Geometry and Stochastic Calculus on Wasserstein spaces

Christian Selinger

PhD Thesis under the supervision of Prof. Anton Thalmaier at
Université du Luxembourg, September 2010

Contents

Chapter 0

Introduction

Damit der Mensch sich irre,
muß er schon mit der Menschheit konform urteilen.
L. Wittgenstein: Über Gewissheit.

The main object of interest in the present thesis is $P(M)$ – the space of probability measures on a Riemannian manifold (M, g) endowed with the Wasserstein distance: Let us fix μ and ν in $P(M)$ and give ourselves a (lower-semicontinuous) cost function $c : M \times M \to \mathbb{R}_+ \cup \{+\infty\}$. We denote by Π the set of all product probability measures π on $M \times M$ such that $\pi(E \times M) = \mu(E)$ resp. $\pi(M \times E) = \nu(E)$ for all Borel measurable sets E, in other words Π denotes the set of all product probability measures with fixed marginals μ and ν. The problem of optimal transport is to find

$$\inf_{\mu \in \Pi} \int_{M \times M} c(x, y)\pi(dx, dy). \tag{1}$$

In the 1940s Kantorovich developed a technique (see [Vil08] and [RR98] for a contemporary account on Kantorovich's contribution) that led to the following conclusion: If M happens to be a Polish space, i.e. a complete separable metric space and c a lower-semicontinuous cost function, then there exists a unique optimal transport plan $\tilde{\pi}$ such that

$$\int_{M \times M} c(x, y)\tilde{\pi}(dx, dy)$$

is minimal. If for some $p \in \mathbb{N}$ we have $c(x, y) = d_g(x, y)^p$, then

$$d_W(\mu, \nu)^p = \inf_{\pi \in \Pi} \int_{M \times M} d_g(x, y)^p \pi(dx, dy)$$

is a metric, the so-called p-Wasserstein[1] distance on the space of probability measures. The space $(P^p(M), d_W)$ -defined as the space of all Borel probability

[1] It was the Soviet mathematician L. N. Vasersthein who used the 1-Wasserstein distance in his 1969 publication on "Markov processes over denumerable products of spaces describing large system of automata"; even though the earliest and most foundational contributions to this subject are due to Kantorovich, we stick to the misspelled and historically inacurate naming which appears throughout the contemporary literature.

measures with finite p-th moment- is called p-Wasserstein space. If not stated otherwise in this work we assume that M is a complete, simply connected Riemannian manifold without boundary. Additionnally we will restrict ourselves to the 2-Wasserstein distance and henceforth omit any superscripts, i.e. if not stated otherwise $(P(M), d_W)$ denotes the 2-Wasserstein space over a complete, simply connected Riemannian manifold without boundary.
If we restrict ourselves to P_{ac}, the space of absolutely continuous probability measures, it was proved by [Bre91] and [McC01] (using Kantorovich's duality to prove a polar factorization of vector valued maps) that the optimal transport plan $\tilde{\pi}$ is actually an optimal transport map, i.e.

$$\tilde{\pi}(dx, dy) = (\mathrm{id} \times T)\#\mu$$

for T being a μ-almost unique map from M to M which has the following form

$$T(x) = \exp_x(-\nabla\varphi(x)),$$

where φ is a μ-almost unqiue convex function on M. As a corollary it is shown that $\tilde{T}(x) = \exp_x(-\nabla\varphi^*(x))$ is the inverse optimal transport map from ν to μ, * means the Legendre transform with respect to the Riemannian metric g.

In the last decade Wasserstein spaces have come in touch with many different areas of mathematics:

- In the field of PDEs it was possible to prove that a huge class of diffusion equations happens to be gradient flow equations (see [AGS08]) on $P(M)$ which made it possible to show contractivity for the corresponding diffusion semigroups and theorems on logarithmic Sobolev and Poincaré inequalities (see [Vil08]).
- In the field of metric geometry the convexity properties of Boltzmann and Renyi entropies along Wasserstein geodesics (where the underlying space M is not any longer a Riemannian manifold but just a geodesically complete length space) led to a consistent generalization of lower Ricci curvature bounds on metric spaces which is stable under Gromov-Hausdorff convergence of families of metric spaces (see [Stu06a], [Stu06b] and [LV09]).
- In the field of infinite differential geometry [Ott01] introduced for the first time the subspace $P^\infty(M)$ of smooth, positive probability densities on $M = \mathbb{R}^d$ as infinite-dimensional Riemannian manifold and calculated geodesic equations and sectional curvature bounds on this space. In the sequal [Lot08] developed formulas for Riemannian curvature on $P^\infty(M)$ in case that M is a complete, simply-connected Riemannian manifold without boundary and [GKP10] made clear how differential forms on Wasserstein spaces look like and proved that the first deRham cohomology group vanishes on $P(\mathbb{R}^d)$.
- In the field of stochastic analysis it was [SvR09] who constructed a continuous time Markov process X_t which is reversible with respect to the so-called entropic measure $\mathbb{P}^\beta$ on $P([0,1))$ resp. $P([0,1])$. Since the transition semi-group $p_t(x, dy)$ satisfies the following short-time asymptotics

$$\lim_{t\downarrow 0} t \log \int_A \int_B p_t(x, dy)\mathbb{P}^\beta(dx) = -\frac{d_W(A,B)^2}{2}$$

for Borel subsets A and B of $P([0,1))$ resp. $P([0,1])$ (with respect to the Wasserstein topology) and the square-field operator with respect to the correponding generator equals the square norm of the Wasserstein gradient, the process X_t is called Wasserstein diffusion. The entire construction of this process relies on Dirichlet forms on the space of non-decreasing functions from the interval to itself: this space and its differentiable structure can be mapped isometrically to the Wasserstein space. The Dirichlet form and the integration by parts formula involve the entropic measure $\mathbb{Q}^\beta$, which is nothing but the measure subject to the Dirichlet process. It defines $\mathbb{P}^\beta$ as the image under the forementioned isometry. The resulting process is then mapped to the space of probability measures.

In this thesis several of the above mentioned areas are treated:
In chapter 1 we give the most basic topological facts on the interplay between weak convergence and Wasserstein distance on P, additionally we introduce a locally convex topology on P^∞ and identify this space as infinite dimensional manifold in the sense of [KM97].
In chapter 2 we develop further the Riemannian calculus on P resp. P^∞ where the different approaches (calculus of variation, Riemannian geometry on spaces of smooth mappings and analysis on metric spaces) are shown to be equivalent on P^∞.
In chapter 3 we restrict ourself to tori as underlying manifolds and give calculations of renormalized (connection) Laplacians on the respective Wasserstein spaces, seen as the Hilbert-Schmidt trace of the Hessian: This stems from the fact that the only locally finite translation invariant measure on any Banach space is the trivial measure, which applies also to $P^\infty \subset P$, consequently we cannot give a proper meaning to neither the Hodge Laplacian nor to the usual connection Laplacian. The forementioned Hilbert-Schmidt trace depends on a real parameter $s > 3/2$. When calculating explicitely this trace for the Wasserstein space above the unit circle we are able to give the analytic continuation of the trace as a function of $s \in \mathbb{C} \setminus \{1\}$. This analytic continuation allows us to calculate the value of the trace at $s = 0$: It is shown that the resulting operator $\Delta_{\mathrm{P}^\infty(\mathbb{T}^1)}$ has a square-field operator which equals the squared Wasserstein gradient times the volume of the unit circle.
In chapter 4 we give an approximation of the Wasserstein space $P([0,1])$ by spaces of box-type measures (i.e. probability densities which are piecewise constant and have constant weight on each set of any n-size partition of the interval). This space is geodesically convex (in the sense of Wasserstein geometry) and can be mapped isometrically (via a mapping m^{-1}) to the $(n-1)$-simplex where a sticky diffusion process X_t^n (with respect to a non-Euclidean metric) in the spirit of Ikeda-Watanabe is constructed. We show that the family of processes $\{m(X_t^n); n \in \mathbb{N}\}$ is tight in $\mathcal{C}(\mathbb{R}_+; P([0,1]))$ with respect to the Skorohod topology.
In the last chapter we restrict ourselves to the space of histograms on the unit interval (i.e. probability densities which are piecewise constant and have varying weights; here the subsets of the partition are of uniform length). This space is not geodesically convex (in the Wasserstein sense) but we can calculate the Wasserstein distances numerically and obtain again a Riemannian metric on the n-simplex. We investigate explosion behaviour of the respective diffusion processes in dimension 1 and 2.

The present thesis lacks in each of its chapters of ground-breaking results – it is still very much work in progress, we mention therefore the most important open questions:
The main open issues in chapter 3 is the question in how far we can generalize the regularization procedure to Wasserstein spaces above arbitrary compact manifolds and if one can use this procedure to give a theory of stochastic flows with interaction as sketched in section 3.3 – an issue which has been worked on by LeJan and Raimond with completely different techniques (see the survey article [LR05] for Brownian flows on the unit circle and [LR04]). Another interesting question is if the regularization procedure can produce explicit formulas for a regularized Ricci curvature on Wasserstein spaces.
In chapter 4 the main issue is the identification of the limiting process: How far is its generator related to the regularized Wasserstein Laplacian?
In chapter 5 the unhandy Riemannian metrics obtained by projecting Wasserstein geodesics to the space of histograms give completely different boundary behaviour of the corresponding Brownian motion when comparing only dimension 1 and dimension 2. Since computational issues are still far behind the achievements of the abstract theory of optimal transport it would be interesting to shed more light on this question.

Acknowledgements. The author would like to thank first of all Anton Thalmaier for constant institutional and mathematical support at the university of Luxembourg. Karl-Theodor Sturm and Josef Teichmann are kindly acknowledged for discussions on the subject at various occasions. Max-Konstantin von Renesse is kindly thanked for sharing unpublished ideas and for collaboration on chapter three and four.
Of course I am indebted to my parents and my sister for constant moral backup.
I would like to thank Svetlana Adjoua Coffi for being with me. Thanks to Stephan Sturm, Kolehe Abdoulaye Coulibaly, Nicolas Juillet, Hendrik Weber, Hans and Thomas Schoiswohl, Johanna Moser, Daniel Kohlmeigner, Andreas Hutterer, Bruno Mounikou, Nicolas Genest, Ninel Kameraz-Kos, Ilona Kryszkiewicz, Fatoumata Faye, Jean-François Grassineau, Yero Bobo Bah and all those I forgot for discussions on mathematics, birds, fish, gods and poetry, for hiking the mountains, for making and teaching me music and for cooking and eating together.

Chapter 1

Topologies of probability measures

In this chapter we will introduce the space of all probability measures on a Polish space and its various subsets, among them the space of smooth, positive probability densities on a compact Riemannian manifold. In addition to the weak topology we put on the latter a coarser topology: the final topology with respect to all smooth curves into the space of smooth, positive densities. In contrast to the weak topology, which makes the space of all probability measures on a compact space itself into a Polish space, the so-called topology of smooth curves is a priori only definied on he space of smooth, positive probability densities; and this subspace is not complete for the weak topology. But from the point of view of infinite-dimensional differential geometry, the topology of smooth curves makes the space of smooth, positive densities into a topological (and even smooth) open submanifold of a locally convex space. The reason for choosing a locally convex space and not a Banach one as modelling space lies in a theorem by [Omo78], which states that any Banach Lie group acting effectively on a finite-dimensional compact manifold is necessarily finite dimensional itself. Hence the group of smooth diffeomorphisms cannot be modelled on a Banach space. As we will see later the space of smooth positive probability densities can be seen as a topological space with a foliation given by the left action of the smooth diffeomorphism group on it - the group action is locally free when we mod out the subgroup of volume preserving diffeomorphism- and the topology of smooth curves on the diffeomorphism group is inherited in this way by the space of smooth, positive probability densities.
Thanks to Urysohn's theorem we know that the space of probability measures on a Polish measurable space $(X, \mathcal{F})$ is metrizable; among the many notions of distance we mention the following ones:

Definition 1.0.1. *For $\mu, \nu \in P(X)$ we define*

- *Total variation distance*

$$d_{TV}(\mu, \nu) = \sup\{|\mu(A) - \nu(A)|; A \in \mathcal{F}\}$$

- *Hellinger distance*

$$d_H(\mu,\nu)^2 = \frac{1}{2}\int\left(\sqrt{\frac{d\mu}{dvol}} - \sqrt{\frac{d\nu}{dvol}}\right)^2 dvol$$

- *Wasserstein distance: Given $\mu, \nu \in P$: $\Pi := \{\pi \in P(X\times X) : \pi(A\times X) = \mu(A); \pi(X \times B) = \nu(B)$ for all Borel sets $A, B\}$.*

$$d_W(\mu,\nu)^2 := \inf_{\pi\in\Pi}\int d(x,y)^2\pi(dx,dy)$$

is called quadratic Wasserstein distance and the metric space (P, d_W) is called Wasserstein space. The existence and uniqueness of the variational problem was proved by Kantorovich introducing a duality technique, for a survey see [Vil03].

Whereas the total variation distance (widely used in statistics) metrizes strong convergence (with respect to the total variation norm) Wasserstein distances can be used more widely since it metrizes weak convergence (by testing against bounded continuous functions). Recently in [MH08] a spectral gap for Markov semigroups in an infinite-dimensional setting has been shown by using Wasserstein distances rather than total variation distances, since usual (Harris) conditions failed. Last but not least, from a geometric point of view the (quadratic) Wasserstein distance can be understood as geodesic distance and calculated as the infimum of the energy over all paths linking two distinct probability measures, see chapter 2 for a precise meaning of this paraphrasing.

Definition 1.0.2. *Given a complete simply connected finite-dimensional Riemannian manifold M without boundary with its geodesic distance d we define:*

- *$P(M) := \{\mu$ Borel probability measure on M such that $\int_M d(x_0,x)^2\mu(dx) < \infty\}$. If we don't need to emphasize the underlying manifold we write shorthand P.*
- *$P_{ac}(M) := \{\mu \in P : \mu \ll vol_M\}$*
- *Define*

$$P^\infty(M) := \{\mu \in P_{ac} : m(x) := \frac{d\mu}{dvol_M}(x) > 0;\ \text{for a.e. } x \in M, m \in \mathcal{C}^\infty(M)\}$$

For all measures in $P_{ac}(M)$ (when not stated otherwise) we use henceforth the symbol μ both for the measure and its density function.

- *$\Sigma_n = \{\underline{\lambda} \in \mathbb{R}^{n+1}_{\geq 0} : \sum_{i=1}^{n+1}\lambda_i = 1\}$*
- *Given a finite partition $\mathfrak{A} = \cup_{i=1}^n A_i$ of M, where each A_i is a measurable set (with respect to the volume measure of M) that has non-empty interior. To each $\underline{\lambda} \in \Sigma_{n-1}$ we associate a probability measure (**'histogram'**) on M having the following density functions with respect to the volume measure:*

$$f(x) := \sum_{i=1}^n \frac{\lambda_i}{vol(A_i)}1_{A_i}(x);\ \ x \in M$$

We denote the set of all histograms (with respect to a fixed partition) by H_n.

- *In the case when* M *is an interval (e.g.* $[0,1]$*) we define a (***'box-type measure'***) by associating to each* $\underline{\lambda} \in \Sigma_{n-1}$ *the following density functions with respect to the volume measure:*

$$g(x) := \sum_{i=1}^{n} \left(\frac{1}{n\lambda_i} \mathbf{1}_{\{\lambda_i>0\}} \mathbf{1}_{[\sum_{k=1}^{i-1}\lambda_k, \sum_{k=1}^{i}\lambda_k)}(x)dx + \frac{1}{n}\mathbf{1}_{\{\lambda_i=0\}}\delta_{\sum_{k=1}^{i}\lambda_k}(x) \right); \ \ x \in M$$

We denote the set of all box-type measures (with n *boxes) by* G_n.

Lemma 1.0.1 (Probabilistic glueing)**.** *[Vil03] Let* μ_1, μ_2, μ_3 *be probability measures supported in the Polish spaces* X_1, X_2, X_3 *respectively and let* $\pi_{12} \in \Pi(\mu_1, \mu_2)$ *and* $\pi_{23} \in \Pi(\mu_2, \mu_3)$. *Then there exists a probability measure* $\pi \in P(X_1 \times X_2 \times X_3)$ *with marginals* π_{12} *on* $X_1 \times X_2$ *resp.* π_{23} *on* $X_2 \times X_3$.

Proof. Consider the probability measures π_{12} and π_{23} with common marginal μ_2. By the disintegration theorem there exist measurable mappings:

$$\pi_{12,2} : X_2 \to P(X_1),$$
$$\pi_{23,2} : X_2 \to P(X_3)$$

such that:

$$\pi_{12} = \int_{X_2} (\pi_{12,2}(x_2) \otimes \delta_{x_2})\, d\mu_2(x_2),$$
$$\pi_{23} = \int_{X_2} (\delta_{x_2} \otimes \pi_{23,2}(x_2))\, d\mu_2(x_2)$$

in the sense that for all measurable $A \subset X_1 \times X_2$:

$$\pi_{12}(A) = \int_{X_2} \pi_{12,2}(x_2)(A^{x_2}) d\mu_2(x_2)$$

with $A^{x_2} = \{x_1 \in X_1 : (x_1, x_2) \in A\}$. In order to construct a probability measure on $X_1 \times X_2 \times X_3$ we set:

$$\pi = \int_{X_2} (\pi_{12,2}(x_2) \otimes \delta_{x_2} \otimes \pi_{23,2}(x_2))\, d\mu_2(x_2)$$

We check that π is indeed a probablity measure: For any measurable set $A \subset X_1 \times X_2 \times X_3$ write pr_{12} resp. pr_{23} for the projection onto $X_1 \times X_2$ resp. $X_2 \times X_3$.

$$\pi(A) = \int_{X_2} \pi_{12,2}(x_2)[(pr_{12}(A))^{x_2}]\, \pi_{23,2}(x_2)[(pr_{23}(A))^{x_2}]\, d\mu_2(x_2).$$

Since for $d\mu_2$-almost every x_2 $\pi_{12,2}(x_2)$ and $\pi_{23,2}(x_2)$ are probability measures the total mass of π equals one.
We check that π has marginal π_{12} on $X_1 \times X_2$: Take any measurable set $A \subset X_1$, then:

$$\begin{aligned} \pi(A \times X_2 \times X_3) &= \int_{X_2} \pi_{12,2}(x_2)[(A \times X_2)^{x_2}]\pi_{23,2}(x_2)(X_3)\, d\mu_2(x_2) \\ &= \int_{X_2} \pi_{12,2}(x_2)(A)\, d\mu_2(x_2) \\ &= \mu_2(A), \end{aligned}$$

for marginals of π on $X_2 \times X_3$ an analogous calculation applies. □

1.1 Weak topology and optimal transport distance

Theorem 1.1.1 (Topological properties)**.** *[Vil03] Let M be a metric space and $P(M)$ be equipped with the Wasserstein distance d_W, then:*

1. *P is a metric space.*
2. *Convergence in Wasserstein distance is equivalent to weak convergence plus convergence of second moments.*
3. *If M is Polish, then the Wasserstein space $P(M)$ is also Polish.*

Proof. **1. (P is a metric space.)** Symmetry and non-negativity is clear by definition and it holds for all $\mu \in P(M)$ that $d_W(\mu,\mu) = 0$. Let $\mu, \nu \in P(M)$ be given and suppose that $d_W(\mu,\nu) = 0$. Let $\pi \in \Pi(\mu,\nu)$ be an optimal transport plan, i.e. $\int_{M\times M} d(x,y)^2\, \pi(dx,dy) = d_W(\mu,\nu)^2 = 0$, it follows that π is supported on the diagonal. From this and by the marginal condition it follows that for all $\varphi \in \mathcal{C}_b(M)$:

$$\begin{aligned}
\int_X \varphi(x)\, \mu(dx) &= \int_{M\times M} \varphi(x)\, \pi(dx,dy) \\
&= \int_{M\times M} \varphi(y)\, \pi(dx,dy) \\
&= \int_M \varphi(y)\, \nu(dy),
\end{aligned}$$

hence $\mu = \nu$.

The triangle inequality: Let μ_1, μ_2, μ_3 and π_{12}, π_{23}, π be as in Lemma 1.0.1 and π_{13} the marginal of π on $X_1 \times X_3$. Then using the Minkowski inequality:

$$\begin{aligned}
d_W(\mu_1,\mu_3) &\le \left(\int_{X_1\times X_3} d(x_1,x_3)^2\, d\pi_{13}(x_1,x_3) \right)^{\frac{1}{2}} \\
&= \left(\int_{X_1\times X_2\times X_3} d(x_1,x_3)^2\, d\pi(x_1,x_2,x_3) \right)^{\frac{1}{2}} \\
&\le \left(\int_{X_1\times X_2\times X_3} (d(x_1,x_2)^2 + d(x_2,x_3))^2\, d\pi(x_1,x_2,x_3) \right)^{\frac{1}{2}} \\
&\le \left(\int_{X_1\times X_2\times X_3} d(x_1,x_2)^2\, d\pi(x_1,x_2,x_3) \right)^{\frac{1}{2}} \\
&\quad + \left(\int_{X_1\times X_2\times X_3} d(x_2,x_3)^2\, d\pi(x_1,x_2,x_3) \right)^{\frac{1}{2}} \\
&= \left(\int_{X_1\times X_2} d(x_1,x_2)^2\, d\pi_{12}(x_1,x_2) \right)^{\frac{1}{2}} \\
&\quad + \left(\int_{X_2\times X_3} d(x_2,x_3)^2\, d\pi_{23}(x_2,x_3) \right)^{\frac{1}{2}} \\
&= d_W(\mu_1,\mu_2) + d_W(\mu_2,\mu_3).
\end{aligned}$$

2. (Wasserstein distance metrizes weak convergence.) Let $(\mu_k)_{k\in N}$ be a sequence of probability measures in $P(M)$. Assume $\mu \in P(M)$. The following statements are equivalent:

1. $\lim_{k\to\infty} d_W(\mu_k,\mu) = 0$
2. $\lim_{k\to\infty} \mu_k = \mu$ weakly, i.e. for all measurable $\varphi \in \mathcal{C}_b(X)$:
$$\lim_{k\to\infty} \int_M \varphi \, d\mu_k = \int_M \varphi \, d\mu.$$
and the sequence $(\mu_k)_{k\in\mathbb{N}}$ satisfies the following tightness condition: for some $x_0 \in M$:
$$\lim_{R\to\infty} \limsup_{k\to\infty} \int_{d(x_0,x)\geq R} d(x_0,x)^2 \, d\mu_k(x) = 0 \tag{1.1}$$
3. $\lim_{k\to\infty} \mu_k = \mu$ weakly and there is convergence of the moment of order 2: for some $x_0 \in M$:
$$\lim_{k\to\infty} \int_M d(x_0,x)^2 \, d\mu_k(x) = \int_M d(x_0,x)^2 \, d\mu(x). \tag{1.2}$$
4. For all $\varphi \in \mathcal{C}(M)$ with $|\varphi(x)| \leq C(1+d(x_0,x)^2)$ for some $x_0 \in M$, $C \in \mathbb{R}$:
$$\lim_{k\to\infty} \int_M \varphi \, d\mu_k = \int_M \varphi \, d\mu.$$

Note that by the triangle inequality for d we can extend the statements in 2. and 3. to any $x_0 \in M$.

