning as in (5.22). We proved (5.20) in (5.21) and (5.22) since $\omega = \pi_S^\mu\xi$. The last identity is by now routine computation: removing π_v for clarity,

$$W_\mu^2(\pi_T^\mu\xi, \pi_S^\mu\xi) = \|\pi_T^\mu\xi\|_\mu^2 - 2\underbrace{\langle \pi_T^\mu\xi, \pi_S^\mu\xi\rangle_\mu^+}_{=0} + \|\pi_S^\mu\xi\|_\mu^2 = \int_{(x,v)\in \mathrm{T}\mathbb{R}^d} |T(x,v)|^2 + |S(x,v)|^2 d\xi$$

$$= \int_{(x,v)\in\mathrm{T}\mathbb{R}^d} |T(x,v) + S(x,v)|^2 - 2\langle T(x,v), S(x,v)\rangle\, d\xi = \|\xi\|_\mu^2 - 2\underbrace{\int_{\mathrm{T}\mathbb{R}^d} \langle T(x,v), S(x,v)\rangle\, d\xi}_{=0}.$$

Hence the result. □

In $L_\mu^2(\mathbb{R}^d; \mathrm{T}\mathbb{R}^d)$, vector fields are uniquely determined by their projections on tangent and solenoidal fields, since any $f \in L_\mu^2$ can be reconstructed from its projections by sum. This uniqueness does not stand in $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, and there might be several measure fields sharing the same projections on $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$.



For instance, let μ be the Hausdorff measure on the unit circle $\mathbb{S}^1 \subset \mathbb{R}^2$. Let $f, g : \mathbb{R}^2 \to \mathbb{R}^2$ be defined as $f(x,y) := (x,y)$ and $g(x,y) := (-y,x)$, and consider

$$\xi_0 := 1/2\,((id, f+g)\#\mu) + 1/2\,((id, -f-g)\#\mu),$$
$$\xi_1 := 1/2\,((id, f-g)\#\mu) + 1/2\,((id, -f+g)\#\mu).$$

Then $\pi_T^\mu\xi_i = \frac{1}{2}(id, f)\#\mu + \frac{1}{2}(id, -f)\#\mu$ for both ξ_i with $i \in \{0,1\}$, and $\pi_S^\mu\xi_i = \frac{1}{2}(id, -g)\#\mu + \frac{1}{2}(id, g)\#\mu$.

In this example, ξ_0 and ξ_1 are both represented as sums of their projections, each for a different transport plan in $\Gamma_\mu(\pi_T^\mu\xi^i, \pi_S^\mu\xi^i)$. This is made precise in the following result: using the notation $\oplus$ for the set-valued sum, the set of plans ξ admitting (η,ζ) as projections on $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$ is exactly $\eta \oplus \zeta$.

Proposition 5.2.13 (Identification by projections). *Let $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, $\eta \in \mathbf{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$ and $\zeta \in \mathbf{Sol}_\mu\mathscr{P}_2(\mathbb{R}^d)$. Then the following statements are equivalent:*

1. *$\pi_T^\mu\xi = \eta$ and $\pi_S^\mu\xi = \zeta$,*
2. *there exists $\alpha \in \Gamma_\mu(\eta,\zeta)$ such that $\xi = (\pi_x, \pi_v + \pi_w)\#\alpha$.*

Proof. Assume that the first point holds for ξ,η,ζ. Under the notations of Theorem 5.2.12, the plan $\alpha := (\pi_x, \pi_v(T(\pi_x,\pi_v)), \pi_v(S(\pi_x,\pi_v)))\#\xi$ satisfies the equality of Point 2. On the other hand, assume Point 2. Let any $\gamma \in \mathbf{Tan}_\mu$. By Lemma 1.1.38 p. 13, any plan in $\Gamma_\mu(\gamma,\xi)$ writes as $(\pi_x, \pi_u, \pi_v + \pi_w)\#\beta$ for some $\beta \in \Gamma_\mu(\gamma,\alpha)$, and any plan in $\Gamma_\mu(\gamma,\eta)$ writes as $(\pi_x, \pi_u, \pi_v)\#\beta$ for some $\beta \in \Gamma_\mu(\gamma,\alpha)$. Then

$$\langle \gamma, \xi\rangle_\mu^+ = \sup_{\beta\in\Gamma_\mu(\gamma,\alpha)} \int_{(x,u,v,w)\in\mathrm{T}^3\mathbb{R}^d} \langle u, v+w\rangle\, d\beta = \sup_{\beta\in\Gamma_\mu(\gamma,\alpha)} \int_{(x,u,v,w)\in\mathrm{T}^3\mathbb{R}^d} \langle u, v\rangle\, d\beta + 0 = \langle \gamma, \eta\rangle_\mu^+.$$

Here the integral of $\langle u, w\rangle$ against β vanishes, since $(\pi_x, \pi_u, \pi_w)\#\beta$ is a transport plan between a tangent measure field and a solenoidal measure field (see Remark 5.2.6). Hence

$$W_\mu^2(\gamma,\xi) = \|\gamma\|_\mu^2 - 2\langle\gamma,\xi\rangle_\mu^+ + \|\xi\|_\mu^2 = \|\gamma\|_\mu^2 - 2\langle \gamma,\eta\rangle_\mu^+ + \|\eta\|_\mu^2 + \|\zeta\|_\mu^2 \geqslant W_\mu^2(\gamma,\eta) + W_\mu^2(\eta,\xi),$$

and the infimum of $W_\mu^2(\gamma,\xi)$ over $\gamma \in \mathbf{Tan}_\mu$ is attained for $\gamma = \eta$. The same argument yields $\zeta = \pi_S^\mu\xi$. □

Remark 5.2.14. *Combining Proposition 5.2.13 and the equality (5.19) in Theorem 5.2.12, we get that for all $\eta \in \mathbf{Tan}_\mu$, $\xi \in \mathbf{Tan}_\mu$, $\zeta \in \mathbf{Sol}_\mu$ and $\alpha \in \Gamma_\mu(\xi,\zeta)$, there holds*

$$\langle \eta, (\pi_x, \pi_v + \pi_w)\#\alpha\rangle_\mu^\pm = \langle \eta, \xi\rangle_\mu^\pm.$$

Remark 5.2.15 (Pythagoras fails)**.** *As a consequence of Proposition 5.2.13, the formula*

$$W_\mu^2(\xi,\zeta) = W_\mu^2(\pi_T^\mu\xi,\pi_T^\mu\zeta) + W_\mu^2(\pi_S^\mu\xi,\pi_S^\mu\zeta)$$

does not hold in general. Indeed, there may be distinct ξ,ζ sharing the same projections on $\mathbf{Tan}_\mu$ *and* $\mathbf{Sol}_\mu$*, so that the left hand-side is positive while the right hand-side vanishes.*

Remark 5.2.16 (Generalization)**.** *The arguments of Theorem 5.2.12 and proposition 5.2.13 rely solely on the fact that* $\mathbf{Tan}_\mu$ *and* $\mathbf{Sol}_\mu$ *are $W_\mu-$closed horizontally convex sets which are orthogonal to each other, and not on the particular structure of tangent measure fields. In consequence, the decomposition and reconstruction formulae extend to any such pair.*

5.2.3 Horizontal spans of centred fields are vertically convex

This section applies to the centred subsets $\mathbf{Tan}_\mu^0 := \mathbf{Tan}_\mu \cap \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu^0$ and $\mathbf{Sol}_\mu^0 := \mathbf{Sol}_\mu \cap \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu^0$, and in general to each $\mathbf{Set}_\mu^0 \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu^0$ that is $W_\mu-$closed, horizontally convex and stable by multiplication by any scalar. Define

$$\mathscr{F} := \left\{ f \in L_\mu^2(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d) \;\middle|\; \frac{1}{2}\left[(-f)\#\mu + f\#\mu\right] \in \mathbf{Set}_\mu^0 \right\}. \tag{5.23}$$

By assumption on $\mathbf{Set}_\mu^0$, $\mathscr{F}$ is stable by multiplication by a scalar. It is convex as a subset of $L_\mu^2(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$, since for any $f_0, f_1 \in \mathscr{F}$ and $t \in [0,1]$, the measure field $\frac{1}{2}\left[(-(1-t)f_0 - tf_1)\#\mu + ((1-t)f_0 + tf_1)\#\mu\right]$ is the horizontal interpolation along $\beta := \frac{1}{2}\left[(\pi_x, -\pi_v(f_0(\pi_x)), -\pi_v(f_1(\pi_x)))\#\mu + (\pi_x, \pi_v(f_0(\pi_x)), \pi_v(f_1(\pi_x)))\#\mu\right]$. It is also closed, since $W_\mu^2\left(\frac{1}{2}\left[(-f_0)\#\mu + f_0\#\mu\right], \frac{1}{2}\left[(-f_1)\#\mu + f_1\#\mu\right]\right) \leqslant \|f_0 - f_1\|_{L_\mu^2}^2$ by vertical convexity of W_μ^2.

Our aim is to represent $\mathbf{Set}_\mu^0$ by a span of the elements of $\mathscr{F}$, in the following sense.

Definition 5.2.17 (Span and closed span). *The horizontal span of $\mathcal{A} \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ is the set*

$$\operatorname{span}\mathcal{A} := \left\{ \left(\pi_x, \sum_{i=1}^N \lambda_i \pi_{v_i}\right)\#\alpha \;\middle|\; \begin{array}{c} N \in \mathbb{N}_*,\ (\lambda_i)_{i=1}^N \subset \mathbb{R},\ \textit{and} \\ \alpha \in \mathscr{P}_2\left(\mathrm{T}^N\mathbb{R}^d\right)_\mu \ \textit{with}\ (\pi_x, \pi_{v_i})\#\alpha \in \mathcal{A}\ \textit{for all}\ i \in [\![1, N]\!]. \end{array} \right\}.$$

The closed horizontal span $\overrightarrow{\operatorname{span}}\mathcal{A}$ is defined as the closure of $\operatorname{span}\mathcal{A}$ *in $\left(\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu, W_\mu\right)$.*

The interest of the form $\frac{1}{2}\left[(-f)\#\mu + f\#\mu\right]$ is that orthogonality with respect to the metric scalar product in $\mathscr{P}_2(\mathbb{R}^d)$ implies orthogonality $\mu-$almost everywhere.

Lemma 5.2.18 (Orthogonality with respect to a simple field). *A centred measure field $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu^0$ is orthogonal to $\zeta := \frac{1}{2}\left[(-f)\#\mu + f\#\mu\right]$ for $f \in L_\mu^2(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ if and only if $\langle v, f(x)\rangle = 0$ for $\xi-$almost all $(x,v) \in \mathrm{T}\mathbb{R}^d$.*

Proof. It is clear that if $\langle v, f(x)\rangle$ vanishes $\xi-$almost everywhere, then $\langle \xi,\zeta\rangle_\mu^+ = 0$. On the other hand, if $\langle \xi,\zeta\rangle_\mu^+ = 0$, then $\langle \xi,\zeta\rangle_\mu^- = 0$ since $-\zeta = (\pi_x, -\pi_v)\#\zeta = \zeta$. Consequently, all transport plans $\alpha \in \Gamma_\mu(\xi,\zeta)$ must satisfy $\int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v, w\rangle\, d\alpha = 0$; our strategy is to construct such a transport plan for which the integral does not vanish.

Let $\alpha \in \Gamma(\xi,\zeta)$, and split it into $\frac{1}{2}((\pi_x, -f(\pi_x), \pi_v)\#\xi_- + (\pi_x, f(\pi_x), \pi_v)\#\xi_+)$, where $\frac{1}{2}[\xi_- + \xi_+] = \xi$. Then

$$\begin{aligned} 0 = 2\int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v, w\rangle\, d\alpha &= \int_{(x,v)\in\mathrm{T}\mathbb{R}^d} \langle -f(x), v\rangle\, d\xi_- + \frac{1}{2}\int_{(x,v)\in\mathrm{T}\mathbb{R}^d} \langle f(x), v\rangle\, d\xi_+ \\ &= \underbrace{\int_{\langle -f(x),v\rangle<0} \langle -f(x), v\rangle\, d\xi_-}_{A_{--}} + \underbrace{\int_{\langle -f(x),v\rangle>0} \langle -f(x), v\rangle\, d\xi_-}_{A_{-+}} + \underbrace{\int_{\langle f(x),v\rangle<0} \langle f(x), v\rangle\, d\xi_+}_{A_{+-}} + \underbrace{\int_{\langle f(x),v\rangle>0} \langle f(x), v\rangle\, d\xi_+}_{A_{++}}. \end{aligned}$$

If one of $A_{--}, A_{+-} \leqslant 0$ is nonzero, then one of $A_{-+}, A_{++} \geqslant 0$ is nonzero as well, since the sum vanishes. We discuss the possible cases. If $A_{+-} = A_{++} = 0$, then ξ_+ is concentrated on the (x,v) such that $\langle v, f(x)\rangle = 0$. In this case, if $A_{--} < 0 < A_{-+}$, construct a plan between ξ and ζ by redirecting some of the mass that ξ_-

puts of $\langle -f(x), v\rangle < 0$ towards f, and the same amount of mass of ξ_+ towards $-f$. The first redirection will strictly increase A_{--}, and the second will produce a null term, and this is absurd. The case where $A_{--} = A_{-+} = 0$ is symmetric. If now compensation occurs between ξ_- and ξ_+, for instance if $A_{--} < 0 < A_{++}$, then both ξ_- and ξ_+ put mass on the half-spaces $\{\langle v, f(x)\rangle > 0\}$. Since ξ is centred, it must put equal mass on $\{\langle v, f(x)\rangle < 0\}$, so that all terms $A_{\pm\pm}$ are nonzero. Then, redirecting some mass that ξ_- puts on $\{\langle v, -f(x)\rangle < 0\}$ to $f(x)$, and the same amount of mass from ξ_+ restricted to $\{\langle v, f(x)\rangle < 0\}$ to $-f(x)$, we strictly increase the values of A_{--} and A_{+-}, and get a contradiction. In consequence, $A_{\pm\pm} = 0$, and $\langle v, f(x)\rangle$ must vanish $\xi-$almost everywhere. □

The key step is to show that any $\zeta \in \mathbf{Set}^0_\mu$ is aligned with an element of $\mathscr{F}$, in the following sense.

Lemma 5.2.19. *Let $\zeta \in \mathbf{Set}^0_\mu$. There exists $f \in \mathscr{F}$ such that for any measurable $A \subset \mathbb{R}^d$ on which $\|\zeta_x\| > 0$ $\mu-$a.e., there holds $|f(x)| > 0$ $\mu-$almost everywhere. Moreover, if $\|\zeta\|_\mu > 0$, then $\langle \zeta, \frac{1}{2}\left[(-f)\#\mu + f\#\mu\right]\rangle^+_\mu > 0$.*

Proof. Disintegrate ζ in $\zeta_x \otimes \mu$ for a measurable family $(\zeta_x)_{x\in\mathbb{R}^d}$, and let $A := \left\{x \in \mathbb{R}^d \mid \|\zeta_x\| > 0\right\}$. We claim that there exists $g \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ such that, denoting $\xi := \frac{1}{2}\left[(-g)\#\mu + g\#\mu\right]$, there holds $\langle \zeta_x, \xi_x\rangle^+_{\delta_x} > 0$ for $\mu-$almost every $x \in A$. Consider $(e_1, \cdots, e_d)$ the basis of $\mathbb{R}^d$, and let $g^j := \frac{1}{2}\left[(\pi_x, -e_j)\#\mu + (\pi_x, e_j)\#\mu\right]$. Let A^1 be a measurable set on which $\langle \zeta_x, g^1_x\rangle^+_{\delta_x} > 0$ for $\mu-$almost every $x \in A^1$, then A^2 be a measurable set such that $\langle \zeta_x, g^2_x\rangle^+_{\delta_x} > 0$ for $\mu-$a.e. $x \in A^2 \setminus A^1$, and so on. Let $B := A \setminus \bigcup_{j=1}^d A^j$. If $\mu(B) > 0$, then by Lemma 5.2.18, $(x, v) = (x, 0)$ for $\zeta-$almost any (x, v) with $x \in B$, in contradiction with the construction of A. So $g := \sum_{i=1}^d e_i \mathbb{I}_{A^i}$ fills the claim.

Let $\alpha = \alpha(dx, dv, dw) \in \Gamma_{\mu,o}(\zeta, \xi)$. Split $\alpha = \frac{1}{2}(\pi_x, \pi_v, -g(\pi_x))\#\zeta_- + \frac{1}{2}(\pi_x, \pi_v, g(\pi_x))\#\zeta_+$, where $\frac{1}{2}\zeta_- + \frac{1}{2}\zeta_+ = \zeta$. There holds $0 = \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\zeta) = \frac{1}{2}\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\zeta_-) + \frac{1}{2}\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\zeta_+)$. Since the barycenters are induced by maps, we must have $f := \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\zeta_+) = -\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\zeta_-)$ in $L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$. Moreover, for $\mu-$a.e. $x \in A$,

$$0 < \langle \zeta_x, \xi_x\rangle = \frac{1}{2}\int_{v\in\mathrm{T}_x\mathbb{R}^d} \langle v, -g(x)\rangle\, d(\xi_-)_x + \frac{1}{2}\int_{v\in\mathrm{T}_x\mathbb{R}^d} \langle v, g(x)\rangle\, d(\xi_+)_x = \frac{1}{2}\langle -f(x), -g(x)\rangle + \frac{1}{2}\langle f(x), g(x)\rangle,$$

so that $\left|f(x)\right|$ is positive for $\mu-$almost every $x \in A$. Denote $\gamma := \frac{1}{2}\left[(-f)\#\mu + f\#\mu\right]$. Using the transport plan $\frac{1}{2}\left[(\pi_x, \pi_v, -f(\pi_x))\#\zeta_- + (\pi_x, \pi_v, f(\pi_x))\#\zeta_+\right]$, there holds $\langle \zeta, \gamma\rangle^+_\mu \geqslant \|f\|^2_{L^2_\mu}$, which is positive as soon as $\mu(A) > 0$. To prove that $\gamma \in \mathbf{Set}^0_\mu$, we show that it is reached as the limit of horizontal convex combinations of ζ, in a way that is similar to the argument of Lemma 5.1.13. Let $\beta \in \Gamma_\mu(\zeta, \zeta)$ be given by $\frac{1}{2}\left[\zeta_- \otimes_x \zeta_- + \zeta_+ \otimes_x \zeta_+\right]$, where the pointwise product measure is defined by $\omega \otimes_x \omega := \int_{x\in\mathbb{R}^d} \omega_x \otimes \omega_x d\mu(x)$ for any $\omega = \omega_x \otimes \mu \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. The field $\left(\pi_x, \frac{\pi_v + \pi_w}{2}\right)\#\beta$ belongs to $\mathbf{Set}^0_\mu$, and its two components are respectively closer to $(-f)\#\mu$ and $f\#\mu$ than ζ_-, ζ_+. Indeed, using that $\left|f(x) - \frac{v+w}{2}\right|^2 = \frac{1}{4}|f(x) - v|^2 + \frac{1}{4}|f(x) - w|^2 - \frac{1}{2}\langle f(x) - v, f(x) - w\rangle$ and the definition of barycenter,

$$\begin{aligned} W^2_\mu\left(f\#\mu, \left(\pi_x, \frac{\pi_v + \pi_w}{2}\right)\#\left[\zeta_+ \otimes_x \zeta_+\right]\right) &\leqslant \int_{(x,v,w)\in\mathrm{T}\mathbb{R}^d} \left|f(x) - \frac{v+w}{2}\right|^2 d\left[\zeta_+ \otimes_x \zeta_+\right] \\ &\leqslant \frac{1}{4}W^2_\mu\left(f\#\mu, \zeta_+\right) + \frac{1}{4}W^2_\mu(f\#\mu, \zeta_+) - 0. \end{aligned}$$

The same estimate holds for $(-f)\#\mu$ and $\left(\pi_x, \frac{\pi_v+\pi_w}{2}\right)\#\left[\zeta_- \otimes_x \zeta_-\right]$. The barycenter of $\left(\pi_x, \frac{\pi_v+\pi_w}{2}\right)\#\left[\zeta_\pm \otimes_x \zeta_\pm\right]$ is again $(\pm f)\#\mu$, so iterating, we obtain a sequence of measure fields of $\mathbf{Set}^0_\mu$ converging with respect to W_μ towards γ, which must belongs to $\mathbf{Set}^0_\mu$. We conclude that $f \in \mathscr{F}$. □

Proposition 5.2.20 (Basis of $\mathbf{Set}^0_\mu$). *There holds $\mathbf{Set}^0_\mu = \overrightarrow{\mathrm{span}}\left\{\frac{1}{2}[(-f)\#\mu + f\#\mu] \mid f \in \mathscr{F}\right\}$, and $\mathbf{Set}^0_\mu$ is vertically convex.*

Proof. Denote $\mathcal{A} := \left\{\frac{1}{2}[(-f)\#\mu + f\#\mu] \mid f \in \mathscr{F}\right\}$. The inclusion $\overrightarrow{\mathrm{span}}\,\mathcal{A} \supset \mathbf{Set}^0_\mu$ holds by assumption on $\mathbf{Set}^0_\mu$. Conversely, let $\zeta \in \mathbf{Set}^0_\mu \setminus \overrightarrow{\mathrm{span}}\,\mathcal{A}$. The set

$$C := \mathbf{Set}^0_\mu \cap \left\{\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \;\middle|\; \langle \xi, \gamma\rangle^+_\mu = 0 \text{ for all } \gamma \in \mathcal{A}\right\}$$

is W_μ−closed by continuity of the metric scalar product. Note that any $\gamma \in \mathcal{A}$ satisfies $-\gamma = (\pi_x, -\pi_v)\#\gamma = \gamma$, so that $0 = \langle \xi, \gamma \rangle_\mu^+ = -\langle \xi, \gamma \rangle_\mu^-$. Hence C is horizontally convex by the convexity of $\langle \cdot, \cdot \rangle_\mu^+$ and the concavity of $\langle \cdot, \cdot \rangle_\mu^-$ given in Lemma 5.1.4. By Proposition 5.2.13 (see Remark 5.2.16), ζ writes as $(\pi_x, \pi_v + \pi_w)\#\alpha$ for some transport plan $\alpha \in \mathscr{P}_2(\mathrm{T}^2\mathbb{R}^d)_\mu$ between the metric projections of ζ on C and $\overrightarrow{\mathrm{span}}\mathcal{A}$. If the metric projection $\pi_C^\mu \zeta$ of ζ on C is not equal to 0_μ, then Lemma 5.2.19 provides an element $f \in \mathscr{F}$ such that $\langle \pi_C^\mu \zeta, \frac{1}{2}\left[(-f)\#\mu + f\#\mu\right]\rangle_\mu^+ > 0$, which is absurd. Hence $\zeta \in \overrightarrow{\mathrm{span}}\mathcal{A}$.

Let $(f_n)_{n\in\mathbb{N}} \subset \mathscr{F}$ be dense in $\mathscr{F}$. The previous point shows that any $\zeta \in \mathbf{Set}_\mu^0$ satisfies $v \in \mathrm{span}\{f_n(x) \mid n \in \mathbb{N}\}$ for ζ−a.e. $(x, v) \in \mathrm{T}\mathbb{R}^d$. Conversely, if some centred ξ does not satisfy the previous equality, let $\alpha := (\pi_x, p(\pi_v), \pi_v - q(\pi_v))\#\xi$, where $p \in L_\mu^2(\mathbb{R}^d; \mathbb{R}^d)$ is the projection on $\mathrm{span}\{f_n(x) \mid n \in \mathbb{N}\}$. By explicit computation, $W_\mu^2(\zeta, \xi) = W_\mu^2(\gamma, (\pi_x, \pi_v)\#\alpha) + \|(\pi_x, \pi_w)\#\alpha\|_\mu^2$ for any $\zeta \in \overrightarrow{\mathrm{span}}\mathcal{A}$. Since $\|(\pi_x, \pi_w)\#\alpha\|_\mu^2 > 0$ by assumption, we get that $\xi \notin \mathbf{Set}_\mu^0$. In consequence, the pointwise condition $v \in \mathrm{span}\{f_n(x)\}_n$ for ζ−a.e. (x, v) characterizes $\zeta \in \mathbf{Set}_\mu^0$, and passes to vertical convex combinations. □

5.2.4 The case of dimension 1

In dimension 1, we may completely describe tangent and solenoidal measure fields. To this aim, we introduce the class of *transport-regular measures*, borrowing a well-chosen terminology from Juillet [Jui11].

Definition 5.2.21 (Transport-regular measure). *A measure $\mu \in \mathscr{P}_2(\mathbb{R})$ is transport-regular if for any $\nu \in \mathscr{P}_2(\mathbb{R})$, the set $\exp_\mu^{-1}(\nu)$ is reduced to a singleton whose element is induced by a map.*

The characterization of transport-regular measures was the central question of the theory of optimal transportation, starting with Monge, the Kantorovich reformulation, and solved in two successive steps: the Brenier-McCann theorem, that states that all absolutely continuous measures in $\mathscr{P}_2(\mathbb{R})$ are transport-regular [Bre91; McC01], and the exact characterization of [Gig11]. Here we state only the one-dimensional version.

Proposition 5.2.22 (Transport-regular measure in dimension 1 [Gig11]). *A measure $\mu \in \mathscr{P}(\mathbb{R})$ is transport-regular if and only if it is atomless.*

As a consequence, we may decompose any measure $\mu \in \mathscr{P}_2(\mathbb{R})$ as a sum of mutually singular components $\mu = m_a\mu^a + m_d\mu^d$, with $0 \leq m_a, m_d$, $m_a + m_d = 1$, μ^a totally atomic and μ^d diffuse (atomless) [AFP00]. Let $A \subset \mathbb{R}$ be the set of atoms of μ, which is at most countable, and $\mathrm{T}A := \{(x, v) \in \mathrm{T}\mathbb{R} \mid x \in A\}$. We adopt the following convention for measure fields with base μ: the notation $\xi = m_a\xi^a + m_d\xi^d$ refers to the decomposition in which $m_a\xi^a(B) = \xi(B \cap \mathrm{T}A)$, and $m_d\xi^d(B) = \xi(B \setminus \mathrm{T}A)$, for any measurable $B \subset \mathrm{T}\mathbb{R}$.

Theorem 5.2.23 (Tangent and solenoidal fields in dimension 1). *Let $\mu = m_a\mu^a + m_d\mu^d \in \mathscr{P}_2(\mathbb{R})$.*

- *$\xi \in \mathbf{Tan}_\mu$ if and only if $\xi = m_a\xi^a + m_d\xi^d$ with $\xi^d = f^d\#\mu^d$ for some $f^d \in L_{\mu^d}^2(\mathbb{R}; \mathrm{T}\mathbb{R})$,*
- *$\zeta \in \mathbf{Sol}_\mu$ if and only if $\zeta = m_a 0_{\mu^a} + m_d\zeta^d$ with $\mathrm{Bary}_{\mathrm{T}\mathbb{R}}\left(\zeta^d\right) = 0$ in $L_{\mu^d}^2(\mathbb{R}; \mathrm{T}\mathbb{R})$.*

Proof. Consider first that $\eta \in \mathbf{Tan}_\mu$ is the velocity of a geodesic issued from μ. By the restriction of optimality [Vil09, Theorem 4.6], $\eta = m_a\eta^a + m_d\eta^d$, with $\eta^a \in \mathscr{P}_2(\mathrm{T}\mathbb{R})_{\mu^a}$ and $\eta^d \in \mathscr{P}_2(\mathrm{T}\mathbb{R})_{\mu^d}$ also velocities of geodesics. The measure μ^d being transport-regular, η^d is induced by a map f^d. If $\xi = \lambda \cdot \eta$ for some $\lambda > 0$, then ξ^d is induced by the map λf^d. Assume now that $(\xi_n)_{n\in\mathbb{N}} \subset \mathbf{Tan}_\mu$ is a Cauchy sequence with respect to W_μ, each ξ_n being optimal on a nontrivial interval. By the previous steps, each ξ_n decomposes as $m_a\xi_n^a + m_d f_n^d\#\mu^d$. The measures μ^a and μ^d being mutually singular, one has

$$W_\mu^2(\xi_n, \xi_m) = m_a W_{\mu^a}^2\left(\xi_n^a, \xi_m^a\right) + m_d W_{\mu^d}^2\left(f_n^d\#\mu^d, f_m^d\#\mu^d\right) = m_a W_{\mu^a}^2\left(\xi_n^a, \xi_m^a\right) + m_d\|f_n^d - f_m^d\|_{\mu^d}^2.$$

Hence, if $m_d > 0$, the sequence $(f_n^d)_{n\in\mathbb{N}}$ is Cauchy in the complete space $L_\mu^2(\mathbb{R}; \mathrm{T}\mathbb{R})$. Since $W_\mu^2(\xi, \xi_n) = m_a W_{\mu^a}(\xi^a, \xi_n^a) + m_d W_{\mu^d}^2(\xi^d, \xi_n^d)$, letting $n \to \infty$, we deduce that ξ^d is induced by the limiting map.

Assume now that $\xi = m_a\xi^a + m_d f^d\#\mu$ for some $\xi^a \in \mathscr{P}_2(\mathrm{T}\mathbb{R})_{\mu^a}$ and $f^d \in L_{\mu^d}^2$. Let $\varepsilon > 0$.