(4. $\Rightarrow$ weak convergence.): Assuming 4. we obtain convergence for all continuous bounded functions, i.e. weak convergence.
(2. $\Rightarrow$ 4.): Assume that 2. is satisfied for some $x_0 \in M$ and take any φ satisfying the growth condition in 4. For $R > 1$ write

$$\varphi_R(x) = \inf\{\varphi(x), C(1+R^2)\} \qquad \psi_R(x) = \varphi(x) - \varphi_R(x),$$

the latter being pointwise bounded by $C\, d(x_0,x)^2\, \mathbf{1}_{d(x_0,x)\geq R}$. It holds that:

$$\begin{aligned} |\int_M \varphi(x)\, d\mu_k(x) - \int_M \varphi(x)\, d\mu(x)| \quad \leq \quad & |\int_M \varphi_R(x)\, d(\mu_k - \mu)| \\ & + C \int_{d(x_0,x)\geq R} d(x_0,x)^2 \, d\mu_k(x) \\ & + C \int_{d(x_0,x)\geq R} d(x_0,x)^2 \, d\mu(x). \end{aligned}$$

By assumption we have weak convergence also for the second moment, we conclude:

$$\limsup_{k\to\infty} |\int_M \varphi(x)\, d\mu_k(x) - \int_M \varphi(x)\, d\mu(x)| \leq C \limsup_{k\to\infty} \int_{d(x_0,x)\geq R} d(x_0,x)^2 \, (d\mu_k + d\mu)(x),$$

letting R go to infinity we obtain by assumption that the right-hand sides goes to zero, which implies convergence in 4.

(3. $\Rightarrow$ 2.): It holds by assumption:

$$\lim_{k\to\infty}\int_M (d(x_0,x)\wedge R)^2\,d\mu_k(x) = \int_M d(x_0,x)\wedge R)^2\,d\mu(x),$$

and by monotone convergence:

$$\lim_{R\to\infty}\int_X (d(x_0,x)\wedge R)^2\,d\mu(x) = \int_X d(x_0,x)\,d\mu(x);$$

which yields:

$$\lim_{R\to\infty}\lim_{k\to\infty}\int_X \left[d(x_0,x)^p - (d(x_0,x)\wedge R)^p\right]\,d\mu_k(x) = 0$$

Assume that $d(x_0,x)^2 \geq 2R$. Then

$$\begin{aligned} d(x_0,x)^2 - R^2 &= d(x_0,x)^2\left(1-\frac{R^2}{d(x_0,x)^2}\right) \\ &\geq d(x_0,x)^2\left(1-\frac{1}{2^2}\right). \end{aligned}$$

It follows that:

$$\lim_{R\to\infty}\lim_{k\to\infty}\int_{d(x_0,x)\geq 2R} d(x_0,x)^2\,d\mu_k(x) = 0.$$

(1. $\Rightarrow$ 3.): We want to show that convergence in the Wasserstein distance implies weak convergence. As a preliminary note that weak convergence implies

$$\begin{aligned} \int_X d(x_0,x)^2\,d\mu(x) &= \lim_{R\to\infty}\lim_{k\to\infty}\int_X (d(x_0,x)\wedge R)^2\,d\mu_k(x) \\ &\leq \liminf_{k\to\infty}\int_X d(x_0,x)^2\,d\mu_k(x), \end{aligned}$$

and

$$\limsup_{k\to\infty}\int_X d(x_0,x)^2\,d\mu_k(x) \leq \int_X d(x_0,x)^2\,d\mu(x), \tag{1.3}$$

is equivalent to convergence of the second moment in 3.

Take a sequence $(\mu_k)_{k\in N}$ in $P(X)$ with:

$$\lim_{k\to\infty} d_W(\mu_k,\mu) = 0$$

and an optimal transference plan π_k transporting μ_k to μ. For any $\epsilon > 0$ there exists a constant $C_\epsilon > 0$ such that for any $x_0, x, y \in X$:

$$d(x_0,x)^2 \leq (1+\epsilon)d(x_0,y)^2 + C_\epsilon d(x,y)^2.$$

By the marginal condition on π_k it follows:

$$\begin{aligned}\int_{X\times X} d(x_0,x)^2\,d\pi_k(x,y) &= \int_X d(x_0,x)^2\,d\mu_k(x)\\ &\leq (1+\epsilon)\int_X d\mu(y) + C_\epsilon \int_{M\times M} d(x,y)^2\,d\pi_k(x,y)\\ &= (1+\epsilon)\int_X d\mu(y) + C_\epsilon d_W(\mu_k,\mu)^2,\end{aligned}$$

Letting k tend to infinity the Wasserstein distance goes to zero and we obtain:

$$\limsup_{k\to\infty}\int_M d(x_0,x)\,d\mu_k(x) \leq (1+\epsilon)\int_M d(x_0,x)^2\,d\mu(x)$$

and with $\epsilon\to 0$ we obtain the claimed convergence of the p-th moment.

Claim B. In order to prove (1. $\Rightarrow$ weak convergence) and (3. $\Rightarrow$ 1.) it is sufficient to prove only the case where d is bounded.
`Proof of claim B.` Define $\hat{d} = d\wedge 1$ and let $\hat{\mathcal{W}}_2$ be the Wasserstein distance associated to $\hat{d}$. By definition $d_W \geq \hat{\mathcal{W}}_2$, hence in order to prove (d_W-convergence $\Rightarrow$ weak convergence) it is sufficient to prove ($\hat{\mathcal{W}}_2$-convergence $\Rightarrow$ weak convergence).
Assume now that 3. holds and that $(\mu_k)_{k\in N}$ converges in $\hat{\mathcal{W}}_2$. We want to show that $(\mu_k)_{k\in N}$ also converges in d_W. By elementary geometric reasoning for all $x,y\in(M,d)$ it holds that for every $R>0$ and $x_0\in M$:

$$d(x,y)\leq d(x,y)\wedge R + 2\,d(x,x_0)\,\mathbf{1}_{d(x,x_0)\geq R/2} + d(x_0,y)\,\mathbf{1}_{d(x_0,y)\geq R/2}.$$

and there exists a constant $C_p>0$:

$$d(x,y)^2 \leq C_2\left([d(x,y)\wedge R]^2 + d(x,x_0)^2\,\mathbf{1}_{d(x,x_0)\geq R/2} + d(x_0,y)^2\,\mathbf{1}_{d(x_0,y)\geq R/2}\right).$$

Let π_k be an optimal transference plan for transporting μ_k to μ with cost function d^p. For $R\geq 1$:

$$\begin{aligned} d_W(\mu_k,\mu)^2 = &\\ &\int_{X\times X} d(x,y)^2\,d\pi_k(x,y)\\ \leq\ & C_p\int_{X\times X}[d(x,y)\wedge R]^2\,d\pi_k(x,y)\\ &+C_p\int_{\{d(x,x_0)\geq R/2\}\times Y} d(x,x_0)^2\,d\pi_k(x,y)\\ &+C_p\int_{\{d(x_0,y)\geq R/2\}\times X} d(x_0,y)^2\,d\pi_k(x,y)\\ \leq\ & R^2\,\hat{\mathcal{W}}_2^2(\mu_k,\mu) + C_2\int_{\{d(x,x_0)\geq R/2\}\times X} d(x,x_0)^2\,d\pi_k(x,y)\\ &+C_p\int_{\{d(x_0,y)\geq R/2\}\times X} d(x_0,y)^2\,d\pi_k(x,y).\end{aligned}$$

Now let $k\to\infty$ and using assumption 3. let $R\to\infty$ we obtain convergence in the d_W sense.

qed. Claim B.

Assume that $d \leq 1$. In this case all distances d_W (with cost-function $d(x,y)^p$) are equivalent, we prove the case $p = 1$. Assume that the sequence $(\mu_k)_{k\in N}$ converges to μ in the 1-Wasserstein distance. Since we are in the case where c is a metric, the Kantorovich-Rubinstein theorem applies and convergence in 1-Wasserstein distance reduces to

$$\lim_{k\to\infty} \sup_{\|\varphi\|_{Lip}\leq 1} \int_M \varphi(x)\, d(\mu_k - \mu)(x) = 0. \tag{1.4}$$

We want to prove weak convergence, that is for all $\varphi \in \mathcal{C}_b(M)$,

$$\lim_{k\to\infty} \int_M \varphi\, d\mu_k(x) = \int_M \varphi\, d\mu(x),$$

which is true (by the Kantorovich-Rubinstein theory) if φ is 1-Lipschitz and replacing φ by $\frac{\varphi}{\|\varphi\|_{Lip}}$ in the case $\varphi \neq 0$ convergence holds even for all Lipschitz functions. For every bounded function on (M,d) there exist $(a_n)_{n\in N}$ and $(b_n)_{n\in N}$ of uniformly Lipschitz functions such that (a_n) resp. (b_n) is pointwise increasing resp. decreasing in n and:

$$\lim_{n\to\infty} a_n = \varphi = \lim_{n\to\infty} b_n.$$

It follows that

$$\begin{aligned}
\limsup_{k\to\infty} \int_M \varphi\, d\mu_k(x) &\leq \liminf_{n\to\infty} \limsup_{k\to\infty} \int_M b_n\, d\mu_k \\
&= \liminf_{n\to\infty} \int_M b_n\, d\mu \\
&= \int_M \varphi(x)\, d\mu(x),
\end{aligned}$$

the last equality follows from dominated convergence. Analogously it holds that

$$\liminf_{k\to\infty} \int_M \varphi(x)\, d\mu_k(x) \geq \int_M \varphi(x)\, d\mu(x),$$

which proves weak convergence.

(3. $\Rightarrow$ 1.): Assume that $(\mu_k)_{k\in N}$ converges in the weak sense towars μ. We want to prove convergence in the sense of Kantorovich-Rubinstein theory, i.e. (1.4). Take any $x_0 \in M$ and denote the space of all Lipschitz functions on M with Lipschitz constant less or equal 1 such that $\varphi(x_0) = 0$ by $Lip_{1;x_0}(M)$, and it suffices to prove:

$$\lim_{k\to\infty} \sup_{\varphi\in Lip_{1;x_0}} \int_M \varphi\, d(\mu_k - \mu) = 0$$

in order to show (1.4). From Prokhorov's theorem we know that $(\mu_k)_{k\in N}$ is a tight family of probability measures on X: Take an increasing sequence of compact sets in M: $(K_n)_{n\in N}$ with $\epsilon = \frac{1}{n}$, i.e. for $n \geq 1$:

$$\mu(K_n^c) \leq \frac{1}{n} \qquad\qquad \sup_{k\in N} \mu_k(K_n^c) \leq \frac{1}{n}.$$

Without loss of generality we may assume $x_0 \in K_1$. Of course for all $n \geq 1$:

$$\{\varphi\,\mathbf{1}_{K_n}; \qquad \varphi \in Lip_{1;x_0}(M)\}$$

is a subset of $Lip_{1;x_0}(M)$, the latter being a pointwise bounded equicontinuous family of functions on M. Due to separability of M Arzela-Ascoli applies: From any sequence (φ_k) in $Lip_{1;x_0}(M)$ we can extract a subsequence (φ_{k_j}) which converges uniformly on $Lip_{1;x_0}(K_n)$ for K_n compact and by taking the diagonal (with index k_j and n) there exists a subsequence which converges uniformly on every K_n towards a measurable function ψ defined on $\bigcup K_n$ which is bounded Lipschitz since (φ_k) is uniformly bounded and uniformly Lipschitz, i.e. the limit ψ is also 1-Lipschitz and can even be extended from $\bigcup K_n$ to the whole M by setting:

$$\tilde{\psi}(x) = \inf_{y\in\bigcup K_n} \{\psi(y) + d(x,y)\}.$$

It remains to show that

$$\lim_{k\to\infty} \int \varphi_k \, d(\mu_k - \mu) :$$

Note that

$$\begin{aligned} \int \varphi_k \, d(\mu_k - \mu) \quad &\leq \quad |\int_{K_n} (\varphi_k - \psi)\, d(\mu_k - \mu)| \\ &\quad + |\int_{K_n^c} (\varphi_k - \psi)\, d(\mu_k - \mu)| \\ &\quad + |\int_M \psi \, d(\mu_k - \mu)|. \end{aligned}$$

We claim that all three terms on the right-hand side go to 0 for $n \to \infty$ and then $k \to \infty$. The first one goes to zero for fixed n and $k \to \infty$ since the φ_k's converge uniformly on each compact K_n towards ψ. The second term: All φ_k's and ψ are uniformly bounded by some constant $c > 0$ and the tightness of (μ_k) and μ yields:

$$c\,(\mu_k(K_n^c) + \mu(K_n^c)) \leq 2c\frac{1}{n}.$$

Taking the limit in n one obtains convergence towards zero uniformly in k. The last term converges to zero since we assumed weak convergence of (μ_k) towards μ.

3. (M Polish entails $P(M)$ Polish)

We have to show that there exist a countable dense subset of $P(M)$. For this purpose we take a dense sequence $Y := \{y_k; k \in \mathbb{N}\} \subset M$ and write K as the set of all probability measure of the form $\sum_i a_i\delta_{x_i}$ where a_i are rational numbers and the x_i's are finitely many elements in Y. We claim that K is the countable dense subset of $P(M)$ in question:
Let $\epsilon > 0$ be given, and let x_0 be an arbitrary element of Y. If μ lies in $P(M)$,

then there exists a compact set $L \subset M$ such that $\int_{M \setminus L} d(x_0, x)^2 d\mu(x) \leq \epsilon^2$. Cover L by a finite family of balls $B(x_k, \epsilon/2), 1 \leq k \leq N$, with centers $x_k \in Y$ and define

$$B_k = B(x_k, \epsilon/2) \setminus \bigcup_{j<k} B(x_j, \epsilon/2).$$

Then all B_k are disjoint and still cover L. Define a function f on M by

$$f(B_k \cap L) = \{x_k\}, \qquad f(M \setminus L) = \{x_0\}.$$

Then, for any $x \in L, d(x, f(x)) \leq \epsilon$. So

$$\begin{aligned} \int d(x, f(x))^2 d\mu(x) &\leq \epsilon^2 \int_L d\mu(x) + \int_{M \setminus L} d(x_0, x)^2 \, d\mu(x) \\ &\leq \epsilon^2 + \epsilon^2 = 2\epsilon^2 \end{aligned}$$

Since (Id, f) is a transport plan from μ to $f\#\mu$, $d_W(\mu, f\#\mu)^2 \leq 2\epsilon^2$. $f\#\mu$ can be written as $\sum a_j \delta_{x_j}, 0 \leq j \leq N$. This shows that μ might be approximated, with arbitrary precision, by a finite combination of Dirac masses. To conclude, it is sufficient to show that the coefficients a_j might be replaced by rational coefficients, up to a very small error in Wasserstein distance. By Theorem 6.15 in [Vil08]

$$d_W(\sum_{j=1}^N a_j \delta_{x_j}, \sum_{j=1}^N b_j \delta_{x_j}) \leq 2^{\frac{1}{p'}} \max_{k,l} d(x_k, x_l) \sum_{j=1}^N |a_j - a_j|^{\frac{1}{2}},$$

and obviously the latter quantity can be made as small as possible for some well-chosen rational coefficients b_j.

Completeness: Let $\{\mu_k; k \in \mathbb{N}\}$ be a Cauchy sequence in $P(M)$. By a consequence of Prokhorovs theorem it admits a subsequence $\{\mu_{k'}\}$ which converges weakly to some measure μ. Then

$$\int d(x_0, x)^2 d\mu(x) \leq \liminf_{k' \to \infty} \int d(x_0, x)^2 d\mu_{k'}(x) < +\infty$$

so μ belongs to $P(M)$. Moreover, by lower semicontinuity of d_W ,

$$d_W(\mu, \mu_{l'}) \leq \liminf_{k' \to \infty} d_W(\mu_{k'}, \mu_{l'}),$$

so in particular

$$\limsup_{k' \to \infty} d_W(\mu, \mu_{l'}) \leq \limsup_{k', l' \to \infty} d_W(\mu_{k'}, \mu_{l'}) = 0,$$

which means that $\mu_{l'}$ converges to μ in the d_W sense. Since μ_k is a Cauchy sequence with a converging subsequence, it follows by that the whole sequence is converging. □

Proposition 1.1.1. *Let M be compact, then*

1. *$\{H_n; n \in \mathbb{N}\} \subset P$ is dense with respect to d_W.*

2. *$\{G_n; n \in \mathbb{N}\} \subset P$ is dense with respect to d_W.*

3. *$P^\infty \subset P$ is dense with respect to d_W, hence $P_{ac} \subset P$ is also dense with respect to d_W.*

4. *None of the above subspaces is complete with respect to d_W. Only $\{H_n; n \in \mathbb{N}\}$ and $\{G_n; n \in \mathbb{N}\}$ are countable.*

Proof. We prove the first statement: Since M is compact we define disjoint finite cover $\{B_k; k = 1, \dots n\}$ as in part three of the preceding theorem:

$$B_k = B(x_k, \epsilon/2) \setminus \bigcup_{j<k} B(x_j, \epsilon/2).$$

We define a mapping $f : M \to M$ such that $f(B_k \cap M) = B_k \cap M$ and $f\#\mu(B_k \cap M) = \mu(B_k \cap M)$, i.e. f is a (not necessarily monotone) rearrangement of each cell $B_k \cap M$ which should additionally verify $\frac{df\#\mu}{dx}(x)|_{B_k \cap M} = \frac{\mu(B_k \cap M)\mathbf{1}_{B_k \cap M}(x)}{\mathrm{vol}(B_k \cap M)}$. This map verifies all criteria of an admissible transport map (although it is difficult to construct it for most concrete examples). Obviously $d(x, f(x)) \leq \epsilon$. So

$$\begin{aligned} d_W(\mu, f\#\mu)^2 \leq \int d(x, f(x))^2 d\mu(x) &\leq \epsilon^2 \int_M d\mu(x) \\ &\leq \epsilon^2 \end{aligned}$$

And by construction $f\#\mu \in H_n$. Similar arguments apply to G_n and to P^∞. The lack of completeness is illustrated for the subspace $(P^\infty, d_W) \subset (P, d_W)$: Convolution of a positive smooth density with rescaled Gaussians converges in Wasserstein distance to a Dirac measure. On the space of histograms we define weights $a_1 = 1 - \frac{1}{2^n}$ and $a_2, \dots a_n$ such that $\sum_{j=1}^n a_j = 1$, then the corresponding sequence of histograms converges in d_W to a Dirac measure. □

Remark 1.1.1. *If M is compact, then $P(M)$ can be made into a compact metric space ('Watanabe compactification') $P(M) \cup \{\infty_W\}$ with respect to a topology defined by*

$$\mu_n \to \mu \in P(M) \Leftrightarrow \int f\mu_n \to \int f\mu$$

and

$$\mu_n \to \infty_W \Leftrightarrow \int 1\mu_n \to \infty.$$

See remark 3.2.2 in [Daw93]

1.2 Topology of smooth curves

Complementary to the weak topology we put another topology on the subspace P^∞ in the flavor of [KM97]:

Definition 1.2.1 (Topology of smooth curves). *Let E be a locally convex (Hausdorff) vector space. We say that a curve $c : \mathbb{R} \to E$ is differentiable if the limit $c'(t) := \lim_{s\to 0} \frac{1}{s}(c(t+s) - c(t))$ exists. The curve is called smooth if all iterated derivatives exist. A set $A \subset E$ is called c^∞-open if for any smooth curve*

$c : \mathbb{R} \to E$ the set $c^{-1}(A) \subset \mathbb{R}$ is open. We write c^∞ for the topology of smooth curves, i.e. for the topology whose basis consists of c^∞-open sets. Let $U \subset E$ bounded. By E_U we denote the linear span of U in E, equipped with the Minkowski functional $p_U(v) := \inf\{\lambda > 0 : v \in \lambda.U\}$. (E_U, p_U) is a normed space. A locally convex vector space E endowed with the c^∞-topology is defined to be complete if any Mackey-Cauchy sequence is convergent in the following sense: A sequence is $\{x_n; n \in \mathbb{N}\}$ in E is called Mackey convergent to x if there exists a bounded, absolutely convex $U \subset E$ such that $\{x_n; n \in \mathbb{N}\}$ converges to x in the normed space E_U. Note that convergence statement in E_U is equivalent to the existence of a real-valued sequence $\{y_n > 0; n \in \mathbb{N}\}$ converging to zero which satisfies $x_n \in y_n.U$. We can thus paraphrase: A sequence $\{x_n; n \in \mathbb{N}\}$ in E is called Mackey-Cauchy if there exists a bounded absolutely convex $U \subset E$ and a net $\{y_{(n,n')}; (n, n') \in \mathbb{N}^2\}$ in $\mathbb{R}$ converging to zero such that $x_n - x_{n'} \in y_{(n,n')}.U$. By Theorem 2.14 in [KM97] we know that a convenient space is also characterized by the following condition: If $c : \mathbb{R} \to E$ is a curve such that $l \circ c : \mathbb{R} \to \mathbb{R}$ is smooth for all continuous linear functionals $l \in E^*$, then c is smooth.

Definition 1.2.2 (Smooth manifolds)**.** *Let X be a set. A chart (U, u) on X is a bijection $u : U \to u(U) \subset E_U$ from a subset $U \subset X$ onto a c^∞-open subset in E_U. For two charts (U_α, u_α) and (U_β, u_β) on X the mapping $u_{\alpha\beta} := u_\alpha \circ u_\beta^{-1} : u_\beta(U_{\alpha\beta}) \to u_\alpha(U_{\alpha\beta})$ for $\alpha, \beta \in A$ is called the chart changing, where $U_{\alpha\beta} := U_\alpha \cap U_\beta$. A family $(U_\alpha, u_\alpha)_{\alpha \in A}$ of charts on X is called an atlas for X , if the U_α form a cover of M and all chart changings $u_{\alpha\beta}$ are defined on c^∞-open subsets. An atlas $(U_\alpha, u_\alpha)_{\alpha \in A}$ for X is said to be a C^∞-atlas, if all chart $u_{\alpha\beta} := u_\alpha \circ u_\beta^{-1} : u_\beta(U_{\alpha\beta}) \to u_\alpha(U_{\alpha\beta})$ are smooth. Two C^∞-atlas are called C^∞-equivalent, if their union is again a C^∞-atlas for X. An equivalence class of C^∞-atlas is sometimes called a C^∞-structure on X. The union of all atlas in an equivalence class is again an atlas, the maximal atlas for this C^∞-structure. A C^∞-manifold X is a set together with a C^∞-structure on it.*

Remark 1.2.1. *Let M be compact. Then $C^\infty(M)$ is a convenient vector space: It is a vector space over $\mathbb{R}$, and with respect to the topology of smooth curves it is Hausdorff. Addition and scalar multiplication are continuous (they are even smooth) and the function 0 has a basis of neighborhoods consisting of convex sets. It is complete with respect to the topology of smooth curves: We have to verify that for each continuous linear functional l on $C^\infty(M)$ the curve $l \circ c$ is smooth from $\mathbb{R}$ to $\mathbb{R}$. It is sufficient to prove that $t \mapsto \int_M \varphi c_t vol$ is smooth (for $\varphi \in \mathcal{C}_c^\infty$) which is true since we can put differentiation with respect to t inside the integral.*

Definition 1.2.3 (Kinematic tangent bundle)**.** *Consider a manifold M with a smooth atlas $(M \supset U_\alpha \xleftarrow{u_\alpha} E_\alpha)_{\alpha \in A}$. On the disjoint union*

$$\bigcup_{\alpha \in A} U_\alpha \times E_\alpha \times \{\alpha\}$$

we define an equivalence relation

$$(x, v, \alpha) \sim (y, w, \beta) \Leftrightarrow x = y \text{ and } d(u_{\alpha\beta})(u_\beta(x))w = v$$

and denote the quotient set by TM which we call kinematic tangent bundle of M which embeds as subbundle into the so-called operational tangent bundle - a *notion, which we will not belabour here but which is equivalent to the kinematic tangent bundle in the case of finite-dimensional manifolds. The name 'kinematic tangent bundle' comes from the fact that there exists a bijection from* TM *to* $\mathcal{C}^\infty(\mathbb{R}; M)/\sim$ *where two curves* $c \sim e$ *if and only if* $c(0) = e(0)$ *and in one chart* (U, u) *with* $c(0) \in U$ *we have* $\frac{d}{dt}|_0(u \circ c)(t) = \frac{d}{dt}|_0(u \circ e)(t)$.