- First let $M > 0$ be large enough so that the bounded measure field $\xi_1 := (\pi_x, \pi_v / \max(1, |\pi_v|/M))\#\xi$ approximates ξ with $W_\mu(\xi, \xi_1) \leqslant \varepsilon/2$. Denote $f_1^d \in L^2_{\mu^d}(\mathbb{R}; \mathrm{T}\mathbb{R})$ the map such that $\xi_1^d = f_1^d \# \mu^d$.
- Let $(x_n)_{n\in\mathbb{N}}$ be the atoms of μ, and pick n_0 large enough so that $\mu(\cup_{n>n_0}\{x_n\}) \leqslant \frac{(\varepsilon/2)^2}{3(2M)^2}$. Relabel the first n_0 atoms in increasing order. Since μ is outer regular, its mass given to a decreasing family of measurable sets with empty intersection converges to 0. Hence there exists $0 < r_{\varepsilon,n_0}$ small enough so that $\sum_{n\leqslant n_0} \mu\left(\mathscr{B}(x_n, r_{\varepsilon,n_0}) \setminus \{x_n\}\right) \leqslant \frac{(\varepsilon/2)^2}{3(2M)^2}$, and $x_n + r_{\varepsilon,n_0} < x_{n+1} - r_{\varepsilon,n_0}$ for $n < n_0$.
- Let $\widetilde{\mu}^d$ be the restriction of μ^d to $\mathbb{R} \setminus \cup_{n\leqslant n_0}\mathscr{B}(x_n, r_{\varepsilon,n_0})$. By classical density results [Bog07, Corollary 4.2.2], the restriction of f_1^d to the same set can be approximated in $L^2_{\widetilde{\mu}^d}(\mathbb{R}; \mathrm{T}\mathbb{R})$ by a Lipschitz function $\varphi_1^d : \mathbb{R} \setminus \bigcup_{n\leqslant n_0} \mathscr{B}(x_n, r_{\varepsilon,n_0}) \to \mathrm{T}\mathbb{R}$, itself bounded by M, with $L^2_{\widetilde{\mu}^d}$−error inferior to $\frac{\varepsilon/2}{\sqrt{3}}$.
- Let $\tau := r_{\varepsilon,n_0}/(2M)$. By construction, $\tau \cdot \varphi_1^d$ is bounded by $r_{\varepsilon,n_0}/2$ on its domain of definition. Extend it to $\mathbb{R} \setminus \cup_{n\leqslant n_0}\{x_n\}$ so as to let $x + \tau\varphi_1^d(x) = x_n - r_{\varepsilon,n_0}/2$ for $x \in (x_n - r_{\varepsilon,n_0}, x_n)$, and $x + \tau\varphi_1^d(x) = x_n + r_{\varepsilon,n_0}/2$ for $x \in (x_n, x_n + r_{\varepsilon,n_0})$, for all $n \leqslant n_0$. The extension is still bounded by M.
- Denote $(\xi_1^a)_{x_n}$ the disintegration of ξ_1^a on each x_n for $n \leqslant n_0$, unambiguous since x_n is an atom. Define
$$\xi_2 := \varphi_1^d \# \Big(\mu^d + \sum_{n>n_0} \mu\{x_n\}\delta_{x_n}\Big) + \sum_{n\leqslant n_0} \mu\{x_n\}(\xi_1^a)_{x_n}.$$
For $0 < s \leqslant \tau$ such that $s \cdot \varphi_1^d$ is 1-Lipschitz, $s \cdot \xi_2$ is the optimal velocity of a geodesic. Indeed, for any $x_1, x_2 \in \mathbb{R}$ and $y_1, y_2 \in \mathbb{R}$ such that $(x_1, y_1), (x_2, y_2) \in \operatorname{supp}(\pi_x, \pi_x + s\pi_v)\#\xi_2$, one has $(x_2 - x_1)(y_2 - y_1) \geqslant 0$, and this is sufficient in dimension 1 [San15, Lemma 2.8]. In addition,
$$W_\mu^2(\xi_1, \xi_2) \leqslant \|f_1^d - \varphi_1^d\|^2_{L^2_{\widetilde{\mu}^d}} + (2M)^2 \sum_{n\leqslant n_0} \mu\left(\mathscr{B}(x_n, r_{\varepsilon,n_0}) \setminus \{x_n\}\right) + (2M)^2 \sum_{n>n_0} \mu\{x_n\} \leqslant (\varepsilon/2)^2.$$
By the triangular inequality, $W_\mu(\xi, \xi_2) \leqslant \varepsilon$, and ε being arbitrary, the measure field ξ is tangent.

We turn to solenoidal measure fields. If $\zeta = m_a\zeta^a + m_d\zeta^d$ for $\zeta^a \in \mathscr{P}_2(\mathrm{T}\mathbb{R})_{\mu^a}$ and $\zeta^d \in \mathscr{P}_2(\mathrm{T}\mathbb{R})_{\mu^d}$, then for any $\xi^a \in \mathscr{P}_2(\mathrm{T}\mathbb{R})_{\mu^a}$ and $f^d \in L^2_{\mu^d}(\mathbb{R}; \mathrm{T}\mathbb{R})$, there holds
$$\langle \zeta, m_a\xi^a + m_d f^d \# \mu^d \rangle_\mu^+ = m_a \langle \zeta^a, \xi^a \rangle^+_{\mu^a} + m_d \left\langle \mathrm{Bary}_{\mathrm{T}\mathbb{R}}\left(\xi^d\right), f^d \right\rangle_{L^2_{\mu^d}}.$$
By the previous step, $\zeta \in \mathbf{Sol}_\mu$ if and only if the above vanishes for all ξ^a and f^d, which happens if and only if $\zeta^a = 0_\mu$ and $\mathrm{Bary}_{\mathrm{T}\mathbb{R}}\left(\xi^d\right) = 0$. □

An important consequence of the previous decomposition is the following closedness result.

Theorem 5.2.24 ($d_{\mathcal{W},\mathrm{T}\mathbb{R}}$−closedness of solenoidal fields in dimension 1). *The set* $\mathbf{Sol}_\mu$ *is closed with respect to the Wasserstein topology on* $\mathscr{P}_2(\mathrm{T}\mathbb{R})$.

This is specific to solenoidal fields: the set $\mathbf{Tan}_\mu$ is not $d_{\mathcal{W},\mathrm{T}\mathbb{R}}$−closed in general. For instance, let μ be the Lebesgue measure on $[0,1]$, and f^n take alternatively the values -1 and 1 on intervals of size e^{-n}. Each $f^n\#\mu$ is tangent, but their $d_{\mathcal{W},\mathrm{T}\mathbb{R}}$−limit $\frac{1}{2}(id, -1)\#\mu + \frac{1}{2}(id, 1)\#\mu$ is solenoidal according to Theorem 5.2.23.

Proof. Let $(\zeta_n)_{n\in\mathbb{N}} \subset \mathbf{Sol}_\mu$ be a Cauchy sequence with respect to $d_{\mathcal{W},\mathrm{T}\mathbb{R}}$, and denote $\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R})_\mu$ its limit. By Theorem 5.2.23, $\mathrm{Bary}_{\mathrm{T}\mathbb{R}}(\zeta_n) = 0$ for all n. Let $\varphi \in \mathcal{C}(\mathrm{T}\mathbb{R}; \mathbb{R})$ be linear with respect to its second argument and satisfy $|\varphi(x, v)| \leqslant C(1 + |x|^2 + |v|^2)$ for some constant C. The convergence with respect to $d_{\mathcal{W},\mathrm{T}\mathbb{R}}$ is equivalent to the convergence of the integrals against continuous and quadratically growing maps [Vil09, Definition 6.8], so that
$$\int_{(x,v)\in\mathrm{T}\mathbb{R}} \varphi(x, v)\, d\zeta = \lim_{n\to\infty} \int_{(x,v)\in\mathrm{T}\mathbb{R}} \varphi(x, v)\, d\zeta_n = \lim_{n\to\infty} \int_{x\in\mathbb{R}} \varphi(x, \mathrm{Bary}_{\mathrm{T}\mathbb{R}}(\zeta_n)(x))\, d\mu = 0.$$
This shows that $\mathrm{Bary}_{\mathrm{T}\mathbb{R}}(\zeta) = 0$. Let now $(x_m)_{m\in I} \subset \mathbb{R}$ be the set of atoms of μ, with $I \subset \mathbb{N}$. Let $\varepsilon > 0$ and $(\ell_m)_{m\in I} \subset [0,1]$ be a sequence summing to 1. For each m, the nested family $(\mathscr{B}(x_m, r) \setminus \{x_m\})_{r>0}$ has

empty intersection, so there exists $r_{\varepsilon,m} > 0$ small enough so that $\mu\left(\mathscr{B}(x_m, r_{\varepsilon,m}) \setminus \{x_m\}\right) \leqslant \varepsilon \ell_m$. Up to taking minima, we may assume that $r_{\varepsilon,m}$ decreases with ε. For any $\lambda \in (0,1)$, the set

$$B_{\varepsilon,\lambda} := \left\{(x,v) \in \mathrm{T}\mathbb{R} \;\middle|\; x \notin \cup_m \mathscr{B}_{\mathbb{R}}(x_m, r_{\varepsilon,m})\right\} \cup \bigcup_{m \in I} \overline{\mathscr{B}}_{\mathrm{T}\mathbb{R}}((x_m, 0), \lambda r_{\varepsilon,m})$$

is closed, and satisfies $\mu\left(B_{\varepsilon,\lambda}\right) \geqslant 1 - \sum_{m \in I} \mu(\mathscr{B}(x_m, r_{\varepsilon,m}) \setminus \{x_m\}) \geqslant 1 - \varepsilon$. Moreover, by Theorem 5.2.23, ζ_n puts mass on (x_m, v) only if $v = 0$. Using that $B_{\varepsilon,\lambda}$ is the union of two disjoint components,

$$\begin{aligned}
\zeta_n(B_{\varepsilon,\lambda}) &= \zeta_n \left\{(x,v) \in \mathrm{T}\mathbb{R} \;\middle|\; x \notin \cup_m \mathscr{B}_{\mathbb{R}}(x_m, r_{\varepsilon,m})\right\} + \zeta_n \left(\bigcup_{m \in I} \overline{\mathscr{B}}_{\mathrm{T}\mathbb{R}}((x_m,0), \lambda r_{\varepsilon,m})\right) \\
&\geqslant \mu\left\{x \notin \cup_m \mathscr{B}(x_m, r_{\varepsilon,m})\right\} + \sum_{m \in I} \zeta_n\{(x_m, 0)\} = 1 - \sum_{m \in I} \mu(\mathscr{B}(x_m, r_{\varepsilon,m}) \setminus \{x_m\}) \geqslant 1 - \varepsilon.
\end{aligned}$$

Since $B_{\varepsilon,\lambda}$ is closed, there holds $\zeta(B_{\varepsilon,\lambda}) \geqslant \limsup_{n\to\infty} \zeta_n(B_{\varepsilon,\lambda}) \geqslant 1-\varepsilon$. Taking first the monotone limit along a vanishing sequence of $(\lambda_j)_j$, then the monotone limit along a vanishing sequence of $(\varepsilon_k)_k$, we deduce

$$\zeta\left(\cap_k \cap_j B_{\varepsilon_k,\lambda_j}\right) = \lim_{k\to\infty} \lim_{j\to\infty} \zeta(B_{\varepsilon_k,\lambda_j}) = 1.$$

The set $\cap_k \cap_j B_{\varepsilon_k,\lambda_j}$ is the complementary in $\mathrm{T}\mathbb{R}$ of the points (x_m, v) for $v \neq 0$, which shows that the disintegration of ζ on atoms of μ is concentrated on $v = 0$. By Theorem 5.2.23, ζ is solenoidal. □

5.3 Classification of $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$ by directional derivatives of $d_{\mathcal{W}}(\mu,\cdot)$

We start this section with the following observation:

$$\xi \in \mathbf{Tan}_\mu \mathscr{P}_2(\mathrm{T}\mathbb{R}^d) \qquad \Longrightarrow \qquad \lim_{h \searrow 0} \frac{d_{\mathcal{W}}(\mu, \exp_\mu(h\cdot\xi))}{h} = \|\xi\|_\mu, \tag{5.24a}$$

$$\zeta \in \mathbf{Sol}_\mu \mathscr{P}_2(\mathrm{T}\mathbb{R}^d) \qquad \Longleftarrow \qquad \lim_{h \searrow 0} \frac{d_{\mathcal{W}}(\mu, \exp_\mu(h\cdot\zeta))}{h} = 0. \tag{5.24b}$$

Indeed, any $\xi \in \mathbf{Tan}_\mu$ is the W_μ−limit of a sequence $(\xi_n)_{n\in\mathbb{N}}$ for which $d_{\mathcal{W}}(\mu, \exp_\mu(h\cdot\xi_n)) = h\|\xi_n\|_\mu$ for all $h \in [0, h_n]$, with $h_n > 0$; since $d_{\mathcal{W}}(\exp_\mu(h\cdot\xi), \exp_\mu(h\cdot\xi_n)) \leqslant h W_\mu(\xi, \xi_n)$, we get

$$\begin{aligned}
\|\xi\|_\mu = \lim_{n\to\infty} \|\xi_n\|_\mu = \lim_{n\to\infty}\lim_{h\searrow 0} \frac{d_{\mathcal{W}}(\mu, \exp_\mu(h\cdot\xi_n))}{h} &\leqslant \lim_{n\to\infty} \liminf_{h\searrow 0} \frac{d_{\mathcal{W}}(\mu, \exp_\mu(h\cdot\xi))}{h} + W_\mu(\xi,\xi_n) \\
= \liminf_{h\searrow 0} \frac{d_{\mathcal{W}}(\mu, \exp_\mu(h\cdot\xi))}{h},
\end{aligned}$$

and as the inequality $\frac{d_{\mathcal{W}}(\mu,\exp_\mu(h\cdot\xi))}{h} \leqslant \|\xi\|_\mu$, we conclude to (5.24a). Conversely, for any $\nu \in \mathscr{P}_2(\mathbb{R}^d)$, there holds

$$\left|D_\mu d^2_{\mathcal{W}}(\cdot,\nu)(\zeta)\right| = \lim_{h\searrow 0} \frac{\left|d^2_{\mathcal{W}}(\exp_\mu(h\cdot\zeta),\nu) - d^2_{\mathcal{W}}(\mu,\nu)\right|}{h} \leqslant \lim_{h\searrow 0} \left(d_{\mathcal{W}}(\exp_\mu(h\cdot\zeta),\nu) + d_{\mathcal{W}}(\mu,\nu)\right) \frac{d_{\mathcal{W}}(\mu, \exp_\mu(h\cdot\xi))}{h}.$$

If ζ satisfies the right hand-side of (5.24b), the limit vanishes, and $\zeta \in \mathbf{Sol}_\mu$ by Lemma 5.2.10. With in mind the interpretation of tangent measure fields as the directions that are infinitesimally optimal, and solenoidal fields as measure fields turning around μ, it is natural to expect some form of converse to (5.24). This is the topic of this section.

We point that [Gig08, Theorem 4.41] is really close in spirit, but it seems to us that it does not imply any of the results in the sequel.

5.3.1 Some convergence tools

If the plan ξ is induced by a map, the convergence with respect to $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}$ can be improved into a strong one, with respect to W_μ. We provide the argument for completeness, but do not claim any originality in this classical-looking statement.

Lemma 5.3.1 (From $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}-$ to $W_\mu-$convergence). *For any $(\xi_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ and $b \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$,*

$$d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(b\#\mu,\xi_n\right) \underset{n\to\infty}{\longrightarrow} 0 \qquad \Longrightarrow \qquad W_\mu(b\#\mu,\xi_n) \underset{n\to\infty}{\longrightarrow} 0.$$

Proof. Denote $b_n := \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi_n) \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$. On the one hand, the convergence with respect to $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}$ implies that $\|\xi_n\|^2_\mu \underset{n\to\infty}{\longrightarrow} \|b\|^2_{L^2_\mu}$. On the other hand, for any $\varepsilon > 0$, let $\varphi \in \mathcal{C}^\infty_c(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ approximate b with $\|\varphi - b\|_{L^2_\mu} \leqslant \varepsilon$. Denote $\mathrm{Lip}(\varphi)$ a Lipschitz constant for φ, and $\beta^n \in \Gamma_o(b\#\mu,\xi_n)$. Then

$$\begin{aligned}\|b\|^2_{L^2_\mu} &\leqslant \int_{x\in\mathbb{R}^d} \langle b(x),\varphi(x)\rangle \, d\mu + \varepsilon\|b\|_{L^2_\mu} = \int_{(x,v),(y,w)\in\mathrm{T}\mathbb{R}^d} \langle v,\varphi(x)\rangle - \langle w,\varphi(y)\rangle \, d\beta^n + \langle b_n,\varphi\rangle_{L^2_\mu} + \varepsilon\|b\|_{L^2_\mu}\\ &\leqslant \max(1,\mathrm{Lip}(\varphi))\, d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(\xi_n,b\#\mu\right)\sqrt{\|\varphi\|^2_{L^2_\mu} + \|\xi_n\|^2_\mu} + \langle b_n,b\rangle_{L^2_\mu} + \varepsilon\left(\|b\|_{L^2_\mu} + \|b_n\|_{L^2_\mu}\right).\end{aligned}$$

Taking the limit inf in n, then the limit in $\varepsilon \searrow 0$, there holds $\|b\|^2_{L^2_\mu} \leqslant \liminf_{n\to\infty} \langle b_n,b\rangle_{L^2_\mu}$. Thus

$$\limsup_{n\to\infty} W^2_\mu\left(\xi_n,b\#\mu\right) \leqslant \lim_{n\to\infty} \|\xi_n\|^2_\mu - 2\liminf_{n\to\infty}\langle b_n,b\rangle_{L^2_\mu} + \|b\|^2_{L^2_\mu} \leqslant \|b\|^2_{L^2_\mu} - 2\langle b,b\rangle_{L^2_\mu} + \|b\|^2_{L^2_\mu} = 0.$$

Here we used that $\langle \xi_n, b\#\mu\rangle^+_\mu = \langle \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi_n), b\#\mu\rangle^+_\mu = \langle b_n,b\rangle_{L^2_\mu}$ since $b\#\mu$ is induced by a map. □

In the previous statement, the measure μ is fixed. A similar result can be deduced for varying base measure, using the application $W_{\mu,\nu}$ in replacement of W_μ. Recall from Definition 1.1.39 p. 13 that

$$W^2_{\mu,\nu}(\xi,\zeta) := \inf\left\{ \int_{(x,v),(y,w)\in\mathrm{T}\mathbb{R}^d} |v-w|^2 \, d\alpha \;\middle|\; \alpha \in \Gamma(\xi,\zeta) \text{ and } (\pi_x,\pi_y)\#\alpha \in \Gamma_o(\mu,\nu) \right\}.$$

Corollary 5.3.2 (Varying base measure). *For $b \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ and $(\xi_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)$ with $\mu_n := \pi_x\#\xi_n$,*

$$d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(b\#\mu,\xi_n\right) \underset{n\to\infty}{\longrightarrow} 0 \qquad \Longrightarrow \qquad W_{\mu,\mu_n}(b\#\mu,\xi_n) \underset{n\to\infty}{\longrightarrow} 0.$$

Proof. For each n, let $\alpha^n \in \Gamma(b\#\mu,\xi_n)$ realize the infimum defining $W_{\mu,\nu}(b\#\mu,\xi_n)$. Define $\widetilde{\xi}_n := (\pi_x,\pi_w)\#\alpha^n \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. Then, using that $\Gamma_\mu(b\#\mu,\widetilde{\xi}_n)$ is reduced to one element,

$$W^2_{\mu,\mu_n}(b\#\mu,\xi_n) = \int_{x\in\mathbb{R}^d,(y,w)\in\mathrm{T}\mathbb{R}^d} |b(x)-w|^2 \, d\alpha^n = \int_{(x,w)} |b(x)-w|^2 \, d(\pi_x,\pi_w)\#\alpha^n = W^2_\mu(b\#\mu,\widetilde{\xi}_n). \tag{5.25}$$

Note that $d_{\mathcal{W}}(\mu,\mu_n) \leqslant d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}(b\#\mu,\xi_n) \underset{n\to\infty}{\longrightarrow} 0$. As

$$d^2_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(\xi_n,\widetilde{\xi}_n\right) \leqslant \int_{(x,v),(y,w)\in\mathrm{T}\mathbb{R}^d} \left|x-y\right|^2 + |v-w|^2 \, d\left(\pi_y,\pi_w,\pi_x,\pi_w\right)\#\alpha^n = d^2_{\mathcal{W}}(\mu,\mu_n) \underset{n\to\infty}{\longrightarrow} 0,$$

the sequence $(\widetilde{\xi}_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ converges towards $b\#\mu$ with respect to $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}$. Applying Lemma 5.3.1, we obtain that $W^2_\mu(b\#\mu,\widetilde{\xi}_n) \to_n 0$, and the conclusion follows from (5.25). □

The following elementary remarks will also be extensively used in the following section. In both, (X,d_X) is a Polish space and o a point of X.

Remark 5.3.3 (Dominated convergence). *Let $(\mu_n)_{n\in\mathbb{N}}$ be contained in a fixed compact of $\big(\mathscr{P}_2(X), d_{\mathcal{W},X}\big)$, and for each n, let ν_n be a submeasure of μ_n such that $\nu_n(X) \to_n 0$. Then $\int_{x\in\mathbb{R}^d} d^2(x,o)d\nu_n \to_n 0$. Indeed, for all radius $R \geqslant 0$,*

$$\begin{aligned}\limsup_{n\to\infty}\int_{x\in X} d^2(x,o)d\nu_n &\leqslant \limsup_{n\to\infty} R^2\nu_n(\overline{\mathscr{B}}_X(o,R)) + \int_{x\in X, d(x,o)\geqslant R} d^2(x,o)d\mu_n\\ &\leqslant R^2\lim_{n\to\infty}\nu_n(X) + \sup_{n\in\mathbb{N}}\int_{x\in X,d(x,o)\geqslant R} d^2(x,o)d\mu_n = \sup_{n\in\mathbb{N}}\int_{x\in X,d(x,o)\geqslant R} d^2(x,o)d\mu_n.\end{aligned}$$

By Theorem 1.1.24, the last term can be made as small as desired by taking R large enough.

Remark 5.3.4 (Submeasures of equal mass are separated). *Let $\mu,\nu \in \mathscr{P}_2(X)$ be mutually singular measures, and $0 < \iota \leqslant 1$. Then $\inf\big\{ d_{\mathcal{W},X}\left(\mu_\iota/\iota, \nu_\iota/\iota\right) \mid \mu_\iota$ (resp. ν_ι) is a submeasure of μ (resp . ν) of mass $\iota\big\} > 0$. Indeed, the set $S^\iota_\mu \subset \mathscr{P}_2(X)$ of μ_ι/ι such that μ_ι is a submeasure of μ of mass ι satisfies*

$$\lim_{R\to\infty}\sup_{\mu_\iota\in S^\iota_\mu}\int_{x\in X,d(x,o)\geqslant R} d^2(x,o)d\frac{\mu_\iota}{\iota} \leqslant \frac{1}{\iota}\lim_{R\to\infty}\int_{x\in X,d(x,o)\geqslant R} d^2(x,o)d\mu = 0,$$

*so is relatively compact by Theorem 1.1.24**p. 9**. It is closed since submeasures are characterized by $\int\varphi d\mu_\iota \leqslant \int \varphi d\mu$ for any $\varphi \in C_b(X;\mathbb{R}^+)$. So S^ι_μ and S^ι_ν are compact subsets of $\big(\mathscr{P}_2(X), d_{\mathcal{W},X}\big)$ with empty intersection, and must be separated.*

5.3.2 What can be said in the general case

We show that if a measure field ξ escapes from μ with speed $\|\xi\|_\mu$, then there is a sequence of reparametrized geodesics converging towards ξ *in the topology induced by the Wasserstein distance on the tangent bundle.* If the convergence were to hold with respect to W_μ instead, this would imply that $\xi \in \mathbf{Tan}_\mu$. Recall that $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}(\cdot,\cdot)$ is the 2−Wasserstein distance on $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)$, for the distance $(x,v),(y,w) \mapsto \sqrt{|x-y|^2+|v-w|^2}$.

Lemma 5.3.5 (Weak converse of (5.24a)). *Let $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ be such that $\lim_{h\searrow 0} h^{-1} d_{\mathcal{W}}(\mu, \exp_\mu(h\cdot\xi)) = \|\xi\|_\mu$. Then there exists a vanishing sequence $(h_n)_{n\in\mathbb{N}} \subset (0,1]$ such that*

$$\lim_{n\to\infty}\sup_{\gamma\in\frac{1}{h_n}\cdot\exp_\mu^{-1}(\exp_\mu(h_n\cdot\xi))} d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(\gamma,\xi\right) = 0. \tag{5.26}$$

Proof. Denote $\mu_s := \exp_\mu(s\cdot\xi)$. For each $s \in [0,1)$ and small $\varepsilon > 0$, there holds $\mu_{s+\varepsilon} = (\pi_x + (s+\varepsilon)\pi_v)\#\xi = \exp_{\mu_s}(\varepsilon\cdot\xi_s)$, where $\xi_s := (\pi_x + s\pi_v, \pi_v)\#\xi$. The Lipschitz curve $s \mapsto d^2_{\mathcal{W}}(\mu,\mu_s)$ is absolutely continuous, so admits a derivative for a.e. s. Hence $\frac{d}{ds}d^2_{\mathcal{W}}(\mu,\mu_s) = D_{\mu_s}d^2_{\mathcal{W}}(\cdot,\mu_s)(\xi_s)$ for such s, and by the formula (1.16),

$$d^2_{\mathcal{W}}(\mu,\mu_h) - 0 = \int_{s\in[0,h]}\frac{d}{ds}d^2_{\mathcal{W}}(\mu,\mu_s)ds = \int_{s\in[0,h]} D_{\mu_s}d^2_{\mathcal{W}}(\mu,\cdot)(\xi_s)ds = \int_{s\in[0,h]}\inf_{\eta\in\exp^{-1}_{\mu_s}(\mu)}\langle -2\cdot\eta,\xi_s\rangle^-_{\mu_s}ds.$$

By assumption, there exists $m : \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim\limits_{s\searrow 0} m(s) = 0$ such that $\frac{d^2_{\mathcal{W}}(\mu,\mu_s)}{s^2} \geqslant \|\xi\|_\mu - m(s)$ for all $s \in [0,h]$.

Writing $h^2 = \int_{s\in[0,h]} 2s\,ds$, we get after rearrangement that

$$\frac{1}{h/2}\int_{s\in[0,h]}\frac{4s}{h}\left[\|\xi\|^2_\mu - \inf_{\eta\in\exp^{-1}_{\mu_s}(\mu)}\langle -s^{-1}\cdot\eta,\xi_s\rangle^-_{\mu_s}\right]ds \leqslant m(h) \qquad \forall h > 0.$$

Denote $\varphi(s) := \|\xi\|^2_\mu - \inf_{\eta\in\exp^{-1}_{\mu_s}(\mu)}\langle -s^{-1}\cdot\eta,\xi_s\rangle^-_{\mu_s}$. Since $\left|\langle -s^{-1}\cdot\eta,\xi_s\rangle^-_{\mu_s}\right| \leqslant s^{-1}d_{\mathcal{W}}(\mu,\mu_s)\|\xi_s\|_{\mu_s}$, with $\|\xi_s\|_{\mu_s} = \|\xi\|_\mu$ for all $s \in [0,1]$ by definition, and $s^{-1}d_{\mathcal{W}}(\mu,\mu_s) \leqslant \|\xi\|_\mu$, the function φ is nonnegative. Sacrificing the integral over $[0,h/2)$, and bounding $4s/h$ from below by 2 on the interval $[h/2,h]$, we get

$$\frac{1}{h/2}\int_{s\in[h/2,h]} 2\varphi(s) \leqslant m(h) \qquad \forall h > 0.$$

In other words, the mean of φ over each interval $[h/2,h]$ is inferior to $m(h)/2$. In particular, in each interval $[2^{-n+1}, 2^{-n}]$, there must exist s_n such that $\varphi(s_n) \leqslant m(s_n)/2$.

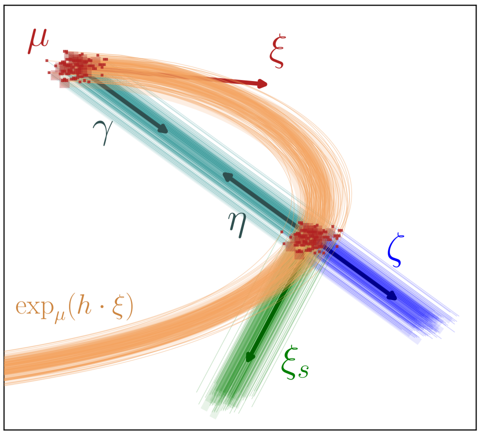
We turn to the estimate (5.26) along the vanishing sequence $(s_n)_n$. For any $\gamma \in \exp_\mu^{-1}(\mu_{s_n})$, denote $\zeta := \left(\pi_x + \pi_v, s_n^{-1}\pi_v\right)\#\gamma$. One has

$$\begin{aligned} & d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(\frac{1}{s_n}\cdot\gamma,\xi\right) \\ & \leqslant d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(\frac{1}{s_n}\cdot\gamma,\zeta\right) + d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(\zeta,\xi_{s_n}\right) + d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(\xi_{s_n},\xi\right) \\ & \leqslant |s_n|\,\|\xi\|_\mu + W_{\mu_{s_n}}\left(\zeta,\xi_{s_n}\right) + |s_n|\,\|\xi\|_\mu, \end{aligned} \tag{5.27}$$

where the estimate between $\frac{1}{s_n}\cdot\gamma$ and ζ uses the plan

$$(\pi_x, s_n^{-1}\pi_v, \pi_x+\pi_v, s_n^{-1}\pi_v)\#\gamma,$$

and the one between ξ_{s_n} and ξ uses $(\pi_x + s_n\pi_v, \pi_v, \pi_x, \pi_v)\#\xi$. On the other hand, using that $\|\zeta\|_{\mu_{s_n}} = s_n^{-1} d_{\mathcal{W}}(\mu,\mu_{s_n}) \leqslant \|\xi\|_\mu$, and $\langle\cdot,\cdot\rangle_\mu^- \leqslant \langle\cdot,\cdot\rangle_\mu^+$, there holds



Notations. Here the exponential is abusively curved to let the optimal plans γ and η appear.

$$W_{\mu_{s_n}}^2\left(\zeta,\xi_{s_n}\right) = \|\zeta\|_{\mu_{s_n}}^2 + \|\xi_{s_n}\|_{\mu_{s_n}}^2 - 2\langle\zeta,\xi_{s_n}\rangle_{\mu_{s_n}}^+ \leqslant 2\|\xi\|_\mu^2 - 2\langle\zeta,\xi_{s_n}\rangle_{\mu_{s_n}}^- .$$

In order to recognise φ, we write $\zeta = -s_n^{-1}\cdot\eta$, where $\eta := -s_n\cdot\zeta = (\pi_x+\pi_v, -\pi_v)\#\gamma$. Since γ is an optimal transport plan between μ and μ_{s_n}, the plan η is optimal between μ_{s_n} and μ. It is admissible in the infimum appearing in the definition of φ, and we get that

$$W_{\mu_{s_n}}^2\left(\zeta,\xi_{s_n}\right) \leqslant 2\|\xi\|_\mu^2 - 2\langle -s_n^{-1}\cdot\eta,\xi_{s_n}\rangle_{\mu_{s_n}}^- \leqslant 2\varphi(s_n) \leqslant m(s_n).$$

Plugging this into (5.27), we obtain an estimate that does not depend on $\gamma \in s_n^{-1}\cdot\exp_\mu^{-1}(\exp_\mu(s_n\cdot\xi))$. □

As a corollary, we can prove – at least in dimension 1 – that a tangent field may be approximated by a sequence of reparametrized geodesics *pointing towards the exponentials.*

Proposition 5.3.6 (Convergence of logarithms in dimension 1). *For any $\xi \in \mathbf{Tan}_\mu$, there exists $(h_n)_{n\in\mathbb{N}} \subset (0,1]$ a vanishing sequence such that any choice of $\eta_n \in \frac{1}{h_n}\cdot\exp_\mu^{-1}(\exp_\mu(h_n\cdot\xi))$ satisfies $W_\mu(\xi,\eta_n) \to_n 0$.*

Proof. Decompose $\mu = m_a\mu^a + m_d\mu^d$ for $\mu^a \in \mathscr{P}_2(\mathbb{R}^d)$ the atomic part of μ, of mass m_a, and μ^d its diffuse part, of mass m_d. By Theorem 5.2.23, $\xi = m_a\xi^a + m_d f^d\#\mu^d$ for some $\xi^a \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_{\mu^a}$, and $f^d \in L^2_{\mu^d}(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$. Let $(h_n)_{n\in\mathbb{N}} \subset (0,1]$ be the vanishing sequence provided by Lemma 5.3.5. Let η_n be as in the statement, which also writes as $m_a\eta_n^a + m_d f_n^d\#\mu^d$.

Separation of atomic and diffuse parts. Let $\xi^d := f^d\#\mu^d$ and $\eta_n^d = f_n^d\#\mu^d$. We claim that

$$\lim_{n\to\infty} m_a d^2_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(\xi^a,\eta_n^a\right) + m_d d^2_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(\xi^d,\eta_n^d\right) = 0.$$

Any $\beta_n \in \Gamma_o(\xi,\eta_n)$ splits as $\beta_n = \beta_n^{a,a} + \beta_n^{a,d} + \beta_n^{d,a} + \beta_n^{d,d}$, where $(\pi_x,\pi_v)\#\beta_n^{b,c}$ is a submeasure of ξ^b and $(\pi_y,\pi_w)\#\beta_n^{b,c}$ a submeasure of η_n^c for $b,c \in \{a,d\}$. In particular,

$$\beta_n^{a,a}(\mathrm{T}\mathbb{R}^d) + \beta_n^{d,a}(\mathrm{T}\mathbb{R}^d) = \eta_n^a(\mathrm{T}\mathbb{R}^d) = \beta_n^{a,a}(\mathrm{T}\mathbb{R}^d) + \beta_n^{a,d}(\mathrm{T}\mathbb{R}^d),$$

so that $\beta_n^{a,d}$ and $\beta_n^{d,a}$ have the same mass. Assume by contradiction that $\limsup_{n\to\infty}\beta_n^{a,d}(\mathrm{T}\mathbb{R}^d) = 2\iota$ for some $\iota > 0$. Then, along some non-relabeled subsequence, some submeasure $\gamma_n^{a,d}$ of $(\pi_x,\pi_y)\#\beta_n^{a,d}$ has mass equal to ι and satisfies

$$d^2_{\mathcal{W}}\left(\frac{\pi_x\#\gamma_n^{a,d}}{\iota}, \frac{\pi_y\#\gamma_n^{a,d}}{\iota}\right) \leqslant \frac{1}{\iota}\int_{x,y\in\mathbb{R}^d}|x-y|^2 d\gamma_n^{a,d} \leqslant \frac{1}{\iota}\int_{x,y\in\mathbb{R}^d}|x-y|^2 d(\pi_x,\pi_y)\#\beta_n^{a,d} \leqslant \frac{1}{\iota}d^2_{\mathcal{W}}(\xi,\xi_n) \to_n 0.$$

The measure $\pi_x\#\gamma_n^{a,d}$ is a submeasure of μ^a, and $\pi_y\#\gamma_n^{a,d}$ a submeasure of μ^d, which are mutually singular. By Remark 5.3.4, this is absurd. So $\beta_n^{a,d}(\mathrm{T}\mathbb{R}^d) \to_n 0$, and the mass of $\beta_n^{b,b}$ converges to m_b for $b \in \{a,d\}$.

Construct a transport plan between $m_a\xi^a$ and $m_a\eta_n^a$ by adding to $\beta_n^{a,a}$ any transport plan between the nonnegative measures $m_a\xi^a-(\pi_x,\pi_v)\#\beta_n^{a,a}$ and $m_a\eta_n^a-(\pi_y,\pi_w)\#\beta_n^{a,a}$. Using $\left|(x,v)-(y,w)\right|^2 \leqslant 2(|(x,v)|^2+|(y,w)|^2)$, we get

$$m_a d^2_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(\xi^a,\eta_n^a\right) \leqslant \int_{(x,v),(y,w)} \left|(x,v)-(y,w)\right|^2 d\beta_n^{a,a}$$
$$+2\left[\int_{(x,v)} |(x,v)|^2\, d\left[m_a\xi^a-(\pi_x,\pi_v)\#\beta_n^{a,a}\right]+\int_{(y,w)} |(y,w)|^2 d\left[m_a\eta_n^a-(\pi_y,\pi_w)\#\beta_n^{a,a}\right]\right].$$

The first term is inferior to $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}(\xi,\eta_n)$ by construction of $\beta_n^{a,a}$, and the second goes to 0 when n goes to ∞ by Remark 5.3.3. The symmetric argument proves that $m_d d^2_{\mathcal{W},\mathrm{T}\mathbb{R}^d}(\xi^d,\eta^d)\to_n 0$.

Improving convergences. By Lemma 5.3.1, the convergence of η_n^d towards the map-induced field ξ^d implies $W_{\mu^d}-$convergence. On the other hand, let $\beta_n^a \in \Gamma_o(\xi^a,\eta_n^a)$. The argument is quite similar to the first step: the plan $(\pi_x,\pi_y)\#\beta_n^a$ writes as $\sum_{i,j\in\mathbb{N}} m^a_{n,i,j}\delta_{(x_i,x_j)}$, with $(x_i)_{i\in\mathbb{N}}$ the countable set of atoms of μ, and satisfies

$$\sum_{i,j\in\mathbb{N}} m^a_{n,i,j}|x_i-x_j|^a = \int_{(x,y)\in\mathbb{R}^{d^2}} |x-y|^2 d(\pi_x,\pi_y)\#\beta_n^a \leqslant d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}(\xi^a,\eta_n^a) \underset{n\to\infty}{\longrightarrow} 0.$$

So for $i\neq j$, one has $m^a_{n,i,j}\to_n 0$. Split $\beta_n^a=\sum_{i,j\in\mathbb{N}}\beta^a_{n,i,j}$ with $(\pi_x,\pi_y)\#\beta^a_{n,i,j}=m^a_{n,i,j}\delta_{(x_i,x_j)}$ the submeasure of β_n^a sending mass from the fiber of x_i to that of x_j. The plan $\sum_i \beta^a_{n,i,i}$ is a transport plan from a submeasure $\check{\xi}_n^a$ of ξ^a to a submeasure $\check{\eta}_n^a$ of η_n^a, both of mass $\sum_i m^a_{n,i,i}\to_n 1$. Additionally, $\sum_i \beta^a_{n,i,i}$ does not move mass between pairs (x,v) and (y,w) with $x\neq y$. We may construct a transport plan of $\Gamma_{\mu^a}(\xi^a,\eta_n^a)$ by summing this plan with any transport plan between the nonnegative measures $\xi^a-\check{\xi}_n^a$ and $\eta_n^a-\check{\eta}_n^a$, and use $|(x,v)-(y,w)|^2\leqslant 2(|(x,v)|^2+|(y,w)|^2)$ to obtain

$$W^2_{\mu^a}\left(\xi^a,\eta_n^a\right)\leqslant \int_{(x,v),(y,w)} |(x,v)-(y,w)|^2 d\left[\sum_{i\in\mathbb{N}}\beta^a_{n,i,i}\right]+2\left[\int_{(x,v)}|(x,v)|^2 d[\xi^a-\check{\xi}_n^a]+\int_{(y,w)}|(y,w)|^2 d[\eta_n^a-\check{\eta}_n^a]\right].$$

The first summand is inferior to $d^2_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(\xi^a,\eta_n^a\right)$, which goes to 0, and the second vanishes when $n\to\infty$ by Remark 5.3.3. So, using that μ^a and μ^d are mutually singular,

$$W^2_\mu(\xi,\eta_n)=m_a W^2_\mu(\xi^a,\eta_n^a)+m_d W^2_\mu(\xi^d,\eta_n^d)\underset{n\to\infty}{\longrightarrow} 0,$$

and the proof is complete. □

Remark 5.3.7. *The sequence $(h_n)_{n\in\mathbb{N}}$ in Proposition 5.3.6 can be any sequence along which $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}(\xi,\eta_n)$ goes to 0 when $n\to\infty$ for some $\eta_n\in\frac{1}{h_n}\cdot\exp_\mu^{-1}(\exp_\mu(h_n\cdot\xi))$. It is not clear to the author whether this holds for any sequence or not.*

5.3.3 The case of measure fields induced by a map

Proposition 5.3.8 (Barycentric case). *If ξ is induced by a map, then the implications in (5.24) are equivalences.*

Proof. Assume first that $f\in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ satisfies

$$\lim_{h\searrow 0}\frac{d_{\mathcal{W}}(\mu,\exp_\mu(h\cdot f\#\mu))}{h}=\|f\|_{L^2_\mu}.$$

By Lemma 5.3.5, there exists a sequence of reparametrized geodesics $(\xi_n)_{n\in\mathbb{N}}\subset\mathbf{Tan}_\mu$ converging towards $f\#\mu$ with respect to $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}$. By Lemma 5.3.1, this sequence converges with respect to W_μ, and $f\#\mu$ is tangent.

Assume now that $\xi := f\#\mu$ is solenoidal for $f\in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$. The beginning of the argument is similar as that of Lemma 5.3.5. Denote $\mu_s:=\exp_\mu(s\cdot\xi)$. For any $s,\tau\geqslant 0$, one has $\mu_{s+\tau}=\exp_{\exp_\mu(s\cdot\xi)}(\tau\cdot\xi_s)$, where $\xi_s := (\pi_x + s\pi_v, \pi_v)\#\xi$. The application $s \mapsto d^2_{\mathcal{W}}(\mu, \exp_\mu(s\cdot\xi))$ is locally Lipschitz, hence absolutely continuous, and its derivative at s is given by the directional derivative of $d^2_{\mathcal{W}}(\mu,\cdot)$ at μ_s in the direction ξ_s whenever it exists. By the estimate of Lemma 1.1.42 p. 14 over the derivatives of the squared Wasserstein distance,

$$d^2_{\mathcal{W}}(\mu,\mu_h) = \int_{s=0}^{h} D_{\mu_s} d^2_{\mathcal{W}}(\mu,\cdot)(\xi_s)\,ds \leqslant \int_{s=0}^{h} -D_\mu d^2_{\mathcal{W}}(\cdot,\mu_s)(\xi) + 2d_{\mathcal{W}}(\mu,\mu_s)W_{\mu,\mu_s}(\xi,\xi_s)\,ds.$$

Since $\xi \in \mathbf{Sol}_\mu$, the term $-D_\mu d^2_{\mathcal{W}}(\cdot,\mu_s)(\xi)$ vanishes by Lemma 5.2.10. Dividing by $h^2 > 0$, and using that $d_{\mathcal{W}}(\mu,\mu_s)/h \leqslant s\|\xi\|_\mu/h \leqslant \|\xi\|_\mu$ for $s \in [0,h]$, we find

$$\frac{d^2_{\mathcal{W}}(\mu,\exp_\mu(h\cdot\xi))}{h^2} \leqslant \frac{1}{h}\int_{s=0}^{h} 2\|\xi\|_\mu W_{\mu,\mu_s}(\xi,\xi_s)\,ds. \tag{5.28}$$

Note that $d^2_{\mathcal{W},\mathrm{T}\mathbb{R}^d}(\xi,\xi_s) \leqslant \int_{(x,v)\in\mathrm{T}\mathbb{R}^d} |x-(x+hv)|^2 + |v-v|^2 d\xi \leqslant h^2\|\xi\|^2_\mu \xrightarrow[h\searrow 0]{} 0$. Since ξ is induced by a map, Corollary 5.3.2 yields that $W_{\mu,\exp_\mu(s\cdot\xi)}(\xi,\xi_s) \xrightarrow[s\searrow 0]{} 0$. Sending h to 0 in (5.28), we conclude. □

In this case, the projection on the tangent cone is the only component of ξ that matters near μ. This result is the original motivation of this study: it allows to say that in the Wasserstein space, any Lipschitz function $\varphi : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ that is directionally differentiable satisfies $D_\mu\varphi(\xi) = D_\mu\varphi(\pi^\mu_T\xi)$ if ξ is induced by a map. It is shown in further sections that (5.29) can fail if ξ is not induced by a map.

Corollary 5.3.9 (Infinitesimal prevalence of the projection). *Let $\xi = f\#\mu$ for some $f \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$. Then*

$$\lim_{h\searrow 0} \frac{d_{\mathcal{W}}\left(\exp_\mu(h\cdot\pi^\mu_T\xi), \exp_\mu(h\cdot\xi)\right)}{h} = 0. \tag{5.29}$$

Proof. By Theorem 5.2.12, the metric projections of ξ on $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$ are themselves induced by maps, denoted $\pi^\mu_T f, \pi^\mu_S f \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$. In consequence,

$$\exp_\mu(h\cdot\xi) = \left(\pi_x + h\pi_v(f(\pi_x))\right)\#\mu = \left(\pi_x + h\pi_v(\pi^\mu_T f(\pi_x)) + h\pi_v(\pi^\mu_S f(\pi_x))\right)\#\mu = \exp_{\exp_\mu(h\cdot\pi^\mu_S\xi)}\left(h\cdot(\pi^\mu_T\xi)_h\right),$$

where $(\pi^\mu_T\xi)_h := \left(\pi_x + h\pi_v(\pi^\mu_S f(\pi_x)), \pi_v(\pi^\mu_T f(\pi_x))\right)\#\mu$. Let $\alpha_h \in \Gamma_{(\mu,\exp_\mu(h\cdot\pi^\mu_S\xi)),o}\left(\pi^\mu_T\xi,(\pi^\mu_T\xi)_h\right)$ realize the infimum in the definition of $W_{\mu,\exp_\mu(h\cdot\pi^\mu_S\xi)}\left(\pi^\mu_T\xi,(\pi^\mu_T\xi)_h\right)$. The plan $\beta := \left(\pi_x + h\pi_v, \pi_y + h\pi_w\right)\#\alpha$ satisfies

$$\pi_x\#\beta = (\pi_x + h\pi_v)\#\alpha = (\pi_x + h\pi_v)\#\pi^\mu_T\xi = \exp_\mu(h\cdot\pi^\mu_T\xi),$$
$$\pi_y\#\beta = (\pi_y + h\pi_w)\#\alpha = (\pi_x + h\pi_v)\#(\pi^\mu_T\xi)_h = (\pi_x + h\pi_v\left(\pi^\mu_S f(\pi_x)\right) + h\pi_v(\pi^\mu_T f(\pi_x)))\#\mu = \exp_\mu(h\cdot\xi).$$

Hence it provides the estimate

$$\begin{aligned} d_{\mathcal{W}}(\exp_\mu(h\cdot\pi^\mu_T\xi),\exp_\mu(h\cdot\xi)) &\leqslant \left(\int_{(x,v),(y,w)\in\mathrm{T}\mathbb{R}^d} \left|(x+hv)-(y+hw)\right|^2 d\alpha\right)^{1/2} \\ &\leqslant \left(\int_{(x,y)\in(\mathbb{R}^d)^2} \left|x-y\right|^2 d\alpha\right)^{1/2} + h\left(\int_{(x,v),(y,w)\in\mathrm{T}\mathbb{R}^d} |v-w|^2\,d\alpha\right)^{1/2} \\ &= d_{\mathcal{W}}(\mu,\exp_\mu(h\cdot\pi^\mu_S\xi)) + hW_{\mu,\exp_\mu(h\cdot\pi^\mu_S\xi)}\left(\pi^\mu_T\xi,(\pi^\mu_T\xi)_h\right). \end{aligned} \tag{5.30}$$

On the one hand, since $\pi^\mu_S\xi$ is induced by a map and solenoidal, the converse of (5.24b) holds by Proposition 5.3.8, and $d_{\mathcal{W}}(\mu,\exp_\mu(h\cdot\pi^\mu_S\xi)) = o(h)$. On the other hand, $\pi^\mu_T\xi$ is also induced by a map, and

$$\begin{aligned} d^2_{\mathcal{W},\mathrm{T}\mathbb{R}^d}\left(\pi^\mu_T\xi,(\pi^\mu_T\xi)_h\right) &\leqslant \int_{(x,v),(y,w)} |x-y|^2 + |v-w|^2 d\left(\pi_x, \pi_v(\pi^\mu_T f(\pi_x)), \pi_x + h\pi_v(\pi^\mu_S f(\pi_x)), \pi_v(\pi^\mu_T f(\pi_x))\right)\#\mu \\ &\leqslant \int_{x\in\mathbb{R}^d} |h\pi_v(\pi^\mu_S f(x))|^2 + 0\,d\mu = h^2\|\pi^\mu_S\xi\|^2_\mu \xrightarrow[h\searrow 0]{} 0. \end{aligned}$$

By Corollary 5.3.2, $W_{\mu,\exp_\mu(h\cdot\pi^\mu_S\xi)}\left(\pi^\mu_T\xi,(\pi^\mu_T\xi)_h\right) = O(h)$. Dividing by $h > 0$ in (5.30) and sending h to 0, we conclude. □

In the argument of Proposition 5.3.8, the condition that the plan is induced by a map intervenes only as a way to improve the convergence with respect to $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}$ in a stronger one. There might be other classes of plans on which this improvement holds. We state as a curiosity the following example – although to our opinion, it bears no interest in itself, since the requirement that the sequence $(\xi_n)_n$ has a particular structure makes it almost impossible to apply. In consequence, we do not detail the computations supporting visual facts.

Definition 5.3.10 (Bubble field). *A measure field $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ is a bubble field if there exists a radius function $\rho \in L^2_\mu(\mathbb{R}^d;\mathbb{R}^+)$ such that*

$$\int_{(x,v)\in\mathrm{T}\mathbb{R}^d} \varphi(x,v)d\xi = \int_{x\in\mathbb{R}^d}\int_{v\in\mathrm{T}_x\mathbb{R}^d,|v|=1} \varphi(x,\rho(x)\cdot v)d\mathcal{H}(v)d\mu(x) \qquad \forall\varphi\in\mathcal{C}_b(\mathrm{T}\mathbb{R}^d;\mathbb{R}).$$

Here $\mathcal{H}$ is the normalized $(d-1)-$Hausdorff measure on the unit ball of the tangent space.

Then ξ puts mass on the "bubble" of radius $\rho(x)$ around the point x. To simplify the notation, denote $\text{⚲} \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ the unit bubble field, i.e. the measure field characterized by

$$\int_{(x,v)\in\mathrm{T}\mathbb{R}^d} \varphi(x,v)d\text{⚲}(x,v) = \int_{x\in\mathbb{R}^d}\int_{v\in\mathrm{T}_x\mathbb{R}^d,|v|=1} \varphi(x,v)d\mathcal{H}(v)d\mu(x).$$

Then the bubble field with radii ρ is given by $\rho\cdot\text{⚲} = (\pi_x,\rho(x)\pi_v)\#\text{⚲}$. We use without demonstration the following key property: the optimal transport plan between $(d-1)-$Hausdorff measures supported on concentric spheres of different radii is unique, and moves mass radially. In consequence, for any $\rho,\rho'\in L^2_\mu(\mathbb{R}^d;\mathbb{R}^+)$,

$$W^2_\mu(\rho\cdot\text{⚲},\rho'\cdot\text{⚲}) = \int_{(x,v)\in\mathrm{T}\mathbb{R}^d,|v|=1} \left|\rho(x)v-\rho'(x)v\right|^2 d[\mu\otimes\mathcal{H}](x,v) = \|\rho-\rho'\|^2_{L^2_\mu},$$
$$\langle\rho\cdot\text{⚲},\rho'\cdot\text{⚲}\rangle^+_\mu = \frac{1}{2}\left[\|\rho\cdot\text{⚲}\|^2_\mu + \|\rho'\cdot\text{⚲}\|^2_\mu - W^2_\mu(\rho\cdot\text{⚲},\rho'\cdot\text{⚲})\right] = \langle\rho,\rho'\rangle_{L^2_\mu}.$$

Lemma 5.3.11 (The bubble case). *Let $(\xi_n)_{n\in\mathbb{N}}\subset\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ be bubble fields converging towards a bubble field $\xi\in\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ with respect to $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}(\cdot,\cdot)$. Then $W_\mu(\xi_n,\xi)\underset{n\to\infty}{\longrightarrow}0$.*

Proof. Decompose $\xi=\rho\cdot\text{⚲}$. For any $\varepsilon>0$, let $\varphi\in\mathcal{C}_c(\mathbb{R}^d;\mathbb{R}^+)$ approximate ρ with $\|\varphi-\rho\|_{L^2_\mu}\leqslant\varepsilon$. Then $W_\mu(\varphi\cdot\text{⚲},\rho\cdot\text{⚲})\leqslant\varepsilon$. Denote $\mathrm{Lip}(\varphi)$ a Lipschitz constant for φ, and $\beta^n\in\Gamma_o(\xi,\xi_n)$. Then

$$\begin{aligned}\|\xi\|^2_\mu &= \int_{(x,v)\in\mathrm{T}\mathbb{R}^d,|v|>0}\langle v,\rho(x)\frac{v}{|v|}\rangle\, d\xi(x,v) \leqslant \int_{(x,v)\in\mathrm{T}\mathbb{R}^d,|v|>0}\langle v,\varphi(x)\frac{v}{|v|}\rangle\, d\xi(x,v)+\varepsilon\|\xi\|_\mu\\ &= \int_{(x,v),(y,w)\in\mathrm{T}\mathbb{R}^d,|v|,|w|>0}\langle v,\varphi(x)\frac{v}{|v|}\rangle - \langle w,\varphi(y)\frac{w}{|w|}\rangle\, d\beta^n + \langle\xi_n,\varphi\cdot\text{⚲}\rangle^+_\mu+\varepsilon\|\xi\|_\mu.\end{aligned}$$

Since $\varphi\geqslant0$, one has

$$\langle v,\varphi(x)\frac{v}{|v|}\rangle - \langle w,\varphi(y)\frac{w}{|w|}\rangle = \langle v,\left(\varphi(x)-\varphi(y)\right)\frac{v}{|v|}\rangle + \varphi(y)\,[|v|-|w|] \leqslant \left|\varphi(x)-\varphi(y)\right||v|+\varphi(y)\,|v-w|.$$

Using that φ is Lipschitz with constant $\mathrm{Lip}(\varphi)$ that we may assume greater than 1, we deduce

$$\begin{aligned}\|\xi\|^2_\mu &\leqslant \int_{(x,v),(y,w)\in\mathrm{T}\mathbb{R}^d}\mathrm{Lip}(\varphi)\left|x-y\right||v|+\varphi(y)\,|v-w|\,d\beta^n + \langle\xi_n,\xi\rangle^+_\mu+\varepsilon\left(\|\xi\|_\mu+\|\xi_n\|_\mu\right)\\ &\leqslant \mathrm{Lip}(\varphi)\sqrt{\int_{(x,v),(y,w)}|v|^2+\varphi^2(y)d\beta^n\int_{(x,v),(y,w)}\left|x-y\right|+|v-w|\,d\beta^n} + \langle\xi_n,\xi\rangle^+_\mu+\varepsilon\left(\|\xi\|_\mu+\|\xi_n\|_\mu\right)\\ &= \mathrm{Lip}(\varphi)\sqrt{\|\xi\|^2_\mu+\|\varphi\|^2_{L^2_\mu}d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}(\xi,\xi_n)} + \langle\xi_n,\xi\rangle^+_\mu+\varepsilon\left(\|\xi\|_\mu+\|\xi_n\|_\mu\right).\end{aligned}$$

Taking the limit inf in n, then in $\varepsilon\searrow0$, there holds $\|\xi\|^2_\mu\leqslant\liminf_{n\to\infty}\langle\xi_n,\xi\rangle^+_\mu$. Thus, using $\|\xi_n\|_\mu\to_n\|\xi\|_\mu$,

$$\limsup_{n\to\infty}W^2_\mu(\xi_n,\xi)\leqslant\lim_{n\to\infty}\|\xi_n\|^2_\mu-2\liminf_{n\to\infty}\langle\xi_n,\xi\rangle^+_\mu+\|\xi\|^2_\mu\leqslant\|\xi\|^2_\mu-2\|\xi\|^2_\mu+\|\xi\|^2_\mu=0.$$

Hence the result on bubble fields. □

Unfortunately, we cannot deduce a converse of (5.24) on bubble fields, since there is no reason that the sequences $(\xi_n)_n$ constructed in the proof of Proposition 5.3.8 belong to bubble fields. Moreover, the following section provides a counterexample.

5.3.4 Counterexamples in the general case

We now show that the converse of (5.24) does not hold in general. Precisely, we construct measures μ in dimension $d = 1$, each supported on a compact interval and transport-regular, so that the centred measure field $F[\mu] := \frac{1}{2}((id,-1)\#\mu + (id,1)\#\mu)$ is solenoidal. On these example, the curve $\mu_h := \exp_\mu(h \cdot F[\mu])$ satisfies the following.

Measure	Behaviour of $\mu_h := \exp_\mu(h \cdot F[\mu])$	Conclusion
Cantor	$0 < \liminf_{h\searrow 0} \frac{d_{\mathcal{W}}(\mu,\mu_h)}{h} < \limsup_{h\searrow 0} \frac{d_{\mathcal{W}}(\mu,\mu_h)}{h} < \Vert F[\mu]\Vert_\mu$	(5.24b) does not admit a converse.
Skinny Cantor	$0 < \liminf_{h\searrow 0} \frac{d_{\mathcal{W}}(\mu,\mu_h)}{h} < \limsup_{h\searrow 0} \frac{d_{\mathcal{W}}(\mu,\mu_h)}{h} = \Vert F[\mu]\Vert_\mu$	Intermediate edge case.
Unbalanced skinny Cantor	$\lim_{h\searrow 0} \frac{d_{\mathcal{W}}(\mu,\mu_h)}{h} = \Vert F[\mu]\Vert_\mu$	(5.24a) does not admit a converse.

The skinny Cantor may be seen as a particular case of the unbalanced skinny Cantor, and we restrict to Remark 5.3.12 on its account.

5.3.4.1 Cantor measure

Here we recall classical facts about the Cantor measure, and introduce the notations in use for the study of the Wasserstein distance around it.