Henceforth when we talk about the tangent bundle of P^∞ we use exclusively the notion of kinematic tangent bundle.

Definition 1.2.4 (Tangent map)**.** *Given a smooth mapping* $f : M \to N$ *between manifolds (i.e. smooth curves in* M *are mapped to smooth curves in* N*), then* f *induces a linear mapping* $T(f)(x) : T_xM \to T_{f(x)}N$*: for each* $g \in \mathcal{C}^\infty(N \supset \{f(x)\}, \mathbb{R})$ *and* $x \in M$

$$(T(f)(x)(X_x))(g) = X_x(g \circ f) = d(g \circ f)(x)(X).$$

The differential is understood in the sense that there exist a smooth curve c *in* M *with initial speed vector* $\dot{c}(0) = X$*. Therefore we also write short-hand*

$$T(f)(x)(\dot{c}(0)).$$

The most prominent example of an infinite-dimensional manifold is the group of all smooth diffeomorphism on a compact manifold modelled on the space of all smooth mappings from the manifold to itself. Its tangent space at identity is the space of all smooth vector fields, hence equipped with the L^2-inner product the group becomes an infinite-dimensional Riemannian manifold. The reason for choosing a locally convex space and not a Banach one as modelling space lies in a theorem by [Omo78], which states that any Banach Lie group acting effectively on a finite-dimensional compact manifold is necessarily finite dimensional itself: no way for the group of smooth diffeomorphisms to be modelled on a Banach space.
We return our attention to the space of smooth probability densities by noting several remarks:

Proposition 1.2.1. *Let* M *be compact. Then* $P^\infty(M)$ *is a smooth manifold with respect to the topology of smooth curves, more precisely* $P^\infty(M)$ *is a* c^∞*-open submanifold of* $C^\infty(M)$*.*

Proof. 1. Modelling space $C^\infty(M)$:
Fix $\mu \in P^\infty(M)$. A smooth chart u on $U_\mu \subset P^\infty(M)$ bounded is given by a bijection that maps each smooth positive density in U_μ to a smooth positive function; $u(U_\mu)$ is a c^∞-open subset of the linear span of U_μ in $C^\infty(M)$.

□

Remark 1.2.2. P^∞ *is contractible, i.e. the identity map on* P^∞ *is homotopic to a constant function; in other words given any probability density* $\frac{d\mu}{dvol} \in P^\infty$*, there exists a family of diffeomorphisms* ϕ_t *(given by a global flow to a smooth, fully supported vector field) such that* $\phi_t \# \mu = vol$ *for some* $t \in [0, +\infty]$*, so* P^∞ *can be shrinked continuously with respect to the* c^∞*-topology to the point vol.*

Remark 1.2.3 (Wasserstein space as stratified manifold)**.** *As remarked in [GKP10]* $P^\infty(M)$ *may also be viewed in another way as infinite-dimensional* *manifold, again M is assumed to be compact. Recall that the push-forward of a probablity density μ by a Borel map ϕ from M to itself is defined as*

$$\phi\#\mu(A) := \mu(\phi^{-1}(A))$$

for any Borel set A. Then the map

$$Diff^\infty(M) \times P^\infty(M) \to P^\infty(M)$$

given by

$$(\phi, \mu) \mapsto \phi\#\mu$$

defines a left action of $Diff^\infty(M)$ on $P^\infty(M)$ and is smooth (in the sense of the topology of smooth curves). We denote by

$$\mathcal{O}_\mu := \{\nu \in P^\infty(M) : \nu = \phi\#\mu \text{ for some } \phi \in Diff^\infty(M)\}$$

the orbit and by

$$Diff^\infty(M)_\mu := \{\phi \in Diff^\infty(M) : \phi\#\mu = \mu\}$$

the stabilizer of any fixed measure μ. Note that the latter is itself a Lie subgroup of the diffeomorphism group - its Lie algebra is the space of vector fields whose divergence with respect to the measure μ equals zero. The quotient space $Diff^\infty(M)\diagup Diff^\infty(M)_\mu$ can be mapped one-to-one to $\mathcal{O}_\mu$ via $j : [\phi] \mapsto \phi\#\mu$. This mapping can be lifted to the respective tangent bundles: By Hodge theory for the $L^2(\mu)$-closure of the space of all vector fields we can decompose each vector field into its μ-divergence free part and its gradient part. For the tangent map this means that

$$T(j) : \mathcal{X}(M)\diagup Ker(div_\mu) \to \mathcal{O}_\mu$$

is a bundle isomorphism. This construction is also valid for the more general case when replacing diffeomorphisms by homeomorphisms and in this way P becomes a stratified manifold, i.e. a topological space with a foliation and a differentiable structure definied on each leaf of the foliation: the foliation is induced by the action of $Diff^\infty(M)$ on P. On the other hand $\mathcal{O}_\mu = P^\infty(M)$ for $\mu \in P^\infty(M)$ shows that for the subspace of smooth positive densities there exists a single leaf and $P^\infty(M)$ becomes a homogenous space with the quotient $Diff^\infty(M)\diagup Diff^\infty(M)_\mu$ acting faithfully on $P^\infty(M)$ for any $\mu \in P^\infty(M)$.

Remark 1.2.4 (Wasserstein space as embedding in the space of linear forms)**.** *Another point of view (as advocated by [Lot08]) is*

$$P^\infty(M) \subset (\mathcal{C}^\infty(M))^*,$$

i.e. for every $\varphi \in \mathcal{C}^\infty(M)$ we define a functional $F_\varphi(\mu) := \int \varphi\mu vol$ on $P^\infty(M)$ which is point-seperating and smooth (in the sense of the topology of smooth curves). The functions $F_\varphi(\mu)$ can be thought of as coordinates of the point μ. For a reason which is explained by Proposition 2.2.1 tangent vectors act on smooth functions on $P^\infty(M)$ by

$$(V_\varphi F)(\mu) = \frac{d}{dt}|_{t=0} F(\mu - t\, div_\mu(\nabla\varphi))$$

for $\varphi \in \mathcal{C}^\infty(M)$ *given. As in the original paper of [Ott01] we obtain a bundle isomorphism*

$$\mathcal{C}^\infty(M)/\mathbb{R} \to TP^\infty(M),$$

where the latter is the kinematic tangent bundle.

Not the following

Remark 1.2.5. *Let* M *be compact. If a sequence* $\{\mu_k; k \in \mathbb{N}\}$ *in* $P^\infty(M)$ *converges weakly to some* $\mu \in P^\infty(M)$ *then* $\{\mu_k; k \in \mathbb{N}\}$ *converges with respect to the topology of smooth curves to* μ *.*

Proof. Since $\mathcal{C}^\infty \subset \mathcal{C}_b$ we know that $\lim_{k\to\infty} \mu_k(\phi) := \lim_{k\to\infty} \int \phi\mu_k = \mu(\phi)\ \forall\phi \in \mathcal{C}_b$ implies $\lim_{k\to\infty} \mu_k(\phi) = \mu(\phi)\ \forall\phi \in \mathcal{C}^\infty$. □

Chapter 2

Riemannian geometry of probability measures

Starting from an energy variation formula for paths in P_{ac} we report results about the characterization of Wasserstein geodesics and its links to partial differential equations. We present a Riemannian geometry on P_{ac} from two viewpoints: At first from a analytical one using calculus of variations and PDEs. Secondly we develop a formal Riemannian geometry on P_{ac} which will be shown to be rigorous when we restrict ourselves to P^∞ equipped with the topology of smooth curves. Within this framework we show new and known formulas for the Levi-Cività connection, Riemannian curvature and parallel transport. Additionnally we mention results from [AGS08] which are rather in the flavor of Lipschitz analysis and geometric measure theory.

2.1 Wasserstein geodesics

As Kantorovich laid the foundations of existence and uniqueness of the optimal transport problem and hence the well-definedness of Wasserstein distances for convex cost functions it was Brenier and McCann who in [Bre91] and [McC01] laid the foundations of the characterization of optimal transport maps, i.e. the map $T : M \to M$ that is actually realizing the infimum in (1). We cite the following

Theorem 2.1.1. *Given $\mu, \nu \in P_{ac}$, then the optimal transport plan π (for the cost function $(x, y) \mapsto d(x, y)^2$) realizing the Wasserstein distance between μ and ν is given by a map $T : M \to M$ such that*

$$d_W(\mu, \nu)^2 = \int_M d(x, T(x))^2 \mu(dx),$$

where $T(x) = \exp_x(-\nabla\varphi(x))$ and φ is a μ-a.s. unique convex function on M.

Proof. For the case $M = \mathbb{R}^d$ the theorem was proved in [Bre91], in the case where M is a connected compact, $\mathcal{C}^3$-smooth Riemannian manifold a proof can be found in [McC01]. For the non-compact case a similar statement can be found in [Fig07]. □

The following theorem showed for the first time a connection between fluid dynamics and optimal transport, i.e. we think of μ_0 and μ_1 as the density of particles in a given region in $\mathbb{R}^d$ at time $t = 0$ and $t = 1$. If we assume that for every $t \in [0,1]$ there exists a smooth resp. uniformly Lipschitz vector field v_t which describes how particles move around in a given area we can describe the time evolution of the particles' position by

$$\frac{dX_t}{dt} = v_t(X_t) \tag{2.1}$$

Under regularity assumption on the vector field v_t we obtain for a given initial value $x_0 \in \mathbb{R}^d$ a unique solution $X_{x_0}(t)$ for (2.1) on the whole time interval $[0,1]$; moreover the map $(t, x_0) \mapsto X_{x_0}(t)$ is globally Lipschitz and one-to-one. Thus $(T_t)_{0\le t\le 1} = (x \mapsto X_x(t))_{0\le t\le 1}$ is a locally Lipschitz family of diffeomorphisms and the characteristics method for the linear transport equation applies:
$\mu_t = T_t \# \mu_0$ is a weak solution to

$$\frac{\partial \mu_t}{\partial t} + \operatorname{div}(\mu_t v_t) = 0 \tag{2.2}$$

The quantity $\operatorname{div}(\mu_t v_t)$ describes the flow density under v_t. The total kinetic energy up to a factor $\frac{1}{2}$ is $E(t) = \int_{\mathbb{R}^d} \mu_t |v_t|^2 dx$. The energy one needs to move particles around from time 0 to time 1 according to v_t is defined as $A[\mu, v] = \int_0^1 E(t)dt$.

Theorem 2.1.2.

$$\inf_{(\mu,v)\in V(\mu_0,\mu_1)} A[\mu, v] = d_W(\mu_0, \mu_1)^2, \tag{2.3}$$

where $V(\mu_0, \mu_1)$ is the set of all pairs $(\mu, v) := (\mu_t, v_t)_{t\in[0,1]}$ satisfying the following conditions:

1. $\mu \in \mathcal{C}([0,1], P_{ac}(\mathbb{R}^d))$ *where $P_{ac}(\mathbb{R}^d)$ -the space of absolutely continuous probability measures- is endowed with the weak-$*$-topology, i.e.* $\lim_{i\to\infty} \nu_i = \nu$ *iff* $\int \varphi \nu_i \longrightarrow \int \varphi \nu$ *for all* $\varphi \in \mathcal{C}_c^\infty$.
2. $v \in L^2(d\mu_t(x)dt)$
3. $\bigcup_{t\in[0,1]} supp(\mu_t)$ *is bounded*
4. $\frac{\partial \mu_t}{\partial t} + div(\mu_t v_t) = 0$, *i.e.*
$\int_{\mathbb{R}^d} (\partial_t \varphi(t,x))\mu_t + \int_{\mathbb{R}^d} \langle \nabla \varphi(t,x), v_t \rangle \mu_t = 0$ *for all* $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}_+ \times \mathbb{R}^d)$.
5. $\mu(t = 0, \cdot) = \mu_0(\cdot), \quad \mu(t = 1, \cdot) = \mu_1(\cdot)$

Proof. See [BB00] □

Formula (2.3) can be seen as geodesic equation on the space of probability measures from two points of view: Firstly, as generalization of action minimizing curves, which are no longer deterministic but random and secondly, as realization of the minimum in the energy variation formula for $\mathcal{C}^1$-curves (known from finite-dimensional Riemannian geometry) if we endow P with a Riemannian metric and can make clear what the tangent bundle above P should be.

Remark 2.1.1 (Displacement interpolation)**.** *As already mentioned in the introduction the Wasserstein distance is realized by the optimal coupling between two probability measures μ_0 and μ_1. The notion of coupling may be extended to time-dependent optimal transport. We call every random curve $\gamma : [0,1] \to M$, such that law$(\gamma_0) = \mu_1$ and law$(\gamma_1) = \mu_2$, a dynamical coupling of μ_0 and μ_1. Any probability measure on $\mathcal{C}([0,1];M)$ is called a dynamical transference plan. In chapter 7 in [Vil08] a fairly general machinery of Lagrangian action functionals with respect to semi-continuous cost functions on Polish spaces (for continuous, not necessarily differentiable curves) is developed. We give only the example of kinetic energy for the case of a finite-dimensional Riemannian manifold: For any continuously differentiable curve γ we define $A(\gamma) := \frac{1}{2}\int_0^1 |\dot\gamma(s)|_{\gamma(s)} ds$ and among all $\mathcal{C}^1$-curves we define the subset Γ of action minimizing curves (note that we do not fix starting and end points). Denote by $e_x : \Gamma \to M$ the evaluation functional $e_x(\gamma) = \gamma(x)$. A dynamical optimal transference plan is a probability measure Π on Γ such that*

$$\pi := (e_0, e_1)\#\Pi$$

is an optimal transport plan from μ_0 to μ_1. The following theorem shows that dynamical optimal transference plans are minimizing curves for Lagrangian action functionals on P.

Theorem 2.1.3. *For a continuous curve $(\mu_t)_{0\le t\le 1}$ in P such that the Wasserstein distance between μ_0 and μ_1 is finite, the following statements are equivalent:*

1) For each $t \in [0,1]$, μ_t is the law of γ_t, where $(\gamma_t)_{0\le t\le 1}$ is a dynamical optimal coupling of (μ,ν).
2) The path $(\mu_t)_{0\le t\le 1}$ is a minimizing curve for the action functional $\mathbb{A}$ defined on P by

$$\mathbb{A}(\mu) := \inf_{\gamma} \mathbb{E}A(\gamma)$$

with law$(\gamma_t) = \mu_t$ for each $t \in [0,1]$. Note that the inf is taken over all random curves (i.e. random variables with values in $\mathcal{C}([0,1];M)$) such that law$(\gamma_\tau) = \mu_\tau$ for $0 \le \tau \le 1$.

A curve $(\mu_t)_{0\le t\le 1}$ in P fulfilling one of the above conditions is called displacement interpolation between μ_0 and μ_1. The displacement interpolation is unique if there is a unique optimal transport plan π between μ_0 and μ_1 and if any two points $x_0, x_1 \in M$ are joined $\pi(dx_0, dx_1)$-almost surely by a unique geodesic.

Proof. [Vil08] □

Definition 2.1.1 (Geodesics)**.** *On any complete, locally compact metric space X we define geodesics γ between two points γ_0 and γ_1 as paths that realize the distance, i.e. paths which attain the minimum in*

$$L(\gamma) := \sup_n \sup_{0=t_1<t_2<\ldots<t_{n+1}=1} \sum_{k=1}^{n} d(\gamma_{t_k}, \gamma_{t_{k+1}}),$$

such that

$$d(\gamma_0, \gamma_1) = \inf_{\gamma\in\mathcal{C}^0([0,1];X)} L(\gamma)$$

There is an important corollary to the preceding theorem:

Corollary 2.1.1. *Given $\mu_0, \mu_1 \in P(M)$, M a compact Riemannian manifold, and a continuous curve $\{\mu_t; t \in [0,1]\}$. With the above definition of geodesics the following statements are equivalent:*
1) $t \mapsto \mu_t$ is a geodesic in P
2) μ_t is the law of γ_t, where γ is a random geodesic on M, such that (γ_0, γ_1) is the optimal coupling from μ_0 to μ_1, i.e. the random variables γ_0 and γ_1 induce a product measure which solves the optimal transport problem from μ_0 to μ_1.

The above characterization of Wasserstein geodesics as laws of random geodesics on the underlying space rises further questions (for instance whether the geodesic between two given measures may be branching), which we will no belabor here, recently there has been developed a variational approach to this issue in the spirit of Benamou-Brenier's theorem (see [BBS10]). We turn our attention to the second point of view, i.e. we endow P with a Riemannian structure and prove formulas which are very much inspired from finite-dimensional Riemannian geometry.

2.2 Otto's Riemannian metric

Definition 2.2.1. *Given $\mu \in P(M)$. We define the tangent space to μ as*

$$T_\mu P := \overline{\{\nabla \varphi; \varphi \in \mathcal{C}_c^\infty(M)\}}^{L^2(\mu)},$$

note that if M is compact, then the compactness requirement on support of smooth functions is omitted.

Lemma 2.2.1. *Let $\mu \in P$. A vector $v \in L^2(\mu)$ belongs to $T_\mu P$ iff*

$$||v + w||_{L^2(\mu)} \geq ||v||_{L^2(\mu)} \tag{2.4}$$

for all $w \in L^2(\mu)$ such that $div(\mu w) = 0$. In particular for every $v \in L^2(\mu)$ there exists a unique $\Pi(v) \in T_\mu P$ in the equivalence class of v modulo divergence-free vector fields, $\Pi(v)$ is the element of minimal $L^2(\mu)$-norm in this class.

Proof. See Lemma 8.4.2. in [AGS08]: Convexity of the $L^2(\mu)$ norm entails that (2.4) holds iff $\int_M \langle v, w\rangle \mu = 0$ for any $w \in L^2(\mu)$ such that $\mathrm{div}(\mu w) = 0$, and this is true iff v is in the $L^2(\mu)$ closure of $\{\nabla \varphi; \varphi \in \mathcal{C}_c^\infty(M)\}$.

□

Proposition 2.2.1. *If $\mu \in P^\infty$, then*

$$T_\mu P^\infty = \{\mu(.) \in \mathcal{C}^\infty(\mathbb{R}_+; P^\infty); \mu(0) = \mu\} / \sim,$$

where for two $\mu(.), \nu(.) \in \mathcal{C}^\infty(\mathbb{R}_+; P^\infty)$ we say that $\mu(.) \sim \nu(.)$ if both $\mu(.)$ and $\nu(.)$ solve the continuity equation (2.2) for the same given vector field v_t, i.e. that the tangent space coincides on P^∞ with the kinematic tangent space defined in definition 1.2.3

Proof. To prove the proposition we make use of theorem 8.3.1 in [AGS08], where it is shown that for every absolutely continuous curve μ there exists a Borel vector field v_t with $L^2(\mu_t)$-norm bounded from above by the metric derivative

$$\lim_{t\to 0} \frac{d_W(\mu_{t+h}, \mu_t)}{|h|}$$

of μ such that the continuity equation is satisfied and this applies in particular to smooth curves with values in P^∞. By a variational selection principle (lemma 2.2.1) it is then shown that there exists a unique projection of v_t to the equivalence class of vector fields modulo divergence-free vector fields with minimal $L^2(\mu_t)$-norm, i.e.

$$\int_M \langle v_t, w_t\rangle \mu_t = 0$$

for any $w_t \in L^2(\mu_t)$ such that $\operatorname{div}(\mu_t w_t) = 0$ which is the case iff v_t belongs to the $L^2(\mu_t)$ closure of $\{\nabla\varphi; \varphi \in \mathcal{C}_c^\infty(M)\}$ since $\operatorname{div}(\mu_t w_t) = 0$ means that $\int \langle \nabla\varphi, w_t\rangle \mu_t = 0$. Since for two curves being in the same equivalence class means to solve the continuity equation for the same vector field (which we proved to be the unique vector field of gradient type with minimal $L^2(\mu_t)$-norm) the proof of

$$T_\mu P^\infty \supseteq \{\mu(.) \in \mathcal{C}^\infty(\mathbb{R}_+; P^\infty); \mu(0) = \mu\} \diagup \sim,$$

is achieved. For the converse inclusion we cite again theorem 8.3.1 in [AGS08], where it is shown that any continuous curve satisfying the continuity equation for some Borel vector field has a metric derivative that is less or equal than the $L^2(\mu_t)$-norm of the vector fields, we then apply again the lemma already cited to conclude. □

Definition 2.2.2. *The (kinematic) tangent bundle TP^∞ is defined as the disjoint union*

$$\bigcup_{\mu\in P^\infty} T_\mu P^\infty.$$

Note that this definition coincides with definition 1.2.3 using the global chart of the embedding $P^\infty \subset \mathcal{C}^\infty$.