Description of the Cantor set. Denote $f_0 : x \mapsto x/3$ and $f_1 : x \mapsto 1-(1-x)/3 = 2/3 + x/3$ two affine transformations. The recursive approximation of the Cantor set is constructed as

$$C^0 := [0,1], \qquad C^{n+1} := f_0(C^n) \sqcup f_1(C^n), \qquad C := \lim_n C^n = \bigcap_n C^n.$$

In particular,

$$f_0(C) \sqcup f_1(C) = f_0\left(\bigcap_n C^n\right) \bigsqcup f_1\left(\bigcap_n C^n\right) = \left(\bigcap_n f_0(C^n)\right) \bigsqcup \left(\bigcap_n f_1(C^n)\right) = \bigcap_n \left(f_0(C^n) \sqcup f_1(C^n)\right) = \bigcap_n C^{n+1} = C.$$

Let $d = \ln_3(2) = \ln(2)/\ln(3)$. The Hausdorff measure on C is defined for any Borel set A by

$$\mu(A) := \lim_{h\searrow 0} \inf_{\substack{(B_m)_{m\in\mathbb{N}} \\ \operatorname{diam} B_m \leq h \\ C\cap A \subset \cup_m B_m}} \sum_{m\in\mathbb{N}} \alpha_d \left(\frac{\operatorname{diam} B_m}{2}\right)^d. \tag{5.31}$$

The limit exists since the inf is nondecreasing with h. Technically, (5.31) is defined for any set $A \subset \mathbb{R}$; the fact that the restriction of μ to Borel sets is a σ−additive measure is a result [AFP00, Proposition 2.49]. The measure μ gives 0 mass to any singleton, hence is transport-regular. By Theorem 5.2.23, $\mathbf{Tan}_\mu = \operatorname{Tan}_\mu$ and $\mathbf{Sol}_\mu = \mathscr{P}_2(\mathrm{T}\mathbb{R})^0_\mu$.

Invariance properties. We claim that $\mu = \frac{f_0\#\mu + f_1\#\mu}{2}$.

Let $A \subset \mathbb{R}$ be Borel. For any countable cover $(B_m)_m$ of $C \cap A = (f_0(C)\cap A) \sqcup (f_1(C)\cap A)$, define $B_m^0 := B_m \cap f_0([0,1])$ and $B_m^1 := B_m \cap f_1([0,1])$. Then $B_m^0 \cap B_n^0 = \varnothing$ for all m,n since $f_0([0,1])$ and $f_1([0,1])$ are disjoint, the diameter of B_m^k is inferior of that of B_m, and $f_k(C)\cap A \subset \cup_m B_m^k$ for $k \in \{0,1\}$. Hence the family $(B_m^k)_{m\in\mathbb{N},k\in\{0,1\}}$ is a better competitor in (5.31), and

$$\mu(A) := \lim_{h\searrow 0} \inf_{\substack{(B_m^0)_{m\in\mathbb{N}} \\ \operatorname{diam} B_m^0 \leq h \\ f_0(C)\cap A \subset \cup_m B_m^0}} \sum_{m\in\mathbb{N}} \alpha_d \left(\frac{\operatorname{diam} B_m^0}{2}\right)^d + \inf_{\substack{(B_m^1)_{m\in\mathbb{N}} \\ \operatorname{diam} B_m^1 \leq h \\ f_1(C)\cap A \subset \cup_m B_m^1}} \sum_{m\in\mathbb{N}} \alpha_d \left(\frac{\operatorname{diam} B_m^1}{2}\right)^d$$

Since one has

$$f_0(C)\cap A=\left\{x \mid \exists y\in C, f_0(y)=x \text{ and } x\in A\right\}=\left\{f_0(y) \mid y\in C \text{ and } y\in f_0^{-1}(A)\right\}=f_0\left(C\cap f_0^{-1}(A)\right),$$

the sets B_m^0 cover $f_0(C)\cap A$ if and only if the sets $D_m^0 := f_0^{-1}(B_m^0)$ cover $C\cap f_0^{-1}(A)$. Recalling that $f_0(x)=x/3$, one has $\operatorname{diam} D_m^0 = 3\operatorname{diam} B_m^0$, and

$$\inf_{\substack{(B_m^0)_{m\in\mathbb{N}}\\ \operatorname{diam} B_m^0\leqslant h\\ f_0(C)\cap A\subset\cup_m B_m^0}} \sum_{m\in\mathbb{N}} \alpha_d\left(\frac{\operatorname{diam} B_m^0}{2}\right)^d = \inf_{\substack{(D_m^0)_{m\in\mathbb{N}}\\ \operatorname{diam} D_m^0\leqslant 3h\\ C\cap f_0^{-1}(A)\subset\cup_m D_m^0}} \sum_{m\in\mathbb{N}} \alpha_d\left(\frac{1/3\operatorname{diam} D_m^0}{2}\right)^d.$$

Here, $3^d=3^{\ln(2)/\ln(3)}=3^{\ln_3(2)}=2$. Applying the same reasoning on f_1, we conclude that

$$\mu(A)=\lim_{h\searrow 0} \inf_{\substack{(D_m^0)_{m\in\mathbb{N}}\\ \operatorname{diam} D_m^0\leqslant 3h\\ C\cap f_0^{-1}(A)\subset\cup_m D_m^0}} \sum_{m\in\mathbb{N}} \frac{\alpha_d}{2}\left(\frac{\operatorname{diam} D_m^0}{2}\right)^d + \inf_{\substack{(D_m^1)_{m\in\mathbb{N}}\\ \operatorname{diam} D_m^1\leqslant 3h\\ C\cap f_1^{-1}(A)\subset\cup_m D_m^1}} \sum_{m\in\mathbb{N}} \frac{\alpha_d}{2}\left(\frac{\operatorname{diam} D_m^1}{2}\right)^d = \frac{\mu\left(f_0^{-1}(A)\right)+\mu\left(f_1^{-1}(A)\right)}{2}.$$

By definition of the pushforward, $\mu=(f_0\#\mu+f_1\#\mu)/2$. For any multi-index $a=(a_n,\cdots,a_1)\in A^n:=\{0,1\}^n$, denote $f_a:=f_{a_n}\circ f_{a_{n-1}}\circ\cdots\circ f_{a_1}$. Iterating the previous invariance, one sees that

$$\mu=\frac{1}{2^n}\sum_{a\in A^n} f_a\#\mu.$$

Each $f_a\#\mu$ is supported in the compact $f_a([0,1])=\left[\frac{2}{3}\sum_{i=1}^n \frac{a_i}{3^{n-i}}, \frac{1}{3^n}+\frac{2}{3}\sum_{i=1}^n \frac{a_i}{3^{n-i}}\right]$. Whenever $a\neq b\in A^n$, the two sets $f_a([0,1])$ and $f_b([0,1])$ are separated by an interval of length at least $\frac{1}{3^n}$, and the sets $f_a([0,1])$ are ordered by the lexicographic order on A^n, i.e. if $a\leqslant b$, then all elements of $f_a([0,1])$ are inferior to all elements of $f_b([0,1])$.

Log-periodicity. For any $\nu\in\mathscr{P}_2(\mathbb{R})$, denote $\nu_h:=\exp_\nu(h\cdot F[\nu])$ the curve obtained by shifting half of the mass of ν towards the left, and half towards the right. As $f_k(x)\pm h=f_k(x\pm 3h)$ for $k\in\{0,1\}$, there holds $f_a(x)\pm h=f_a(x\pm 3^n h)$. The pushforward being linear, one has for any $h>0$ that

$$\mu_h=\left(\sum_{a\in A^n}\frac{1}{2^n} f_a\#\mu\right)_h=\sum_{a\in A^n}\frac{1}{2^n}(f_a\#\mu)_h=\sum_{a\in A^n}\frac{1}{2^n} f_a\#(\mu_{3^n h}).$$

Let $h_0\in[1/6,1/2)$ and $h_n:=3^{-n}h_0$. The support of $(f_a\#\mu)_{h_n}$ is contained in

$$\operatorname{supp} f_a\#\mu\pm 3^{-n}h_0\subset\left[\frac{2}{3}\sum_{i=1}^n\frac{a_i}{3^{n-i}}-\frac{1}{2\cdot 3^n}, \frac{1}{3^n}+\frac{2}{3}\sum_{i=1}^n\frac{a_i}{3^{n-i}}+\frac{1}{2\cdot 3^n}\right],$$

so that the supports of the shifted measures $(f_a\#\mu)_{h_n}$ for $a\in A^n$ do not overlap. Using Lemma 1.1.43, there holds

$$d_{\mathcal{W}}^2\left(\mu,\mu_{h_n}\right)=d_{\mathcal{W}}^2\left(\sum_{a\in A^n}\frac{1}{2^n}f_a\#\mu, \sum_{a\in A^n}\frac{1}{2^n}(f_a\#\mu)_{h_n}\right)=\sum_{a\in A^n}\frac{d_{\mathcal{W}}^2\left(f_a\#\mu,(f_a\#\mu)_{h_n}\right)}{2^n}=\sum_{a\in A^n}\frac{d_{\mathcal{W}}^2\left(f_a\#\mu, f_a\#(\mu_{3^n h_n})\right)}{2^n}.$$

Recalling that $f_a(x)=x/3^n+c$ for some constant $c=c_a$, one has $d_{\mathcal{W}}^2\left(f_a\#\mu,f_a\#\mu_{3^n h_n}\right)=3^{-2n}d_{\mathcal{W}}^2\left(\mu,\mu_{3^n h_n}\right)$ independently of $a\in A^n$. Taking square roots and dividing by $h_n=3^{-n}h_0$, we get

$$\frac{d_{\mathcal{W}}\left(\mu,\mu_{h_n}\right)}{h_n}=\frac{3^{-n}d_{\mathcal{W}}\left(\mu,\mu_{3^n h_n}\right)}{3^{-n}h_0}=\frac{d_{\mathcal{W}}\left(\mu,\mu_{h_0}\right)}{h_0}\qquad\forall n\in\mathbb{N}. \tag{5.32}$$

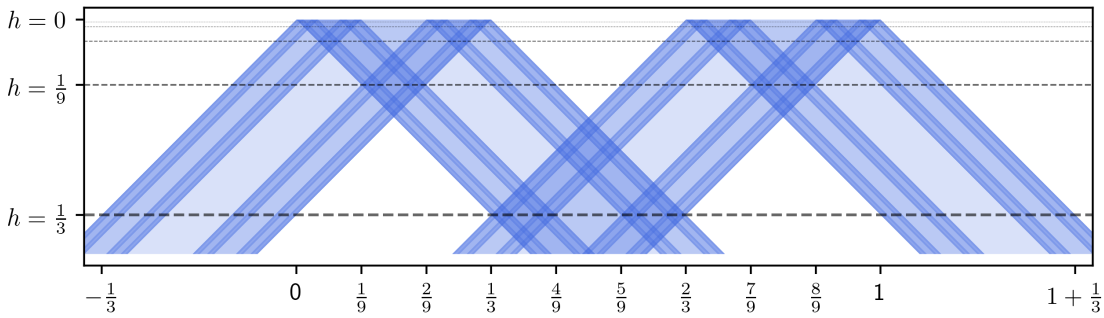


Figure 5.3: Evolution of $h \mapsto \mu_h = \exp_\mu(h \cdot F[\mu])$ for $h \in [0, 1/3]$.

Exploiting self-similarly, one sees that for $h = 1/9$, the optimal transport plan between μ and μ_h is the sum of two rescaled copies of the optimal plan for $h = 1/3$. The rescaling factor and the repartition of mass cancel precisely as to obtain $d_{\mathcal{W}}(\mu, \mu_{1/9})/(1/9) = d_{\mathcal{W}}(\mu, \mu_{1/3})/(1/3)$. The equality (5.32) generalizes this to any $h_n = 3^{-n}h_0$ for $h_0 \in [1/3, 1/2)$.

Behaviour in $h \in [1/6, 1/2)$**.** Since $h \mapsto d_{\mathcal{W}}(\mu, \mu_h)/h$ is log-periodic in h according to (5.32), it admits a limit when $h \searrow 0$ if and only if it is constant. This is not the case, as we show by computing at hand the Wasserstein distance for some well-chosen times h. For the sake of curiosity, we provide a numerical approximation of the curve for the remaining times.

The exact computation relies on the fact that the optimal map can be seen to behave piecewise polynomially. We first compute the integrals against μ of the canonical basis of second-order polynomials. One trivially has that $\int_{\mathbb{R}} 1 d\mu = 1$ and $\int_{\mathbb{R}} x d\mu = 1/2$. On the other hand, using the invariance property,

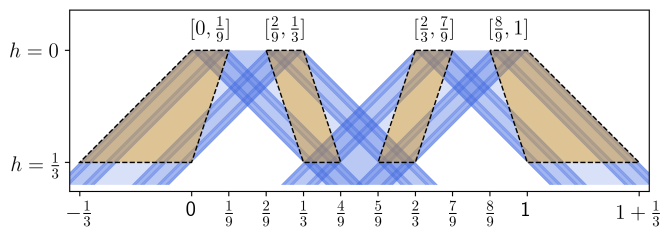
$$\int_{x\in\mathbb{R}} x^2 d\mu(x) = \int_{x\in\mathbb{R}} x^2 d\left[\frac{f_0\#\mu + f_1\#\mu}{2}\right](x) = \int_{x\in\mathbb{R}} \frac{(x/3)^2 + (x/3+2/3)^2}{2} d\mu(x) = \frac{1}{9}\int_{x\in\mathbb{R}} x^2 d\mu + \frac{1}{9} + \frac{2}{9},$$

and $\int_{\mathbb{R}} x^2 d\mu = 3/8$. Consequently, denoting $p_{a,b} : x \mapsto (ax+b)^2$, there holds

$$\int_{x\in\mathbb{R}} p_{a,b}(x) d\mu(x) = a^2 \int_{x\in\mathbb{R}} x^2 d\mu(x) + 2ab\int_{x\in\mathbb{R}} x d\mu(x) + b^2 = \frac{3}{8}a^2 + ab + b^2.$$

Let us consider $h = 1/3$. The measure μ is transport-regular since nonatomic, and there exists a unique transport map between μ and $\mu_h = \exp_\mu(h \cdot F[\mu])$. In dimension 1, this map can be computed explicitly from the pseudo-inverse of the distribution function. In the particular case of $h = 1/3$, it sends submeasures on affine transformations of themselves, and is given by

$$T(x) := \begin{cases} -1/3 + 3x & x \in [0, 1/9], \\ x + 1/9 & x \in [2/9, 1/3], \\ x - 1/9 & x \in [2/3, 7/9], \\ 1 + 3(x - 8/9) & x \in [8/9, 1]. \end{cases}$$



As $\mu(\cdot \cap [0, 1/9]) = \frac{1}{4} f_0\#f_0\#\mu$, one can compute

$$\int_{x\in[0,1/9]} |T(x) - x|^2 d\mu = \frac{1}{4}\int_{x\in[0,1]} \left|T(f_0(f_0(x))) - f_0(f_0(x))\right|^2 d\mu,$$

which is explicit since the integrand is a second-order polynomial in x. Proceeding in the same way on the three remaining pieces of the definition of T and summing up, one gets the exact value. This reasoning can be carried out for other times, and we collect some exact values in Figure 5.4 for $h \in \{\frac{4}{18}, \frac{5}{18}, \frac{1}{3}, \frac{7}{18}, \frac{4}{9}, \frac{1}{2}\}$.

One can also proceed by computing the exact transport map by the formula presented in [San15, Definition 2.3] in dimension 1, that is, $\gamma^{\text{opt}} = (F_\mu^{[-1]}, F_{\mu_h}^{[-1]})\#\mathcal{L}_{[0,1]}$, where $F_\nu : x \mapsto \mu((-\infty, x])$ is the distribution function of ν, and $F_\nu^{[-1]}$ is pseudo-inverse. The distribution function of μ is the Cantor staircase, and the distribution function of μ_h is the sum of two shifted staircases, that can be computed exactly for the approximation of μ by a sum of renormalized Lebesgue measures on the segments of C^n. This approximation is displayed in Figure 5.4 in blue for $n = 9$. We mention that the numerical approximations are visually undistinguishable from each other already for $n \geqslant 5$.

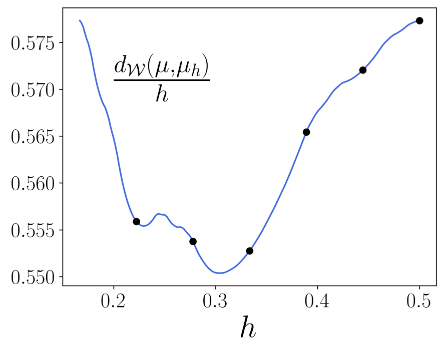


Figure 5.4: Approximation of $h \mapsto \frac{d_{\mathcal{W}}(\mu,\mu_h)}{h}$ for $h \in [1/6, 1/2)$.

The black dots are exact computations at times $h \in \left\{\frac{4}{18}, \frac{5}{18}, \frac{1}{3}, \frac{7}{18}, \frac{4}{9}, \frac{1}{2}\right\}$. The blue curve is a numerical approximation using renormalized Lebesgue measures on the segments of C^n, and the explicit formula of the transport map in 1D.

Conclusion of the example. The measure field $F[\mu]$ is centred, hence solenoidal for the non-atomic measure μ. However, the value of $h \mapsto \frac{d_{\mathcal{W}}(\mu, \exp_\mu(h \cdot F[\mu]))}{h}$ oscillates in an interval strictly contained in $[0,1] = [0, \|F[\mu]\|_\mu]$. Therefore (5.24b) does not admit a converse in general.

5.3.4.2 Unbalanced skinny Cantor measure

We turn to the construction of a measure on which $h \mapsto d_{\mathcal{W}}(\mu, \mu_h)/h = \|F[\mu]\|_\mu$, with $F[\mu]$ the measure field sending half of the mass to the left and half to the right, defined as

$$F[\mu] := \frac{(id, -1)\#\mu + (id, 1)\#\mu}{2},$$

and $\mu_h := \exp_\mu(h \cdot F[\mu]) = (id - h)\#\mu + (id + h)\#\mu$. Before entering the details, let us briefly describe how the example works. For a fixed positive $h > 0$, the class of measures μ such that $d_{\mathcal{W}}(\mu, \exp_\mu(h \cdot F[\mu])) = h\|F[\mu]\|_\mu$ is reduced to purely atomic measures with atoms separated by at least $2h$. We construct μ in a way that at time h_{n_k}, it is approximated by such an atomic measure with error $o(h_{n_k})$.

Construction of the measure. Let $(h_n)_{n\in\mathbb{N}} \subset (0,1]$ be a decreasing sequence converging to 0 such that $(h_{n+1}/h_n)_{n\in\mathbb{N}}$ is itself a decreasing sequence that converges to 0, for instance $h_n = \exp(-n^2)$. Up to a shift of index, we may assume that $h_{n_k+1}/h_{n_k} < 1/2$ for all k. Let $\mu^0 := \delta_0$. For each $k \in \mathbb{N}$, assume μ^k is already constructed, and given by $\mu^k = \frac{1}{2^k}\sum_{j=1}^{2^k} \delta_{x_j^k}$. Denote $n_k := \sum_{m=0}^{k} 2^m = 2^{k+1} - 1$. Define

$$\mu^{k+1} := \frac{1}{2^k}\sum_{j=1}^{2^k} \frac{\delta_{x_j^k - h_{n_k+j}} + \delta_{x_j^k + h_{n_k+j}}}{2}.$$

In words, each atom x_j^k is split in two children with an increasingly small distance h_{n_k+j}. Since $h_{n_k+j} \leqslant h_{n_k+1}$, the subsequent atoms issued from a given x_j^k stay at distance

$$\sum_{\ell \geqslant k} h_{n_\ell+1} = h_{n_k+1} \sum_{\ell \geqslant k} \prod_{m=k}^{\ell-1} \frac{h_{n_{j+1}+1}}{h_{n_j+1}} \leqslant h_{n_k+1} \sum_{\ell \geqslant k} (1/2)^{m-k} \leqslant 2h_{n_k+1}.$$

This implies that $d_{\mathcal{W}}(\mu^k, \mu^\ell) \leqslant 2h_{n_k+1}$ for all $\ell \geqslant k$, and the sequence $(\mu^k)_{k\in\mathbb{N}}$ is Cauchy. Denote $\bar{\mu}$ its limit. Each point of the support of $\bar{\mu}$ is the limit of a subsequence $(x_{j_k}^k)_{k\in\mathbb{N}}$, hence contained in the intersection of the sets $\mathcal{B}(x_{j_k}^k, 2h_{n_k+1})$. Since $\bar{\mu}\left(\mathcal{B}(x_{j_k}^k, 2h_{n_k+1})\right) = 2^{-k}$ for all k, $\bar{\mu}$ does not have atom. Consequently, $F[\bar{\mu}]$ belongs to $\mathbf{Sol}_{\bar{\mu}}$ by Theorem 5.2.23.

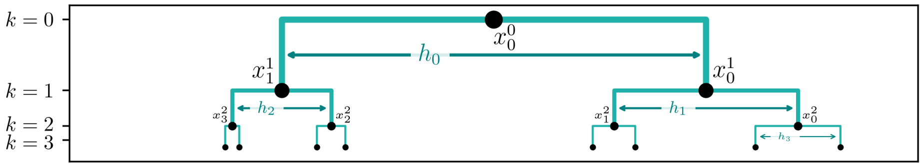


Figure 5.5: Schematic representation of the first iterations of the construction.

Atoms are represented by the black dots, and split at each iteration k. The term "unbalanced" refers to the fact that h_1 is larger than h_2, h_3 the largest distance between two children created at step $k=3$, and so on.

The solenoidal field. For any $\eta \in \Gamma_o(\mu,\nu)$, the plan $\frac{1}{2}(\pi_x - h, \pi_y - h)\#\eta + \frac{1}{2}(\pi_x + h, \pi_y + h)\#\eta \in \Gamma(\exp_\mu(h\cdot F[\mu]), \exp_\nu(h\cdot F[\nu]))$ provides the estimate $d_{\mathcal{W}}\left(\exp_\mu(h\cdot F[\mu]), \exp_\nu(h\cdot F[\nu])\right) \leqslant d_{\mathcal{W}}(\mu,\nu)$. From the second triangular inequality, one deduces for all $h>0$ and $k\in\mathbb{N}$ that

$$\frac{d_{\mathcal{W}}\left(\bar\mu, \exp_{\bar\mu}(h\cdot F[\bar\mu])\right)}{h} \geqslant \frac{d_{\mathcal{W}}\left(\mu^k, \exp_{\mu^k}(h\cdot F[\mu^k])\right)}{h} - \frac{4h_{n_k+1}}{h}.$$

For each h, consider k such that $h\in(h_{n_k}/2, h_{n_{k-1}}/2]$. We first show that $d_{\mathcal{W}}\left(\mu^k, \exp_{\mu^k}(h\cdot F[\mu^k])\right)$ can be computed by summing local contributions around the atoms of μ^{k-1}, then estimate each of these contributions from below.

Localization. The measure μ^k writes as $\frac{1}{2^{k-1}}\sum_{j=1}^{2^{k-1}}\mu^{k,j}$, where $\mu^{k,j} := \frac{1}{2}\left[\delta_{x_j - h_{n_{k-1}+j}} + \delta_{x_j + h_{n_{k-1}+j}}\right]$, and the points $(x_j)_{j\in[\![1,2^{k-1}]\!]}$ are the atoms of μ^{k-1} sorted in increasing order. Each $\exp_{\mu^{k,j}}(h\cdot F[\mu^{k,j}])$ is supported in the interval $[x_j - h_{n_{k-1}+j} - h, x_j + h_{n_{k-1}+j} + h]$. The distance between x_{j+1} and x_j is at least the smallest splitting distance at step $k-2$, that is, $2h_{n_{k-1}}$, so that for $h\leqslant h_{n_{k-1}}/2$,

$$\left(x_{j+1} - h_{n_{k-1}+j+1} - h\right) - \left(x_j + h_{n_{k-1}+j} + h\right) \geqslant 2h_{n_{k-1}} - 2h_{n_{k-1}+1} - 2h \geqslant 2h_{n_{k-1}}\left(\frac{1}{2} - \frac{h_{n_{k-1}+1}}{h_{n_{k-1}}}\right) \geqslant 0.$$

So, for such h, the measures μ^k and $\exp_{\mu^k}(h\cdot F[\mu^k])$ give the same mass to the mutually disjoint intervals $[x_j - h_{n_{k-1}+j} - h, x_j + h_{n_{k-1}+j} + h]$. By Lemma 1.1.43, transport happens only within these intervals, and

$$d_{\mathcal{W}}^2\left(\mu^k, \exp_{\mu^k}(h\cdot F[\mu^k])\right) = \frac{1}{2^{k-1}}\sum_{j=1}^{2^{k-1}} d_{\mathcal{W}}^2\left(\mu^{k,j}, \exp_{\mu^{k,j}}(h\cdot F[\mu^{k,j}])\right). \tag{5.33}$$

Estimate of each contribution. For $a\in\mathbb{R}$ and $d\geqslant 0$, the Wasserstein distance between $\nu := \frac{1}{2}\delta_{a-d} + \delta_{a+d}$ and $\exp_\nu(h\cdot F[\nu])$ is given by

$$d_{\mathcal{W}}^2\left(\nu, \exp_\nu(h\cdot F[\nu])\right) = \begin{cases} h^2 & h\leqslant d, \\ \frac{h^2}{2} + \frac{|h-2d|^2}{2} & h\geqslant d. \end{cases} \tag{5.34}$$

If $h\in(h_{n_k}/2, 2h_{n_k-1}]$, all terms in the sum (5.33) are equal to h^2. Otherwise, we must distinguish the cases where h is larger than the distance d separating the atoms issued from x_j, with $d = 2h_{n_{k-1}+j}$. Let $j^*\in[\![1,2^{k-1}]\!]$ be such that $2h_{n_{k-1}+j^*} < h \leqslant 2h_{n_{k-1}+j^*-1}$. If $j\leqslant j^*-1$, then $h\leqslant 2h_{n_{k-1}+j}$, and the corresponding j^{th} term of the sum is exactly equal to h^2. If $j\geqslant j^*$, then the corresponding term is computed using the second case in (5.34). Hence (5.33) may be estimated from below by

$$\frac{1}{2^{k-1}}\sum_{j=1}^{2^{k-1}} d_{\mathcal{W}}^2\left(\mu^{k,j}, \exp_{\mu^{k,j}}(h\cdot F[\mu^{k,j}])\right) \geqslant \frac{1}{2^{k-1}}\sum_{j=1}^{j^*-1} h^2 + 0 + \frac{1}{2^{k-1}}\sum_{j=j^*+1}^{2^{k-1}}\left[\frac{h^2}{2} + \frac{\left|h - 4h_{n_{k-1}+j}\right|^2}{2}\right].$$

For each $j\geqslant j^*+1$, one has $h > 2h_{n_{k-1}+j^*} \geqslant 4h_{n_{k-1}+j}$ since $h_{m+1}/h_m \leqslant 1/2$. Using the monotonicity of $(h_m)_m$ and $(h_{m+1}/h_m)_m$, we deduce that

$$\left|h - 4h_{n_{k-1}+j}\right| = h\left(1 - 4\frac{h_{n_{k-1}+j}}{h}\right) \geqslant h\left(1 - 2\frac{h_{n_{k-1}+j}}{h_{n_{k-1}+j^*}}\right) \geqslant h\left(1 - 2\frac{h_{n_{k-1}+j^*+1}}{h_{n_{k-1}+j^*}}\right) \geqslant h\left(1 - 2\frac{h_{n_{k-1}+1}}{h_{n_{k-1}}}\right).$$

Adding $\frac{h^2-h^2}{2^{k-1}}$ at the missing index j^* and coarsely estimating $h^2 \geqslant h^2\left(1-2\frac{h_{n_{k-1}+1}}{h_{n_{k-1}}}\right)^2$ for all j, we get

$$d_{\mathcal{W}}^2\left(\mu^k, \exp_{\mu^k}(h\cdot F[\mu^k])\right) \geqslant h^2\left(1-2\frac{h_{n_{k-1}+1}}{h_{n_{k-1}}}\right)^2 - \frac{h^2}{2^{k-1}}.$$

We can conclude that for $h \in (h_{n_k}/2, h_{n_{k-1}}/2]$,

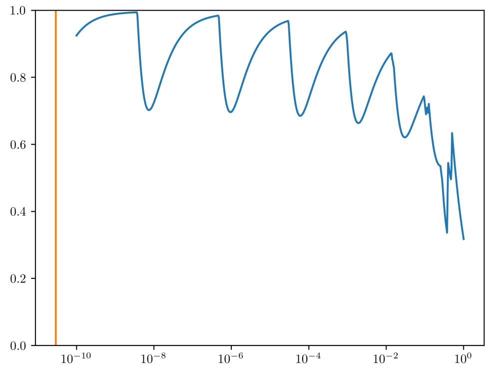
$$\frac{d_{\mathcal{W}}\left(\mu, \exp_\mu(h\cdot F[\mu])\right)}{h} \geqslant \frac{d_{\mathcal{W}}\left(\mu^k, \exp_{\mu^k}(h\cdot F[\mu^k])\right)}{h} - \frac{4h_{n_k+1}}{h} \geqslant \sqrt{\left(1-2\frac{h_{n_{k-1}+1}}{h_{n_{k-1}}}\right)^2 - \frac{1}{2^{k-1}}} - \frac{4h_{n_k+1}}{h_{n_k}/2}.$$

By assumption, $h_{m+1}/h_m \to_m 0$. When h goes to 0, the lower bound converges to $1 = \|F[\bar\mu]\|_{\bar\mu}$.

Conclusion of the example. The solenoidal measure field $F[\bar\mu]$ satisfies $\lim_{h\searrow 0}\frac{d_{\mathcal{W}}(\bar\mu, \exp_{\bar\mu}(h\cdot F[\bar\mu]))}{h} = \|F[\bar\mu]\|_{\bar\mu}$, which shows at once that both lines of (5.24) do not admit converses.

Remark 5.3.12 (Why unbalanced)**.** *In the above example, the atoms of a given μ^k all split with different radii. It would be simpler to consider a family $(\mu^k)_k$ such that μ^{k+1} splits all atoms of μ^k in two children separated by h_k.*

With this choice, $h \mapsto d_{\mathcal{W}}(\bar\mu, \exp_{\bar\mu}(h\cdot F[\bar\mu]))/h$ is very close to the sum of the contributions of each pairs of atoms around $h = h_k$, as in (5.33). However, the distance d in (5.34) is now common to all pairs, and for $h = 2d$, the value of $d_{\mathcal{W}}(\mu, \mu_h)/h$ decreases to $1/\sqrt{2} \sim 0.707$. The figure on the right shows a numerical approximation of $h \mapsto d_{\mathcal{W}}(\bar\mu, \bar\mu_h)/h$ in this case; the limit sup indeed goes to $1 = \|F[\bar\mu]\|_{\bar\mu}$, but not the limit inf.



5.4 $\mathscr{P}_2(\mathbb{R}^d)$ as a convex subset of the Banach space of measures

This section is concerned with operations in the Banach sense of measures, instead of horizontal interpolation. We first state an approximation result, then comment on a representation of solenoidal measure fields by vertical superposition of loops.

5.4.1 Vertical superpositions

Our first aim is to provide a representation of tangent measure fields as vertical sums of elements of the regular tangent space. Recall that the latter is defined as

$$\mathrm{Tan}_\mu := \overline{\left\{x \mapsto (x, \nabla\varphi(x)) \;\middle|\; \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^d;\mathbb{R})\right\}}^{L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)}. \tag{5.35}$$

By [Gig08, Theorem 4.15], the regular tangent space is precisely given as

$$\mathrm{Tan}_\mu = \left\{\mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi) \;\middle|\; \xi \in \mathbf{Tan}_\mu\right\} = \left\{f \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d) \;\middle|\; f\#\mu \in \mathbf{Tan}_\mu\right\}.$$

By definition, any element of the regular tangent space can be approximated by a smooth gradient. We focus on one possible generalization to the geometric tangent space $\mathbf{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$. The definition of $\mathbf{Tan}_\mu$ provides an approximation by reparametrized geodesics, but does not guarantee any regularity in space. At the opposite, we show in Proposition 5.4.4 below that any $\xi \in \mathbf{Tan}_\mu$ writes as $\int_{b\in\mathrm{Tan}_\mu} b\#\mu d\omega(b)$ for some measure $\omega \in \mathscr{P}_2(\mathrm{Tan}_\mu)$, in that

$$\int_{(x,v)\in\mathrm{T}\mathbb{R}^d} \varphi(x,v)d\xi(x,v) = \int_{b\in\mathrm{Tan}_\mu}\int_{x\in\mathbb{R}^d} \varphi(x,b(x))d\mu(x)d\omega(b) \qquad \forall\varphi \in \mathcal{C}_b(\mathrm{T}\mathbb{R}^d;\mathbb{R}). \tag{5.36}$$

Using Lemma 5.4.2, this provides a way to approximate $\xi \in \mathbf{Tan}_\mu$ by a vertical superposition of gradient functions, all sharing the same Lipschitz bound. It should be noted that the approximations may get out of $\mathbf{Tan}_\mu$.

Recall that $\mathrm{T}\mathbb{R}^d$ is the set of (x,v) such that $x \in \mathbb{R}^d$ and $v \in \mathrm{T}_x\mathbb{R}^d$, endowed with the Euclidean distance $(x,v),(y,w) \mapsto \sqrt{|x-y|^2+|v-w|^2}$. We consider the space $L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ equipped with the classical L^2_μ norm, and the set $\mathscr{P}_2(L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d))$ of measures over this space with finite second moment.

Lemma 5.4.1 (Existence). *Let $\xi \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$. There exists a superposition measure $\omega \in \mathscr{P}_2(L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d))$ such that $\xi = \int_{b\in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)} b\#\mu d\omega(b)$ in the sense of (5.36).*

Proof. The set $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ is a convex subset of the topological vector space of signed Borel measures on $\mathrm{T}\mathbb{R}^d$, endowed with the narrow topology. Since the latter is defined by duality against $\mathcal{C}_b(\mathrm{T}\mathbb{R}^d;\mathbb{R})$, it is a locally convex topological vector space. The extreme points of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ are exactly the elements that are induced by maps, i.e. of the form $f\#\mu$ for some $f \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$. Any intersection of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ with a Wasserstein ball is narrowly compact and vertically convex, so by Krein-Milman, each of its elements belong to the convex hull of the extreme points. The latter convex hull is described by barycenter of measures. Taking the union over the radii of balls, we obtained the desired result. □

The following approximation result follows quite directly. We state it for an abstract approximating class, with for instance $\mathcal{D}_C$ being the set of gradients of functions $\varphi \in \mathcal{C}^1(\mathbb{R}^d;\mathbb{R})$ supported in $\overline{\mathscr{B}}(0,C)$, Lipschitz with constant C and with gradient Lipschitz with constant C.

Lemma 5.4.2 (Approximation). *Let $(\mathcal{D}_C)_{C\geqslant 0} \subset L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ be an nondecreasing family of nonempty compact sets. Denote $A := \overline{\bigcup_{C\geqslant 0}\mathcal{D}_C}^{L^2_\mu} \subset L^2_\mu$. Then for any $C \geqslant 0$, there exists a measurable application $R_C : L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d) \to \mathcal{D}_C$ such that for all $\omega \in \mathscr{P}_2(L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d))$ with $\operatorname{supp}\omega \subset A$, there holds*

$$d_{\mathcal{W},L^2_\mu}(\omega, R_C\#\omega) \underset{C\to\infty}{\longrightarrow} 0.$$

Proof. By [AB06, Theorem 18.19], the minimization problem

$$\text{Find } \overline{b} \in \mathcal{D}_C \text{ such that } \|b-\overline{b}\|_{L^2_\mu} = \inf_{\beta\in\mathcal{D}_C} \|b-\beta\|_{L^2_\mu}$$

admits a measurable selection $R_C : L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d) \to \mathcal{D}_C$ of minimizers. Let $b_0 \in \mathcal{D}_0$: since $(\mathcal{D}_C)_C$ is an increasing family, $b_0 \in \mathcal{D}_C$ for all C, and there holds

$$\|b - R_C(b)\|_{L^2_\mu} \leqslant \|b-b_0\|_{L^2_\mu} \qquad \forall b \in L^2_\mu \text{ and } C \geqslant 0.$$

Moreover for a fixed $b \in A$, the application $C \mapsto \|b-R_C(b)\|_{L^2_\mu}$ is nonincreasing and goes to 0 when C goes to ∞. Hence, by Lebesgue's dominated convergence,

$$\limsup_{C\to\infty} d^2_{\mathcal{W}}(\omega, R_C\#\omega) \leqslant \limsup_{C\to\infty} \int_{b\in A} \|b-R_C(b)\|_{L^2_\mu} d\omega(b) = 0$$

for any $\omega \in \mathscr{P}_2(L^2_\mu)$ such that $\operatorname{supp}\omega \subset A$. □

Remark 5.4.3 (Estimate on the vertical superpositions). *Let $\xi,\xi' \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$, and $\omega,\omega' \in \mathscr{P}_2(L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d))$ such that $\xi = \int_{b\in L^2_\mu} b\#\mu d\omega$ and $\xi' = \int_{b\in L^2_\mu} b\#\mu d\omega'$ in the sense of (5.36). Then*

$$W_\mu\left(\xi,\xi'\right) \leqslant d_{\mathcal{W},L^2_\mu}\left(\omega,\omega'\right).$$

Indeed, let $P : \left(L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)\right)^2 \to \mathscr{P}_2(\mathrm{T}^2\mathbb{R}^d)_\mu$ be the application gluing two vector fields in a plan between both, i.e. $P(f,g) := \left(\pi_x, \pi_v(f(\pi_x)), \pi_v(g(\pi_x))\right)\#\mu$. Let $\beta = \beta(db,db') \in \Gamma_o(\omega,\omega')$. The transport plan $\int_{b,b'\in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)} P(b,b')d\beta$ belongs to $\Gamma_\mu(\xi,\xi')$, and provides the estimate

$$W^2_\mu\left(\xi,\xi'\right) \leqslant \int_{b,b'\in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)} \int_{(x,v,w)} |v-w|^2 \, d[P(b,b')](x,v,w) d\beta(b,b') = \int_{b,b'} \|b-b'\|^2_{L^2_\mu} d\beta = d^2_{\mathcal{W},L^2_\mu}(\omega,\omega').$$

We now come back to the original aim of the decomposition of $\mathbf{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$. Using the vertical convexity of the squared distance, we will prove that whenever $\xi \in \mathbf{Tan}_\mu$, the support of any representation by superposition is contained in $\mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$. This provides an application of the characterization of $\mathrm{Tan}_\mu = \mathbf{Tan}_\mu \cap L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)\#\mu$ by the directional derivative of the squared Wasserstein distance.

Proposition 5.4.4 (The case of $\mathbf{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$). *Let $\xi \in \mathbf{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$ and $\omega \in \mathscr{P}_2(L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d))$ such that $\xi = \int_{b\in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)} b\#\mu d\omega$ in the sense of (5.36). Then*

$$\operatorname{supp} \omega \subset \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d). \tag{5.37}$$

Proof. Let ξ, ω be as in the statement, and $h > 0$. Using that the pushforward is vertically linear, there holds $\exp_\mu(h\cdot\xi) = \mathrm{Bary}_{\mathscr{P}_2(\mathbb{R}^d)}\left([\exp_\mu(h\cdot\pi_b\#\mu)]\#\omega\right)$, in the sense that

$$\begin{aligned}\int_{y\in\mathbb{R}^d} \varphi(y) d\exp_\mu(h\cdot\xi)(y) &= \int_{(x,v)\in\mathbb{R}^d} \varphi(x+hv) d\xi(x,v) = \int_{b\in L^2_\mu}\int_{x\in\mathbb{R}^d} \varphi(x+hb(x)) d\mu(x) d\omega(b) \\ &= \int_{b\in L^2_\mu}\int_{(x,v)\in\mathrm{T}\mathbb{R}^d} \varphi(x,v) d[\exp_\mu(h\cdot b\#\mu)] d\omega(b) \qquad \forall \varphi \in C_b(\mathbb{R}^d;\mathbb{R}).\end{aligned}$$

So $\exp_\mu(h\cdot\xi)$ writes as a convex combination of measures. By the vertical convexity of $d^2_{\mathcal{W}}$ [Vil09, Th. 4.8],

$$\|\xi\|^2_\mu = \lim_{h\searrow 0} \frac{d^2_{\mathcal{W}}(\mu, \exp_\mu(h\cdot\xi))}{h^2} \leq \int_{b\in L^2_\mu} \limsup_{h\searrow 0} \frac{d^2_{\mathcal{W}}(\mu,\exp_\mu(h\cdot b\#\mu))}{h^2} d\omega(b) \leq \int_{b\in L^2_\mu} \|b\|^2_{L^2_\mu} d\omega(b) = \|\xi\|^2_\mu.$$

Hence equality holds. Since $\limsup_{h\searrow 0} \frac{d^2_{\mathcal{W}}(\mu,\exp_\mu(h\cdot b\#\mu))}{h^2} \leq \|b\|_{L^2_\mu}$ for all $b \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$, equality must hold for $\omega-$almost all $b \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$. By Proposition 5.3.8, that we can apply since $b\#\mu$ is induced by a map, this implies $b \in \mathrm{Tan}_\mu$. Since Tan_μ is closed in L^2_μ, we conclude. □

5.4.2 On the validity of Smirnov representations for solenoidal measure fields

Let us start this section with a counterpart of Proposition 5.4.4 for $\mathbf{Sol}_\mu$, that is not satisfied by $\mathbf{Tan}_\mu$.

Proposition 5.4.5 (Vertical convex combinations). *The set of solenoidal measure fields is vertically convex.*

Proof. Denote $\mathcal{G} \subset L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ the subset of vector fields such that $\frac{1}{2}\left[(-g)\#\mu + g\#\mu\right] \in \mathbf{Sol}^0_\mu$. By Proposition 5.2.20, $\zeta \in \mathbf{Sol}^0_\mu$ if and only if for any countable dense set $(g_n)_{n\in\mathbb{N}} \subset \mathcal{G}$, there holds $v \in \mathrm{span}\{g_n(x) \mid n \in \mathbb{N}\}$ for $\zeta-$a.e. $(x,v) \in \mathrm{T}\mathbb{R}^d$. Fix such a sequence $(g_n)_{n\in\mathbb{N}}$, and let us show that the same pointwise condition is satisfied by all solenoidal fields.

Consider first $f\#\mu \in \mathbf{Sol}_\mu$ for some $f \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$. By Proposition 5.2.9, $(-f)\#\mu$ is solenoidal. By the vertical convexity of the squared Wasserstein distance [Vil09, Theorem 4.8],

$$\frac{d^2_{\mathcal{W}}\left(\mu, \exp_\mu\left(h\cdot\frac{1}{2}\left[(-f)\#\mu + f\#\mu\right]\right)\right)}{h^2} \leq \frac{1}{2}\frac{d^2_{\mathcal{W}}\left(\mu,\exp_\mu(h\cdot f\#\mu)\right)}{h^2} + \frac{1}{2}\frac{d^2_{\mathcal{W}}\left(\mu,\exp_\mu(h\cdot(-f)\#\mu)\right)}{h^2}. \tag{5.38}$$

Since $(\pm f)\#\mu$ is map-induced, the limit in $h \searrow 0$ of the left-hand side goes to 0 by Proposition 5.3.8. By (5.24b), the centred field $\frac{1}{2}\left[(-f)\#\mu + f\#\mu\right]$ is solenoidal, so $f(x) \in \mathrm{span}\{g_n(x) \mid n \in \mathbb{N}\}$ for $\mu-$a.e. $x \in \mathbb{R}^d$. If now $\zeta \in \mathbf{Sol}_\mu$ is a general measure field, it is the unique element of $b_\zeta \oplus \zeta^0$ by Proposition 5.2.3. By Lemma 5.1.13, the barycenter field $b_\zeta\#\mu$ is solenoidal. The centred field $\zeta^0 \in \zeta \oplus (-b_\zeta\#\mu)$ is also solenoidal, and centred. Hence both of them satisfy $v \in \mathrm{span}\{g_n(x) \mid n \in \mathbb{N}\}$ almost everywhere, and so does ζ. Since this condition is related to the support of ζ, it passes to vertical convex combinations.

To conclude, let $\zeta_0, \zeta_1 \in \mathbf{Sol}_\mu$ and $\lambda \in [0,1]$. The barycenter $b_\lambda \in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ of $\zeta_\lambda := (1-\lambda)\zeta_0 + \lambda\zeta_1$ writes as $(1-\lambda)b_0 + \lambda b_1$, thus $b_\lambda\#\mu$ is solenoidal by horizontal convex combinations). The centred part $\zeta^0_\lambda = (\pi_x, \pi_v - b_\lambda(\pi_x))\#\zeta_\lambda$ satisfies $v \in \mathrm{span}\{g_n(x) | n \in \mathbb{N}\}$ almost everywhere, so that Proposition 5.2.20 ensures that $\zeta^0_\lambda \in \mathbf{Sol}^0_\mu$. Writing $\zeta_\lambda = b_\lambda\#\mu \oplus \zeta^0_\lambda$, we get that $\zeta_\lambda \in \mathbf{Sol}_\mu$. □

Proposition 5.4.5 implies that any superposition of map-induced solenoidal fields stays solenoidal. The interesting question is the converse: can any solenoidal measure field be decomposed into a superposition of map-induced solenoidal measure fields? Formulated in this way, the answer is negative; in dimension one, the only map-induced solenoidal field is 0_μ by Theorem 5.2.23, but there may be plenty of nonzero solenoidal measure fields. The result that comes closest to a vertical superposition representation is, to our opinion, the Smirnov decomposition [Smi94]. This very nice theorem states that a divergence-free field can be represented as the vertical superposition of "elementary solenoids", that are currents generalizing parametrized curves with unit speed winding around tori.

The situation for general solenoidal fields is not that clear. On the one hand, map-induced solenoidal fields generate a "flow" in some sense, which sends the particles along the Lipschitz curves used to define elementary solenoids in [Smi94, Def. p. 847]. It could seem intuitive to extend this definition by allowing self-intersection; for instance, taking μ as the Lebesgue measure on $[0,1]$, the solenoidal field $\frac{(id,-1)\#\mu+(id,1)\#\mu}{2}$ could be represented with one single loop going from 0 to 1 and back from 1 to 0 at speed one. On the other hand, the same construction is solenoidal for the Hausdorff measure on the Cantor set, which has a totally disconnected support. This seems to prevent the existence of a flow parametrized by continuous curves... Note that for $\mathcal{L}_{[0,1]}$ or the Cantor measure, the only barycentric solenoidal field is 0_μ.

Here, we focus on a simpler representation of solenoidal measure fields, that holds only for the measures μ on which the converse of (5.24b) stands. In this case, $\mathbf{Sol}_\mu$ writes as the closed cone over a particular subset of plans, in the same way that $\mathbf{Tan}_\mu$ is the closed cone over velocities of geodesics. This subset if made of the measure fields in $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ that pass through μ on a short time.

Lemma 5.4.6 (Regularity of $\mathbf{Sol}_\mu$). *The following conditions are equivalent:*

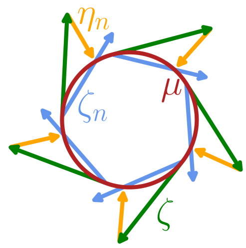
1. *for any $\zeta \in \mathbf{Sol}_\mu$, there holds $d_{\mathcal{W}}(\mu, \exp_\mu(h\cdot\zeta)) = o(h)$,*

2. *for any $\zeta \in \mathbf{Sol}_\mu$ and any vanishing sequence $(h_n)_{n\in\mathbb{N}} \subset (0,1]$, there exists $(\zeta_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ such that $\exp_\mu(h_n\cdot\zeta_n) = \mu$, and $W_\mu(\zeta,\zeta_n) \to_n 0$.*

The link with a Smirnov representation is that a superposition on curves could be constructed by gluing trajectories following the fields ζ_n, which end and start from the same measure, and pass to the limit. This would define a flow of ζ that leaves μ invariant.

Proof. If Point 1 holds, let $\zeta \in \mathbf{Sol}_\mu$, and $(h_n)_{n\in\mathbb{N}}$ a vanishing sequence. We build ζ_n by composing ζ with an optimal transport plan projecting back $\exp_\mu(h_n\cdot\zeta)$ on μ. For each n, let $\eta_n \in \exp^{-1}_{\exp_\mu(h_n\cdot\zeta)}(\mu)$ and $\alpha^n \in \Gamma_{\exp_\mu(h_n\cdot\zeta)}\left(\eta_n, (\pi_x + h_n\pi_v, -\pi_v)\#\zeta\right)$. Let

$$\zeta_n := \left(\pi_x + h_n\pi_w, \frac{\pi_x+\pi_v-(\pi_x+h_n\pi_w)}{h_n}\right)\#\alpha^n$$



Then $\pi_x\#\zeta_n = (\pi_x + h_n\pi_w)\#\alpha^n = (\pi_x + h_n\pi_v + h_n(-\pi_v))\#\zeta = \mu$, and

$$\begin{aligned}\exp_\mu(h_n\cdot\zeta_n) &= (\pi_x + h_n\pi_v)\#\zeta_n = (\pi_x + h_n\pi_w + (\pi_x+\pi_v-(\pi_x+h_n\pi_w)))\#\alpha^n\\ &= (\pi_x+\pi_v)\#\eta_n = \mu.\end{aligned}$$

This implies that

$$\begin{aligned}W_\mu^2(\zeta,\zeta_n) &\leqslant \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} |v-w|^2\, d\left(\pi_x + h_n\pi_w, -\pi_w, \frac{\pi_x+\pi_v-(\pi_x+h_n\pi_w)}{h_n}\right)\#\alpha^n\\ &= \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \left|-w-\frac{v-h_nw}{h_n}\right|^2 d\alpha^n = \frac{1}{h_n^2}\int_{(x,v)\in\mathrm{T}\mathbb{R}^d} |v|^2 d\eta_n = \frac{d^2_{\mathcal{W}}\left(\mu,\exp_\mu(h_n\cdot\zeta)\right)}{h_n^2}.\end{aligned}$$

By assumption, the latter quantity goes to 0 when n goes to $+\infty$, proving Point 2.

Assume now Point 2. Let $\zeta \in \mathbf{Sol}_\mu$, and $(h_n)_{n\in\mathbb{N}}$ realize the limit sup of $d_{\mathcal{W}}(\mu,\exp_\mu(h\cdot\zeta))/h$ when $h \searrow 0$. By assumption, there exists $(\zeta_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ with $\exp_\mu(h_n\cdot\zeta_n) = \mu$ and $W_\mu(\zeta,\zeta_n) \to_n 0$, so that

$$\lim_{n\to\infty} \frac{d_{\mathcal{W}}(\mu,\exp_\mu(h_n\cdot\zeta))}{h_n} = \lim_{n\to\infty} \frac{d_{\mathcal{W}}(\exp_\mu(h_n\cdot\zeta_n),\exp_\mu(h_n\cdot\zeta))}{h_n} \leqslant \lim_{n\to\infty} W_\mu(\zeta_n,\zeta) = 0.$$

Hence Point 1. □

If μ is purely atomic in dimension 1, then Lemma 5.4.6 holds trivially, since $\mathbf{Sol}_\mu = 0_\mu$. On the other hand, we prove that any absolutely continuous measure also satisfies Lemma 5.4.6. The examples in Section 5.3.4 show that Lemma 5.4.6 does not always hold; we conjecture that it fails as soon as μ has a Cantor part.

Lemma 5.4.7 (Absolutely continuous measures in dimension 1). *Let $\mu = \rho d\mathcal{L}$ for some $\rho \in L^1(\mathbb{R};\mathbb{R}^+)$ of integral 1. For any $\zeta \in \mathbf{Sol}_\mu$, there holds $d_\mathcal{W}\left(\mu, \exp_\mu(h\cdot\zeta)\right) = o(h)$.*

The core of the proof lies in (5.39), which uses the invariance by translation of the Lebesgue measure.

Proof. For any compact interval $I \subset \mathbb{R}$, denote $\mathcal{L}_I \in \mathscr{P}_2(\mathbb{R})$ the normalized Lebesgue measure on I.

Elementary case. Let $\mu = \mathcal{L}_{[0,1]}$ and $\zeta := \sum_{i=1}^N \alpha_i (id, v_i)\#\mu$ for a finite sequence $(\alpha_i)_{i\in[\![1,N]\!]} \subset [0,1]$ summing to one, and a finite collection of vectors $v_i \in \mathbb{R}$. Let $M := \max_{i\in[\![1,N]\!]} |v_i|$. For $0 < h < 1/(2M)$, there holds

$$\exp_\mu(h\cdot\zeta) = \sum_{i=1}^N \alpha_i \mathcal{L}_{[hv_i, 1+hv_i]} = hM\nu_h^0 + \mathcal{L}_{[hM,1-hM]} + hM\nu_h^1, \tag{5.39}$$

where ν_h^0 and ν_h^1 have mass one, and are respectively supported in $[-hM, hM]$ and $[1-hM, 1+hM]$. Hence

$$d_\mathcal{W}^2\left(\mu, \exp_\mu(h\cdot\zeta)\right) \leqslant hM d_\mathcal{W}^2\left(\mathcal{L}_{[0,hM]}, \nu_h^0\right) + 0 + hM d_\mathcal{W}^2(\mathcal{L}_{[1-hM,1]}, \nu_h^1) \leqslant 2hM(hM)^2,$$

and $d_\mathcal{W}(\mu, \exp_\mu(h\cdot\zeta)) = o(h)$. By density, this extends to any centred measure field that is constant in the space variable. If $\nu = \mathcal{L}_{[a,b]}$, let $\tau : x \in [0,1] \mapsto a + (b-a)x$, so that $\nu = \tau\#\mu$. Any $\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R})_\nu$ centred and constant in the space variable writes as $(\tau(\pi_x), (b-a)\pi_v)\#\xi$ for some centred constant field in $\mathscr{P}_2(\mathrm{T}\mathbb{R})_\mu$. Taking any $\alpha \in \Gamma_o(\mu, \exp_\mu(h\cdot\xi))$, we get

$$d_\mathcal{W}^2(\nu, \exp_\nu(h\cdot\zeta)) \leqslant \int_{(x,v)\in\mathrm{T}\mathbb{R}} |\tau(x) - (\tau(x) + h(b-a)\pi_v)|^2\, d\alpha = (b-a)^2 d_\mathcal{W}^2(\mu, \exp_\mu(h\cdot\xi)) = o(h).$$

Approximation on the measure field. By Lemma 5.4.1, there exists $\omega \in \mathscr{P}_2(L^2(0,1;\mathrm{T}\mathbb{R}))$ such that $\zeta = \int_{b\in L^2(0,1;\mathrm{T}\mathbb{R})} b\#\mu d\omega$ in the vertical sense, that is, $\int_{\mathrm{T}\mathbb{R}} \varphi(x,v) d\zeta = \int_{b\in L^2(0,1;\mathrm{T}\mathbb{R})} \int_{x\in\mathbb{R}} \varphi(x, b(x)) d\mu(x) d\omega(b)$ for any $\varphi \in \mathcal{C}_b(\mathrm{T}\mathbb{R};\mathbb{R})$. Let $D_N \subset L^2(0,1;\mathrm{T}\mathbb{R})$ be the set of vector fields that are piecewise constant with at most N pieces. Each D_N is compact in L^2, and $\overline{\bigcup_{N\in\mathbb{N}} D_N} = L^2(0,1;\mathrm{T}\mathbb{R})$, where the closure is taken in L^2. By Lemma 5.4.2, there exists a sequence $\omega_N^{(0)} \subset \mathscr{P}_2(L^2(0,1;\mathrm{T}\mathbb{R}))$, converging towards ω with respect to $d_{\mathcal{W},L^2(0,1;\mathrm{T}\mathbb{R})}$ when $N \to \infty$, and such that $\operatorname{supp} \omega_N^{(0)} \subset D_N$. The measure field $\zeta_N^{(0)} := \int_{b\in L^2(0,1;\mathrm{T}\mathbb{R})} b\#\mu d\omega_N^{(0)}$ may not be centred; to obtained a centred measure field, define $\omega_N^{(1)} := \left(\pi_b - \mathrm{Bary}_{\mathrm{T}\mathbb{R}}\left(\zeta_N^{(0)}\right)\right)\#\omega_N^{(0)}$. The centred measure field $\zeta_N^{(1)} := \int_{b\in L^2(0,1;\mathrm{T}\mathbb{R})} b\#\mu d\omega_N^{(1)}$ satisfies

$$W_\mu\left(\zeta, \zeta_N^{(1)}\right) \leqslant W_\mu\left(\zeta, \zeta_N^{(0)}\right) \leqslant d_{\mathcal{W},L^2(0,1;\mathrm{T}\mathbb{R})}\left(\omega, \omega_N^{(0)}\right) \underset{N\to\infty}{\longrightarrow} 0.$$

Here we used respectively the Pythagoras estimate Lemma 5.2.2 for barycentric/centred decompositions, and the estimate of Remark 5.4.3. The measure field $\zeta_N^{(1)}$ writes as $\sum_{m\in[\![1,M]\!]} \alpha_m \zeta_{N,m}^{(1)}$, where $\pi_x\#\zeta_{N,m}^{(1)} = \mathcal{L}_{I_m}$ for some mutually disjoint intervals I_m covering $[0,1]$, the masses $\alpha_m \in [0,1]$ are summing to 1 and each $\zeta_{N,m}^{(1)}$ is constant in space. The pushforward being linear, $\exp_\mu(h\cdot\zeta_N^{(1)}) = \sum_{m\in[\![1,M]\!]} \alpha_m \exp_{\mathcal{L}_{I_m}}(h\cdot\zeta_{N,m}^{(1)})$, and

$$d_\mathcal{W}^2\left(\mu, \exp_\mu(h\cdot\zeta_N^{(1)})\right) \leqslant \sum_{m\in[\![1,M]\!]} \alpha_m d_\mathcal{W}^2\left(\mathcal{L}_{I_m}, \exp_{\mathcal{L}_{I_m}}(h\cdot\zeta_{N,m}^{(1)})\right) = o(h)$$

by the previous step. We might now conclude: for any $\varepsilon > 0$, there exists N such that $W_\mu\left(\zeta, \zeta_N^{(1)}\right) \leqslant \varepsilon$, and

$$d_\mathcal{W}(\mu, \exp_\mu(h\cdot\zeta)) \leqslant d_\mathcal{W}(\mu, \exp_\mu(h\cdot\zeta_N^{(1)})) + d_\mathcal{W}(\exp_\mu(h\cdot\zeta_N^{(1)}), \exp_\mu(h\cdot\zeta)) \leqslant o(h) + hW_\mu\left(\zeta, \zeta_N^{(1)}\right) = o(h) + h\varepsilon,$$

so dividing by h, we get that $\limsup_{h\searrow 0} \frac{d_{\mathcal{W}}(\mu,\exp_\mu(h\cdot\zeta))}{h} \leqslant \varepsilon$ for $\varepsilon > 0$ arbitrary. This concludes the case of $\mu = \mathcal{L}_{[0,1]}$. The case of $\mathcal{L}_I$ for some nontrivial bounded interval I is deduced as in the previous step.