Definition 2.2.3. *We define on the kinematic tangent space $T_\mu P^\infty$ a Riemannian metric denoted by*

$$\langle \dot\mu, \dot\mu\rangle_\mu := \int_M |v_0|^2 \, d\mu$$

where $\mu(.)$ is in the equivalence class of smooth curves satisfying $\mu(0) = \mu$ and the continuity equation (at $t = 0$) with respect to the gradient-type vector field v_0, i. e.

$$\dot\mu := \frac{d}{dt}|_0 \mu = -div(\mu v_0)$$

Proposition 2.2.2 (Wasserstein Gradient formula)**.** *Let $F : P^\infty \to \mathbb{R}\cup\{\infty\}$ such that*

$$F(\mu) = \int_M f(\mu(x)) vol$$

for a twice differentiable function $f : \mathbb{R}_+ \to \mathbb{R}$ and let $\mu(.)$ be a smooth curve in P^∞ such that $\mu(0) = \mu$ and $\dot{\mu} = -div(\mu v)$. Then

$$\langle \nabla^{P^\infty} F(\mu), v\rangle_\mu := \frac{d}{dt}|_0 F(\mu(t)) = \langle \nabla \frac{\delta}{\delta\mu} F(\mu), v\rangle_\mu$$

Proof. In the spirit of the pioneering work [Ott01] we know that

$$\frac{d}{dt}|_0 F(\mu(t)) = \int_M \langle (f' \circ \mu), \dot{\mu}\rangle \mathrm{vol} = -\int_M (f' \circ \mu)\mathrm{div}(\mu v)\mathrm{vol}$$

which equals by integration by parts

$$\int_M \langle \nabla (f' \circ \mu), v\rangle \mu \mathrm{vol}.$$

□

Definition 2.2.4. *Vector fields on P^∞ are defined as smooth (in the sense of c^∞-topology, see [KM97]) sections of the kinematic tangent bundle TP^∞, i.e. $V \in \Gamma(TP^\infty \leftarrow P^\infty)$ if $V : \mu \mapsto \dot{\mu}$ such that $pr_{P^\infty} \circ V(\mu) = \mu$. By proposition 2.2.1 we know that each equivalence class of curves corresponds to some element in the $L^2(\mu)$-closure of the space of gradient-type vector fields. We say that a smooth functions v defined on the underlying manifold determines a vector field V on the space of smooth, positive probability densities if there exists some representative of a smooth curve $c : \mathbb{R} \to P^\infty$ which passes at zero in μ and which verifies the continuity equation $\dot{c} = -div(\mu\nabla v)$ at time zero for a smooth function v. This means that V might be seen as (regular) distribution acting on test functions in the following way:*
For all $\varphi \in \mathcal{C}_c^\infty(M)$, a function $v \in \mathcal{C}^\infty(M)$ and $m_0 = \frac{d\mu}{dvol}$:

$$(V(\mu)|\varphi) = \int_M \langle \nabla v,\, \nabla\varphi\rangle_x m_0(x) vol(dx).$$

We write short-hand

$$V(\mu) = -div(\mu\,\nabla v),$$

We emphasize that the smooth function $v : M \ni x \mapsto \bar{v}(m_0(x)) \in \mathbb{R}$ for $\bar{v} : \mathbb{R}_+ \to \mathbb{R}$ is a possible choice, i.e. take $\bar{v}(x) = log(x)+1$, then $\nabla\, v(x) = \nabla\, (log(m_0(x))+1)$ gives the vector field associated to the entropy via the Wasserstein gradient:

$$\nabla^{P^\infty} \int_M m_0\, log(m_0)\, vol(dx) = div(\mu\,\nabla\, v)$$

2.3 Levi-Cività connection

Recall the notion of tangent map. For a smooth mapping $F : P^\infty \to \mathbb{R}$ and any smooth curve $c : (-a, a) \to P^\infty$ such that $c(0) = \mu$ and $\dot{c}(0) = -\mathrm{div}(\mu\,\nabla\, u)$ the tangent map

$$\begin{aligned} \text{T(F):} \quad & TP^\infty \to \mathbb{R} \times \mathbb{R} \\ & (\mu, \dot{c}(0)) \mapsto (F(\mu), T(F)(\mu).\dot{c}(0)) := (F(\mu), \tfrac{d}{dt}|_0 (F \circ c)(t)) \end{aligned}$$

Proposition 2.3.1 (Smooth Lie bracket [Sel06])**.** *Given* $U, V \in \Gamma(TP^\infty \leftarrow P^\infty)$*. Since vector fields are not complete we have to construct their respective flows explicitely: For* $0 < a \ll \epsilon$ *we define* $Fl_t^U, Fl_t^V : (-a, a) \times P^\infty \to P^\infty$ *by*

$$\begin{aligned}\frac{\partial}{\partial t} Fl_t^U(\mu) &= U(Fl_t^U(\mu)) \\ &= -div(Fl_t^U(\mu)\nabla\bar{u}(Fl_t^U(\mu)))\end{aligned}$$

resp.

$$\begin{aligned}\frac{\partial}{\partial t} Fl_t^V(\mu) &= V(Fl_t^V(\mu)) \\ &= -div(Fl_t^U(\mu)\nabla\bar{v}(Fl_t^V(\mu)))\end{aligned}$$

Then the Lie bracket reads as follows:

$$\begin{aligned}[U,V](\mu) &= div(V(\mu)\,\nabla u) - div(U(\mu)\,\nabla v) + \\ &\quad + div(\mu\,\nabla T(u)(\mu).V(\mu)) - div(\mu\,\nabla T(v)(\mu).U(\mu))\end{aligned}$$

Here $T(u)(\mu)$ *is the tangent map of* u *at* μ*, since* u *is a real-valued function on* P^∞ *it is the differential of* u *at* μ*.*

Proof. As a prerequisite we calculate

$$\frac{\partial}{\partial t} T\left(\mathrm{Fl}_{-t}^U\right)(\mu),$$

i.e. the expression we differentiate is the tangent map of Fl_{-t}^U at μ. By the product rule applied to the flow equation:

$$\begin{aligned}\frac{\partial}{\partial t} T(\mathrm{Fl}_{-t}^U)(\mu) &= T\left(\frac{\partial}{\partial t}\mathrm{Fl}_{-t}^U\right)(\mu) \\ &= -T\left(-\mathrm{div}(\mathrm{Fl}_{-t}^U\nabla\left\{u\circ\mathrm{Fl}_{-t}^U\right\})\right)(\mu) \\ &= \mathrm{div}\left[T(\mathrm{Fl}_{-t}^U)(\mu)\nabla\left\{u\circ\mathrm{Fl}_{-t}^U\right\}(\mu)\right] + \\ &\qquad + \mathrm{div}\left[(\mathrm{Fl}_{-t}^U)(\mu)\nabla\left\{T(u\circ\mathrm{Fl}_{-t}^U)\right\}(\mu)\right] \\ &= \mathrm{div}\left[T(\mathrm{Fl}_{-t}^U)(\mu)\nabla\left\{u\circ\mathrm{Fl}_{-t}^U\right\}(\mu)\right] + \\ &\qquad + \mathrm{div}\left[(\mathrm{Fl}_{-t}^U)(\mu)\nabla\left\{T(u)(\mathrm{Fl}_{-t}^U(\mu))T(\mathrm{Fl}_{-t}^U)(\mu)\right\}\right]\end{aligned}$$

By definition

$$
\begin{aligned}
[U,V](\mu) &= \frac{\partial}{\partial t}|_0(\mathrm{Fl}_t^U)^* V(\mu) \\
&= \frac{\partial}{\partial t}|_0(T(\mathrm{Fl}_{-t}^U)\circ V\circ \mathrm{Fl}_t^U)(\mu) \\
&= \left(\frac{\partial}{\partial t}|_0 T(\mathrm{Fl}_{-t}^U)(V\circ \mathrm{Fl}_t^U|_0)\right)(\mu) + T(\mathrm{Fl}_{-t}^U)|_0 \circ \left(\frac{\partial}{\partial t}|_0 V\circ \mathrm{Fl}_t^U\right)(\mu) \\
&= \frac{\partial}{\partial t}|_0 T(\mathrm{Fl}_{-t}^U)(V(\mu)) + \frac{\partial}{\partial t}|_0 V(\mathrm{Fl}_t^U(\mu)) \\
&= \mathrm{div}\,[V(\mu)\nabla\, u(\mu)] + \mathrm{div}\,[\mu\,\nabla\,\{T(u)(\mu).V(\mu)\}] + \\
&\qquad\qquad + \frac{\partial}{\partial t}|_0\left(-\mathrm{div}\left[\mathrm{Fl}_t^U(\mu)\nabla\, v(\mathrm{Fl}_t^U(\mu))\right]\right) \\
&= \mathrm{div}\,[V(\mu)\nabla\, u(\mu)] + \mathrm{div}\,[\mu\,\nabla\,\{T(u)(\mu).V(\mu)\}] + \\
&\qquad -\mathrm{div}\,[U(\mu)\nabla\, v(\mu)] - \mathrm{div}\,[\mu\,\nabla\,\{T(v)(\mu).U(\mu)\}]
\end{aligned}
$$

□

This formula generalizes formulas obtained by [Lot08] (where the functions $\bar{v}$ do not depend on the density). In view of the Lie bracket we define the Levi-Civitá connection on P^∞ and show that it is Riemannian and torsion-free.

Proposition 2.3.2 (Smooth Levi-Civitá connection)**.**

$$
\begin{aligned}
\widetilde{\nabla}_U V(\mu) &:= -div\,[U(\mu)\nabla\, v(\mu)] - div\,[\mu\,\nabla(\,T(v)(\mu).U(\mu))] \\
&= div\,[div(\mu\,\nabla\, u(\mu))\nabla\, v(\mu)] - div\,[\mu\,\nabla(\,T(v)(\mu).U(\mu))]\,,
\end{aligned}
$$

i.e. for all $\varphi \in \mathcal{C}_c^\infty(M)$:

$$
(\widetilde{\nabla}_U V(\mu)|\varphi) \;:=\; \int_M \langle\nabla\langle\nabla\varphi,\nabla v\rangle_x,\nabla u\rangle_x \mu(dx) + \int_M \langle\nabla(T(v)(\mu).U(\mu)),\nabla\varphi\rangle_x \mu(dx)
$$

Proof. We have to show that

$$
U\langle V,W\rangle_\mu = \langle\widetilde{\nabla}_U V, W\rangle_\mu + \langle V, \widetilde{\nabla}_U W\rangle_\mu,
$$

i.e. for $I(\mu) = \langle V(\mu), W(\mu)\rangle_\mu$

$$\begin{aligned}
U\langle V, W\rangle_\mu = & \\
& T(I).U \\
= & \frac{d}{dt}|_0 \int_M \{\langle \nabla v((\mathrm{id} + t\nabla u(\mu))\#\mu), \nabla w((\mathrm{id} + t\nabla u(\mu))\#\mu)\rangle_x \times \\
& \times (\mathrm{id} + t\nabla u(\mu))\#\mu)\} \\
= & \int_M \left\langle \frac{d}{dt}|_0 \nabla v((\mathrm{id} + t\nabla u(\mu))\#\mu), \nabla w(\mu)\right\rangle_x \mu(dx) \\
& + \int_M \left\langle \nabla v(\mu), \frac{d}{dt}|_0 \nabla w((\mathrm{id} + t\nabla u(\mu))\#\mu)\right\rangle_x \mu(dx) \\
& + \int_M \langle \nabla(\langle \nabla v(\mu), \nabla w(\mu)\rangle_x), \nabla u(\mu)\rangle_x \\
= & \int_M \langle \nabla T(v).U, \nabla w(\mu)\rangle_x \, \mu(dx) + \int_M \langle \nabla v(\mu), \nabla T(w).U\rangle_x \, \mu(dx) \\
& + \int_M \langle \nabla\nabla v, \nabla w\rangle_x, \nabla u\rangle_x \, \mu + \int_M \langle\langle \nabla v, \nabla\nabla w\rangle_x, \nabla u\rangle_x \, \mu \\
= & -\langle \mathrm{div}(\mu \nabla T(v).U), w\rangle_\mu - \langle \mathrm{div}(\mu \nabla T(w).U), v\rangle_\mu \\
& + \langle \mathrm{div}(\mathrm{div}\, \mu \nabla u)\nabla v), w\rangle_\mu + \langle \mathrm{div}(\mathrm{div}\, \mu \nabla v)\nabla u), v\rangle_\mu \\
= & \langle \widetilde{\nabla}_U V, W\rangle_\mu + \langle V, \widetilde{\nabla}_U W\rangle_\mu.
\end{aligned}$$

Taking some Riemannian connection $\overline{\nabla}$ definied in terms of the Koszul formula

$$\begin{aligned}
2\langle \overline{\nabla}_U V, W\rangle_\mu = & \; U\langle V, W\rangle_\mu + V\langle W, U\rangle_\mu - W\langle U, V\rangle_\mu + \langle W, [U, V]\rangle_\mu \quad (2.5) \\
& -\langle V, [U, W]\rangle_\mu - \langle U, [V, W]\rangle_\mu \quad (2.6)
\end{aligned}$$

and substituting the Lie bracket and the calculations of $U\langle V, W\rangle_\mu$ into this formula shows that $\overline{\nabla} = \widetilde{\nabla}$. It is the Levi-Cività connection since $\widetilde{\nabla}$ is torsion-free by definition. □

Remark 2.3.1. *Note that $\widetilde{\nabla}_U V(\mu) \in T_\mu P^\infty$ since*

$$\begin{aligned}
(\widetilde{\nabla}_U V(\mu)|\varphi) = & \int_M \langle \nabla\varphi, \nabla(G_\mu d^*_\mu(\nabla_i \nabla_j v \nabla^j u dx^i))\rangle_x \mu(dx) \\
& + \int_M \langle \nabla\varphi, \nabla(T(v)(\mu).U(\mu))\rangle_x \mu(dx)
\end{aligned}$$

by lemma 4.14 in [Lot08]. Here G_μ denotes the Green operator for $d^*_\mu d$ on $L^2(\mu)$.

Remark 2.3.2. *[Gig09] developed notions in order to generalize covariant derivatives to the case of vector fields on P_{ac} by introducing parallel transport of absolutely continuous vector fields along regular curves. Regular curves $c : [0,1] \to P_{ac}$ are those whose velocity vector field v_t (given by the solution of the continuity equation) satisfy the Lipschitz condition $\int_0^1 Lip(v_t)\, dt < \infty$ and $\int_0^1 |v_t|^2_{c(t)} dt < \infty$. Absolutely continuous vector fields u_t are those for which the translation $\tau_t^s(u_t)$ from $L^2(c(t))$ to $L^2(c(s))$ are absolutely continuous in t for any s. It is shown the the angle between tangent spaces varies smoothly along* *the regular curves, i.e. the translation of a vector field from $L^2(c(t))$ to $L^2(c(s))$ along such curves is almost in the tangent space to P_{ac} when s and t are close. With the help of the parallel transport $\mathcal{T}_{t+h}^t : T_{\mu_{t+h}}P \to T_{\mu_t}P$ the covariant derivative is defined as*

$$\frac{\mathbf{D}}{dt}u_t := \lim_{h\to 0} \frac{\mathcal{T}_{t+h}^t(u_{t+h})) - u_t}{h},$$

It is shown in [Gig09] that on the space of smooth positive densities this notion of covariant derivative and the formulas obtained in proposition are the same.

Let us return to the case of smooth positive densities on a compact manifold:

Proposition 2.3.3 (Parallel transport)**.**

Proof. With the help of the Levi-Cività connection on P^∞ we are able to formulate parallel transport reminding finite dimensional Riemannian geometry. Given a curve $c(t) \in P^\infty$ and a vector field V along this curve in the sense that there exists a smooth time-depending function v_t such that

$$\frac{dc}{dt} = -\text{div}(c(t)\nabla v_t) = V_t(c(t)).$$

To transport a vector field parallely along a given curve $c(t)$ is equivalent to asking for a time-depending vector field W_t on P^∞ (given by another smooth time-depending function w_t) such that

$$(\tilde{\nabla}_{V_t} W_t)(c(t)) = 0.$$

By a basis $\{E_\alpha; \alpha \in \mathbb{N}\}$ of $\mathcal{C}^\infty$ we obtain a global basis of TP^∞ given by a family of vector fields $\{E_\alpha; \alpha \in \mathbb{N}\}$ on TP^∞. We write

$$W_t = w_t^\alpha E_\alpha|_{c(t)}$$

By the derivation rule for the covariant derivative this gives the following formula evaluated at $c(t)$:

$$0 = \tilde{\nabla}_{V_t} W_t = \tilde{\nabla}_{V_t}(w_t^\alpha E_\alpha) = \langle \nabla^{\mathrm{P}^\infty} w^\alpha, V_t\rangle_{c(t)} E_\alpha + w_t^\alpha \tilde{\nabla}_{V_t} E_\alpha$$

which is equivalent to

$$\langle \frac{d}{dt} w^\alpha(c(t)), V_t(c(t))\rangle_{c(t)} E_\alpha + w_t^\alpha \tilde{\nabla}_{V_t} E_\alpha = 0,$$

i.e.

$$\text{div}\left(c(t)\left(\nabla \frac{d}{dt} w_t^\alpha + \langle \nabla v_t, \nabla^2 w_t^\alpha\rangle\right)\right) = 0$$

with respect to the basis E_α. See also [Lot08]. □

This formula will be applied to obtain a

Proposition 2.3.4 (Geodesic equation)**.** *[Lot08] The curve $c(t) \in P^\infty$ is a geodesic if*

$$\tilde{\nabla}_{\dot{c}_t} \dot{c}_t = 0,$$

i.e. for a non-constant time-depending function $\phi \in \mathcal{C}^\infty$ which solves

$$\frac{\partial \phi}{\partial t} + \frac{1}{2}|\nabla \phi|^2 = 0$$

such that

$$\dot{c}_t = -div(c(t)\nabla \phi_t)$$

for $t \in [0,1]$. Note that this formula recovers a result that was known by calculus of variation, see [Ott01].

Proof. Put $W_t = V_t = -\mathrm{div}(c(t)\nabla\phi_t)$ in the above formula. □

We turn our attention to the Riemannian curvature operator on P^∞:

2.4 Riemannian curvature

Definition 2.4.1. *For smooth functions ϕ and ψ we define Π_μ as the orthogonal projection onto $\overline{Im(d)}$ in $\Omega^1_{L^2(\mu)}$ and $T_{\phi\psi} := (id - \Pi_\mu)(\phi' \psi'' dx)$. We use the letter R for the Riemannian curvature operator on the underlying manifold, i.e.*

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$$

Theorem 2.4.1 (Riemannian curvature operator)**.** *[Lot08] Let $\phi_1, \phi_2, \phi_3, \phi_4 \in \mathcal{C}^\infty$ determine respective vector fields $V_1, V_2, V_3, V_4 \in \Gamma(TP^\infty)$. The curvature operator $\bar{R}$ is given in $\mu \in P^\infty$ by*

$$\begin{aligned}\langle \bar{R}(V_1,V_2)V_3, V_4\rangle_\mu &= \int_M \langle R(\phi_1,\phi_2)\phi_3, \phi_4\rangle \mu(dx) - \\ &\quad -2\langle T_{\phi_1\phi_2}, T_{\phi_3\phi_4}\rangle_\mu + \langle T_{\phi_2\phi_3}, T_{\phi_1\phi_4}\rangle_\mu - \langle T_{\phi_1\phi_3}, T_{\phi_2\phi_4}\rangle_\mu\end{aligned}$$

Proof. Iterate Koszul's formula (see (2.5)) to obtain

$$\begin{aligned}\langle \bar{R}(V_1,V_2)V_3, V_4\rangle_\mu &= V_1\langle \overline{\nabla}_{V_2} V_3, V_4\rangle - \langle \overline{\nabla}_{V_2} V_3, \overline{\nabla}_{V_1} V_4\rangle - \\ &\quad -V_2\langle \overline{\nabla}_{V_1} V_3, V_4\rangle + \langle \overline{\nabla}_{V_1} V_3, \overline{\nabla}_{V_2} V_4\rangle - \\ &\quad -\langle \overline{\nabla}_{[V_1,V_2]} V_3, V_4\rangle\end{aligned}$$

The rest of the calculations follow directly from Remark 2.3.1. □

Chapter 3

Zeta function regularized Laplacian

In this chapter we continue the Riemannian calculus on P^∞ by calculating formulas for the Hessian of a functional. For the example of $P^\infty(\mathbb{T}^d)$ we calculate explicitely the trace of the Hessian by intertwining a Hilbert-Schmidt operator on $\Gamma(TP^\infty)$ in order to make the trace (which depends on an additional parameter) convergent: this is called renormalization. By a procedure known from mathematical physics we consider the analytical continuation (in the parameter variable) of the trace to the complex plane and obtain an expression for the trace when taking the parameter to zero (with the help of the residue of this function at zero). The resulting operator is called zeta function regularized Laplacian: Its iterated square field operator (see [BE85]) is calculated. Relations to the generator of Sturm-von Renesse's Wasserstein diffusion are shown.

3.1 Second order calculus

In [Ott01] the Hessian of the entropy functional $\mathrm{Ent}(\mu) = \int_{\mathbb{R}^n} \mu \log(\mu)\,\mathrm{vol}(dx)$ with respect to Kantorovich-Rubinstein metric was calculated by second order variation of the entropy functional along constant speed geodesics. We will calculate the Hessian with respect to the Levi-Cività connection on P^∞ for any smooth functional $E : P^\infty \to \mathbb{R}$ of the type

$$E(\mu) = \int_M e(m(x))\,\mathrm{vol}(dx), \quad \frac{d\mu}{d\mathrm{vol}}(x) = m(x), \quad e : \mathbb{R}_+ \to \mathbb{R}.$$

This will be done in normal coordinates, i.e. covariant derivatives are calculated in directions $U \in \Gamma(TP^\infty)$ giving rise to geodesics: $U(\mu) = -\mathrm{div}(\mu \nabla u)$ for some $u \in \mathcal{C}^\infty(M)$ depending not on m.

Proposition 3.1.1. *[The Hessian: a variational approach] Given a functional $E : P_{ac}(M) \to \mathbb{R}$ of the type*

$$E(\mu) = \int_M e(\mu(x))vol(dx)$$

where $e : \mathbb{R}_+ \to \mathbb{R}$ is twice differentiable we define

$$\begin{aligned} p(\mu) &= \mu e'(\mu) - e(\mu) \\ p_2(\mu) &= \mu p'(\mu) - p(\mu). \end{aligned}$$

By $Hess^{var} E(\dot\mu, \dot\mu)$ we denote the second order variation of E along a geodesic path $t \mapsto \mu_t$ in P_{ac} of the form

$$\begin{cases} \partial_t \mu + \nabla.(\mu \nabla\varphi) = 0 \\ \partial_t \varphi + \frac{|\nabla\varphi|^2}{2} = 0. \end{cases}$$

By Γ_2 we denote the iterated square field operator with respect to Δ: Define $\Gamma(f,g) = \Delta(fg) - g\Delta f - f\Delta g$, then $\Gamma_2(f) \equiv \Gamma_2(f,f) := \Delta\Gamma(f,f) - \Gamma(f,\Delta f) - \Gamma(f,\Delta f)$.

Then

$$Hess^{var} E(\dot\mu,\dot\mu) = \int_M \Gamma_2(\varphi_0)\, p(\mu)\, e^{-V}\, vol + \int_M (L\varphi_0)^2\, p_2(\mu)\, e^{-V}\, vol. \tag{3.1}$$

Proof. See [Vil08], p441f.

By the formula for the Wasserstein gradient we have for first order variation that

$$\begin{aligned} \frac{d}{dt} E(\mu_t) &= \int_M \langle \nabla\varphi_t, \nabla e'(\mu_t)\rangle \mu_t \mathrm{vol} \\ &= \int_M \langle \nabla\varphi_t, \nabla p(\mu_t)\rangle \mathrm{vol} \\ &= -\int_M (\Delta\varphi_t) p(\mu_t) \mathrm{vol} \end{aligned}$$

Differentiating once again

$$\begin{aligned} \frac{d^2}{dt^2} E(\mu_t) &= -\int_M (\Delta\partial_t\varphi_t) p(\mu_t)\mathrm{vol} - \int_M (\Delta\varphi_t) p'(\mu_t)\partial_t\mu_t \mathrm{vol} \\ &= \int_M \Delta\left(\frac{|\nabla\varphi_t|^2}{2}\right) p(\mu_t)\mathrm{vol} - \int_M (\Delta\varphi_t) p'(\mu_t)\partial_t\mu_t\mathrm{vol} \end{aligned}$$

Note that

$$\begin{aligned} -\int_M (\Delta\varphi_t)p'(\mu_t)\partial_t\mu_t\mathrm{vol} &= \int_M (\Delta\varphi_t)p'(\mu_t)\nabla.(\mu_t\nabla\varphi_t)\mathrm{vol} \\ &= -\int_M \langle \nabla((\Delta\varphi_t)p'(\mu_t)), \nabla\varphi_t\rangle \mu_t\mathrm{vol} \\ &= -\int_M \langle \nabla(\Delta\varphi_t), \nabla\varphi_t\rangle p'(\mu_t)\mu_t\mathrm{vol} - \int_M (\Delta\varphi_t)p''(\mu_t)\mu_t\langle\nabla\mu_t, \nabla\varphi_t\rangle\mathrm{vol} \\ &= -\int_M \langle \nabla(\Delta\varphi_t), \nabla\varphi_t\rangle p'(\mu_t)\mu_t\mathrm{vol} - \int_M (\Delta\varphi_t)\langle\nabla p_2(\mu_t), \nabla\varphi_t\rangle\mathrm{vol} \end{aligned}$$

and using integration by parts

$$\begin{aligned} -\int_M (\Delta\varphi_t)\langle\nabla p_2(\mu_t),\nabla\varphi_t\rangle\mathrm{vol} &= -\left(\int_M \langle\nabla((\Delta\varphi_t)p_2(\mu_t)),\nabla\varphi_t\rangle\mathrm{vol} - \int_M \langle\nabla(\Delta\varphi_t),\nabla\varphi_t\rangle p_2(\mu_t)\mathrm{vol}\right) \\ &= \int_M (\Delta\varphi_t)^2 p_2(\mu_t)\mathrm{vol} + \int_M \langle\nabla(\Delta\varphi_t),\nabla\varphi_t\rangle p_2(\mu_t)\mathrm{vol} \end{aligned}$$

Collecting all terms we get

$$\begin{aligned}
\frac{d^2}{dt^2}E(\mu_t) &= \int_M \Delta\left(\frac{|\nabla\varphi_t|^2}{2}\right)p(\mu_t)\mathrm{vol} + \int_M (\Delta\varphi_t)^2 p_2(\mu_t)\mathrm{vol} \\
&\quad + \int_M \langle\nabla(\Delta\varphi_t), \nabla\varphi_t\rangle p_2(\mu_t)\mathrm{vol} - \int_M \langle\nabla(\Delta\varphi_t), \nabla\varphi_t\rangle p'(\mu_t)\mu_t\mathrm{vol} \\
&= \int_M \Delta\left(\frac{|\nabla\varphi_t|^2}{2}\right)p(\mu_t)\mathrm{vol} + \int_M (\Delta\varphi_t)^2 p_2(\mu_t)\mathrm{vol} + \\
&\quad \int_M \langle\nabla(\Delta\varphi_t), \nabla\varphi_t\rangle \underbrace{\{p_2(\mu_t) - p'(\mu_t)\mu_t\}}_{-p(\mu_t)}\mathrm{vol} \\
&= \int_M \underbrace{\left(\Delta\left(\frac{|\nabla\varphi_t|^2}{2}\right) - \langle\nabla(\Delta\varphi_t), \nabla\varphi_t\rangle\right)}_{\Gamma_2(\varphi_t)} p(\mu_t)\mathrm{vol} + \int_M (\Delta\varphi_t)^2 p_2(\mu_t)\mathrm{vol}
\end{aligned}$$

Remark 3.1.1. *In Villani's book this proposition is referred to as "formula" in order to caution the reader against the so called "formal Riemannian calculus" on P_{ac}. Indeed Otto's Wasserstein gradient formula requires a restriction to the (non-complete with respect to the weak topology) subspace $P^\infty \subset P$ the above calculations are rigorous due to the formalism developed in chapter 1.*

□

3.2 Zeta function regularized Laplacian on $\mathbf{P}^\infty(\mathbb{T}^1)$

The *trace* of a symmetric bilinear form B on a Riemannian manifolds $(M, \langle\cdot,\cdot\rangle_x)$ at a point $x \in M$ for a chosen orthonormal basis $\{e_i\}_{i=1,\ldots,\dim(M)} \subset T_xM$ is definied as

$$\mathrm{tr}(B)(x) := \sum_{i=1}^{\dim(M)} \langle Be_i, e_i\rangle_x.$$

The functional *tr* is by definition invariant under change of the basis by any orthogonal matrix $O \in O(\dim(M))$. In order to make the definition a global one one has to clarify how an element e_x of O_x, the set of all orthonormal bases of T_xM, changes in dependence on the basepoint x: Any basis e_x will be moved by parallel transport along a smooth curve to a point $e_y \in O_y$. In infinite dimension two questions arise immediately: Firstly, how can one make the series $\sum_{i=1}^\infty \langle Be_i, e_i\rangle_x$ converge, and secondly, what should be meant by invariance of the trace under some group $O(\infty)$? In the case of Hilbert manifolds M modelled on H^s-Sobolev completions of $\Gamma(TM \leftarrow M)$ for sufficiently large $s \in \mathbb{R}$ and Hilbert-Schmidt operator A acting on $\Gamma(TM \leftarrow M)$ one can remedy the convergence question in defining the *A-trace* by $\mathrm{tr}^A(B)(p) = \sum_{i=1}^m \langle A^*BAe_i, e_i\rangle_p$ with $p \in M$ and $\{e_i\}_{i=1}^\infty$ a complete orthornomal system of T_pM. Since the operator B is bounded and the Hilbert-Schmidt norm $||A||_{\mathrm{HS}}^2 := \sum_{i=1}^\infty \langle Ae_i, Ae_i\rangle_p$ is finite, the *A-trace* is convergent. In the sequal we adopt a similar point of view in the case of P^∞.