Approximation on the measure. Let now $\mu = \rho d\mathcal{L}$ and $\zeta \in \mathbf{Sol}_\mu$. Assume by contradiction that

$$\limsup_{h\searrow 0} \frac{d_{\mathcal{W}}\left(\mu,\exp_\mu(h\cdot\zeta)\right)}{h} > 0. \tag{5.40}$$

If ρ is not piecewise constant on finitely many intervals, approximate it by such functions. Let $(\rho_n)_{n\in\mathbb{N}} \subset L^1(\mathbb{R};\mathbb{R}+)$ be a nondecreasing sequence of simple functions such that $\rho = \lim_n \rho_n$ in L^1. Up to removing terms, we may assume that $\int \rho_{n+1} d\mathcal{L} > \int \rho_n d\mathcal{L} > \cdots \int \rho_0 d\mathcal{L} > 0$. Then, defining

$$\mu_0 = \frac{\rho_0 \mathcal{L}}{\int_{x\in\mathbb{R}} \rho_0(x) d\mathcal{L}(x)} \qquad \text{and} \qquad \mu_{n+1} := \frac{(\rho_{n+1}-\rho_n)\mathcal{L}}{\int_{x\in\mathbb{R}} (\rho_{n+1}(x)-\rho_n(x)) d\mathcal{L}(x)} \quad \forall n \in \mathbb{N},$$

we obtain a countable family of probability measures such that $\mu = \sum_{n\in\mathbb{N}} m_n \mu_n$ for a sequence $m_n \in (0,1]$ summing to 1, each writing as a finite sum of renormalized Lebesgue measures on intervals. Disintegrate $\zeta = \zeta_x \otimes \mu$ for some measurable family $(\zeta_x)_{x\in\mathbb{R}}$, and define $\zeta_n := \int_{x\in\mathbb{R}} \zeta_x d\mu_n(x)$. Then

$$d^2_{\mathcal{W}}\left(\mu,\exp_\mu(h\cdot\zeta)\right) \leqslant \sum_{n\in\mathbb{N}} m_n d^2_{\mathcal{W}}\left(\mu_n,\exp_{\mu_n}(h\cdot\zeta_n)\right).$$

As $n \mapsto m_n d^2_{\mathcal{W}}\left(\mu_n,\exp_{\mu_n}(h\cdot\zeta_n)\right) \leqslant m_n\|\zeta_n\|_{\mu_n}$ is in ℓ^1, there must be at least one $n \in \mathbb{N}$ such that (5.40) holds with μ_n,ζ_n in place of μ,ζ. Writing $\mu_n = \sum_{m\in[\![0,M]\!]} m_{n,m}\mu_{n,m}$ for $\mu_{n,m} = \mathcal{L}_{I_m}$, and denoting $\zeta_{n,m} := \zeta_n|_{I_m\times\mathbb{R}} = \zeta_n(\cdot\cap(I_m\times\mathbb{R}))/\zeta_n(I_m\times\mathbb{R})$ a centred field, we get

$$0 < \limsup_{h\searrow 0} \frac{d^2_{\mathcal{W}}\left(\mu_n,\exp_{\mu_n}(h\cdot\zeta_n)\right)}{h^2} \leqslant \sum_{m\in[\![0,M]\!]} m_{n,m} \limsup_{h\searrow 0} \frac{d^2_{\mathcal{W}}\left(\mu_{n,m},\exp_{\mu_{n,m}}(h\cdot\zeta_{n,m})\right)}{h^2},$$

which is absurd by the previous step. □

5.5 Decomposition in submeasures with tangent cones of a given dimension

This section exploits two simple facts: the Chasles relation, and the fact that the metric scalar product between centred fields is nonnegative. We start by Chasles. For any measurable $A \subset \mathbb{R}^d$, denote $\mathrm{T}A := \{(x,v) \mid x \in A,\ v \in \mathrm{T}_x\mathbb{R}^d\}$, and $\xi|_{\mathrm{T}A} := \xi(\cdot\cap\mathrm{T}A)/\xi(\mathrm{T}A)$ (with the convention $\xi|_{\mathrm{T}A} = 0$ if $\xi(\mathrm{T}A) = 0$). There holds for any $\xi,\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ that

$$\langle\xi,\zeta\rangle^+_\mu = \int_{x\in A} \langle\xi_x,\zeta_x\rangle^+_{\delta_x} d\mu + \int_{x\in A^c} \langle\xi_x,\zeta_x\rangle^+_{\delta_x} d\mu = \mu(A)\,\langle\xi|_{\mathrm{T}A},\zeta|_{\mathrm{T}A}\rangle^+_{\mu|_A} + \mu(A^c)\,\langle\xi|_{\mathrm{T}A^c},\zeta|_{\mathrm{T}A^c}\rangle^+_{\mu|_{A^c}}. \tag{5.41}$$

If $\mu(A) = 0$, the corresponding term can be omitted. This is implicit in the sequel. Secondly, if $\xi,\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ are centred, then we may construct a transport plan $\alpha \in \Gamma_\mu(\xi,\zeta)$ by taking pointwise product measures between the disintegrations, and obtain

$$\langle\xi,\zeta\rangle^+_\mu \geqslant \int_{(x,v,w)\in\mathrm{T}^2\mathbb{R}^d} \langle v,w\rangle\, d\alpha = \int_{x\in\mathbb{R}^d} \langle \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi)(x), \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\zeta)(x)\rangle\, d\mu = 0.$$

The combination of these identities allows us to show that $\mathbf{Sol}^0_\mu$ and $\mathbf{Tan}^0_\mu$ depend essentially from the local structure of the measure μ. This is very different from $\mathbf{Sol}_\mu$ and $\mathbf{Tan}_\mu$, which behave more like affine spaces with vector components given by their centred subsets. The main result is the decomposition provided in Theorem 5.5.9.

5.5.1 Locality of $\mathbf{Tan}^0_\mu$ and $\mathbf{Sol}^0_\mu$

As seen in Propositions 5.2.8 and 5.2.9, the sets $\mathbf{Tan}_\mu$ and $\mathbf{Sol}_\mu$ are stable by multiplication $\cdot$ by a scalar, defined as $\lambda\cdot\xi = (\pi_x, \lambda\pi_v)\#\xi$. Extend this operation into $\cdot : L^\infty_\mu(\mathbb{R}^d;\mathbb{R}) \times \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu \to \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ as

$$\lambda\cdot\xi := (\pi_x, \lambda(\pi_x)\cdot\pi_v)\#\xi.$$

Then it may happen that an element ζ is solenoidal, but $\lambda\cdot\zeta$ is not for some $\lambda \in L^\infty_\mu(\mathbb{R}^d;\mathbb{R})$.

Let μ be the Hausdorff measure on the unit square in dimension $d = 2$. Let $\zeta := f\#\mu$, where

$$f(x_1, 0) = (-1, 0), \quad f(0, x_2) = (0, 1), \quad f(x_1, 1) = (1, 0), \quad f(1, x_2) = (0, -1).$$

The exponential $h \mapsto \exp_\mu(h \cdot \zeta)$ slides each side of the square in the associated direction. For each $h > 0$, construct a transport plan α_h between μ and $\exp_\mu(h \cdot \zeta)$ by folding the part of the sides that slid out of the square back on it, as depicted in dotted line in Figure 5.6. Since most of the mass is not moved by this plan, we get the estimate

$$d_{\mathcal{W}}^2 \left(\mu, \exp_\mu(h \cdot \zeta) \right) \leqslant 4 \times \frac{1}{4} \int_{s=0}^{h} \left| \sqrt{2} s \right|^2 ds = 2 \frac{h^3}{3}.$$

By (5.24b), ζ is solenoidal. Consider now $\lambda(x) = \mathbb{1}_{x_1 \geqslant 1}$, for which only the right side of the square moves downwards. Then we might compute that $\mathrm{Bary}_{\mathbb{R}^d} \left(\exp_\mu(h\lambda \cdot \zeta) \right) = \mathrm{Bary}_{\mathbb{R}^d}(\mu) - (0, h/4)$, so that

$$\frac{h}{4} = d_{\mathcal{W}} \left(\delta_{\mathrm{Bary}_{\mathbb{R}^d}(\mu)}, \delta_{\mathrm{Bary}_{\mathbb{R}^d}\left(\exp_\mu(h\lambda \cdot \zeta)\right)} \right) \leqslant d_{\mathcal{W}} \left(\mu, \exp_\mu(h\lambda \cdot \zeta) \right).$$

As the converse of (5.24b) holds for map-induced fields by Proposition 5.3.8, $\lambda \cdot \zeta$ is not solenoidal.

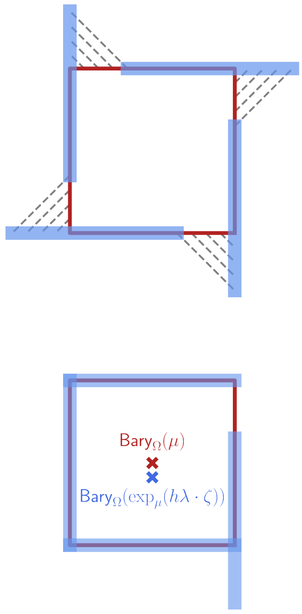


Figure 5.6: $\zeta \in \mathrm{Sol}_\mu$, whereas $\lambda \cdot \zeta \notin \mathrm{Sol}_\mu$.

However, the *centred* solenoidal measure fields are stable by such pointwise renormalizations.

Lemma 5.5.1 (Stability by pointwise renormalisation). *For any $\zeta \in \mathbf{Sol}_\mu^0$ (resp. $\xi \in \mathbf{Tan}_\mu^0$) and $\lambda \in L_\mu^\infty(\mathbb{R}^d; \mathbb{R})$, there holds $\lambda \cdot \zeta \in \mathbf{Sol}_\mu^0$ (resp. $\lambda \cdot \xi \in \mathbf{Tan}_\mu^0$).*

Proof. Let $\zeta \in \mathbf{Sol}_\mu^0$, $\lambda \in L_\mu^\infty(\mathbb{R}^d; \mathbb{R})$ and $\xi \in \mathbf{Tan}_\mu$. Denoting ξ^0 the centred component of ξ, there holds by Proposition 5.2.3 that $\langle \xi, \lambda \cdot \zeta \rangle_\mu^+ = \langle \xi^0, \lambda \cdot \zeta \rangle_\mu^+$. We first note that on centred fields, orthogonality with respect to $\langle \cdot, \cdot \rangle_\mu^+$ implies orthogonality of the disintegrations μ−almost everywhere. Indeed,

$$0 = \langle \xi^0, \zeta \rangle_\mu^+ = \int_{x \in \mathbb{R}^d} \langle \xi_x^0, \zeta_x \rangle_{\delta_x}^+ d\mu. \tag{5.42}$$

Using the product measure between ξ_x^0 and ζ_x, there holds $\langle \xi_x^0, \zeta_x \rangle_{\delta_x}^+ \geqslant \langle \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\xi_x^0), \mathrm{Bary}_{\mathrm{T}\mathbb{R}^d}(\zeta_x) \rangle = 0$ μ−almost everywhere, so that equality must hold μ−a.e. as well.

Let now $\alpha \in \Gamma_{\mu,o}(\xi^0, \lambda \cdot \zeta)$, and $\beta \in \Gamma_\mu(\xi^0, \zeta)$ such that $\alpha = (\pi_x, \pi_v, \lambda(\pi_x)\pi_w)\#\beta$. The plan β can be constructed as $(\pi_x, \pi_v, \lambda^{-1}(\pi_x) \cdot \pi_w)$ on any set where $\lambda \neq 0$ μ−almost everywhere, and arbitrarily otherwise. Then, disintegrating $\beta = \beta_x \otimes \mu$ for a measurable family $(\beta_x)_{x \in \mathbb{R}^d}$ whose marginals provide disintegrations of ξ^0 and ζ,

$$\langle \xi^0, \lambda \cdot \zeta \rangle_\mu^+ = \int_{(x,v,w) \in \mathrm{T}^2\mathbb{R}^d} \langle v, w \rangle \, d\alpha = \int_{x \in \mathbb{R}^d} \lambda(x) \int_{(v,w) \in \mathrm{T}_x^2\mathbb{R}^d} \langle v, w \rangle \, d\beta_x d\mu \leqslant \int_{x \in \mathbb{R}^d} |\lambda(x)| \langle \xi_x^0, \zeta_x \rangle_{\delta_x}^+ d\mu = 0.$$

By Proposition 5.2.9, this is enough to get that $\lambda \cdot \zeta \in \mathbf{Sol}_\mu$. The reasoning on $\mathbf{Tan}_\mu^0$ is symmetric. □

To go on, we show that the centred solenoidal spaces are stable by restriction to a subset, in the sense that $\mathbf{Sol}_{\mu|_A}^0 = \mathbf{Sol}_\mu^0|_{\mathrm{T}A}$ for any measurable set $A \subset \mathbb{R}^d$. Our strategy is to characterize solenoidal measure fields by orthogonality against a smaller subset of $\mathbf{Tan}_\mu^0$, namely the velocities of geodesics targeting compactly supported measures. These fields can be "extended" as follows.

Lemma 5.5.2 (Extension of optimal plan). *Let $\mu = (1-\lambda)\mu_1 + \lambda\mu_2 \in \mathscr{P}_2(\mathbb{R}^d)$, for $\mu_i \in \mathscr{P}_2(\mathbb{R}^d)$ and $\lambda \in [0,1]$. Let $\eta \in \exp_{\mu_1}^{-1}(\nu)$, where $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ is compactly supported. Then there exists $\gamma \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_{\mu_2}$ such that*

$$\xi = (1-\lambda)\eta + \lambda\gamma \tag{5.43}$$

is the velocity of a geodesic issued from μ.

Proof. Let $M \geqslant 0$ be such that $\operatorname{supp} \nu \subset \overline{\mathscr{B}}(0, M)$, and $(x_0, v_0) \in \operatorname{supp}(\pi_x, \pi_x + \pi_v)\#\eta$. The formula

$$\varphi(x) := \sup\left\{ \sum_{i=0}^{n} |x_i - y_i|^2 - \sum_{i=0}^{n-1} |x_{i+1} - y_i|^2 - |x - y_n|^2 \;\middle|\; n \in \mathbb{N},\ (x_i, y_i)_{i=1}^{n} \subset \operatorname{supp}(\pi_x, \pi_x + \pi_v)\#\eta \right\}$$

defines a semiconvex function from $\mathbb{R}^d$ to $\mathbb{R} \cup \{\infty\}$ with the property that $\varphi(x) + \varphi^c(y) = |x - y|^2$ for $(\pi_x, \pi_x + \pi_v)\#\eta$–almost (x, y) [Vil09, Theorem 5.10]. The support of $(\pi_x, \pi_x + \pi_v)\#\eta$ is cyclically monotone, so taking $x = x_0$, we get that $\varphi(x_0) \leqslant 0$. Since each y_i appearing in the supremum is contained in $\operatorname{supp}(\pi_x + \pi_v)\#\eta = \operatorname{supp} \nu \subset \overline{\mathscr{B}}(0, M)$, one has

$$\begin{aligned}
\varphi(x) - \varphi(x_0) &\leqslant \sup\left\{ -|x - y_n|^2 + |x_0 - y_n|^2 \;\middle|\; \exists x_n \in \mathbb{R}^d \text{ with } (x_n, y_n) \in \operatorname{supp}(\pi_x, \pi_x + \pi_v)\#\eta \right\} \\
&\leqslant |x_0|^2 + 2M|x_0 - x| - |x|^2.
\end{aligned}$$

The function φ is lower bounded by $-|x - y_0|^2 + |y_0 - x_0|^2$ by definition, so locally bounded, hence locally Lipschitz since it is semiconvex. Therefore the set-valued subdifferential application $x \mapsto \partial^+ \left(\varphi(x) - \frac{|x|^2}{2}\right)$ is compact-valued, and upper semicontinuous in the set-valued sense by [Roc70, Corollary 24.5.1]. By classical selection theorems, for instance [AB06, p. 18.13], it admits a measurable selection $f : \mathbb{R}^d \to \mathrm{T}\mathbb{R}^d$. Define $\xi := (1 - \lambda)\eta + \lambda(f\#\mu_2)$. By construction, ξ is still supported on the subdifferential of $\varphi - |\cdot|^2/2$, so $(\pi_x, \pi_x + \pi_v)\#\xi$ is cyclically monotone, hence optimal. □

In the above result, the measure ν is compactly supported. This is sharp; if $\mu_1 = \delta_0$ and $\mu_2 = \delta_1$ in dimension one, consider ν the Gaussian measure. The only $\eta \in \exp_\mu^{-1}(\nu)$ puts mass on velocities passing though 1 with any magnitude, and there cannot be any $\gamma \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_{\mu_2}$ satisfying (5.43); for any w on which γ puts mass, there exists $r > 0$ large enough such that $(1 - 0)((1 + w) - (0 + r)) < 0$. In consequence, to be able to use Lemma 5.5.2, we characterize $\mathbf{Sol}_\mu$ by orthogonality with respect to velocities going towards compactly supported measures.

Lemma 5.5.3 (Finer characterization of $\mathbf{Sol}_\mu$). *Let $\zeta \in \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ satisfy $\langle \zeta, (\pi_x, \pi_y - \pi_x)\#\gamma \rangle_\mu^+ = 0$ for any $\gamma \in \Gamma_o(\mu, \nu)$ with $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ compactly supported. Then $\zeta \in \mathbf{Sol}_\mu$.*

Proof. From Lemma 5.2.10, we know that it is sufficient that $\langle \zeta, (\pi_x, \pi_y - \pi_x)\#\eta \rangle_\mu^+ = 0$ for any $\eta \in \Gamma_o(\mu, \nu)$ with $\nu \in \mathscr{P}_2(\mathbb{R}^d)$. Stays to approximate any such η by a family of velocities of optimal plans with compactly supported second marginal. Since we need a convergence with respect to W_μ, and not only $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}$, we construct explicitly this approximation instead of using stability of optimality.

Let $\eta \in \Gamma_o(\mu, \nu)$ for some $\nu \in \mathscr{P}_2(\mathbb{R}^d)$. Let $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ be a proper c-convex Kantorovich potential for the pair (μ, ν), and φ^c its c-transform, such that $\varphi(x) = \sup_{y \in \mathbb{R}^d} \varphi^c(y) - |x - y|^2$. For each $R > 0$, define

$$\varphi_R(x) := \sup_{y \in \overline{\mathscr{B}}(0,R)} \varphi^c(y) - |x - y|^2.$$

The function φ_R is c-convex by definition, inferior to φ, lower bounded by a quadratic polynomial, and for any $x_0 \in \operatorname{dom} \varphi$,

$$\varphi_R(x) - \varphi_R(x_0) \leqslant \sup_{y \in \overline{\mathscr{B}}(0,R)} |x_0 - y|^2 - |x - y|^2 \leqslant |x_0|^2 + 2R|x_0 - x| - |x|^2.$$

Hence φ_R is locally bounded, and locally Lipschitz since semiconvex. Let $\Gamma_R \subset (\mathbb{R}^d)^2$ be the set of (x, y) such that $\varphi_R(x) + \varphi^c(y) = |x - y|^2$. For each $R > 0$, the correspondence $x \mapsto \left\{ y \;\middle|\; (x, y) \in \Gamma_R \right\}$ is upper semicontinuous with compact images. By [AB06, p. 18.13], it admits a measurable selection $f_R : \mathbb{R}^d \to \mathbb{R}^d$, that satisfies $|f_R(x)| \leqslant R$ for all $x \in \mathbb{R}^d$. Define

$$\gamma_R := \eta\left(\cdot \cap (\mathbb{R}^d \times \overline{\mathscr{B}}(0, R))\right) + (\pi_x, f_R(\pi_x))\#\eta\left(\cdot \cap (\mathbb{R}^d \times \overline{\mathscr{B}}(0, R))^c\right).$$

This is a probability measure concentrated on the cyclically monotone set Γ_R, hence an optimal transport plan between its marginals. The measure $\pi_y\#\gamma_R$ is supported on $\overline{\mathscr{B}}(0, R)$ by construction, and since $|f_R(x)| \leqslant R$, one has

$$W_\mu^2\left((\pi_x, \pi_y - \pi_x)\#\eta, (\pi_x, \pi_y - \pi_x)\#\gamma_R\right) \leqslant \int_{x,y \in \mathbb{R}^d, |y| > R} \left|f_R(x) - y\right|^2 d\eta \leqslant \int_{y \in \mathbb{R}^d, |y| > R} (R + |y|)^2 d\nu \xrightarrow[R \to \infty]{} 0.$$

As $\langle \zeta, (\pi_x, \pi_y - \pi_x)\#\gamma_R \rangle_\mu^+ = 0$ for any R, we get that $\langle \zeta, (\pi_x, \pi_y - \pi_x)\#\eta \rangle_\mu^+ = 0$, and ζ is solenoidal. □

Proposition 5.5.4 (Restriction of centred solenoidal spaces). *Let $A_1,\cdots,A_N\subset\mathbb{R}^d$ be measurable partition of $\mathbb{R}^d$. Let $\mu_k:=\mu|_{A_k}$. Then $\zeta\in\mathbf{Sol}^0_{\mu_k}$ if and only if $\zeta|_{\mathrm{T}A_k}\in\mathbf{Sol}^0_{\mu_k}$ for all $k\in[\![1,N]\!]$.*

Proof. We first prove that for all measurable set A, there holds $\mathbf{Sol}^0_{\mu_A}=\mathbf{Sol}^0_{\mu}|_{\mathrm{T}A}$, with $\mu_A:=\mu|_A$.

Let $m:=\mu(A)$, and $\mu_{A^c}:=\mu|_{A^c}$ be the measure such that $\mu=m\mu_A+(1-m)\mu_{A^c}$. Assume first that $\zeta\in\mathbf{Sol}^0_{\mu_A}$, and let $\zeta:=m\zeta+(1-m)0_{\mu_{A^c}}$. The measure field ζ is centred. Let $\eta\in\exp_\mu^{-1}(\nu)$ for some $\nu\in\mathscr{P}_2(\mathbb{R}^d)$. By restriction of optimality [Vil09, Theorem 4.6], $\eta|_{\mathrm{T}A}$ is the velocity of a geodesic issued from μ_A, so belongs to $\mathbf{Tan}_{\mu_A}$. Hence, using Chasles as in (5.41),

$$\langle\zeta,\eta\rangle_\mu^+=m\,\langle\zeta,(\eta|_{\mathrm{T}A})\rangle_{\mu_A}^+ + (1-m)\,\langle 0_{\mu_{A^c}},(\eta|_{\mathrm{T}A^c})\rangle_{\mu_{A^c}}^+=0.$$

By Proposition 5.2.9, this characterizes $\zeta\in\mathbf{Sol}_\mu$. Since $\zeta=\zeta|_{\mathrm{T}A}$, we deduce that $\mathbf{Sol}^0_{\mu_A}\subset\mathbf{Sol}^0_\mu|_{\mathrm{T}A}$.

Let now $\zeta\in\mathbf{Sol}^0_\mu$ be arbitrary. Define $\zeta_A:=\zeta|_{\mathrm{T}A}$, which is centred. Let $\eta_A\in\exp^{-1}_{\mu_A}(\nu)$ for some compactly supported measure $\nu\in\mathscr{P}_2(\mathbb{R}^d)$. By Lemma 5.5.2 applied to the mutually singular measures μ_A and μ_{A^c}, there exists $\eta\in\exp^{-1}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ such that $\eta|_{\mathrm{T}A}=\eta_A$. Therefore, using that metric scalar products of centred measure fields are nonnegative,

$$0=\langle\zeta,\eta\rangle_\mu^+=m\,\langle\zeta_A,\eta_A\rangle_{\mu_A}^+ + (1-m)\,\langle\zeta|_{\mathrm{T}A^c},\eta|_{\mathrm{T}A^c}\rangle_{\mu_{A^c}}^+\geqslant m\,\langle\zeta_A,\eta_A\rangle_{\mu_A}^+\geqslant 0, \tag{5.44}$$

and $\langle\zeta_A,\eta_A\rangle^+_{\mu_A}=0$. By Lemma 5.5.3, this characterizes $\zeta_A\in\mathbf{Sol}^0_{\mu_A}$, and the second inclusion is proved.

We now come back to the case of sets $A_1,\cdots,A_N$ forming a measurable partition of $\mathbb{R}^d$. The previous case shows that if $\zeta\in\mathbf{Sol}^0_\mu$, then $\zeta|_{\mathrm{T}A_k}\in\mathbf{Sol}^0_{\mu_k}$ for all $k\in[\![1,N]\!]$. On the other hand, let $\zeta\in\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)^0_\mu$ such that $\zeta_k:=\zeta|_{\mathrm{T}A_k}$ belongs to $\mathbf{Sol}^0_{\mu_k}$. Consider $\eta\in\exp^{-1}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ a velocity of a geodesic. By restriction of optimality, each $\eta_k:=\eta|_{\mathrm{T}A_k}$ is the velocity of a geodesic, hence belongs to $\mathbf{Tan}_{\mu_k}$. Consequently, by the Chasles relation (5.41) applied to the partition $(A_k)_{k\in[\![1,N]\!]}$,

$$\langle\zeta,\eta\rangle_\mu^+=\sum_{k\in[\![0,d]\!]}\mu(A_k)\,\langle\zeta_k,\eta_k\rangle^+_{\mu_k}=0.$$

By Lemma 5.2.10, $\zeta\in\mathbf{Sol}^0_\mu$, which concludes the proof. □

Corollary 5.5.5 (Restriction of centred tangent spaces). *Under the notations of Proposition 5.5.4, a measure field ξ belongs to $\mathbf{Tan}^0_\mu$ if and only if $\xi|_{\mathrm{T}A_k}\in\mathbf{Tan}^0_{\mu_k}$ for all $k\in[\![1,N]\!]$.*

Proof. As for solenoidal fields, we start by showing that for any measurable A, there holds $\mathbf{Tan}^0_{\mu_A}=\mathbf{Tan}^0_\mu|_{\mathrm{T}A}$. Let $\xi_A\in\mathbf{Tan}^0_{\mu_A}$, and construct $\xi:=m\xi_A+(1-m)0_{\mu_{A^c}}$. For any $\zeta\in\mathbf{Sol}^0_\mu$, the measure field $\zeta_A:=\zeta|_{\mathrm{T}A}$ belongs to $\mathbf{Sol}^0_{\mu_A}$ by Proposition 5.5.4, and

$$\langle\xi,\zeta\rangle^+_\mu=m\,\langle\xi|_{\mathrm{T}A},\zeta|_{\mathrm{T}A}\rangle^+_{\mu_A}+0=m\,\langle\xi_A,\zeta_A\rangle^+_{\mu_A}=0.$$

Hence $\xi\in\mathbf{Tan}^0_\mu$, and as $\xi_A=\xi|_{\mathrm{T}A}$, we deduce that $\mathbf{Tan}^0_{\mu_A}\subset\mathbf{Tan}^0_\mu|_{\mathrm{T}A}$. On the other hand, let $\xi\in\mathbf{Tan}^0_\mu$, and define $\xi_A:=\xi|_{\mathrm{T}A}$. For any $\zeta_A\in\mathbf{Sol}^0_{\mu_A}$, let $\zeta\in\mathbf{Sol}^0_\mu$ such that $\zeta_A=\zeta|_{\mathrm{T}A}$. Then, by Chasles,

$$0=\langle\xi,\zeta\rangle^+_\mu\geqslant m\,\langle\xi_A,\zeta_A\rangle^+_{\mu_A}+0,$$

and as ζ_A is arbitrary in $\mathbf{Sol}^0_{\mu_A}$, we deduce that $\xi_A\in\mathbf{Tan}^0_{\mu_A}$. The case of a measurable partition follows verbatim the argument of Proposition 5.5.4, with $\mathbf{Sol}^0_\mu$ in place of $\exp^{-1}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, and the restriction of solenoidal fields in place of the restriction of optimality. □

5.5.2 Decomposition of centred tangent and solenoidal fields

The result that we target is greatly facilitated by working with maps instead of plans. Consider the set

$$\mathcal{G}:=\left\{g\in L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)\;\middle|\;\frac{1}{2}\left[(-g)\#\mu+g\#\mu\right]\in\mathbf{Sol}^0_\mu\right\}. \tag{5.45}$$

By Lemma 5.4.1, the set of barycentric solenoidal fields is a subset of $\mathcal{G}$. It may be a strict subset: in dimension one and for a nonatomic measure, the only barycentric solenoidal field is 0_μ, but $\mathcal{G} = L^2_\mu$ by Theorem 5.2.23. The same example shows that the same property is not satisfied by $\mathbf{Tan}^0_\mu$, since all nonzero tangent fields are barycentric; however, by Proposition 5.4.4, if $f \in L^2_\mu$ is such that $\frac{1}{2}\left[(-f)\#\mu + f\#\mu\right] \in \mathbf{Tan}_\mu$, then $f\#\mu \in \mathbf{Tan}_\mu$.