Proposition 3.2.1 (Renormalized Laplacian on $\mathrm{P}^\infty(\mathbb{T}^1)$)**.** *Given a functional $E : P^\infty(\mathbb{T}^1) \to \mathbb{R}$ of the type*

$$E(\mu) = \int_{\mathbb{T}^1} e(\mu(x))vol(dx),$$

where $e : \mathbb{R}_+ \to \mathbb{R}$ is $\mathcal{C}^3$. For an orthonormal system $\{e_k(\mu)\}_{k\in\mathbb{N}}$ of

$$T_\mu P^\infty(\mathbb{T}^1) := \overline{\mathcal{C}^\infty(\mathbb{T}^1)/\mathbb{R}}^{L^2(\mu)}$$

we define an operator A on $T_\mu P^\infty(\mathbb{T}^1)$ by diagonalization in the basis $\{e_i(\mu)\}_{k\in\mathbb{N}}$:

$$A : e_k(\mu) \mapsto, \lfloor k/2 \rfloor^{-a} e_k(\mu);\ k \in \{2, 3, \dots\},\ a > \frac{3}{2}.$$

For the first mode we define $A : e_1(\mu) \mapsto 2\pi\sqrt{2}e_1(\mu)$. Let $\widetilde{HessE}$ be the Hessian operator associated to the the (variational) Hessian $Hess^{var}E(.,.)(\mu)$. The **renormalized Wasserstein Laplacian** *in an open neighbourhood of μ as defined below is finite:*

$$\Delta^a_{P^\infty(\mathbb{T}^1)} E(\mu) := \sum_{k=1}^{\infty} \langle \widetilde{HessE} A e_k(\mu), A e_k(\mu)\rangle_\mu < \infty$$

Proof. For the inner product $\langle e_k, e_k\rangle_{\mathrm{vol}} \equiv \langle e_k, e_k\rangle_{H^1(\mathrm{vol})} := \frac{1}{(2\pi)^2}\langle e'_k, e'_k\rangle_{L^2}$ on $T_{\mathrm{vol}}\mathrm{P}^\infty(\mathbb{T}^1)$, we are given a complete orthonormal system on $T_{\mathrm{vol}}\mathrm{P}^\infty(\mathbb{T}^1)$ by

$$\begin{cases} e_{2k}(x) = \sqrt{2}\,k^{-1}\sin 2\pi k x, & k \in \mathbb{N} \\ e_{2k+1}(x) = \sqrt{2}\,k^{-1}\cos 2\pi k x, & k \in \mathbb{N} \\ e_1(x) = 1. \end{cases}$$

Likewise by

$$\begin{cases} e_{2k}(\mu)(x) \text{ such that} & \frac{d}{dx}e_{2k}(\mu)(x) = \frac{1}{\sqrt{\mu(x)}}\frac{d}{dx}e_{2k}(x), & k \in \mathbb{N} \\ e_{2k+1}(\mu)(x) \text{ such that} & \frac{d}{dx}e_{2k+1}(\mu)(x) = \frac{1}{\sqrt{\mu(x)}}\frac{d}{dx}e_{2k+1}(x), & k \in \mathbb{N} \\ e_1(\mu)(x) \text{ such that} & \frac{d}{dx}e_1(\mu)(x) = \frac{1}{\sqrt{\mu(x)}}. \end{cases}$$

with inital data

$$\begin{cases} e_{2k}(\mu)(0) = 0, & k \in \mathbb{N} \\ e_{2k+1}(\mu)(0) = 0, & k \in \mathbb{N} \end{cases}$$

we are given a complete orthonormal system of $T_\mu\mathrm{P}^\infty(\mathbb{T}^1)$: On the torus we can solve the defining differential equation by integration and orthonormality of $\{e_k(\mu)\}_{k\in\mathbb{N}}$ is given by definition. To show that $\{e_k(\mu)\}_{k\in\mathbb{N}} \subset T_\mu\mathrm{P}^\infty(\mathbb{T}^1)$ we consider a vector field u such that $\mathrm{div}_\mu u = 0$. We have to show that $e_k(\mu) \perp u$

with respect to $\langle .,. \rangle_\mu$ for all $k \in \mathbb{N}$:

$$\begin{aligned}
\int_{\mathbb{T}^1} e_k'(\mu).u\mu &= \int_{\mathbb{T}^1} e_k'.u\sqrt{\mu} \\
&= -\int_{\mathbb{T}^1} \underbrace{(\sqrt{\mu}/\mu)e_k}_{\varphi_1 \in \mathcal{C}_c^\infty(\mathbb{T}^1)}.(u\mu)' + \int_{\mathbb{T}^1} (e_k u\mu)\frac{\mu'}{2\sqrt{\mu^3}} \\
&= 0 + \int_{\mathbb{T}^1} \underbrace{\left(e_k \frac{\mu'}{2\sqrt{\mu^3}}\right)}_{\varphi_2' \text{ for } \varphi_2 \in \mathcal{C}_c^\infty(\mathbb{T}^1)} u\mu = 0
\end{aligned}$$

since $\int \mu u.\varphi' = 0$ for any $\varphi \in \mathcal{C}_c^\infty(S^1)$. Note that at this place it is crucial to deal with differentiable densities with full support. The function φ_2 is obtained by integration.

Given a functional $E : \mathrm{P}^\infty \to \mathbb{R}$ and a distribution $U \in T\mathrm{P}^\infty$ such that $(U(\mu)|\varphi) = \int_{\mathbb{T}^1} u'\varphi'\mu$ for smooth, compactly supported functions u and φ. According to ([Vil08]):

$$\mathrm{Hess}^{\mathrm{var}} E(U,U)(\mu) = \int_{\mathbb{T}^1} \Gamma_2^\Delta(u)(\mu e'(\mu) - e(\mu))\mathrm{vol} + \int_{\mathbb{T}^1} (\Delta u)^2(\mu p'(\mu) - p(\mu))\mathrm{vol},$$

with

$$p(x) = xe'(x) - e(x) \text{ and } p'(x) = xe''(x) + e'(x) - e'(x) = xe''(x)$$

and Γ_2^Δ the iterated carré du champ operator with respect to $\Delta = \Delta_{\mathbb{T}^1} = \frac{d^2}{dx^2}$. Then

$$\begin{aligned}
\mathrm{Hess}^{\mathrm{var}} E(U,U)(\mu) &= \int_{\mathbb{T}^1} (u'')^2(\mu e'(\mu) - e(\mu))\mathrm{vol} + \int_{\mathbb{T}^1} (u'')^2(\mu^2 e''(\mu) - \mu e'(\mu) + e(\mu))\mathrm{vol} \\
&= \int_{\mathbb{T}^1} (u'')^2 \mu^2 e''(\mu)\mathrm{vol},
\end{aligned}$$

and

$$\begin{aligned}
\Delta^a_{\mathrm{P}^\infty(\mathbb{T}^1)} E(\mu) &= \int_{\mathbb{T}^1} 2(2\pi)^2((e_1(\mu))'')^2 \mu^2 e''(\mu)\mathrm{vol}(dx) \\
&\quad + \sum_{k=2}^\infty \int_{\mathbb{T}^1} \lfloor k/2 \rfloor^{-2a}((e_k(\mu))'')^2 \mu^2 e''(\mu)\mathrm{vol}(dx) \\
&= \sum_{k=1}^\infty \int_{\mathbb{T}^1} k^{-2a}(((\mu^{-1/2}e_{2k}')')^2 + ((\mu^{-1/2}e_{2k+1}')')^2)\mu^2 e''(\mu)\mathrm{vol}(dx) \\
&\quad + 2(2\pi)^2 \int_{\mathbb{T}^1} \frac{((\log\mu)')^2}{4\mu}\mu^2 e''(\mu)\mathrm{vol}(dx)
\end{aligned}$$

Since $e_{2k}'' = 2\pi k e_{2k+1}'$ resp. $e_{2k+1}'' = -2\pi k e_{2k}'$ and $(e_{2k}')^2 + (e_{2k+1}')^2 = 2(2\pi)^2$ it follows that

$$\begin{aligned}
((\mu^{-1/2}e_{2k}')')^2 + ((\mu^{-1/2}e_{2k+1}')')^2 &= ((e_{2k}')^2 + (e_{2k+1}')^2)(1/4\mu^{-3}(\mu')^2) + \\
&\quad ((e_{2k}')^2 + (e_{2k+1}')^2)(\mu^{-1}(2\pi k)^2) + \\
&\quad e_{2k}'e_{2k+1}(-\mu^{-3/2}\mu'\mu^{-1/2}2\pi k + \mu^{-3/2}\mu'\mu^{-1/2}2\pi k) \\
&= 2(2\pi)^2\{1/4\mu^{-3}(\mu')^2 + \mu^{-1}(2\pi k)^2\}
\end{aligned}$$

Consequently

$$\begin{aligned}
\Delta^a_{\mathrm{P}^\infty(\mathbb{T}^1)} E(\mu) &= \sum_{k=1}^{\infty} \int_{\mathbb{T}^1} k^{-2a} 2(2\pi)^2 \left\{1/4\mu^{-3}(\mu')^2 + \mu^{-1}(2\pi k)^2\right\} \mu^2 e''(\mu)\mathrm{vol}(dx) \\
&\quad + 2(2\pi)^2 \int_{\mathbb{T}^1} \frac{((\log\mu)')^2}{4\mu} \mu^2 e''(\mu)\mathrm{vol}(dx) \\
&= 2(2\pi)^2 \sum_{k=1}^{\infty} k^{-2a} \int_{\mathbb{T}^1} \left\{1/4((\log\mu)')^2 + (2\pi k)^2\right\} \mu e''(\mu)\mathrm{vol}(dx) \\
&\quad + 2(2\pi)^2 \int_{\mathbb{T}^1} \frac{((\log\mu)')^2}{4} \mu e''(\mu)\mathrm{vol}(dx) \\
&< \infty
\end{aligned}$$

since

$$\begin{aligned}
&||((\log\mu)')^2 \mu e''(\mu)||_\infty < +\infty \\
&||\mu e''(\mu)||_\infty < +\infty,
\end{aligned}$$

which is guaranteed since the densities are supposed to have full support and to be sufficiently regular.

□

For the Riemann zeta function $\zeta^R(s) = \sum_{k=1}^{\infty} \frac{1}{k^s}, \Re(s) > 1$ there exists a meromorphic continuation to the complex plane with single pole at $s = 1$ which was proved by Riemann himself in 1859 by the following functional equation:

$$\zeta^R(s) = 2^s \pi^{s-1} \sin\frac{s\pi}{2}\Gamma(1-s)\zeta^R(1-s); \;\; s \in \mathbb{C}\setminus\{1\}$$

which enables us to calculate a specific value:

$$\begin{aligned}
\zeta^R(0) &= \frac{1}{\pi}\lim_{s\to 0} \sin\frac{s\pi}{2}\Gamma(1-s)\zeta^R(1-s) \\
&= \frac{1}{\pi}\lim_{s\to 0}\left(\frac{s\pi}{2} - \frac{s^3\pi^3}{48} + \dots\right)\left(-\frac{1}{s} + \dots\right) = -\frac{1}{2}.
\end{aligned}$$

We used that $\mathrm{Res}(\zeta, 1) = \lim_{s\to 1}(s-1)\zeta^R(s) = 1 = a_{-1}$ and the Laurent series reads $\zeta^R(s) = \sum_{n=-1}^{\infty} a_n(s-1)^n$ i.e. $\zeta^R(1-s) = -\frac{1}{s} + \dots$.

Definition 3.2.1 (Zeta function regularized Laplacian)**.**

$$\Delta_{P^\infty(\mathbb{T}^1)} E(\mu) := \lim_{a\to 0} \Delta^a_{P^\infty(\mathbb{T}^1)} E(\mu)$$

is called (Zeta function) regularized Laplacian.

Proposition 3.2.2. *Given a functional $E : P^\infty(\mathbb{T}^1) \to \mathbb{R}$ of the type*

$$E(\mu) = \int_{\mathbb{T}^1} e(\mu(x)) vol(dx),$$

where $e : \mathbb{R}_+ \to \mathbb{R}$ is $\mathcal{C}^3$. Then

$$\begin{aligned}\Delta_{P^\infty(\mathbb{T}^1)} E(\mu) &= 2(2\pi)^2(\zeta^R(0)+1)\int_{\mathbb{T}^1}\left\{((\log\mu)'/2)^2\}\mu e''(\mu)\right\} vol(dx)\\ &= \pi^2\int_{\mathbb{T}^1}\left\{(\log\mu)'\right\}^2 \mu e''(\mu) vol(dx).\end{aligned}$$

We used additionally the fact that for the analytical continuation of the Zeta function $\zeta^R(-2) = 0$ holds.

Example 3.2.1. *For $Ent(\mu) = \int_{\mathbb{T}^1}\mu(x)\log\mu(x)\, vol(dx)$ we have*

$$\Delta_{P^\infty(\mathbb{T}^1)} Ent(\mu) = \pi^2 ||(\log\mu)'||^2_{L^2(vol)}$$

Example 3.2.2. *For functionals $E(\mu) = \int_{\mathbb{T}^1} f(x) vol(dx)$ with f a measurable function on $\mathbb{T}^1$ we have $\Delta_{P^\infty(\mathbb{T}^1)} E(\mu) = 0$ for all $\mu \in P^\infty$.*

Remark 3.2.1. *Set $E(\mu) = \frac{1}{2}\int_{\mathbb{T}^1}\mu^2 vol$, then*

$$\Delta_{P^\infty} E(\mu) = \pi^2 ||\nabla^{P^\infty} Ent(\mu)||^2_\mu$$

for all $\mu \in P^\infty$.

Proposition 3.2.3. *Given a functional $F : P^\infty(\mathbb{T}^1) \to \mathbb{R}$ of the type*

$$F(\mu) = \Phi\left(\langle f, \mu\rangle\right),$$

where $f \in \mathcal{C}_b(\mathbb{T}^1)$ and $\Phi \in \mathcal{C}_b(\mathbb{R})$. Then

$$\Delta_{P^\infty} F(\mu) = 2(2\pi)^2\Phi''(\langle f, \mu\rangle)||f'\sqrt{\mu}||^2_{L^2}$$

and the square-field operator with respect to Δ_{P^∞} applied to functionals F reads:

$$\Gamma(F) = 2(2\pi)^2||\nabla^{P^\infty} F(\mu)||^2_\mu$$

Proof. We denote the $L^2(\mu)$ inner product by $\langle , \rangle_\mu$, if no measure is specified we consider the inner product on $L^2(\text{vol})$. Following ([Lot08]) a geodesic $(\mu_t)_{t\in[0,T]}$ in P^∞ starting at $\mu_0 = \mu$ satisfies

$$\dot{\mu}_t = -\text{div}(\mu_t \nabla v_t)$$

where the smooth function v_t satisfies

$$\dot{v}_t = \frac{-|\nabla v_t|^2}{2}.$$

The second order variation of F along $(\mu_t)_{t\in[0,T]}$ reads

$$\begin{aligned}\frac{d^2}{dt^2}\Phi(\langle f, \mu_t\rangle) &= \frac{d}{dt}(\Phi'(\langle f, \mu_t\rangle)\langle f, \dot{\mu}_t\rangle)\\ &= \Phi''(\langle f, \mu_t\rangle)\langle f, \dot{\mu}_t\rangle^2 + \Phi'(\langle f, \mu_t\rangle)\frac{d}{dt}\langle f, \dot{\mu}_t\rangle\end{aligned}$$

Remark since $\dot{v}_t' = -v_t' v_t''$ and $\dot{\mu}_t = -(\mu_t v_t')'$

$$\begin{aligned} \frac{d}{dt}\langle f, \dot{\mu}_t\rangle &= \frac{d}{dt}\int f' v_t' \mu_t \mathrm{vol} \\ &= \langle f', -v_t' v_t'' \mu_t + v_t' \dot{\mu}_t\rangle \\ &= -\langle f', ((v_t')^2)' \mu_t + \mu_t' (v_t')^2\rangle \\ &= -\langle f', ((v_t')^2 \mu_t)'\rangle \end{aligned}$$

$$\begin{aligned} \mathrm{Hess}(F)(v', v')(\mu) &= \frac{d^2}{dt^2}_{t=0} \Phi(\langle f, \mu_t\rangle) \\ &= \Phi''(\langle f, \mu\rangle)\langle f', v'\rangle_\mu^2 + \Phi'(\langle f, \mu\rangle)\langle f'', (v')^2 \mu\rangle \end{aligned}$$

In this formula at the place of v' we plug in (here for $s > 1/2$) $k^{-s} e_k(\mu)'$ as in the proof of Proposition 3.2.1 in order to calculate

$$\Delta^s_{\mathrm{P}^\infty(\mathbb{T}^1)} F(\mu) = \sum_{k=1}^{\infty} \Phi''(\langle f, \mu\rangle)\langle f', k^{-s} e_k(\mu)'\rangle_\mu^2 + \Phi'(\langle f, \mu\rangle)\langle f'', (k^{-s} e_k(\mu)')^2 \mu\rangle$$

which equals

$$\Phi''(\langle f, \mu\rangle) \sum_{k=1}^{\infty} \langle f', k^{-s} \mu^{-1/2} e_k'\rangle_\mu^2 + 2(2\pi)^2 \Phi'(\langle f, \mu\rangle)\langle f'', 1\rangle \zeta(2s)$$

Since $\zeta(2s)$ is finite for $2s > 1$ and we now that $\langle f'', 1\rangle = 0$ the second term vanishes and by the functional equation for ζ we define again

$$\Delta_{\mathrm{P}^\infty(\mathbb{T}^1)} F(\mu) := \lim_{s \to 0} \Delta^s_{\mathrm{P}^\infty(\mathbb{T}^1)} F(\mu)$$

which equals

$$\lim_{s \to 0} \Phi''(\langle f, \mu\rangle) 2(2\pi)^2 ||f' \mu^{1/2}||^2_{H^{-s}} = \Phi''(\langle f, \mu\rangle) 2(2\pi)^2 ||f' \mu^{1/2}||^2_{L^2}.$$

Note that the limit is taken for $s \in \mathbb{C}$.

The square-field operator $\Gamma^s(F)$ with respect to $\Delta^s_{\mathrm{P}^\infty}$ is defined by

$$\frac{1}{2}\Delta^s_{\mathrm{P}^\infty}(F^2) - F \Delta^s_{\mathrm{P}^\infty}(F).$$

In a first step we remark that

$$\frac{1}{2}\frac{d^2}{dt^2}(F(\mu_t))^2 = (\frac{d}{dt} F(\mu_t))^2 + F(\mu_t) \frac{d^2}{dt^2} F(\mu_t)$$

and so

$$\frac{1}{2}\frac{d^2}{dt^2}(F(\mu_t))^2 - F \frac{d^2}{dt^2}(F(\mu_t)) = (\frac{d}{dt} F(\mu_t))^2 = (\Phi'(\langle f, \mu\rangle)\langle f', v'\rangle_\mu)^2$$

which entails

$$\Gamma^s(F) = \sum_{k=1}^{\infty} (\Phi'(\langle f, \mu\rangle)\langle f', k^{-s} e_k(\mu)'\rangle_\mu)^2 = 2(2\pi)^2 (\Phi'(\langle f, \mu\rangle))^2 ||f'||^2_{H_\mu^{-s}}$$

But

$$\lim_{s\to 0} ||f'||^2_{H_\mu^{-s}} = ||f'||^2_{L^2(\mu)}$$

and consequently

$$\lim_{s\to 0} \Gamma^s(F) = 2(2\pi)^2 ||\nabla^{\mathrm{P}^\infty} F(\mu)||^2_\mu$$

□

Remark 3.2.2. *By the chain rule the formulas for the regularized Wasserstein Laplacian can be extended to the set of test functions*

$$\mathfrak{Z} = \left\{ F(\mu) \equiv \Phi\left(\langle \underline{f}, \mu\rangle\right); \Phi \in \mathcal{C}^2(\mathbb{R}^d),\ \ \underline{f} = (f_1, \dots, f_d) \in \mathcal{C}^2(\mathbb{T}^1; \mathbb{R}^d); \mu \in P^\infty(\mathbb{T}^1) \right\},$$

i.e.

$$\Delta_{P^\infty} F(\mu) = \sum_{i,j=1}^{d} \partial_i \partial_j \Phi\left(\langle \underline{f}, \mu\rangle\right) \int_0^1 f_i' f_j' \mu$$

Remark 3.2.3. *Observe that within the class $\mathfrak{Z}$ there are test functions $\Phi(x) = e^{ix}$ for $d = 1$ such that*

$$\Delta_{P^\infty} e^{i \int_{\mathbb{T}^1} f\mu} = -e^{i \int_{\mathbb{T}^1} f\mu} \int_{\mathbb{T}^1} (f')^2 \mu$$

Proposition 3.2.4 (Iterated square-field operator)**.** *Let $F \in \mathfrak{Z}$ with $d = 1$:*

$$\Gamma_2(F) := \frac{1}{2} \Delta_{P^\infty} ||\nabla^{P^\infty} F||^2_\mu - \langle \nabla^{P^\infty} \Delta_{P^\infty} F, \nabla^{P^\infty} F\rangle_\mu$$

Then $\Gamma_2(F)$ equals

$$2(2\pi)^2 \left\{ (\Phi'')^2 ||f'||^4_\mu + \Phi'\Phi''\langle f', ((f')^2)'\rangle_\mu \right\}$$

(compare to [BE85]).

Proof. Let μ_t be a geodesic and consider at first test functions in $\mathfrak{Z}$ of type $F(\mu) = \Phi(\langle f, \mu\rangle) \equiv \Phi$. Remember that

$$||\nabla^{\mathrm{P}^\infty} F||^2_\mu = ||\Phi'\langle f, \dot\mu\rangle||^2_\mu = (\Phi')^2 ||f'||^2_\mu$$

Then

$$\begin{aligned}
\frac{d^2}{dt^2} ||\nabla^{\mathrm{P}^\infty} F||^2_\mu &= \frac{d}{dt}(2\Phi'\Phi''\langle f, \dot\mu\rangle ||f'||^2_\mu + (\Phi')^2 \langle (f')^2, \dot\mu\rangle) \\
&= 2(\Phi'')^2 \langle f, \dot\mu\rangle^2 ||f'||^2_\mu + 2\Phi'\Phi'''\langle f, \dot\mu\rangle^2 ||f'||^2_\mu + 2\Phi'\Phi''\langle f, \ddot\mu\rangle ||f'||^2_\mu \\
&\quad + 4\Phi'\Phi''\langle f, \dot\mu\rangle\langle (f')^2, \dot\mu\rangle + (\Phi')^2 \langle (f')^2, \ddot\mu\rangle
\end{aligned}$$

Taking the regularized trace we obtain

$$\begin{aligned}\Delta_{\mathrm{P}^\infty}||\nabla^{\mathrm{P}^\infty}F||_\mu^2 &= 2(\Phi'')^2 2(2\pi)^2||f'||_\mu^4 + 2\Phi'\Phi''' 2(2\pi)^2||f'||_\mu^4 + 2\Phi'\Phi''\langle f'', 1\rangle||f'||_\mu^2 2(2\pi)^2 \\ &\quad +2(2\pi)^2 4\Phi'\Phi''\langle f', ((f')^2)'\rangle_\mu + (\Phi')^2\langle((f')^2)'', 1\rangle 2(2\pi)^2 \\ &= 2(2\pi)^2\left\{2(\Phi'')^2||f'||_\mu^4 + 2\Phi'\Phi'''||f'||_\mu^4 + 4\Phi'\Phi''\langle f', ((f')^2)'\rangle_\mu\right\}\end{aligned}$$

On the other hand

$$\langle\nabla^{\mathrm{P}^\infty}\Delta_{\mathrm{P}^\infty}F, \nabla^{\mathrm{P}^\infty}F\rangle_\mu = 2(2\pi)^2\Phi'\Phi'''\langle f', f'\rangle_\mu||f'||_\mu^2 + 2(2\pi)^2\Phi'\Phi''\langle f', ((f')^2)'\rangle_\mu$$

So

$$\frac{1}{2}\Delta_{\mathrm{P}^\infty}||\nabla^{\mathrm{P}^\infty}F||_\mu^2 - \langle\nabla^{\mathrm{P}^\infty}\Delta_{\mathrm{P}^\infty}F, \nabla^{\mathrm{P}^\infty}F\rangle_\mu$$

equals

$$2(2\pi)^2\left\{(\Phi'')^2||f'||_\mu^4 + \Phi'\Phi''\langle f', ((f')^2)'\rangle_\mu\right\}.$$

The formula for the iterated square field operator generalizes by the chain rule to any $F \in \mathfrak{Z}$.

□

Remark 3.2.4. *Let $\mathcal{G}$ be the space of non-decreasing functions on $\mathbb{T}^1$. Remark that $\mathcal{G} \subset L^2(vol)$ is convex. We know (see [SvR09]) that there exists an isometry ι between $\mathcal{G}$ and the space of all probability measures on $\mathbb{T}^1$ equipped with the Wasserstein distance given by $\iota(g) = (g^{-1})'$. Denote by $\mathcal{H} \subset \mathcal{G}$ the dense subspace of strictly increasing smooth functions on the unit sphere. On $\mathcal{H}$ we define test funtions $F(h) := \Phi(\int f \circ h vol)$ where Φ resp. f are smooth function on $\mathbb{R}$ resp. on $\mathbb{T}^1$. We can easily calculate the L^2-Hessian:*

$$\begin{aligned}HessF(h)(\xi,\xi) &= \frac{d^2}{dt^2}|_0 F(h+t\xi) = \frac{d}{dt}|_0(\Phi'(\int f\circ h)\int f'\circ(h+t\xi)\xi) \\ &= \Phi''(\int f\circ h)(\int(f'\circ h)\xi)^2 + \Phi'(\int f\circ h)\int(f''\circ h)\xi^2\end{aligned}$$

The Zeta function regularized $L^2(vol)$ Laplacian on $\mathcal{H}$ is given by

$$\begin{aligned}\Delta_{L^2}F(h) &= \lim_{s\to 0}\sum_k k^{-2s} HessF(h)(e_k', e_k') \\ &= \Phi''(\int f\circ h)\langle f'\circ h, f'\circ h\rangle - (2\pi)^2\Phi'(\int f\circ h)\int(f''\circ h)\end{aligned}$$

By $\iota_*(F)(\mu) := F(\iota^{-1}(\mu))$ we obtain test functions on P^∞ and can see

$$\iota_*(\Delta_{L^2}F(h)) - \Delta_{P^\infty}\iota_*(F)(\mu) = \langle\nabla^{P^\infty}\iota_*F(\mu), \nabla^{P^\infty}\iota_*Ent(\mu)\rangle_\mu,$$

Remark 3.2.5. *[SvR09] construct a Markov process with values in $P([0,1])$ via a Dirichlet form (using the Wasserstein gradient) with respect to the so called entropic measure $\mathbb{P}^\beta$. This dubbing stems from the heuristic approach of defining a Gibbs type measure on the space of probability measures*

$$d\mathbb{P}^\beta = \frac{1}{Z_\beta}e^{-\beta Ent(\mu)}d\mathbb{P}$$

for some normalizing constant Z_β and some non-existing uniform distribution $\mathbb{P}$. Since every probability measure on the unit interval can be mapped isometrically (for the Wasserstein distance) to $\mathcal{G}$, the space of non-decreasing functions from the unit interval to itself, the above measure may be considered as a probability measure on this path space. In analogy to Feynman's construction of the Wiener measure on the space of all continuous functions from the unit interval to the real numbers two differences appear:
Firstly Feynman chooses as discretized Hamiltonian the free energy of a path, the anology in [SvR09] fixes a finite partition $0 = t_0 < t_1 < \cdots < t_n = 1$ and define the discretized Hamiltonian for $g \in \mathcal{G}$ by

$$H(g) = -\sum_i \log \frac{g_{t_i} - g_{t_{i-1}}}{t_i - t_{i-1}}(t_i - t_{i-1}).$$

From optimal transport one learns that a solution to the heat equation is also solution to a gradient flow equation on the space of probability measures and the functional for which the heat flow realizes its steepest descent is the Boltzmann entropy. Since [SvR09] want to construct a Wasserstein diffusion as a stochastically perturbed heat flow the choice of the Boltzmann entropy for the Hamiltonian reveals to be the right one.
Secondly Feynman chooses as reference measure for the finite-dimensional distribution of his Gibbs type measure on the path space the uniform distribution. This finite-dimensional distributions constitute a family of consistent probability measures and by Kolmogorov's extension theorem Feynman can show that the limiting measures equals the Wiener measure. In analogy [SvR09] choose

$$q_n(dx_1, \ldots, dx_n) = C^n \frac{dx_1 \ldots dx_n}{x_1(x_2 - x_1)\ldots(x_n - x_{n-1})(1 - x_n)}$$

as finite-dimensional reference measure. This measure (it is not a probability measure!) turns out to be the only one on $\mathcal{G}$ which is invariant under rescaling of any subset of the partition $\{x_k < \cdots < x_l\}$ by $x \mapsto (x_l - x_k)x + x_k$ and which has continuous density.
Combining the two ingredients one obtains as consistent family of finite-dimensional distribution the Dirichlet-Poisson measure $\mathbb{Q}^\beta(g_{t_1} \in dx_1, \ldots, g_{t_n} \in dx_n)$ which equals

$$\frac{1}{Z_{\beta,n}} \prod_{i=1}^{n+1} (x_i - x_{i-1})^{\beta(t_i - t_{i-1})} \frac{dx_1, \ldots, x_n}{x_1(x_2 - x_1)\ldots(x_n - x_{n-1})(1 - x_n)}$$

and the measure $\mathbb{P}^\beta$ is defined as the push-forward of $\mathbb{Q}^\beta$ under the isometry between $\mathcal{G}$ and P. The Dirichlet form built with this measure gives a generator $\mathbb{L}^\beta$ of a continuous Markov process with the Wasserstein distance as intrinsic metric for $\beta > 0$. Define

$$\mathfrak{Z}_1 = \left\{ F(\mu) \equiv \Phi\left(\langle \underline{f}, \mu \rangle\right); \Phi \in \mathcal{C}^2(\mathbb{R}^d),\ \ \underline{f} = (f_1, \ldots, f_d) \in \mathcal{C}^2([0,1]; \mathbb{R}^d); f_i'(0) = f_i'(1) = 0 \right\}.$$

For $F \in \mathfrak{Z}_1$ the generator $\mathbb{L}^\beta$ equals

$$\mathbb{L}^\beta F = \mathbb{L}_1 F + \mathbb{L}_2 F + \beta \mathbb{L}_3 F,$$

with

$$\mathbb{L}_1 F = \sum_{i,j=1}^{d} \partial_i \partial_j \Phi(\int \underline{f} d\mu) \int f_i' f_j' d\mu$$

$$\mathbb{L}_2 F = \partial_i \Phi(\int \underline{f} d\mu) \left(\sum_{gaps(\mu)} \left[\frac{f_i''(I_-) + f_i''(I_+)}{2} - \frac{f_i'(I_-) - f_i'(I_+)}{|I|} \right] - \frac{f_i''(0) + f_i''(0)}{2} \right)$$

$$\mathbb{L}_3 F = \sum_{i=1}^{d} \partial_i \Phi(\int \underline{f} d\mu) \int f_i'' d\mu$$

gaps denotes the set of intervals $I = (I_-, I_+) \subset [0,1]$ of maximal length with $\mu(I) = 0$ and $|I|$ denotes the length of such an interval. From this it follows that the regularized Wasserstein Laplacian is equal to the generator of Sturm-von Renesse's Wasserstein diffusion with inverse temperatur $\beta = 0$ and periodic boundary conditions.

3.3 Zeta function regularized stochastic flows on the torus

Let $\{B_t^k, k \in \mathbb{N}\}$ be a family of independent Brownian motions on $\mathbb{R}$ and s a positive real number. For every $(\mu, x, t) \in \mathrm{P}^\infty \times \mathbb{T}^1 \times \mathbb{R}_+$ we define a random field (compare to Proposition 3.2.1):

$$F(\mu, x, t) \quad := \quad \frac{2\pi\sqrt{2}}{\sqrt{\mu(x)}} B_t^1 + \sum_{k=1}^{\infty} \frac{\sqrt{2} 2\pi k^{-s}}{\sqrt{\mu(x)}} \left\{ B_t^{2k} \sin 2\pi k x + B_t^{2k+1} \cos 2\pi k x \right\}.$$

The process $t \mapsto F(\mu(.), ., t)$ is a continuous local martingale with values in $\mathcal{C}^j(\mathbb{T}^1; \mathbb{R})$ for $j < s - 1$.

The quadratic variation $V^s(\mu, x, y, t)$ of $F(\mu, x, t) - F(\mu, y, t)$ equals

$$\begin{aligned}
& \mathbb{E}\left[(F(\mu,x,t)-F(\mu,y,t))^2\right] = \\
= \;& t2(2\pi)^2\left(\frac{1}{\sqrt{\mu(x)}}-\frac{1}{\sqrt{\mu(y)}}\right)^2 + \\
+ \;& t\sum_{k=1}^{\infty} 2(2\pi)^2 k^{-2s}\mathbb{E}\left[\frac{1}{\sqrt{\mu(x)}}\left\{B_t^{2k}\sin 2\pi kx + B_t^{2k+1}\cos 2\pi kx\right\} - \right. \\
& \left. \frac{1}{\sqrt{\mu(y)}}\left\{B_t^{2k}\sin 2\pi ky + B_t^{2k+1}\cos 2\pi ky\right\}\right]^2 \\
= \;& t2(2\pi)^2\left(\frac{1}{\sqrt{\mu(x)}}-\frac{1}{\sqrt{\mu(y)}}\right)^2 + t2(2\pi)^2\sum_{k=1}^{\infty}k^{-2s}\left\{\frac{1}{\mu(x)}+\frac{1}{\mu(y)} - \right. \\
- \;& \left. 2\frac{1}{\sqrt{\mu(x)}}\frac{1}{\sqrt{\mu(y)}}\left(\sin 2\pi kx\sin 2\pi ky + \cos 2\pi kx\cos 2\pi ky\right)\right\} \\
= \;& t2(2\pi)^2\left(\frac{1}{\sqrt{\mu(x)}}-\frac{1}{\sqrt{\mu(y)}}\right)^2 + t2(2\pi)^2\sum_{k=1}^{\infty}k^{-2s}\left\{\frac{1}{\mu(x)}+\frac{1}{\mu(y)} - \right. \\
- \;& \left. 2\frac{1}{\sqrt{\mu(x)}}\frac{1}{\sqrt{\mu(y)}}\left(1-2\sin^2(2\pi k\frac{x-y}{2})\right)\right\}
\end{aligned}$$

which is finite and Lipschitz continuous for $s > 3/2$ (since the densities are smooth and positive). As [Fan02] remarks one can construct with this random fields (with $\mu \equiv 1$ or not) Brownian motion on $\mathrm{Diff}^1(\mathbb{T}^1)$ by using the theory of stochastic flows (see [Bax84] and [Kun90]). The border case $s = 3/2$ could be treated by looking into [Fan02], where it is proven that there exists a constant $c > 0$ such that for all $0 < \theta := x - y \leq \frac{1}{2}$

$$\sum_{k=1}^{\infty} k^{-2s}\sin^2(2\pi k\frac{\theta}{2}) \leq c\theta^2\log\theta.$$

As was already shown in [Mal99] by means of heat kernel regularization (and for $\mu \equiv 1$) the random field F in the case of $s = 3/2$ gives rise to Brownian motion on the space of homeomorphisms of $\mathbb{T}^1$. The parameter $s = 1$ would correspond to the metric used in Wasserstein geometry; but this case cannot be handled by the regularization techniques used in [Mal99].
Let us consider the analytic continuation of the quadratic variation $s \mapsto V^s(\mu, x, y, t)$ to the complex plane:

Note that

$$\sin^2 2\pi kx = \frac{2-\exp(4\pi ikx)-\exp(-4\pi ikx)}{4}.$$

The polylogarithm

$$s \mapsto \mathrm{Li}_s(z) := \sum_{k=1}^{\infty}\frac{z^k}{k^s} \qquad |z| < 1$$

has an analytic continuation (in s) to the complex plane and satisfies the following well-known identity (see [AS70] and [Apo76])

$$\mathrm{Li}_n(e^{2\pi ix}) + (-1)^n \mathrm{Li}_n(e^{-2\pi ix}) = -\frac{(2\pi i)^n}{n!} B_n(x)$$

for $x \in \mathbb{R}$, $n = 0, 1, 2, \dots$ and $B_n(x)$ being the Bernoulli polynomial; in particular

$$\mathrm{Li}_0(e^{2\pi ix}) + \mathrm{Li}_0(e^{-2\pi ix}) = -1$$

$$\mathrm{Li}_2(e^{2\pi ix}) + \mathrm{Li}_2(e^{-2\pi ix}) = 2\pi^2 \left(x^2 - x + \frac{1}{6} \right)$$

which entails

$$\lim_{s\to 0} \sum_{k=1}^{\infty} \frac{\sin^2(2\pi k \frac{\theta}{2})}{k^s} = \frac{1}{2}\zeta(0) + \frac{1}{4} = 0.$$

$$\lim_{s\to 2} \sum_{k=1}^{\infty} \frac{\sin^2(2\pi k \frac{\theta}{2})}{k^s} = \frac{1}{2}\zeta(2) - \frac{1}{4}2\pi^2 \left(\theta^2 - \theta + \frac{1}{6} \right) = \frac{1}{2}\frac{\pi^2}{6} - \frac{1}{4}2\pi^2 \left(\theta^2 - \theta + \frac{1}{6} \right).$$

For the analytic continuation of the infinitesimal covariance $a_s(\mu, x, y) := \lim_{t\searrow 0}(V(\mu, x, y, t)/t)$ at $s \in \mathbb{C}$ we define $a_0(\mu, x, y, t) = a_s(\mu, x, y)|_{s=0}$ resp. $a_2(\mu, x, y, t) = as(\mu, x, y)|_{s=2}$ and conclude

$$a_0(\mu, x, y) \quad = \quad (2\pi)^2 \left(\frac{1}{\sqrt{\mu(x)}} - \frac{1}{\sqrt{\mu(y)}} \right)^2$$

Note that $a_0(\mu, x, x) = 0$. Resp.

$$a_2(\mu, x, y) = ((2\pi)^2 + (2\pi)^2 \frac{\pi^2}{6}) \left(\frac{1}{\sqrt{\mu(x)}} - \frac{1}{\sqrt{\mu(y)}} \right)^2 + 8(2\pi)^2 (-\frac{2\pi^2}{4} \left(\theta^2 - \theta \right))$$

Unfortunately both $a_1(\mu, x, y)$ and $a_2(\mu, x, y)$ are not positive definite for any fixed $x, y \in \mathbb{T}^1$.

Definition 3.3.1. *Given the random field F as above with $s > 3/2$ (for fixed $\mu \in P^\infty(\mathbb{T}^1)$ it is a random field with local characteristics $(0, a)$ in the sense of [Kun90]), we define a generalized Kunita stochastic differential equation with interaction to be a process*

$$t \mapsto (\varphi_t, \mu_t) \in \mathit{Diff}^\infty(\mathbb{T}^1) \times P^\infty(\mathbb{T}^1)$$

which is a simultaneous solution of

$$d\varphi_t = F(\mu_t, \varphi_t, dt) \tag{3.2}$$

$$\mu_t = \varphi_t \# \mu_0. \tag{3.3}$$

The first equation is a Kunita SDE with values in $\mathit{Diff}^\infty(\mathbb{T}^1)$ and the solution of the second equation is a $P^\infty(\mathbb{T}^1)$-valued stochastic process.

We say that the process $t \mapsto \mu_t$ solves (3.2) and (3.3) iff the process $t \mapsto \mu_t$ is a fix point of the mapping $\kappa : \mathcal{C}(\mathbb{R}_+; P^\infty(\mathbb{T}^1)) \mapsto \mathcal{C}(\mathbb{R}_+; P^\infty(\mathbb{T}^1))$, where $(\kappa\mu)_t := \varphi_t \# \mu_0$ for φ_t solution to $d\varphi_t = F(\mu_t, \varphi_t, dt)$.

Proposition 3.3.1. *Let $t \mapsto \mu_t$ be a solution to (3.2) and (3.3) with $\mu_0 \in P^\infty(\mathbb{T}^1)$. Then $t \mapsto \mu_t$ is a diffusion process on $P^\infty(\mathbb{T}^1)$ with generator*

$$\begin{aligned}\mathcal{L}(\Psi(\langle f,\mu_0\rangle)) &= \frac{1}{2}\Psi''(\langle f,\mu_0\rangle)\int_{\mathbb{T}^1}\int_{\mathbb{T}^1} a_s(\mu,x,y)\nabla f(x)\nabla f(y)\mu_0(dx)\mu_0(dy)\\ &\quad +\frac{1}{2}\Psi'(\langle f,\mu_0\rangle)\int_{\mathbb{T}^1} a_s(\mu,x,x)\Delta f(x)\mu_0(dx)\end{aligned}$$

for $s > 3$.

Proof. Use the generalized Ito formula in [Kun90] p92f. □

3.4 Renormalized Laplacian on $\mathbf{P}^\infty(\mathbb{T}^d)$

Recall ([Mal08]) that the topological dual of $\mathbb{T}^d$ is $\mathbb{Z}^d$, where the coupling between $k \in \mathbb{Z}^d$ and $x \in \mathbb{T}^d$ is given by $(k,x) := \exp(i\langle k,x\rangle)$, with $\langle k,x\rangle = \sum_{i=1}^d k_i x_i$. The Fourier transform of a complex function on $\mathbb{T}^d$ is given by $\hat{f}(k) = \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} f(x)(-k,x)\,dx$ and any $f \in L^2(\frac{1}{(2\pi)^d}\mathrm{vol})$ can be written as $f(x) = \sum_{k\in\mathbb{Z}^d}\hat{f}(k)(k,x)$. The function f is real if and only if $\hat{f}(-k) = \bar{\hat{f}}(k)$. Denote $\tilde{\mathbb{Z}}^d \subset \mathbb{Z}^d$ such that each equivalence class of the equivalence relation defined by $k \sim k'$ if $k = -k'$ has a unique representative in $\tilde{\mathbb{Z}}^d$. Note that in contrast to the one-dimensional case the choice of $\tilde{\mathbb{Z}}^d$ is not unique. The Fourier expansion for real valued functions then reads $f(x) = 2\sum_{k\in\tilde{\mathbb{Z}}^d}\Re(\hat{f}(k))\cos\langle k,x\rangle - \Im(\hat{f}(k))\sin\langle k,x\rangle$.