Remark 5.5.6 (Pointwise vector subspace)**.** *By Section 5.2.3, $\mathcal{G}$ is a closed vector subspace of L^2_μ. However, the stability of $\mathbf{Sol}^0_\mu$ by horizontal convexity in Proposition 5.2.9, and by pointwise renormalization in Lemma 5.5.1, implies a much stronger property; if $g_1, \cdots, g_N \in \mathcal{G}$ and $g \in L^2_\mu(\mathbb{R}^d; \mathrm{T}\mathbb{R}^d)$ is such that $g(x) \in \operatorname{span}\{g_1(x), \cdots, g_N(x)\}$ for μ−a.e. $x \in \mathbb{R}^d$, then g also belongs to $\mathcal{G}$. We deduce that for any $g_1, g_2 \in \mathcal{G}$ and measurable $A \subset \mathbb{R}^d$, the gluing $g_3 := g_1 \mathbb{I}_A + g_2 \mathbb{I}_{A^c}$ belongs to $\mathcal{G}$, since $g_3(x) \in \operatorname{span}\{g_1(x), g_2(x)\}$ μ−a.e..*

Lemma 5.5.7 (Decomposition by basis). *Let $\mathcal{G} \subset L^2_\mu(\mathbb{R}^d; \mathrm{T}\mathbb{R}^d)$ be defined in (5.45). By Proposition 5.2.20, $\mathbf{Sol}^0_\mu = \overrightarrow{\operatorname{span}}\left\{\frac{1}{2}\left[(-g)\#\mu + g\#\mu\right] \,\middle|\, g \in \mathcal{G}\right\}$, where the closed horizontal span is given in Definition 5.2.17. There exist 2-by-2 disjoint measurable sets $A_0, \cdots, A_d \subset \mathbb{R}^d$ covering $\mathbb{R}^d$ such that*

1. *if $k < d$, any $k+1$ elements in $\mathcal{G}$ are linked μ−almost everywhere in A_k,*
2. *if $k > 0$, there exist $g_1, \cdots, g_k \in \mathcal{G}$ such that $g_1(x), \cdots, g_k(x)$ is an orthonormal family μ−a.e. in A_k.*

Proof. For $k \in [\![1, d]\!]$, let $D_k : \mathrm{T}^k\mathbb{R}^d \to \mathbb{R}^+$ be the k−dimensional volume, i.e.

$$D_k(v_1, \cdots, v_k) = \operatorname{Vol}_k\left\{\sum_{i=1}^{k} \lambda_i v_i \,\middle|\, \lambda_i \in [0,1] \text{ for } i \in [\![1,k]\!]\right\}.$$

Each D_k is continuous and positively homogeneous, so Lipschitz as a function of $\mathrm{T}^k\mathbb{R}^d$ to $\mathbb{R}^+$.

Let $(g^n)_{n\in\mathbb{N}}$ be a countable L^2_μ−dense subset of $\mathcal{G}$. For each n, let A^n_0 be a measurable set such that $|g^n(x)| = 0$ for μ−a.e. $x \in A^n_0$, and $|g^n(x)| > 0$ for μ−a.e. $x \notin A^n_0$. The countable intersection $A_0 := \bigcap_{n\in\mathbb{N}} A^n_0$ stays measurable, and all g^n vanish μ−a.e. in A_0. Assume now that the set A_k is defined. Let $A^{n_1,n_2,\cdots,n_{k+1}}_k$ be a measurable set such that $D_{k+1}(g^{n_1}(x), \cdots, g^{n_{k+1}}(x))$ vanishes μ−almost everywhere in $A^{n_1,\cdots,n_{k+1}}_k$, and $D_{k+1}(g^{n_1}(x), \cdots, g^{n_{k+1}}(x)) > 0$ for μ−a.e. $x \notin A^{n_1,\cdots,n_{k+1}}_k$. The set $A_{k+1} := \bigcap_{n_1,\cdots,n_{k+1}\in\mathbb{N}} A^{n_1,\cdots,n_{k+1}}_k \setminus A_k$ is measurable, and any $k+1$ vector fields g^n are linked μ−almost everywhere in A_k. Let $A_d := \mathbb{R}^d \setminus \bigcup_{k=0}^{d-1} A_k$.

If $k < d$, consider $g_1, \cdots, g_{k+1} \in \mathcal{G}$. Each g_i can be approximated by some g^{n_i} with L^2_μ−norm as small as desired, with $\int_{x\in\mathbb{R}^d} \mathbb{I}_{A_k}(x) D_{k+1}(g^{n_1}(x), \cdots, g^{n_{k+1}}(x)) d\mu(x) = 0$ by construction of the set A_k. As D_{k+1} is Lipschitz, we may pass to the limit to get $\int_{x\in\mathbb{R}^d} \mathbb{I}_{A_k}(x) D_{k+1}(g_1(x), \cdots, g_{k+1}(x)) d\mu(x) = 0$, so that Point 1 holds.

We turn to Point 2. Let $(\sigma_m)_{m\in\mathbb{N}}$ be the countable set of all injective applications from $[\![1,k]\!]$ to $\mathbb{N}$. Let $m = 1$, and $A^1 \subset A_k$ be a measurable set such that $D_k(g^{\sigma_1(1)}(x), \cdots, g^{\sigma_1(k)}(x)) > 0$ for μ−a.e. $x \in A^1$, and vanishes μ−a.e. in the complementary of A^1. By Remark 5.5.6, we can orthogonalize the family $g^{\sigma_1(i)}$ to produce vector fields $f^{1,i} \in \mathcal{G}$ forming an orthonormal basis of $\operatorname{span}(g^{\sigma_1(1)}(x), \cdots, g^{\sigma_1(k)}(x))$ for μ−almost every $x \in A^1$. Up to multiplication with $\mathbb{I}_{A^1}$, we can assume that $f^{1,i}(x) = 0$ for $x \notin A^1$. Assume now that $(f^{m,i})_{i\in[\![1,k]\!]}$ and A^m are given. Let A^{m+1} be a measurable set on which $D_{k+1}(g^{\sigma_{m+1}(1)}, \cdots, g^{\sigma_{m+1}(k)})$ is nonzero μ−a.e. in A^{m+1}, and vanish μ−a.e. outside of A^{m+1}. Define $f^{m+1,i}$ on $A^{m+1} \setminus A^m$ as to obtain an orthonormal basis of $\operatorname{span}\{g^{\sigma_{m+1}(1)}(x), \cdots, g^{\sigma_{m+1}(k)}(x)\}$ for μ−a.e. $x \in A^{m+1} \setminus A^m$, and by $f^{m+1,i} = f^{m,i}$ on A^m. By Remark 5.5.6, $f^{m+1,i} \in \mathcal{G}$. Let $B := A_k \setminus \bigcup_{q\in\mathbb{N}} A^q$. By construction, all choices of k vector fields in $\{g_n \mid n \in \mathbb{N}\}$ are linked a.e. in B. This implies that $\mu(B \setminus A_{k-1}) = 0$, and as A_k is disjoint from A_{k+1}, the intersection $B \cap A_k$ must be μ−negligible. So the pointwise limits of $f^{m,i}$ in $m \to \infty$ are well-defined, μ−almost nowhere vanishing, belong to $\mathcal{G}$ and satisfy Point 2. □

We get to the central result of this section.

Definition 5.5.8 (Subset of measures of dimension k). *A subset $\mathcal{A} \subset \mathscr{P}_2(\mathrm{T}\mathbb{R}^d)^0_\mu$ is of dimension k if*

$$\mathcal{A} = \overrightarrow{\operatorname{span}}\left\{\frac{1}{2}\left[(-g_j)\#\mu + g_j\#\mu\right] \,\middle|\, j \in [\![1,k]\!]\right\}$$

for a family $g_1, \cdots, g_k \in L^2_\mu(\mathbb{R}^d; \mathrm{T}\mathbb{R}^d)$ such that $g_1(x), \cdots, g_k(x)$ forms an orthonormal family μ−a.e..

By convention, $\mathcal{A}$ is of dimension 0 if it is equal to $\{0_\mu\}$. The definition of $\overrightarrow{\text{span}}$ is local, so that the dimension of any $\mathbf{Set}^0_\mu$ is the same as the dimension of $\mathbf{Set}^0_\mu|_{\mathrm{T}A}$ for any measurable $A \subset \mathbb{R}^d$.

Theorem 5.5.9 (Decomposition of centred tangent and solenoidal spaces). *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, and $(A_k)_{k\in[\![0,d]\!]}$ be the sets given by Lemma 5.5.7. Denote $\mu_k := \mu|_{A_k}$. Then for any $k \in [\![0,d]\!]$,*

1. *$\zeta \in \mathbf{Sol}^0_\mu$ if and only if $\zeta|_{\mathrm{T}A_k} \in \mathbf{Sol}^0_{\mu_k}$ for all $k \in [\![0,d]\!]$,*

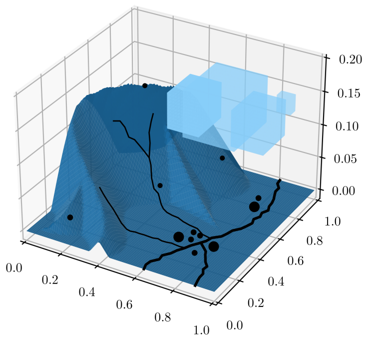
2. *$\xi \in \mathbf{Tan}^0_\mu$ if and only if $\xi|_{\mathrm{T}A_k} \in \mathbf{Tan}^0_{\mu_k}$ for all $k \in [\![0,d]\!]$,*

3. *$\mathbf{Sol}^0_{\mu_k}$ is of dimension k in the sense of Definition 5.2.17,*

4. *$\mathbf{Tan}^0_{\mu_k}$ is of dimension $d-k$ in the sense of Definition 5.2.17.*

Proof. Points 1 and 2 follow from Proposition 5.5.4 and Corollary 5.5.5, since the sets $(A_k)_k$ form a measurable partition of $\mathbb{R}^d$. Point 3 is deduced from Lemma 5.5.7; if $k=0$, then $\mathbf{Sol}^0_{\mu_0} = \{0_{\mu_0}\}$ is of dimension 0 by convention, and otherwise, it provides a family $(g_j)_{j\in[\![1,k]\!]}$ satisfying Definition 5.5.8. In particular, for $k=d$, all centred measure fields are solenoidal, and $\mathbf{Tan}^0_{\mu_d} = \{0_{\mu_d}\}$. If $k<d$, let $f_{k+1},\cdots,f_d \subset L^2_\mu(\mathbb{R}^d;\mathrm{T}\mathbb{R}^d)$ be such that $g_1(x),\cdots,g_k(x),f_{k+1}(x),\cdots,f_d(x)$ forms an orthonormal basis of $\mathbb{R}^d$ $\mu-$almost everywhere. Denote $\xi_i := \frac{1}{2}\left[(-f_i)\#\mu + f_i\#\mu\right]$ for $i \in [\![k+1,d]\!]$. Any element of $\overrightarrow{\text{span}}\{\xi_{k+1},\cdots,\xi_d\}$ is orthogonal to $\mathbf{Sol}^0_\mu$, so belongs to $\mathbf{Tan}^0_\mu$. Moreover, by Lemma 5.2.18, any tangent field of the form $\xi := \frac{1}{2}[(-f)\#\mu + f\#\mu]$ satisfies $\langle f(x), g_j(x)\rangle = 0$ for $j \in [\![1,k]\!]$ and $\mu-$almost every x, so that $f(x) \in \text{span}\{f_{k+1}(x),\cdots,f_d(x)\}$ for $\mu-$a.e. $x \in \mathbb{R}^d$. Using a measurable selection argument, there exists $\lambda_{k+1},\cdots,\lambda_d \in L^2_\mu(\mathbb{R}^d;\mathbb{R})$ such that $f(x) = \sum_{i=k+1}^d \lambda_i(x) f_i(x)$ for $\mu-$a.e. x, and $\xi \in \text{span}\{\xi_{k+1},\cdots,\xi_k\}$. Hence $\overrightarrow{\text{span}}\{\xi_{k+1},\cdots,\xi_d\}$ contains the closed horizontal span of such fields, which equals $\mathbf{Tan}^0_\mu$ by Proposition 5.2.20. This concludes Point 4. □

Let us summarize this part. Any measure μ splits in submeasures that have centred tangent/solenoidal spaces of "uniform" dimension summing to d. This statement avoids any reference to the support of the measures μ_k, but this is the core of the topic.



It is direct that μ_d is transport-regular, and that all other μ_k for $k<d$ are not. It is reasonable to conjecture that μ_0 is the atomic part of μ. In this sense, Theorem 5.5.9 generalises Theorem 5.2.23 for centred fields. For the general case, the guiding intuition is that μ_k is responsible for the mass that μ puts on $k-$dimensional hypersurfaces given by differences of convex functions, and a bold conjecture would be that μ_k has a support covered by countably many such surfaces. We wish the reader to be wiser than us on this.

Perspectives

The previous chapters open in several possible directions, which we discuss here.

Hamilton-Jacobi-Bellman equations in the viscosity sense. In Chapters 2 and 3, the Hamiltonians of the HJB equations are assumed to be continuous. To go on, one could consider a discontinuous Hamiltonian, as for instance in control problems in which the dynamic changes from one region to another. Discontinuities of the Hamiltonian with respect to the space variable are a vast topic, starting from the work of Ishii [Ish85]; the problem is that the viscosity solution might no longer be unique. In the case of a structured discontinuity, the control point of view allows to build Hamiltonians by selecting the dynamics that are taken by the trajectories, as done by Barnard and Wolenski [BW13], Rao and Zidani [RZ13], Jerhaoui and Zidani [JZ23a], with similar techniques in the monograph of Barles and Chasseigne [BC24]. The optimal control framework of Chapter 2 could be used to extend these results, provided the structure of the discontinuity is defined in a proper way. Another interesting direction is the line of flux-limited solutions developed by Imbert and Monneau [IM17], arising from the connection with conservation laws, and in which the Hamiltonian at the point of discontinuity is modified to bound the flux by a particular constant. At the time of writing, the ANR COSS is boiling with new results showing that with one road in and one out, this simple condition arises as a universal limit in homogenization, characterizes contractive semigroups, and that the case of several roads is a vast new world to explore.

In Chapter 3, we proposed a notion of viscosity solutions relying on the semiconcave functions, that are available in any complete geodesic CBB(0) space. This definition is not known to be stable, and it seems to me that the inequalities that are going "in the right direction" in CAT(0) spaces are lost when passing to CBB(0) spaces. However, in the Wasserstein space, other definitions were proposed, which are stable – taking L-differentiable test functions, for instance. I do not know when both definitions are equivalent. The difference lies in the fact that semiconcave test functions may have a non-vanishing directional derivative along centred measure fields. These directions are not seen by the equation $H(\mu, D_\mu u) = 0$ if H depends only on scalar products against map-induced measure fields, but they do intervene in other simple cases, such as the Eikonal equation $H(\mu, D_\mu u) = [\text{Metric Slope at } \mu](u) = \sup_{\xi \in \mathbf{Tan}_\mu, \|\xi\|_\mu = 1} \left|D_\mu u(\xi)\right|$. However, on a large class of measures, the geometric tangent cone reduces to the regular one. If one could prove that the (geometric) superdifferential of u can be attained by a convex hull, in a certain sense, of the regular semidifferentials on the surrounding points, then even for these Hamiltonians, the equivalence may stand.

Finer classification of measures. Chapter 5 brings a negative answer to the question "Is it true that the geometric tangent cone is the quotient of $\mathscr{P}_2(\mathrm{T}\mathbb{R}^d)_\mu$ by the equivalence relation $\xi \sim \xi'$ if $d_{\mathcal{W}}(\exp_\mu(h \cdot \xi), \exp_\mu(h \cdot \xi')) = o(h)$?". This is a deception, since it means that one has to pay attention when comparing a curve driven by a measure field ξ, and the curve driven by the projection of ξ on the tangent cone. However, from the construction of the counterexample, I wonder if the set of measures with this disturbing behaviour could not be identified, and quantified as small, for instance by the measures of Von Renesse and Sturm.

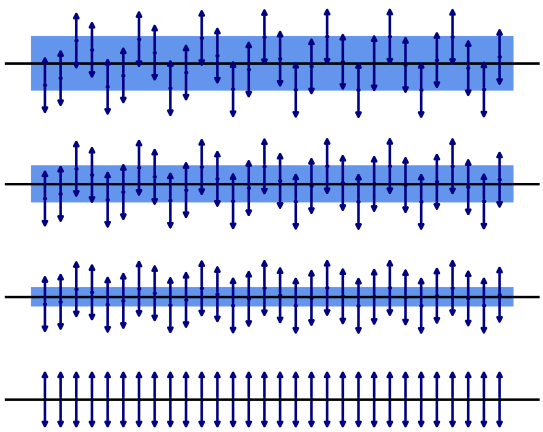
In a few words, Von Renesse and Sturm [VRS09] build a family of measures $(\mathbb{P}_\beta)_{\beta>0}$ on $\mathscr{P}(\Omega)$ that resembles a diffusion semigroup, where Ω is a compact manifold without boundary to start with. To the best of my understanding, it follows this idea (in dimension 1): a set of measures is converted in the set of nondecreasing "paths" γ such that $\mu = \gamma \# \mathcal{L}$, and this set is given the probability that it would have as Brownian paths, except that one does not work with the Brownian motion that induces a gradient flow of the H^1–seminorm $|\gamma|^2_{H^1} = \int_s |\gamma'_s|^2 ds$, but with a construction following the same idea with $|\gamma|_{\text{not } H^1} =$

$\int_s -\log(\gamma'_s)ds = \mathcal{E}(\gamma\#\mathcal{L})$.

It turns out that discontinuous and absolutely continuous curves are both negligible. In $\mathscr{P}_2$ terms, the set of measures with an atomic part, or an absolutely continuous part, is of $\mathbb{P}_\beta-$measure 0. *This means that from the point of view of these semigroup, the Wasserstein space is filled with purely Cantor measures.* The latter are transport-regular, so their tangent cone is a Hilbert space, and all nasty problems of mass-splitting disappear $\mathbb{P}_\beta-$almost everywhere. (Note that this is not equivalent to an entropic penalization, since absolutely continuous measures are $\mathbb{P}_\beta-$negligible.) Dello Schiavo [DS20] proved that $\mathbb{P}_\beta$ supports a Rademacher theorem, in the sense that all $d_{\mathcal{W}}-$Lipschitz maps admit a Wasserstein gradient $\mathbb{P}_\beta-$almost everywhere. This settles once and for all the said Wasserstein gradient as natural, at least for $\mathbb{P}_\beta$.

In the construction made in Section 5.3.4, the path corresponding to the measure must be very well approximated by stair functions: vaguely speaking, the error of approximation by a step function with jump size h should be of order $o(h)$. This makes it a quite rigid class, and perhaps it could be proved to have $\mathbb{P}_\beta-$measure 0. This would complete the discussion on this problem.

Decomposition of measures in function of the dimension of the tangent cone. Prior to addressing these rather difficult points, I want to generalize the results of Chapter 5 formulated in dimension one to higher dimension. The entry point in this would be the decomposition of μ as $\mu_0 + \mu_1 + \cdots + \mu_d$ provided in Theorem 5.5.9. In this sum, each μ_k has centred solenoidal measure fields that are concentrated on the span of k vector fields, and centred tangent fields concentrated on the span of $d-k$ vector fields, the latter being orthogonal to the former in L^2_μ. This quite weak statement is more precise in dimension one, where I can use the fact that an optimal plan can split mass only at the atoms, and the atoms are countable. In higher dimension, one expects that a similar characterization of each μ_k can be done, but this stays to be written down. If this holds, then the techniques of dimension one could probably be extended to show that solenoidal measure fields are closed with respect to $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}$, that any $\xi \in \mathbf{Tan}_\mu$ can be approximated by some sequence in $\frac{1}{h}\cdot\exp_\mu^{-1}(\exp_\mu(h\cdot\xi))$, and to give conditions on which the solenoidal fields are described by the $W_\mu-$cone over some particular plans. I think that these results would demystify the structure of the tangent cone to a given measure, and one could start to work with finer convergences, for instance imposing that μ_k converges to ν_k without transfer of mass between components. This is tightly linked to the subsequent topic.



The tangent and solenoidal sets are discontinuous as functions of μ. For instance, the measure fields ζ_n depicted on the right, which are solenoidal for their base measures $\mu_n = \mathcal{H}^2|_{[-1,1]\times[-1/n,1/n]}$, converge with respect to both $W_{\mu,\nu}$ and $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}$ to a limit that is tangent to its base measure $\mu = \mathcal{H}^1|_{[-1,1]\times\{0\}}$. Here mass is concentrated, and it seems moral to introduce equivalence classes between measures, for instance as $\mu \sim \nu$ if there exists a continuous curve from μ to ν along which the tangent cone changes continuously. In dimension one, the measures that are equivalent to δ_0 shall be the other Dirac masses δ_x. Could this be a way to formalize the "stratified structure" of the Wasserstein space, alluded to by Gangbo, Kim, and Pacini [GKP11], completely exposed for Gaussians by Takatsu [Tak11], perhaps perceived as folklore knowledge, but to my understanding, not written anywhere in rigorous form? Pragmatically, I would like to obtain sufficient conditions on a converging sequence $(\mu_n)_n$ so that the limit of tangent fields stays a tangent field to the limit, and accordingly for solenoidal fields. In the second case, it seems even possible to work with $d_{\mathcal{W},\mathrm{T}\mathbb{R}^d}-$convergence only.

Flows of solenoidal measure fields. The discussion of Smirnov-type decompositions in Section 5.4.2 is very preliminary. It is proved there that if a solenoidal measure field satisfies $d_{\mathcal{W}}(\mu, \exp_\mu(h\cdot\zeta)) = o(h)$, then there is a sequence $(\zeta_n)_{n\in\mathbb{N}}$ converging to ζ with respect to W_μ and such that $\exp_\mu(h_n\cdot\zeta_n) = \mu$ for some vanishing sequence $(h_n)_{n\in\mathbb{N}}$. Since ζ_n comes back on μ at time h_n, it can be "composed with itself" to furnish a superposition measure $\gamma^n \in \mathscr{P}_2\left(\mathrm{AC}([0,T];\mathbb{R}^d)\right)$ on curves γ^n such that $\mu = e_0\#\gamma^n = e_{h_n}\#\gamma^n = e_{2h_n}\#\gamma^n = \cdots$. Assuming the support of ζ is bounded, one can extract a limit of γ^n, and obtain $\gamma \in \mathscr{P}_2\left(\mathrm{AC}([0,T];\mathbb{R}^d)\right)$ such that $e_h\#\gamma = \mu$ for all h. However, the curve γ strongly depends on the plans used to glue ζ^n to itself.

For instance, consider $\Omega = \mathbb{S}^1$, μ the uniform measure, and ζ sending half of the mass along the clockwise rotation, and half along the anticlockwise rotation. In this case, one directly has $\mu = \exp_\mu(h\cdot\zeta)$ for all h. For a fixed $h_0 > 0$, one can produce several superposition measures by gluing $([0,h_0] \ni s \mapsto \exp_{\pi_x}(s\cdot\pi_v))\#\zeta$ to itself: either glue clockwise-clockwise and anticlockwise-anticlockwise, or the opposite, or any convex combination. In the first case, the resulting γ puts mass on exactly two curves issued from any x, the one turning clockwise, and the one turning anticlockwise. However, if the direction of the rotation is reversed at each gluing step, the selected curves will oscillate with increasing frequency near their starting point, and the limit γ will put mass only on static curves. Perhaps the limit points that can be obtained by this procedure are all superposition measures with "velocity", in a sense to define, in the horizontally convex hull of ζ? It might be interesting to look at the span of the derivatives of the curves on which γ puts mass. In higher dimension, these are expected to be "tangent" to the support of μ, at least in weak senses.

To conclude, I only looked at $p = 2$. It seems moral to think that the picture is the same for $p \in]1,\infty[$, although this has to be written. It surely changes completely for $p = 1$, and I do not dare to imagine the monstrosities of $p = \infty$. So, what about $p = 1$? Then one could look at $p \in]0,1[$, other costs, and so on – but the Monge case is already a challenge.

Bibliography

[Abo+03] H. Abou-Kandil, G. Freiling, V. Ionescu, and G. Jank. *Matrix Riccati Equations in Control and Systems Theory*. Systems & Control: Foundations & Applications. Basel: Birkhäuser, 2003.

[Ach+13] Y. Achdou, F. Camilli, A. Cutrì, and N. Tchou. “Hamilton–Jacobi Equations Constrained on Networks”. In: *Nonlinear Differential Equations and Applications* 20.3 (June 2013), pp. 413–445.

[AKA10] B. Ahmadi Kakavandi and M. Amini. “Duality and Subdifferential for Convex Functions on Complete CAT(0) Metric Spaces”. In: *Nonlinear Analysis: Theory, Methods & Applications* 73.10 (Nov. 2010), pp. 3450–3455.

[ACP11] G. Alberti, M. Csörnyei, and D. Preiss. “Differentiability of Lipschitz Functions, Structure of Null Sets, and Other Problems”. In: *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010)*. Hyderabad, India: Hindustan Book Agency, June 2011, pp. 1379–1394.

[AM16] G. Alberti and A. Marchese. “On the Differentiability of Lipschitz Functions with Respect to Measures in the Euclidean Space”. In: *Geometric and Functional Analysis* 26.1 (Feb. 2016), pp. 1–66.

[AKP23] S. Alexander, V. Kapovitch, and A. Petrunin. *Alexandrov Geometry: Foundations*. Graduate Studies in Mathematics 236. Providence, Rhode Island: American Mathematical Society, 2023.

[AB06] C. Aliprantis and K. Border. *Infinite Dimensional Analysis*. Berlin/Heidelberg: Springer-Verlag, 2006.

[ABS21] L. Ambrosio, E. Brué, and D. Semola. *Lectures on Optimal Transport*. Vol. 130. UNITEXT. Cham: Springer International Publishing, 2021.

[AC14] L. Ambrosio and G. Crippa. “Continuity Equations and ODE Flows with Non-Smooth Velocity”. In: *Proceedings of the Royal Society of Edinburgh: Section A Mathematics* 144.6 (Dec. 2014), pp. 1191–1244.

[AF14] L. Ambrosio and J. Feng. “On a Class of First Order Hamilton–Jacobi Equations in Metric Spaces”. In: *Journal of Differential Equations* 256.7 (Apr. 2014), pp. 2194–2245.

[AFP00] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford ; New York: Clarendon Press, 2000.

[AG08] L. Ambrosio and W. Gangbo. “Hamiltonian ODEs in the Wasserstein Space of Probability Measures”. In: *Communications on Pure and Applied Mathematics* 61.1 (2008), pp. 18–53.

[AGS05] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows*. Lectures in Mathematics ETH Zürich. Basel: Birkhäuser-Verlag, 2005.

[ANN18] A. Andoni, A. Naor, and O. Neiman. “Snowflake Universality of Wasserstein Spaces”. In: *Annales scientifiques de l’École normale supérieure* 51.3 (2018), pp. 657–700.

[Aub99] J. P. Aubin. *Mutational and Morphological Analysis*. Systems & Control: Foundations & Applications. Birkhäuser Boston, MA, 1999.

[AC86] J.-P. Aubin and A. Cellina. “Differential Inclusions Set-Valued Maps and Viability Theory”. In: *Acta Applicandae Mathematicae* 6.2 (June 1986), pp. 215–217.

[AF90] J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*. Boston: Birkhäuser, 1990.

[Aum65] R. J. Aumann. "Integrals of Set-Valued Functions". In: *Journal of Mathematical Analysis and Applications* 12.1 (Aug. 1965), pp. 1–12.

[Aus25] A. Aussedat. "On the Structure of the Geometric Tangent Cone to the Wasserstein Space". In: *Journal of Differential Equations* 442 (2025), p. 113520.

[AH24] A. Aussedat and C. Hermosilla. *A Minimality Property of the Value Function in Optimal Control over the Wasserstein Space*. Preprint, available at https://hal.science/hal-04427139. Jan. 2024.

[AJZ24] A. Aussedat, O. Jerhaoui, and H. Zidani. "Viscosity Solutions of Centralized Control Problems in Measure Spaces". In: *ESAIM Control Optimisation and Calculus of Variations* 30 (Oct. 2024), pp. 1–37.

[AZ25] A. Aussedat and H. Zidani. *A Cauchy-Lipschitz Setting for Control Problems in Complete CAT(0) Spaces*. 2025.

[BF22a] Z. Badreddine and H. Frankowska. "Solutions to Hamilton–Jacobi Equation on a Wasserstein Space". In: *Calculus of Variations and Partial Differential Equations* 61.1 (Feb. 2022), p. 9.

[BF22b] Z. Badreddine and H. Frankowska. "Viability and Invariance of Systems on Metric Spaces". In: *Nonlinear Analysis* 225 (Dec. 2022), p. 113133.

[BC97] M. Bardi and I. Capuzzo-Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Boston, MA: Birkhäuser Boston, 1997.

[BBC14] G. Barles, A. Briani, and E. Chasseigne. "A Bellman Approach for Regional Optimal Control Problems in $\mathbb{R}^N$". In: *SIAM Journal on Control and Optimization* 52.3 (Jan. 2014), pp. 1712–1744.

[Bar94] G. Barles. *Solutions de Viscosité Des Équations de Hamilton-Jacobi*. Mathématiques & Applications 17. Springer-Verlag, 1994.

[BBC13] G. Barles, A. Briani, and E. Chasseigne. "A Bellman Approach for Two-Domains Optimal Control Problems in $\mathbb{R}^N$". In: *ESAIM: Control, Optimisation and Calculus of Variations* 19.3 (July 2013), pp. 710–739.

[BC24] G. Barles and E. Chasseigne. *On Modern Approaches of Hamilton-Jacobi Equations and Control Problems with Discontinuities*. Vol. 104. Progress in Nonlinear Differential Equations and Their Applications. Cham: Springer Nature Switzerland, 2024.

[BW13] R. C. Barnard and P. R. Wolenski. "Flow Invariance on Stratified Domains". In: *Set-Valued and Variational Analysis* 21.2 (June 2013), pp. 377–403.

[BJ90] E. N. Barron and R. Jensen. "Semicontinuous Viscosity Solutions For Hamilton–Jacobi Equations With Convex Hamiltonians". In: *Communications in Partial Differential Equations* 15.12 (Jan. 1990), pp. 293–309.

[BEZ24] E. Bayraktar, I. Ekren, and X. Zhang. *Comparison of Viscosity Solutions for a Class of Second Order PDEs on the Wasserstein Space*. Preprint, available at http://arxiv.org/abs/2309.05040. Oct. 2024.

[Bel54] R. Bellman. "The Theory of Dynamic Programming". In: *Bulletin of the American Mathematical Society* 60.6 (1954), pp. 503–515.