Proposition 3.4.1. *Given a functional $E : P^\infty(\mathbb{T}^d) \to \mathbb{R}$ of the type*

$$E(\mu) = \int_{\mathbb{T}^d} e(\mu(x))vol(dx),$$

where $e : \mathbb{R}_+ \to \mathbb{R}$ is $\mathcal{C}^3$. For an orthonormal system $\{e_k(\mu),\bar{e}_k(\mu)\}_{k\in\tilde{\mathbb{Z}}^d}$ of $T_\mu P^\infty(\mathbb{T}^d)$ we define an operator A on $T_\mu P^\infty(\mathbb{T}^d)$ by diagonalization in its basis:

$$A : e_k(\mu) \mapsto |k|^{-a}e_k(\mu);\ k\in\tilde{\mathbb{Z}}^d,\ a > \frac{3}{2}+d,\ \ d>1$$

Let $\widetilde{HessE}$ be the Hessian operator associated to the the (variational) Hessian $Hess^{var}E(.,.)(\mu)$. The **renormalized Wasserstein Laplacian** *in an open neighbourhood of μ as defined below is finite:*

$$\Delta^a_{P^\infty(\mathbb{T}^d)}E(\mu) := \sum_{k\in\tilde{\mathbb{Z}}^d}\langle\widetilde{HessE}Ae_k(\mu),Ae_k(\mu)\rangle_\mu < \infty$$

Proof. We define an orthonormal system of $\mathcal{C}^\infty(\mathbb{T}^d)$ with respect to the inner product $H^1(\frac{1}{(2\pi)^d}dx)$ by

$$\begin{cases} e_k(x) := 2^{d/2}\,|k|^{-1}\sin\langle 2\pi k,x\rangle, & k\in\tilde{\mathbb{Z}}^d\setminus\{|k|\le 1\}\\ \bar{e}_k(x) := 2^{d/2}\,|k|^{-1}\cos\langle 2\pi k,x\rangle, & k\in\tilde{\mathbb{Z}}^d\setminus\{|k|\le 1\}\end{cases}$$

and define smooth functions $e_k(\mu)$ resp. $\bar{e}_k(\mu)$ by integration

$$\begin{cases} \nabla e_k(\mu)(x) = \frac{1}{\sqrt{\mu}}\nabla\, e_k(x) & k\in\tilde{\mathbb{Z}}^d\setminus\{|k|\le 1\}\\ \nabla \bar{e}_k(\mu)(x) = \frac{1}{\sqrt{\mu}}\nabla\, \bar{e}_k(x) & k\in\tilde{\mathbb{Z}}^d\setminus\{|k|\le 1\}\end{cases}$$

with inital data
$$\begin{cases} e_k(\mu)(\mathbf{0}) = 0, & k \in \tilde{\mathbb{Z}}^d \setminus \{|k| \le 1\} \\ \bar{e}_k(\mu)(\mathbf{0}) = 0, & k \in \tilde{\mathbb{Z}}^d \setminus \{|k| \le 1\} \end{cases}$$
The ingredients to calculate for the Hessian formula are again $\Gamma_2(e_k)$ and $(\Delta_{\mathbb{T}^d}(e_k))^2$.

$$\begin{aligned}
\Gamma_2(e_k(\mu)) &= \sum_{ij} (\partial_i \partial_j e_k(\mu))^2 \\
&= \sum_{i,j} \frac{2^d (2\pi)^2 k_j^2}{|k|^2} \left\{ \frac{(\partial_i \mu)^2}{4c^3} \cos^2\langle 2\pi k, x\rangle + \frac{2\pi k_i \partial_i \mu}{\mu^2} \sin\langle 2\pi k, x\rangle \cos\langle 2\pi k, x\rangle + \right. \\
&\quad \left. \frac{(2\pi k_i)^2}{\mu} \sin^2\langle 2\pi k, x\rangle \right\} \\
&= 2^d (2\pi)^2 \left\{ \frac{\langle \nabla\mu, \nabla\mu\rangle}{4\mu^3} \cos^2\langle 2\pi k, x\rangle + \right. \\
&\quad \left. \frac{2\pi\langle k, \nabla\mu\rangle}{\mu^2} \sin\langle 2\pi k, x\rangle \cos\langle 2\pi k, x\rangle + \frac{(2\pi)^2 |k|^2}{\mu} \sin^2\langle 2\pi k, x\rangle \right\}
\end{aligned}$$

and

$$\begin{aligned}
(\Delta_{\mathbb{T}^d}(e_k(\mu)))^2 &= \frac{2^d (2\pi)^2}{|k|^2} \left\{ \frac{|\langle \nabla\mu, k\rangle|^2}{4\mu^3} \cos^2\langle 2\pi k, x\rangle + \right. \\
&\quad \left. \frac{2\pi \langle \nabla\mu, k\rangle |k|^2}{\mu^2} \sin\langle 2\pi k, x\rangle \cos\langle 2\pi k, x\rangle + \frac{(2\pi)^2 |k|^4}{\mu} \sin^2\langle 2\pi k, x\rangle \right\} \\
&\le \Gamma_2(e_k(\mu))
\end{aligned}$$

by Cauchy-Schwarz. Given a distribution $E_k \in T\mathrm{P}^\infty(\mathbb{T}^d)$ such that $(E_k(\mu)|\varphi) = \int_{\mathbb{T}^d} \langle \nabla e_k(\mu), \nabla\varphi\rangle \mu$ for smooth, compactly supported functions φ we have

$$\begin{aligned}
\mathrm{Hess}^{\mathrm{var}} E(E_k, E_k)(\mu) &= \int_{\mathbb{T}^d} \Gamma_2(e_k(\mu))) \{\mu e'(\mu) - e(\mu)\} \mathrm{vol} + \\
&\quad \int_{\mathbb{T}^d} (\Delta(e_k(\mu)))^2 \{\mu^2 e''(\mu) - \mu e'(\mu) + e(\mu)\} \mathrm{vol} \\
&\le \int_{\mathbb{T}^d} \Gamma_2(e_k(\mu)) \mu^2 e''(\mu) \mathrm{vol} \\
&= 2^d (2\pi)^4 |k|^2 \int_{\mathbb{T}^d} \sin^2\langle 2\pi k, x\rangle \mu e''(\mu) \mathrm{vol} + \\
&\quad \frac{2^d (2\pi)^2}{4} \int_{\mathbb{T}^d} |\mathrm{grad} \log\mu|^2 \cos^2\langle 2\pi k, x\rangle \mu\, e''(\mu) \mathrm{vol} + \\
&\quad 2^d (2\pi)^3 \int_{\mathbb{T}^d} \langle \mathrm{grad} \log\mu, k\rangle \sin\langle 2\pi k, x\rangle \cos\langle 2\pi k, x\rangle \mu e''(\mu) \mathrm{vol} \\
&\le |k|^2 (2\pi)^4 ||\mu e''(\mu)||_\infty + \frac{(2\pi)^2}{4} ||\mu e''(\mu) |\nabla \log\mu|^2||_\infty \\
&< +\infty
\end{aligned}$$

if we are able to control

$$||\mu e''(\mu) |\mathrm{grad} \log\mu|^2||_\infty < +\infty \tag{3.4}$$
$$||\mu e''(\mu) |\mathrm{grad} \log\mu|\, ||_\infty < +\infty \tag{3.5}$$
$$||\mu e''(\mu)||_\infty < +\infty, \tag{3.6}$$

which is the case since $\mu \in \mathrm{P}^\infty(\mathbb{T}^d)$. Passing from $e_k(\mu)$ to $\bar{e}_k(\mu)$ we remark that $\Gamma_2(\bar{e}_k(\mu))$ is obtained by exchanging sin and cos and by a change of sign of the the term involving both sin and cos in the formulae of $\Gamma_2(e_k(\mu))$ and $(\Delta(e_k(\mu)))^2$. Again we have $(\Delta(\bar{e}_k(\mu)))^2 \leq \Gamma_2(\bar{e}_k(\mu))$ and we can prove finiteness of $\mathrm{Hess}^{\mathrm{var}} E(\bar{E}_k, \bar{E}_k)$ with the same arguments.
In order to show finiteness of the renormalized Wasserstein Laplacian we introduce a d-dimensional analogue of the zeta function: For $d = 2, 3, \ldots$ and $s \in \mathbb{R}$

$$\zeta_d(s) := \sum_{k \in \tilde{\mathbb{Z}}^d \setminus \{|k| \leq 1\}} \frac{1}{|k|^s}$$

and for $j \in \{2, 3, 4, \ldots\}$ we estimate very roughly

$$h_j := \# \left\{ k \in \tilde{\mathbb{Z}}^d ; j - 1 < |k| \leq j \right\} \leq j^{2d}.$$

Consequently

$$\zeta_d(s) < \sum_{j=2}^{\infty} \frac{h_j}{j^s} \leq \zeta(s - 2d),$$

and so

$$\begin{aligned} &\Delta^a_{\mathrm{P}^\infty(\mathbb{T}^d)} E(\mu) \leq \\ &\qquad \sum_{k \in \tilde{\mathbb{Z}}^d \setminus \{|k| \leq 1\}} |k|^{-2a+2} (2\pi)^4 ||\mu e''(\mu)||_\infty + \frac{(2\pi)^2}{4} ||\mu e''(\mu) |\nabla \log \mu|^2||_\infty \\ &= \zeta_d(2a - 2)(2\pi)^4 ||\mu e''(\mu)||_\infty + \frac{(2\pi)^2}{4} ||\mu e''(\mu) |\nabla \log \mu|^2||_\infty \\ &< \zeta(2a - 2 - 2d) 2^d (2\pi)^4 ||\mu e''(\mu)||_\infty + \frac{(2\pi)^2}{4} ||\mu e''(\mu) |\nabla \log \mu|^2||_\infty \\ &< \infty \end{aligned}$$

if

$$2a - 2d - 2 > 1 \Leftrightarrow a > (3 + 2d)/2.$$

□

Remark 3.4.1. *Estimating the number $N(r)$ of lattice points in $\mathbb{Z}^2$ inside the boundary of a circle with given radius r is known to number theorists as Gauss's circle problem. Gauss showed that*

$$N(r) = \pi r^2 + E(r)$$

with $|E(r)| \leq 2\sqrt{2}\pi r$. Today's best known bounds ([Hux03]) for the error term are

$$|E(r)| \leq C r^\theta$$

with

$$\frac{1}{2} < \theta \leq \frac{131}{208}.$$

In the case of $d = 2$ we can obtain a better upper bound for $\zeta_d(s)$ as in the above proposition:

$$\begin{aligned} h_j &:= \#\left\{k \in \tilde{\mathbb{Z}}^2; j-1 < |k| \le j\right\} \\ &= \frac{1}{2}\left\{N(j) - N(j-1)\right\} \\ &\le \frac{1}{2}\left\{(2j-1)\pi + C\max\left\{j^{\frac{1}{2}} - (j-1)^{\frac{131}{208}}, j^{\frac{131}{208}} - (j-1)^{\frac{1}{2}}\right\}\right\} \\ &\le \frac{1}{2}\left\{(2j-1)\pi + Cj\right\} = C'j + C'' \end{aligned}$$

Consequently

$$\begin{aligned} &\Delta^a_{P^\infty(\mathbb{T}^2)} E(\mu) \le \\ &\quad \sum_{k \in \tilde{\mathbb{Z}}^2 \setminus \{|k| \le 1\}} |k|^{-2a+2}(2\pi)^4 ||\mu e''(\mu)||_\infty + \frac{(2\pi)^2}{4}||\mu e''(\mu)|\nabla \log \mu|^2||_\infty \\ &\le (2\pi)^4 ||\mu e''(\mu)||_\infty (C'\zeta(2a-3) + C''\zeta(2a-2))\frac{(2\pi)^2}{4}||\mu e''(\mu)|\nabla \log \mu|^2||_\infty \\ &< \infty \end{aligned}$$

Remark 3.4.2 (Zeta function as Dirichlet series with respect to spectral representations of operators)**.** *Given an operator in its (purely discrete) spectral representation $A = \sum_{n\in\mathbb{N}} \lambda_n P_n$ with eigenvalues $\{\lambda_n\}$ having multiplicities $\{g_n\}$ and projection operators P_n one defines the associated zeta function by*

$$\zeta_A(s) := \sum_{n \in \mathbb{N}} \frac{g_n}{\lambda_n^s}.$$

For the above mentioned operator

$$A : e_k(\mu) \mapsto |k|^{-a} e_k(\mu);\ k \in \tilde{\mathbb{Z}}^d$$

each eigenvalue $|k|$ has by definition multiplicity $g_{|k|} = \#\{j \in \tilde{\mathbb{Z}}^d : |j| = |k|\}$ and hence

$$\zeta_d(s) = \zeta_A(s)$$

Open question 3.4.1 (Exact calculation of the Laplacian by measured zeta functions)**.** *By definition*

$$\begin{aligned} &\Delta^a_{P^\infty(\mathbb{T}^d)} E(\mu) = \\ &\quad \sum_{k \in \tilde{\mathbb{Z}}^d \setminus \{|k| \le 1\}} |k|^{-2a}\left\{Hess^{var} E(E_k, E_k)(\mu) + Hess^{var} E(\bar{E}_k, \bar{E}_k)(\mu)\right\} \\ &= \sum_{k \in \tilde{\mathbb{Z}}^d \setminus \{|k| \le 1\}} |k|^{-2a} \int_{\mathbb{T}^d} \{\mu e'(\mu) - e(\mu)\} \frac{2^d(2\pi)^2}{4\mu}\left\{|\nabla \log \mu|^2 - \left(\frac{\langle \nabla \log \mu, k\rangle}{|k|}\right)^2\right\} vol \\ &\quad + \sum_{k \in \tilde{\mathbb{Z}}^d \setminus \{|k| \le 1\}} |k|^{-2a} \int_{\mathbb{T}^d} \mu^2 e''(\mu) \frac{2^d(2\pi)^2}{\mu}\left\{\frac{|\langle \nabla \log \mu, k\rangle|^2}{4|k|^2} + 2^d(2\pi)^2|k|^2\right\} vol \end{aligned}$$

Hence we need explicit expressions for the term called measured zeta function:

$$\zeta_d(s,E,\mu) = \sum_{k\in\tilde{\mathbb{Z}}^d\setminus\{|k|\leq 1\}} |k|^{-2s}\int_{\mathbb{T}^d} F(e,\mu)\left\{\frac{\langle\nabla\log\mu, k\rangle}{|k|}\right\}^2 vol$$

for

$$F(e,\mu) = \frac{1}{\mu}\{e(\mu) - \mu e'(\mu) + \mu^2 e''(\mu)\}.$$

In the case of the Boltzmann entropy functional $F = 0$ *and:*

$$\begin{aligned}\Delta^a_{P^\infty(\mathbb{T}^d)} Ent(\mu) &= 2^{d-2}(2\pi)^2\zeta_d(2a)||\nabla\log\mu||^2_{L^2(vol)}\\ &+2^{2d}(2\pi)^4\zeta_d(2a-2)\end{aligned}$$

Chapter 4

Approximation of a Wasserstein diffusion

4.1 Riemannian metrics on the space of box-type measures

Definition 4.1.1 (Box-type measures)**.** *Fix $n \in \mathbb{N}$. We denote the space of all sequence $\underline{s} = (s_i)_{i=0}^n$ with $0 =: s_0 \leq s_1 \leq s_2 \dots s_n \leq s_n := 1$ by $\mathcal{S}_n$. The set of sequences in $\mathcal{S}_n$ which are strictly increasing is denoted by $\overset{\circ}{\mathcal{S}}_n$.*
We write $\Sigma_{n-1} := \{\underline{\lambda} \in \mathbb{R}^n : \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1\}$.
For $x \in [0,1]$ we define probability measures

$$m(\underline{s})(x) := \sum_{i=1}^n \left(\frac{1}{n(s_i - s_{i-1})} \mathbf{1}_{\{s_i - s_{i-1} > 0\}} \mathbf{1}_{[s_{i-1}, s_i)}(x) dx + \frac{1}{n} \mathbf{1}_{\{s_i - s_{i-1} = 0\}} \delta_{s_i}(x) \right).$$

We write

$$G_n := m(\mathcal{S}_n) \qquad \textit{resp.} \qquad \overset{\circ}{G}_n := m(\overset{\circ}{\mathcal{S}}_n)$$

for the space of box-type measures with n boxes. Both spaces are dense in P with respect to the quadratic Wasserstein distance. $\overset{\circ}{G}_n$ is a totally geodesic subspace of P equipped with the Wasserstein distance. We may consider the bijection $\mathcal{S}_n \simeq \Sigma_{n-1}$ given by $s_i - s_{i-1} \mapsto \lambda_i$ which in turn let us associate to every element $\underline{\lambda} \in \Sigma_{n-1}$ a probability measure

$$m(\underline{\lambda}) := \sum_{i=1}^n \left(\frac{1}{n\lambda_i} \mathbf{1}_{\{\lambda_i > 0\}} \mathbf{1}_{[\sum_{k=1}^{i-1} \lambda_k, \sum_{k=1}^{i} \lambda_k)}(x) dx + \frac{1}{n} \mathbf{1}_{\{\lambda_i = 0\}} \delta_{\sum_{k=1}^{i} \lambda_k}(x) \right).$$

Lemma 4.1.1. *G_n is geodesically convex.*

Proof. Each measure $m(\underline{\lambda}) \in G_n$ can be written in terms of quantile functions q in the following way

$$q(x) := \inf\{t \in [0,1] : \int_0^t m(\underline{\lambda}) > x\},$$

i.e.

$$q(x) := \sum_{i\geq 1} g_i(x)\mathbf{1}_{[\frac{i-1}{n},\frac{i}{n})}(x)dx$$

where

$$g_i(x) = n\lambda_i x + \sum_{j=1}^{i} \lambda_j - i\lambda_i$$

For two given quantile functions q^1 and q^2 (with weights $\underline{\lambda}$ and $\underline{\tilde{\lambda}}$) the Wasserstein geodesic γ_t linking q^1 to q^2 is given by

$$\gamma_t := (1-t)q^1 + tq^2.$$

(see Theorem 7.2.8 in [AGS08]).

But for fixed $n \in \mathbb{N}$ the curve γ_t can again be written as

$$\gamma_t(x) := \sum_{i\geq 1} f_i(t,x)\mathbf{1}_{[\frac{i-1}{n},\frac{i}{n})}(x)dx$$

where

$$\begin{aligned} f_i(t,x) &= n((1-t)\lambda_i + t\tilde{\lambda}_i)x + \sum_{j=1}^{i}((1-t)(\lambda_j - i\lambda_i) + t(\tilde{\lambda}_j - i\tilde{\lambda}_i)) \\ &= n((1-t)\lambda_i + t\tilde{\lambda}_i)x + \sum_{j=1}^{i}((1-t)\lambda_j + t\tilde{\lambda}_j) - i((1-t)\lambda_i + t\tilde{\lambda}_i), \end{aligned}$$

i.e. for every $t \in [0,1]$ we obtain a box-type measure with weight

$$(1-t)\underline{\lambda} + t\underline{\tilde{\lambda}}.$$

□

Lemma 4.1.2 (Wasserstein distance between two box-type measures.)**.** *Given two measures $\mu = m(\underline{\lambda})$ and $\tilde{\mu} = m(\underline{\tilde{\lambda}})$ for $\underline{\lambda}, \underline{\tilde{\lambda}} \in \Sigma_{n-1}$ then*

$$d_W(\mu,\tilde{\mu})^2 = \frac{1}{3n}||\underline{\lambda} - \underline{\tilde{\lambda}}||^2_{\mathbb{R}^n} + \frac{1}{n}\sum_{k=1}^{n}\sum_{i\leq k; j\leq k-1}(\lambda_i - \tilde{\lambda}_i)(\lambda_j - \tilde{\lambda}_j).$$

Proof. For probability measures on the unit interval the quadratic Wasserstein distance equals the L^2 distance of the respective quantile functions, hence

$$d_W(\mu,\tilde{\mu})^2 = \sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left(n(\lambda_k - \tilde{\lambda}_k)x + \sum_{j=1}^{k}\lambda_j - \tilde{\lambda}_j - k(\lambda_k - \tilde{\lambda}_k)\right)^2 dx.$$

But the integrand equals

$$\frac{(\frac{k}{n})^3-(\frac{k-1}{n})^3}{3}n^2(\lambda_k-\tilde{\lambda}_k)^2+((\frac{k}{n})^2-(\frac{k-1}{n})^2)n(\lambda_k-\tilde{\lambda}_k)\left(\sum_{j=1}^{k}\lambda_j-\tilde{\lambda}_j-k(\lambda_k-\tilde{\lambda}_k)\right)+$$

$$+\frac{1}{n}\left(\sum_{j=1}^{k}\lambda_j-\tilde{\lambda}_j-k(\lambda_k-\tilde{\lambda}_k\right)^2=$$

$$=\frac{3k^2-3k+1}{3n}(\lambda_k-\tilde{\lambda}_k)^2+\frac{2k-1}{n}(\lambda_k-\tilde{\lambda}_k)\left(\sum_{j=1}^{k}(\lambda_j-\tilde{\lambda}_j)-k(\lambda_k-\tilde{\lambda}_k)\right)+$$

$$+\frac{1}{n}\left(\sum_{j=1}^{k}\lambda_j-\tilde{\lambda}_j-k(\lambda_k-\tilde{\lambda}_k\right)^2=$$

$$=\frac{1}{3n}(\lambda_k-\tilde{\lambda}_k)^2-\frac{\lambda_k-\tilde{\lambda}_k}{n}\sum_{j=1}^{k}(\lambda_j-\tilde{\lambda}_j)+\frac{1}{n}\left(\sum_{j=1}^{k}(\lambda_j-\tilde{\lambda}_j)\right)^2$$

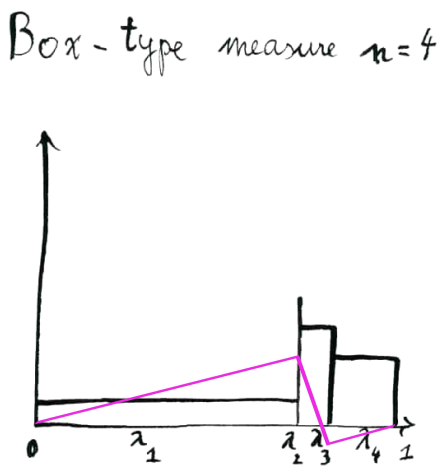
which entails the result. □

Definition 4.1.2 (Tangent vectors)**.** *For $i=0,\dots,n+1$ we are given $V_i\in\mathbb{R}$ such that $V_0=0$ and $V_n=0$. Consider a sequence $\underline{s}\in\mathcal{S}_n$ resp. $\underline{\lambda}\in\Sigma_{n-1}$*

$$V_{\underline{s}}(x) := \sum_{i=1}^{n}\left(\frac{s_i-x}{s_i-s_{i-1}}V_{i-1}+\frac{x-s_{i-1}}{s_i-s_{i-1}}V_i\right)\mathbf{1}_{[s_{i-1},s_i)}(x)$$

$$V_{\underline{\lambda}}(x) := \sum_{i=1}^{n+1}\left(\frac{\sum_{j=1}^{i}\lambda_j-x}{\lambda_i}V_{i-1}+\frac{x-\sum_{j=1}^{i-1}\lambda_j}{\lambda_i}V_i\right)\mathbf{1}_{[\sum_{j=1}^{i-1}\lambda_j,\sum_{j=1}^{i}\lambda_j)}(x)$$

and set $V_{\underline{\lambda}}(1):=0$.



The heuristics behind this definition of tangent vectors is as follows: Each box-type measures has a quantile function that is piecewise linear on each of the *intervals $[i/n,(i+1)/n)$. Displacement interpolation between two box-type measures by Wasserstein geodesics translates to quantile function as convex combination of two functions which are each piecewise linear on intervals $[i/n,(i+1)/n)$. Consequently moving mass of box shape according to optimal transportation consists in moving the parameters λ_i horizontally (as in the above graphics) with the only constraint that no mass should be moved on the boundary.*

Define the tangent space

$$T_{\mu(\underline{\lambda})}G_n = \{V_{\underline{\lambda}}(x); V_i \in \mathbb{R}, i = 0,\dots,n\}$$

Lemma 4.1.3. *With the above definition we obtain a Riemannian metric on the $n-1$-simplex:*

Proof.

$$\begin{aligned}
||V_{\underline{s}}(x)||^2_{dx} &= \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} V_{\underline{s}}(x)^2\,dx \\
&= \sum_{i=1}^{n} \int_0^{s_i-s_{i-1}} \left(\frac{s_i-s_{i-1}-y}{s_i-s_{i-1}}V_{i-1} + \frac{y}{s_i-s_{i-1}}V_i\right)^2 dy \\
&= \sum_{i=1}^{n} \frac{1}{\lambda_i^2}\int_0^{\lambda_i} \left((\lambda_i-y)V_{i-1}+yV_i\right)^2 dy \\
&= \sum_{i=1}^{n} \frac{1}{\lambda_i^2}\int_0^{\lambda_i} y^2\left(V_i^2+V_{i-1}^2-2V_{i-1}V_i\right) + \\
&\quad +y\left(2V_{i-1}V_i\lambda_i - 2\lambda_i V_{i-1}^2\right) + V_{i-1}^2\lambda_i^2\,dy \\
&= \sum_{i=1}^{n} \frac{\lambda_i}{3}\left(V_i^2+V_{i-1}^2+V_{i-1}V_i\right) \\
&= V^t A^n(\underline{\lambda})V
\end{aligned}$$

for the $n \times n$ matrix

$$A^n_{i,j}(\underline{\lambda}) := \delta_{i,j}\frac{\lambda_i+\lambda_{i+1}}{3} + \delta_{j,i+1}\frac{\lambda_{i+1}}{6} + \delta_{j+1,i}\frac{\lambda_{i+1}}{6}$$

Transposition of V is to be understood in the sense of the Euclidean scalar product restricted to the $(n-1)$-simplex.