[BN08] I. D. Berg and I. G. Nikolaev. "Quasilinearization and Curvature of Aleksandrov Spaces". In: *Geometriae Dedicata* 133.1 (Apr. 2008), pp. 195–218.

[Ber59] C. Berge. *Espaces topologiques: fonctions multivoques*. Paris: Dunod, 1959.

[Ber24] C. Bertucci. "Stochastic Optimal Transport and Hamilton–Jacobi–Bellman Equations on the Set of Probability Measures". In: *Annales de l'Institut Henri Poincaré C* (Nov. 2024).

[BL24] C. Bertucci and P. L. Lions. *An Approximation of the Squared Wasserstein Distance and an Application to Hamilton-Jacobi Equations*. Preprint, available at http://arxiv.org/abs/2409.11793. Sept. 2024.

[Bha07] R. Bhatia. *Positive Definite Matrices*. Princeton Series in Applied Mathematics. Princeton (N.J.): Princeton university press, 2007.

[Bil99] P. Billingsley. *Convergence of Probability Measures*. 2nd ed. Wiley Series in Probability and Statistics. Probability and Statistics Section. New York: Wiley, 1999.

[Bog07] V. I. Bogachev. *Measure Theory*. Berlin, Heidelberg: Springer, 2007.

[BF21] B. Bonnet and H. Frankowska. “Differential Inclusions in Wasserstein Spaces: The Cauchy-Lipschitz Framework”. In: *Journal of Differential Equations* 271 (Jan. 2021), pp. 594–637.

[BF24] B. Bonnet-Weill and H. Frankowska. “Carathéodory Theory and a Priori Estimates for Continuity Inclusions in the Space of Probability Measures”. In: *Nonlinear Analysis* 247 (Oct. 2024), p. 113595.

[BZ05] J. M. Borwein and Q. J. Zhu. *Techniques of Variational Analysis*. CMS Books in Mathematics. New York: Springer-Verlag, 2005.

[Bre01] Y. Brenier. “A Monge-Kantorovich Approach to the Maxwell Equations”. In: *Hyperbolic Problems: Theory, Numerics, Applications*. International Series of Numerical Mathematics. Basel: Birkhäuser, 2001, pp. 179–186.

[Bre91] Y. Brenier. “Polar Factorization and Monotone Rearrangement of Vector-valued Functions”. In: *Communications on Pure and Applied Mathematics* 44.4 (June 1991), pp. 375–417.

[Bre+14] A. Bressan, S. Canic, M. Garavello, M. Herty, and B. Piccoli. “Flows on Networks: Recent Results and Perspectives”. In: *EMS Surveys in Mathematical Sciences* 1 (Jan. 2014).

[Bré10] H. Brézis. *Analyse fonctionnelle: théorie et applications*. Nouvelle présentation, [Nachdr.] Mathématiques appliquées pour le master. Paris: Dunod, 2010.

[BH99] M. R. Bridson and A. Haefliger. *Metric Spaces of Non-Positive Curvature*. Springer, 1999.

[BGP92] Y. Burago, M Gromov, and G Perel’man. “A.D. Alexandrov Spaces with Curvature Bounded Below”. In: *Russian Mathematical Surveys* 47.2 (Apr. 1992), pp. 1–58.

[Bur+20] M. Burzoni, V. Ignazio, A. M. Reppen, and H. M. Soner. “Viscosity Solutions for Controlled McKean–Vlasov Jump-Diffusions”. In: *SIAM Journal on Control and Optimization* 58.3 (Jan. 2020), pp. 1676–1699.

[Caf+96] L. Caffarelli, M. J. Crandall, M. Kocan, and A. Swiech. “On Viscosity Solutions of Fully Nonlinear Equations with Measurable Ingredients”. In: *Communications on Pure and Applied Mathematics* 49.4 (Apr. 1996), pp. 365–398.

[Caf89] L. A. Caffarelli. “Interior a Priori Estimates for Solutions of Fully Non-Linear Equations”. In: *Annals of Mathematics* 130.1 (1989), pp. 189–213.

[Cam+21] F. Camilli, G. Cavagnari, R. De Maio, and B. Piccoli. “Superposition Principle and Schemes for Measure Differential Equations”. In: *Kinetic & Related Models* 14.1 (2021), p. 89.

[CDMT17] F. Camilli, R. De Maio, and A. Tosin. “Transport of Measures on Networks”. In: *Networks & Heterogeneous Media* 12.2 (2017), pp. 191–215.

[CM13] F. Camilli and C. Marchi. “A Comparison among Various Notions of Viscosity Solution for Hamilton–Jacobi Equations on Networks”. In: *Journal of Mathematical Analysis and Applications* 407.1 (Nov. 2013), pp. 112–118.

[CSM13] F. Camilli, D. Schieborn, and C. Marchi. “Eikonal Equations on Ramified Spaces”. In: *Interfaces and Free Boundaries* 15.1 (2013), pp. 121–140.

[CS03] F. Camilli and A. Siconolfi. “Hamilton-Jacobi Equations with Measurable Dependence on the State Variable”. In: *Advances in Differential Equations* 8.6 (Jan. 2003), pp. 733–768.

[CS05] F. Camilli and A. Siconolfi. “Time-Dependent Measurable Hamilton–Jacobi Equations”. In: *Communications in Partial Differential Equations* 30.5-6 (May 2005), pp. 813–847.

[CS04] P. Cannarsa and C. Sinestrari. *Semiconcave Functions, Hamilton—Jacobi Equations, and Optimal Control*. Boston, MA: Birkhäuser, 2004.

[CQ08] P. Cardaliaguet and M. Quincampoix. “Deterministic Differential Games under Probability Knowledge of Initial Condition”. In: *International Game Theory Review* 10.01 (Mar. 2008), pp. 1–16.

[Car13] P. Cardaliaguet. “Notes on Mean Field Games”. Available at https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf. 2013.

[Car+19] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions. *The Master Equation and the Convergence Problem in Mean Field Games: (AMS-201)*. Princeton University Press, Aug. 2019.

[Car+24] P. Cardaliaguet, N. Forcadel, T. Girard, and R. Monneau. “Conservation Laws and Hamilton-Jacobi Equations on a Junction: The Convex Case”. In: *Discrete and Continuous Dynamical Systems* 44.12 (2024), pp. 3920–3961.

[CP20] P. Cardaliaguet and A. Porretta. “An Introduction to Mean Field Game Theory”. In: *Mean Field Games: Cetraro, Italy 2019*. Lecture Notes in Mathematics. Springer International Publishing, 2020, pp. 1–158.

[CD18a] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I*. Vol. 83. Probability Theory and Stochastic Modelling. Springer International Publishing, 2018.

[CD18b] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications II*. Vol. 84. Probability Theory and Stochastic Modelling. Springer International Publishing, 2018.

[Cav+18] G. Cavagnari, A. Marigonda, K. T. Nguyen, and F. S. Priuli. “Generalized Control Systems in the Space of Probability Measures”. In: *Set-Valued and Variational Analysis* 26.3 (Sept. 2018), pp. 663–691.

[CSS23a] G. Cavagnari, G. Savaré, and G. E. Sodini. *A Lagrangian Approach to Totally Dissipative Evolutions in Wasserstein Spaces*. Preprint (arXiv:2305.05211). May 2023.

[CSS23b] G. Cavagnari, G. Savaré, and G. E. Sodini. “Dissipative Probability Vector Fields and Generation of Evolution Semigroups in Wasserstein Spaces”. In: *Probability Theory and Related Fields* 185.3-4 (Apr. 2023), pp. 1087–1182.

[Chi17] L. Chizat. “Transport Optimal de Mesures Positives : Modèles, Méthodes Numériques, Applications”. These de Doctorat. Paris Sciences et Lettres (ComUE), Nov. 2017.

[Cla13] F. Clarke. *Functional Analysis, Calculus of Variations and Optimal Control*. Graduate Texts in Mathematics 264. London ; New York: Springer, 2013.

[CPS23] R. M. Colombo, V. Perrollaz, and A. Sylla. “Conservation Laws and Hamilton–Jacobi Equations with Space Inhomogeneity”. In: *Journal of Evolution Equations* 23.3 (July 2023), p. 50.

[Con+24] G. Conforti, R. C. Kraaij, L. Tamanini, and D. Tonon. *Hamilton-Jacobi Equations for Wasserstein Controlled Gradient Flows: Existence of Viscosity Solutions*. Jan. 2024.

[CKT23a] G. Conforti, R. C. Kraaij, and D. Tonon. *Hamilton-Jacobi Equations for Controlled Gradient Flows: Cylindrical Test Functions*. Preprint, available at https://arxiv.org/abs/2302.06571. 2023.

[CKT23b] G. Conforti, R. C. Kraaij, and D. Tonon. “Hamilton-Jacobi Equations for Controlled Gradient Flows: The Comparison Principle”. In: *Journal of Functional Analysis* 284.9 (2023), pp. 1–53.

[Cos+24] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham, and M. Rosestolato. “Master Bellman Equation in the Wasserstein Space: Uniqueness of Viscosity Solutions”. In: *Transactions of the American Mathematical Society* 377.01 (Jan. 2024), pp. 31–83.

[CM23] A. Cosso and M. Martini. “On Smooth Approximations in the Wasserstein Space”. In: *Electronic Communications in Probability* 28.none (Jan. 2023), pp. 1–11.

[CIL92] M. G. Crandall, H. Ishii, and P.-L. Lions. “User’s Guide to Viscosity Solutions of Second Order Partial Differential Equations”. In: *Bulletin of the American Mathematical Society* 27.1 (1992), pp. 1–67.

[CL71] M. G. Crandall and T. M. Liggett. “Generation of Semi-Groups of Nonlinear Transformations on General Banach Spaces”. In: *American Journal of Mathematics* 93.2 (1971), pp. 265–298.

[CL85] M. G. Crandall and P.-L. Lions. "Hamilton-Jacobi Equations in Infinite Dimensions I. Uniqueness of Viscosity Solutions". In: *Journal of Functional Analysis* 62.3 (July 1985), pp. 379–396.

[CL86a] M. G. Crandall and P.-L. Lions. "Hamilton-Jacobi Equations in Infinite Dimensions. II. Existence of Viscosity Solutions". In: *Journal of Functional Analysis* 65.3 (Feb. 1986), pp. 368–405.

[CL86b] M. G. Crandall and P.-L. Lions. "Hamilton-Jacobi Equations in Infinite Dimensions, III". In: *Journal of Functional Analysis* 68.2 (Sept. 1986), pp. 214–247.

[CL83] M. G. Crandall and P.-L. Lions. "Viscosity Solutions of Hamilton-Jacobi Equations". In: *Transactions of the American Mathematical Society* 277.1 (1983), pp. 1–42.

[CL90] M. G. Crandall and P.-L. Lions. "Viscosity Solutions of Hamilton-Jacobi Equations in Infinite Dimensions. IV. Hamiltonians with Unbounded Linear Terms". In: *Journal of Functional Analysis* 90.2 (May 1990), pp. 237–283.

[CL91] M. G. Crandall and P.-L. Lions. "Viscosity Solutions of Hamilton-Jacobi Equations in Infinite Dimensions. V. Unbounded Linear Terms and B-continuous Solutions". In: *Journal of Functional Analysis* 97.2 (May 1991), pp. 417–465.

[CL94] M. G. Crandall and P.-L. Lions. "Viscosity Solutions of Hamilton-Jacobi Equations in Infinite Dimensions. VII. The HJB Equation Is Not Always Satisfied". In: *Journal of Functional Analysis* 125.1 (Oct. 1994), pp. 111–148.

[D'A+10] C. D'Apice, S. Göttlich, M. Herty, and B. Piccoli. *Modeling, Simulation, and Optimization of Supply Chains*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, Jan. 2010.

[Dau23] S. Daudin. "Optimal Control of the Fokker-Planck Equation under State Constraints in the Wasserstein Space". In: *Journal de Mathématiques Pures et Appliquées* 175 (July 2023), pp. 37–75.

[DS24] S. Daudin and B. Seeger. "A Comparison Principle for Semilinear Hamilton–Jacobi–Bellman Equations in the Wasserstein Space". In: *Calculus of Variations and Partial Differential Equations* 63.4 (Apr. 2024), p. 106.

[DS20] L. Dello Schiavo. "A Rademacher-type Theorem on L^2-Wasserstein Spaces over Closed Riemannian Manifolds". In: *Journal of Functional Analysis* 278.6 (Apr. 2020), p. 108397.

[DU77] J. Diestel and J. J. Uhl. *Vector Measures*. Vol. 15. Mathematical Surveys and Monographs. Providence, Rhode Island: American Mathematical Society, June 1977.

[DS82] R. L. Dobrushin and Ra. Siegmund-Schultze. "The Hydrodynamic Limit for Systems of Particles with Independent Evolution". In: *Mathematische Nachrichten* 105.1 (1982), pp. 199–224.

[Dud76] R. M. Dudley. *Probabilities and Metrics: Convergence of Laws on Metric Spaces, with a View to Statistical Testing*. Lecture Notes Series 45. Aarhus Universitet: Matematik Institut, 1976.

[Eke74] I. Ekeland. "On the Variational Principle". In: *Journal of Mathematical Analysis and Applications* 47.2 (Aug. 1974), pp. 324–353.

[ETZ16] I. Ekren, N. Touzi, and J. Zhang. "Viscosity Solutions of Fully Nonlinear Parabolic Path Dependent PDEs: Part I". In: *The Annals of Probability* 44.2 (Mar. 2016), pp. 1212–1253.

[Eva98] L. C. Evans. *Partial Differential Equations*. Graduate Studies in Mathematics v. 19. Providence, R.I: American Mathematical Society, 1998.

[FGŚ17] G. Fabbri, F. Gozzi, and A. Święch. *Stochastic Optimal Control in Infinite Dimension*. Vol. 82. Probability Theory and Stochastic Modelling. Cham: Springer International Publishing, 2017.

[FK09] J. Feng and M. Katsoulakis. "A Comparison Principle for Hamilton–Jacobi Equations Related to Controlled Gradient Flows in Infinite Dimensions". In: *Archive for Rational Mechanics and Analysis* 192.2 (May 2009), pp. 275–310.

[FN12] J. Feng and T. Nguyen. “Hamilton–Jacobi Equations in Space of Measures Associated with a System of Conservation Laws”. In: *Journal de Mathématiques Pures et Appliquées* 97.4 (Apr. 2012), pp. 318–390.

[Fil63] A. F. Filippov. “Differential Equations with Multivalued Discontinuous Right-Hand Side”. In: *Dokl. Akad. Nauk SSSR* 151.1 (1963). Accessible at `https://www.mathnet.ru/links/f213a6e48d91f81f201db1664a2622af/dan28255.pdf`, pp. 65–68.

[FS06] W. H. Fleming and H. M. Soner. *Controlled Markov Processes and Viscosity Solutions*. Vol. 25. Stochastic Modelling and Applied Probability. New York: Springer-Verlag, 2006.

[Fra65] S. P. Franklin. “Spaces in Which Sequences Suffice”. In: *Fundamenta Mathematicae* 57.1 (1965), pp. 107–115.

[Fra93] H. Frankowska. “Lower Semicontinuous Solutions of Hamilton–Jacobi–Bellman Equations”. In: *SIAM Journal on Control and Optimization* 31.1 (Jan. 1993), pp. 257–272.

[FL23] H. Frankowska and T. Lorenz. “Filippov’s Theorem for Mutational Inclusions in a Metric Space”. In: *Annali Scuola Normale Superiore - Classe di Scienze* (June 2023), pp. 1053–1094.

[GNT22] T. Gallouët, A. Natale, and G. Todeschi. “From Geodesic Extrapolation to a Variational BDF2 Scheme for Wasserstein Gradient Flows”. In: (2022).

[GKP11] W. Gangbo, H. Kim, and T. Pacini. *Differential Forms on Wasserstein Space and Infinite-Dimensional Hamiltonian Systems*. Vol. 211. Memoirs of the American Mathematical Society. American Mathematical Society, May 2011.

[GNT08] W. Gangbo, T. Nguyen, and A. Tudorascu. “Hamilton-Jacobi Equations in the Wasserstein Space”. In: *Methods and Applications of Analysis* 15.2 (2008), pp. 155–184.

[GŚ15a] W. Gangbo and A. Święch. “Existence of a Solution to an Equation Arising from the Theory of Mean Field Games”. In: *Journal of Differential Equations* 259.11 (Dec. 2015), pp. 6573–6643.

[GŚ15b] W. Gangbo and A. Święch. “Metric Viscosity Solutions of Hamilton–Jacobi Equations Depending on Local Slopes”. In: *Calculus of Variations and Partial Differential Equations* 54.1 (Sept. 2015), pp. 1183–1218.

[GŚ14] W. Gangbo and A. Święch. “Optimal Transport and Large Number of Particles”. In: *Discrete and Continuous Dynamical Systems* 34.4 (2014), pp. 1397–1441.

[GT19] W. Gangbo and A. Tudorascu. “On Differentiability in the Wasserstein Space and Well-Posedness for Hamilton–Jacobi Equations”. In: *Journal de Mathématiques Pures et Appliquées* 125 (May 2019), pp. 119–174.

[GP06] M Garavello and B. Piccoli. “Traffic Flow on Networks: Conservation Laws Model”. In: Jan. 2006, pp. xvi+243.

[Gib02] J. W. Gibbs. *Elementary Principles of Statistical Mechanics*. Charles Scribner’s Sons, 1902.

[GHN15] Y. Giga, N. Hamamuki, and A. Nakayasu. “Eikonal Equations in Metric Spaces”. In: *Transactions of the American Mathematical Society* 367.1 (Jan. 2015), pp. 49–66.

[Gig08] N. Gigli. “On the Geometry of the Space of Probability Measures Endowed with the Quadratic Optimal Transport Distance”. PhD thesis. Pisa: Scuola Normale Superiore di Pisa, 2008.

[Gig11] N. Gigli. “On the Inverse Implication of Brenier-McCann Theorems and the Structure of $(P_2(M), W_2)$”. In: *Methods and Applications of Analysis* 18.2 (2011), pp. 127–158.

[GR91] E. Godlewski and P. Raviart. *Hyperbolic Systems of Conservation Laws*. Mathématiques & Applications. Ellipses, 1991.

[HK15] R. Hynd and H. K. Kim. “Value Functions in the Wasserstein Spaces: Finite Time Horizons”. In: *Journal of Functional Analysis* 269.4 (Aug. 2015), pp. 968–997.

[Hyt+16] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. *Analysis in Banach Spaces*. Cham: Springer International Publishing, 2016.

[IM17] C. Imbert and R. Monneau. “Flux-Limited Solutions for Quasi-Convex Hamilton-Jacobi Equations on Networks”. In: *Annales scientifiques de l’École normale supérieure* 50.2 (2017), pp. 357–448.

[IMZ13] C. Imbert, R. Monneau, and H. Zidani. “A Hamilton-Jacobi Approach to Junction Problems and Application to Traffic Flows”. In: *ESAIM: Control, Optimisation and Calculus of Variations* 19.1 (Jan. 2013), pp. 129–166.

[Ish85] H. Ishii. “Hamilton-Jacobi Equations with Discontinuous Hamiltonians on Arbitrary Open Sets”. In: *Bulletin of the Faculty of Science and Engineering.* I 28 (1985), pp. 33–77.

[Ish89] H. Ishii. “On Uniqueness and Existence of Viscosity Solutions of Fully Nonlinear Second-Order Elliptic PDE’s”. In: *Communications on Pure and Applied Mathematics* 42.1 (1989), pp. 15–45.

[Jam64] R. C. James. “Weakly Compact Sets”. In: *Transactions of the American Mathematical Society* 113.1 (1964), pp. 129–140.

[Jen88] R. Jensen. “The Maximum Principle for Viscosity Solutions of Fully Nonlinear Second Order Partial Differential Equations”. In: *Archive for Rational Mechanics and Analysis* 101.1 (Mar. 1988), pp. 1–27.

[Jer22] O. Jerhaoui. “Viscosity Theory of First Order Hamilton Jacobi Equations in Some Metric Spaces”. PhD thesis. Paris: Institut Polytechnique de Paris, 2022.

[JZ23a] O. Jerhaoui and H. Zidani. “A General Comparison Principle for Hamilton Jacobi Bellman Equations on Stratified Domains”. In: *ESAIM: Control, Optimisation and Calculus of Variations* 29 (2023), p. 9.

[JZ23b] O. Jerhaoui and H. Zidani. “Viscosity Solutions of Hamilton-Jacobi Equations in Proper CAT(0) Spaces”. In: *The Journal of Geometric Analysis* 34.2 (Dec. 2023), p. 47.

[Jim24] C. Jimenez. “Equivalence between Strict Viscosity Solution and Viscosity Solution in the Space of Wasserstein and Regular Extension of the Hamiltonian in $L^2_{\mathbb{P}}$”. In: *Journal of Convex Analysis* 31.2 (Apr. 2024), p. 619.

[JMQ23] C. Jimenez, A. Marigonda, and M. Quincampoix. “Dynamical Systems and Hamilton-Jacobi-Bellman Equations on the Wasserstein Space and Their L^2 Representations”. In: *SIAM Journal on Mathematical Analysis* 55.5 (Oct. 2023), pp. 5919–5966.

[JMQ20] C. Jimenez, A. Marigonda, and M. Quincampoix. “Optimal Control of Multiagent Systems in the Wasserstein Space”. In: *Calculus of Variations and Partial Differential Equations* 59 (Mar. 2020), pp. 1–45.

[Jui11] N. Juillet. “On Displacement Interpolation of Measures Involved in Brenier’s Theorem”. In: *Proceedings of the American Mathematical Society* 139.10 (Oct. 2011), pp. 3623–3623.

[Kan42] L. Kantorovich. “On the Translocation of Masses”. In: *Dokl. Akad. Nauk SSSR* 37.7-8 (1942), pp. 227–229.

[KS87] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequalities and Their Applications.* Classics in Applied Mathematics 31. Philadelphia: Society for Industrial and Applied Mathematics, 1987.

[KL99] C. Kipnis and C. Landim. *Scaling Limits of Interacting Particle Systems.* Vol. 320. Grundlehren Der Mathematischen Wissenschaften. Berlin, Heidelberg: Springer, 1999.

[KM97] V. N. Kolokoltsov and V. P. Maslov. *Idempotent Analysis and Its Applications.* Dordrecht: Springer Netherlands, 1997.

[Kru70] S. N. Kružkov. “First-Order Quasilinear Equations in Several Independent Variables”. In: *Mathematics of the USSR-Sbornik* 10.2 (Feb. 1970), pp. 217–243.

[Lio82] P.-L. Lions. *Generalized Solutions of Hamilton-Jacobi Equations.* Research Notes in Mathematics 69. Boston: Pitman, 1982.

[Lio06] P.-L. Lions. *Jeux à Champ Moyen*. Conférences au Collège de France. 2006/2007.

[LS84] P.-L. Lions and P. Souganidis. *Differential Games, Optimal Control and Directional Derivatives of Viscosity Solutions of Bellman's and Isaacs' Equations*. Tech. rep. University of Winsconsin-Madison: Mathematics Research Center, 1984. Chap. Technical Reports.

[LS17] P.-L. Lions and P. Souganidis. "Well Posedness for Multi-Dimensional Junction Problems with Kirchoff-type Conditions". In: *Rendiconti Lincei - Matematica e Applicazioni* 28 (Apr. 2017).

[Lor10] T. Lorenz. *Mutational Analysis: A Joint Framework for Cauchy Problems in and Beyond Vector Spaces*. Vol. 1996. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2010.

[Lot16] J. Lott. "On Tangent Cones in Wasserstein Space". In: *Proceedings of the American Mathematical Society* 145.7 (Dec. 2016), pp. 3127–3136.

[LN19] A. Lytchak and K. Nagano. "Geodesically Complete Spaces with an Upper Curvature Bound". In: *Geometric and Functional Analysis* 29.1 (Feb. 2019), pp. 295–342.

[MQ18] A. Marigonda and M. Quincampoix. "Mayer Control Problem with Probabilistic Uncertainty on Initial Positions". In: *Journal of Differential Equations* 264.5 (Mar. 2018), pp. 3212–3252.

[May98] U. F. Mayer. "Gradient Flows on Nonpositively Curved Metric Spaces and Harmonic Maps". In: *Communications in Analysis and Geometry* 6.2 (1998), pp. 199–253.

[McC01] R. McCann. "Polar Factorization of Maps on Riemannian Manifolds". In: *Geometric & Functional Analysis* 11.3 (Aug. 2001), pp. 589–608.

[Mor55] C. B. Morrey. "On the Derivation of the Equations of Hydrodynamics from Statistical Mechanics". In: *Communications on Pure and Applied Mathematics* 8.2 (1955), pp. 279–326.

[MZ24] C. Mou and J. Zhang. *Wellposedness of Second Order Master Equations for Mean Field Games with Nonsmooth Data*. Vol. 302. Memoirs of the American Mathematical Society. Providence: American Mathematical Society, 2024.

[OS94] Y. Otsu and T. Shioya. "The Riemannian Structure of Alexandrov Spaces". In: *Journal of Differential Geometry* 39.3 (Jan. 1994).

[Ott01] F. Otto. "The Geometry of Dissipative Evolution Equations: The Porous Medium Equation". In: *Communications in Partial Differential Equations* 26.1-2 (2001), pp. 101–174.

[Pat23] E. Patterson. *Categorical Probability Theory*. 2023.

[Pet95] A. Petrunin. "Quasigeodesics and Gradient Curves in Alexandrov Spaces". In: (1995). Unpublished note, available at `https://anton-petrunin.github.io/papers/qg_ams.pdf`.

[PW18] H. Pham and X. Wei. "Bellman Equation and Viscosity Solutions for Mean-Field Stochastic Control Problem". In: *ESAIM: Control, Optimisation and Calculus of Variations* 24.1 (Jan. 2018), pp. 437–461.

[Pic19] B. Piccoli. "Measure Differential Equations". In: *Archive for Rational Mechanics and Analysis* 233.3 (Sept. 2019), pp. 1289–1317.

[Pon+62] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. Mishchenko. *The Mathematical Theory of Optimal Processes*. New York: Interscience Publishers, Wiley and Sons, 1962.

[Pro56] Yu. V. Prokhorov. "Convergence of Random Processes and Limit Theorems in Probability Theory". In: *Theory of Probability & Its Applications* 1.2 (Jan. 1956), pp. 157–214.

[RZ13] Z. Rao and H. Zidani. "Hamilton–Jacobi–Bellman Equations on Multi-domains". In: *Control and Optimization with PDE Constraints*. Basel: Springer, 2013, pp. 93–116.

[Roc70] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Dec. 1970.

[RV91] J. D. L. Rowland and R. B. Vinter. "Construction of Optimal Feedback Controls". In: *Systems & Control Letters* 16.5 (May 1991), pp. 357–367.

[San15] F. Santambrogio. *Optimal Transport for Applied Mathematicians*. Vol. 87. Progress in Nonlinear Differential Equations and Their Applications. Springer International Publishing, 2015.

[SC13] D. Schieborn and F. Camilli. "Viscosity Solutions of Eikonal Equations on Topological Networks". In: *Calculus of Variations and Partial Differential Equations* 46.3-4 (Mar. 2013), pp. 671–686.

[Smi94] S. K. Smirnov. "Decomposition of Solenoidal Vector Charges into Elementary Solenoids and the Structure of Normal One-Dimensional Currents". In: *St. Petersburg Math Journal* 5.4 (1994).

[SY24] H. M. Soner and Q. Yan. *Viscosity Solutions of the Eikonal Equation on the Wasserstein Space*. Preprint, available at http://arxiv.org/abs/2308.04097. Jan. 2024.

[Son86] H. M. Soner. "Optimal Control with State-Space Constraint. II". In: *SIAM Journal on Control and Optimization* 24.6 (Nov. 1986), pp. 1110–1122.

[Sto12] I. Stojkovic. "Approximation for Convex Functionals on Non-Positively Curved Spacesand the Trotter–Kato Product Formula". In: *Advances in Calculus of Variations* 5.1 (Jan. 2012), pp. 77–126.

[Stu03] K.-T. Sturm. "Probability Measures on Metric Spaces of Nonpositive Curvature". In: *Contemporary Mathematics*. Vol. 338. American Mathematical Society, 2003, pp. 357–390.

[Tak11] A. Takatsu. "Wasserstein Geometry of Gaussian Measures". In: *Osaka Journal of Mathematics* 48.4 (Dec. 2011), pp. 1005–1026.

[TY12] A. Takatsu and T. Yokota. "Cone Structure of L^2-Wasserstein Spaces". In: *Journal of Topology and Analysis* 04.02 (June 2012), pp. 237–253.

[Vas69] L. N. Vaserstein. "Markov Processes over Denumerable Products of Spaces, Describing Large Systems of Automata". In: *Problemy Peredači Informacii* 5.3 (1969), pp. 64–72.

[Vil09] C. Villani. *Optimal Transport*. Vol. 338. Grundlehren Der Mathematischen Wissenschaften. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[Vil03] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics ; v. 58. Providence, R.I: American Mathematical Society, 2003.

[Vin10] R. Vinter. *Optimal Control*. Boston: Birkhäuser, 2010.

[VRS09] M.-K. Von Renesse and K.-T. Sturm. "Entropic Measure and Wasserstein Diffusion". In: *The Annals of Probability* 37.3 (May 2009), pp. 1114–1191.

[War72] J. Warga. *Optimal Control of Differential and Functional Equations*. New York: Academic Press, 1972.

[Yan+08] E. B. Yanovskaya, I. Irle, S. Simons, J. Von Neumann, and S Kakutani. "Minimax Theorems". In: *Encyclopedia of Optimization*. Springer, Boston, MA, 2008, pp. 2087–2093.