□

We write $g(n)$ for the metric tensor on the $(n-1)$-simplex induced by the matrix $A^n_{i,j}$, i.e. in a global chart we have $g(n)^{ij} = A^n_{i,j}$.

Lemma 4.1.4. *Given a probability measure μ on $[0,1]$ and a sequence of box-type measures μ_n which converges weakly to μ and given a non-constant function $f \in T_\mu P := \overline{\mathcal{C}^\infty([0,1])}^{L^2(\mu)}$ then there exists a sequence of functions $f_n \in T_{\mu_n}P \subset T_\mu P$ such that $||f_n||_{\mu_n} \to ||f||_\mu$.*

Proof. Since the set of box-type measures is dense in the Wasserstein space over $[0,1]$ we know that there exists a sequence of box-type measures $\mu_n \equiv \mu(\underline{\lambda})$

which converges weakly to any given μ. The sequence f_n arises as projection of f to $T_{\mu_n}G_n \subset T_{\mu_n}P$: For a given orthonormal basis $\{e_k^n; k = 1,\dots,n\}$ of $T_{\mu_n}G_n = m^*(T_{\underline{\lambda}}\Sigma_n)$ we obtain $f_n := \sum_{k=1}^n \langle f, e_k^n\rangle_{\mu_n} e_k^n$. More precisely, each $e_k^n(x)$ is given by a vector $(V_0,\dots,V_{n+1}) \in \{0\}\times\mathbb{R}^n\times\{0\}$ such that $e_k^n(x) = V_{\underline{\lambda}}(x)$, normalization of e_k^n amounts to dividing e_k^n by $V^tA^n(\underline{\lambda})V$. By Schmidt orthogonalization with respect to the inner product $\langle .,.\rangle_{\mu_n}$ we obtain the orthonormal basis. By the definition of f_n we know that there exists a vector $\tilde{V} \in \{0\}\times\mathbb{R}^n\times\{0\}$ such that$||f_n||^2_{\mu_n} = \sum_{k=1}^n \langle f, e_k^n\rangle^2_{\mu_n} = ||\tilde{V}_{\underline{\lambda}}||^2$. For each $n \in \mathbb{N}$ the inner product $\langle .,.\rangle_{\mu_n}$ is a strong Riemannian metric on the subspace G_n (modeled on $\mathbb{R}^n$), in addition this spaces are geodesically convex, i.e. the notion of action minimizing curves between any two points in this subspace is well-defined, hence there exists $\lim_{n\to\infty}||f_n||^2_{\mu_n}$. Since the family $\{(TG_n, \langle .,.\rangle_{\mu_n})\}$ is dense in TP we have $\lim_{n\to\infty}||f_n||^2_{\mu_n} = ||f||^2_{\mu}$. □

Remark 4.1.1. *By the above lemma and Theorem 7.2.8 in [AGS08] it follows that the geodesic distance induced by the (strong) Riemannian metric $g(n)$ on G_n equals the Wasserstein distance when restricted to G_n.*

4.2 Sticky diffusion processes on the simplex

Denote the Laplacian with respect to the metric $g(n)$ by

$$\mathcal{L}^n f := \frac{1}{\sqrt{\det(g(n))}}\partial_j(g(n)^{-1}\sqrt{\det(g(n))}\partial_i f) = \sum_{i,j=1}^n a^{ij}\partial_i\partial_j f + \sum_{k=1}^n b^k\partial_k f$$

for any $f \in \mathcal{C}^2(\Sigma_{n-1})$ and functions $a^{ij}, b^k \in \mathcal{C}_b(\Sigma_{n-1})$, i.e.

$$a^{ij} = A^n_{i,j}$$

and

$$b^k = \frac{1}{\sqrt{\det(g(n))}}\partial_j\left(g(n)^{jk}\sqrt{\det(g(n))}\right).$$

Note that

$$\begin{aligned} b^k &= \frac{1}{\sqrt{\det(g(n))}}\frac{\partial_j g(n)^{jk}\frac{1}{\sqrt{\det(g(n))}} - g(n)^{jk}\partial_j\frac{1}{\sqrt{\det(g(n))}}}{(\sqrt{\det(g(n))})^{-2}} \\ &= \partial_j g(n)^{jk} + g(n)^{jk}\frac{1}{2}\mathrm{tr}(g(n)^{-1}\partial_j g(n)) \\ &= \frac{1}{3} + \frac{2}{6} + g(n)^{jk}\frac{1}{2}\mathrm{tr}(g(n)^{-1}\partial_j g(n)). \end{aligned}$$

Define for any $r = 1,\dots,n$ the projection

$$\pi_r : \Sigma_{n-1} \ni \underline{\lambda} \mapsto (\lambda_1,\dots,\lambda_{r-1},\lambda_{r+1},\lambda_{r+2},\dots\lambda_{n+1}) \in \partial_r\Sigma_{n-1}.$$

On the face $\partial_r\Sigma_n$ we obtain by the above lemma a restricted Riemannian metric: $\pi_r^*g(n) = A^n_{i,j}(\pi(\underline{\lambda}))$ for $i,j = 1\dots,r-1,r+1,\dots,n$. Note that the matrices $\pi_r^*g(n)$ are positive definite in $(\partial_r\Sigma_{n-1})^\circ$.

Example 4.2.1. *Riemannian metric on the 3-simplex:*

$$A^4(\underline{\lambda}) = \begin{pmatrix} \frac{\lambda_1+\lambda_2}{3} & \frac{\lambda_2}{6} & 0 & 0 \\ \frac{\lambda_2}{6} & \frac{\lambda_2+\lambda_3}{3} & \frac{\lambda_3}{6} & 0 \\ 0 & \frac{\lambda_3}{6} & \frac{\lambda_3+\lambda_4}{3} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{\lambda_1+\lambda_2}{3} & \frac{\lambda_2}{6} & 0 & 0 \\ \frac{\lambda_2}{6} & \frac{\lambda_2+\lambda_3}{3} & \frac{\lambda_3}{6} & 0 \\ 0 & \frac{\lambda_3}{6} & \frac{1-\lambda_1-\lambda_2}{3} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Example 4.2.2. *Riemannian metric: Projecting from the 3-simplex to $\partial_1\Sigma_3$ and $\partial_4\Sigma_3$:*

$$\begin{pmatrix} \frac{\lambda_1+\lambda_2}{3} & \frac{\lambda_2}{6} & 0 & 0 \\ \frac{\lambda_2}{6} & \frac{\lambda_2+\lambda_3}{3} & \frac{\lambda_3}{6} & 0 \\ 0 & \frac{\lambda_3}{6} & \frac{1-\lambda_1-\lambda_2}{3} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \overset{\pi_1*}{\to} \begin{pmatrix} \frac{\lambda_2+\lambda_3}{3} & \frac{\lambda_3}{6} & 0 \\ \frac{\lambda_3}{6} & \frac{1-\lambda_2}{3} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

resp.

$$\begin{pmatrix} \frac{\lambda_1+\lambda_2}{3} & \frac{\lambda_2}{6} & 0 & 0 \\ \frac{\lambda_2}{6} & \frac{\lambda_2+\lambda_3}{3} & \frac{\lambda_3}{6} & 0 \\ 0 & \frac{\lambda_3}{6} & \frac{1-\lambda_1-\lambda_2}{3} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \overset{\pi_4*}{\to} \begin{pmatrix} \frac{\lambda_1+\lambda_2}{3} & \frac{\lambda_2}{6} & 0 \\ \frac{\lambda_2}{6} & \frac{\lambda_2+\lambda_3}{3} & \frac{\lambda_3}{6} \\ 0 & \frac{\lambda_3}{6} & \frac{\lambda_3}{3} \end{pmatrix}$$

In the same manner as before we write the $\pi_r^* g(n)$-Laplacian (on $\partial_r\Sigma_{n-1}$) in global coordinates:

$$\mathcal{L}_r^n f := \sum_{i,j=1}^{n-1} \alpha_r^{ij}\partial_i\partial_j f + \sum_{k=1}^{n-1} \beta_r^k \partial_k f$$

for $f \in \mathcal{C}^2(\Sigma_{n-1})$ and $\alpha_r^{ij}, \beta_r^k \in \mathcal{C}_b(\partial\Sigma_{n-1})$. Since $\pi_r^* g(n)$ is a Riemannian metric on $\partial_r\Sigma_{n-1}$ we know that the matrix $(\alpha_r^{ij})_{i,j=1}^{n-1}$ is symmetric and non-negative definite on $(\partial_r\Sigma_{n-1})^\circ$. We assume additionally the existence of $\delta_r, \rho_r \in \mathcal{C}_b(\partial\Sigma_{n-1})$ which satisfy $\delta_r(\partial\Sigma_{n-1}) > 0$ and $\rho_r(\partial\Sigma_{n-1}) \geq 0$. We define ad hoc another operator (which turns out to be a boundary operator of Wentzell-type)

$$\mathcal{B}_r^n f := \mathcal{L}_r^n f + \delta_r\partial_r f - \rho_r \mathcal{L}^n f.$$

Note that $\partial_n = -\sum_{i=1}^{n-1}\partial_i$ to obtain normal vector on $\partial_n\Sigma_{n-1}$ pointing into the interior of the $(n-1)$-simplex, whereas for $\partial_1, \dots \partial_{n-1}$ we have the usual partial derivatives in $\mathbb{R}^n$.

Definition 4.2.1. *We say that the tupel $(\mathcal{L}^n, \mathcal{B}_1^n, \dots, \mathcal{B}_n^n)$ generates a diffusion measure if there exists a family $\{\mathbb{P}_x; x \in \Sigma_{n-1}\}$ of strongly Markovian probability measures on $(\mathcal{W}(\Sigma_{n-1}), \mathcal{B}(\mathcal{W}(\Sigma_{n-1})))$ such that the following conditions hold*

- $\mathbb{P}_x(\omega : \omega(0) = x) = 1$

- *For $r = 1, \dots, n+1$ there exists a function $\phi_r(t,\omega)$ on $\mathbb{R}_+ \times \mathcal{W}(\Sigma_{n-1})$ such that for a.a. ω, $\phi_r(0,\omega) = 0, t \mapsto \phi_r(t,\omega)$ is continuous and non-decreasing and*

$$\sum_{r=1}^{n} \int_0^t \mathbf{1}_{\partial_r \Sigma_n}(\omega(s)) d\phi_r(s,\omega) = \phi_r(t,\omega)$$

  Additionally we require that $\omega \mapsto \phi_r(t,\omega)$ is $\mathcal{B}_t(\mathcal{W}(\Sigma_{n-1}))$-measurable for all $t \geq 0$.

- *Furthermore*

$$f(\omega(t)) - f(\omega(0)) - \int_0^t (\mathcal{L}^n f)(\omega(s))ds - \sum_{r=1}^{n} \int_0^t (\mathcal{B}_r^n f)(\omega(s)) d\phi_r(s,\omega)$$

  is a $(\mathbb{P}_x, \mathcal{B}_t(\mathcal{W}(\Sigma_{n-1})))$-martingale for every $f \in \mathcal{C}^2(\Sigma_{n-1})$

- *and*

$$\int_0^t \mathbf{1}_{\partial \Sigma_{n-1}}(\omega(s))ds := \sum_{r=1}^{n} \int_0^t \mathbf{1}_{\partial_r \Sigma_{n-1}}(\omega(s))ds = \sum_{r=1}^{n} \int_0^t \rho_r(\omega(s)) d\phi_r(s,\omega)$$

Proposition 4.2.1 (Stochastic differential equations with stickiness and reflection on the boundary of the simplex)**.** *For $i, k = 1, \dots, n$ we choose $\sigma_k^i \in \mathcal{C}(\Sigma_{n-1})$ such that*

$$a^{ij}(x) = \sum_{k=1}^{n} \sigma_k^i(x) \sigma_k^j(x).$$

Likewise, for $i, k = 1, \dots, n-1$ and $r = 1, \dots, n$ we choose $\tau_k^i(r) \in \mathcal{C}(\partial_r \Sigma_{n-1})$ such that

$$\alpha^{ij}(x) = \sum_{k=1}^{n-1} \tau_k^i(r)(x) \tau_k^j(r)(x)$$

For any vector $X \in \Sigma_{n-1}$ we denote by $X^{\hat{r}} = (X_1 \dots, X_{r-1}, X_{r+1}, \dots, X_n)$. Given $\{B_t^i, i = 1, \dots n\}$ standard n-dimensional Brownian motion on the filtered probability space $(\Omega^n, \mathcal{F}^n, \mathcal{F}_t^n, \mathbb{P})$ we denote by $\{\tilde{B}_t^k, k = 1, \dots n-1\}$ time changed Brownian motion (i.e. $d\tilde{B}_t^i d\tilde{B}_t^j = \delta_{ij} d\phi_r(t)$) which is mutually independent from $\{B_t^i, i = 1, \dots n\}$.

Then the following system of stochastic differential equations is a $(\mathcal{L}^n, \mathcal{B}_1^n, \dots, \mathcal{B}_n^n)$-diffusion:

$$(I) \quad dX_t^{\hat{r}} = \sum_{k=1}^{n} \sigma_k^{\hat{r}}(X_t) \mathbf{1}_{\overset{\circ}{\Sigma}_n}(X_t) dB_t^k + b^{\hat{r}}(X_t) \mathbf{1}_{\overset{\circ}{\Sigma}_n}(X_t)dt +$$
$$+ \sum_{k=1}^{n-1} \tau_k^{\hat{r}}(r)(X_t) \mathbf{1}_{\partial_r \Sigma_{n-1}}(X_t) d\tilde{B}_t^k + \beta^{\hat{r}}(X_t) \mathbf{1}_{\partial_r \Sigma_{n-1}}(X_t)dt$$

$$(II) \quad dX_t^r = \sum_{k=1}^{n} \sigma_k^r(X_t) \mathbf{1}_{\overset{\circ}{\Sigma}_n}(X_t) dB_t^k + b^r(X_t) \mathbf{1}_{\overset{\circ}{\Sigma}_n}(X_t)dt +$$
$$+ \delta_r(X_t) d\phi_r(t)$$

$$(III) \quad \mathbf{1}_{\partial_r \Sigma_{n-1}}(X_t)dt = \rho_r(X_t)d\phi_r(t)$$

The function δ_r indicates the magnitude of reflection on $\partial_r\Sigma_{n-1}$ whereas ρ_r indicates the time of sojourn of X_t in $\partial_r\Sigma_{n-1}$ (stickiness).

Proof. The proof follows the one of Theorem 7.2. in [IW89] p222ff. We need to ensure that σ and b are bounded and Lipschitz continuous on Σ_{n-1}: In a first step we verify that by the choice of the Riemannian metric $g(n)$ it follows that on $\overset{\circ}{\Sigma}_n$ the functions a^{ij} resp. b^k are positive polynomial resp. rational functions in $\lambda_1, \ldots, \lambda_n$, consequently both σ and b are bounded and Lipschitz on the interior of the simplex. By the same reasoning for $\pi^r_* g(n)$ we conclude that τ and β are bounded Lipschitz on $(\partial_r\Sigma_{n-1})^\circ$. Since we want the process to be reflecting (and possibly sticky) we may set without loss of generality $\delta \equiv 1$. Without detailing we assume for the moment that ρ is bounded continuous on the boundary of the simplex. As a last condition we have to guarantee that there exists a positive constant C such that $a^{rr} \geq C$ on the boundary of the simplex which is the case since by definition $a^{rr} = \frac{1}{3}(\lambda_r + \lambda_{r+1}) = \frac{1}{3}\lambda_{r+1} > 0$ for $\underline{\lambda} \in (\partial_r\Sigma_n)^\circ$. To conclude we remark that with probability one boundary elements of dimension less than $n-1$ will not be hit when starting the diffusion in $\overset{\circ}{\Sigma}_n$. This justifies verification of boundary conditions only on $(\partial_r\Sigma_n)^\circ$. □

Lemma 4.2.1. *Let X^x_t denote the solution of the stochastic differential equation (4.2.1). Then X^x_t is a Σ_{n-1}-valued continuous Feller process.*

Proof. For each $n \in \mathbb{N}$ define $\mathfrak{A}_n$ as the algebra of bounded functions $\mathcal{C}_b(\Sigma_n)$ over $\mathcal{C}^\infty(\Sigma_n)$. Since X^x_t is an elliptic diffusion process on a bounded domain (the interior of the simplex)the semigroup $P_tf(x) = \mathbb{E}(f(X^x_t)|X_0 = x \in \overset{\circ}{\Sigma}_n)$ has the following property

$$||P_tf(x)||_{L^p} \leq ||f(X_t)||_{L^p}$$

for some $p \geq 1$ and all $f \in \mathfrak{A}_n$, i.e. the semigroup is contractive and obviously strongly continuous and positive up to the first hitting time τ_r of X^x_t at a boundary $\partial_r\Sigma_{n-1}$. Then we use again ellipticity of the diffusion process subject to the boundary operator $\mathcal{L}^n_r$ and obtain that the semigroup $P^r_tf(x) = \mathbb{E}(f(X^x_t)|X_{\tau_r} = x \in \partial_r\Sigma_{n-1})$ for $\tau_r \leq t < e_r$ (where e_r denotes the exit time of the process X_t leaving $\partial_r\Sigma_{n-1}$ to the interior of the simplex) is again strongly continuous, positive and contractive on $\mathfrak{A}_n$.

□

4.3 Tightness

Starting from a diffusion process $\underline{\lambda}$ solution to the stochastic differential equation (4.2.1) on the simplex we obtain a stochastic process $m(\underline{\lambda})$ on G_n by the continuous (and almost everywhere differentiable) mapping

$$\Sigma_{n-1} \ni \underline{\lambda} \mapsto m(\underline{\lambda}) \in G_n,$$

where G_n is equipped with the weak topology.

We define $\mathfrak{Z}$ as the space of all functions

$$F(\mu(\underline{\lambda})) \equiv F(\underline{\lambda}) := \Phi(\langle f^1, \mu(\underline{\lambda})\rangle, \ldots, \langle f^d, \mu(\underline{\lambda})\rangle) \equiv \Phi(\langle \underline{f}, \mu(\underline{\lambda})\rangle)$$

where $\Phi \in \mathcal{C}^2(\mathbb{R}^d)$, $f^1, \dots, f^d \in L^2(dx)$ and $\mu(\underline{\lambda}) \equiv m(\underline{\lambda})$.

Lemma 4.3.1. *The process $t \mapsto \mu_t \equiv m(\underline{\lambda}_t)$ is a G_n-valued continuous Feller Markov process.*

Proof. We verify in a first step that $\mu(\underline{\lambda})$ is Feller. For $F \in \mathfrak{Z}_n$

$$||(P_t F)(\mu_0)||_{L^p(\Sigma_{n-1})} = ||\mathbb{E}(F(\mu_t))|m(\underline{\lambda}_0) = \mu_0)||_{L^p(\Sigma_{n-1})} \leq ||F(\mu_t)||_{L^p(\Sigma_{n-1})}$$

since $\underline{\lambda} \mapsto F(\underline{\lambda})$ is in $\mathfrak{A}_n$ and by Lemma 4.2.1. Continuity follows since m is a continuous function and the process on the simplex has continuous trajectories. The Markov property follows directly by Dynkin's criterion: Given two $\underline{\lambda}_0, \tilde{\underline{\lambda}}_0 \in \Sigma_{n-1}$ such that $m(\underline{\lambda}_0) = m(\tilde{\underline{\lambda}}_0)$ it follows that the law of $m \circ \underline{\lambda}$ is the same under $\mathbb{P}_{\underline{\lambda}_0}$ and $\mathbb{P}_{\tilde{\underline{\lambda}}_0}$ since m is one-to-one and $\underline{\lambda}$ is Markov. □

Lemma 4.3.2 (Generator of the measure-valued process)**.** *Let $\underline{\lambda}_t$ be a solution to the stochastic differential equation in Proposition 4.2.1. Let $F \in \mathfrak{Z}$ and define*

$$(\mathcal{A}_n F)(\mu(\underline{\lambda}_0)) := \lim_{t \to 0} \frac{1}{t} \mathbb{E}(F(\mu(\underline{\lambda}_t)) - F(\mu(\underline{\lambda}_0)))$$

Then if $\underline{\lambda}_0 \in \overset{\circ}{\Sigma}_n$

$$\begin{aligned}(\mathcal{A}_n F)(\mu(\underline{\lambda}_0)) &= \partial_k \partial_l \Phi(\langle \underline{f}, \mu(\underline{\lambda}_0)\rangle) \partial_i \langle f^l, \mu(\underline{\lambda}_0)\rangle \partial_j \langle f^k, \mu(\underline{\lambda}_0\rangle a^{ij}(\underline{\lambda}_0) + \\ &+ \partial_k \Phi(\langle \underline{f}, \mu(\underline{\lambda}_0)\rangle) \partial_i \partial_j \langle f^k, \mu(\underline{\lambda}_0)\rangle a^{ij}(\underline{\lambda}_0) + \\ &\quad + \partial_k \Phi(\langle \underline{f}, \mu(\underline{\lambda}_0)\rangle) \partial_j \langle f^k, \mu(\underline{\lambda}_0)\rangle b^j(\underline{\lambda}_0)\end{aligned}$$

We use Einstein's summation for $i, j = 1, \dots, n$ and $k, l = 1, \dots, d$. By ∂_i resp. ∂_j we mean $\frac{\partial}{\partial \lambda_i}$ resp. $\frac{\partial}{\partial \lambda_j}$ whereas ∂_k and ∂_l are partial derivatives on $\mathbb{R}^d$.

If $\underline{\lambda}_0 \in \partial_r \Sigma_{n-1}$

$$\begin{aligned}(\mathcal{A}_n F)(\mu(\underline{\lambda}_0)) &= \partial_k \partial_l \Phi(\langle \underline{f}, \mu(\underline{\lambda}_0)\rangle) \partial_i \langle f^l, \mu(\underline{\lambda}_0)\rangle \partial_j \langle f^k, \mu(\underline{\lambda}_0\rangle \alpha^{ij}(\underline{\lambda}_0) + \\ &+ \partial_k \Phi(\langle \underline{f}, \mu(\underline{\lambda}_0)\rangle) \partial_i \partial_j \langle f^k, \mu(\underline{\lambda}_0)\rangle \alpha^{ij}(\underline{\lambda}_0) + \\ &\quad + \partial_k \Phi(\langle \underline{f}, \mu(\underline{\lambda}_0)\rangle) \partial_j \langle f^k, \mu(\underline{\lambda}_0)\rangle [(1 - \delta_{jr}) \beta^j(\underline{\lambda}_0) \rho_r(\underline{\lambda}_0) + \delta_{jr} \delta_r(\underline{\lambda}_0)]\end{aligned}$$

Again we use Einstein's summation for $i, j = 1, \dots, r-1, r+1, \dots, n$ and $k, l = 1, \dots, d$. By ∂_i resp. ∂_j we mean $\frac{\partial}{\partial \lambda_i}$ resp. $\frac{\partial}{\partial \lambda_j}$ whereas ∂_k and ∂_l are partial derivatives on $\mathbb{R}^d$.

Proof. Using Ito's formula (on open sets of $\mathbb{R}^n$)

$$dF(\underline{\lambda}_t) = \nabla F(\underline{\lambda}_t)\, d\underline{\lambda}_t + \frac{1}{2} \nabla^2 F(\underline{\lambda}_t)\, d\underline{\lambda}_t d\underline{\lambda}_t$$

in $\overset{\circ}{\Sigma}_n$ it holds

$$d\underline{\lambda}_t^i d\underline{\lambda}_t^j = \sum_{k=1}^n \sigma_k^i(\underline{\lambda}_t) \sigma_k^j(\underline{\lambda}_t) dt = a^{ij}(\underline{\lambda}_t)\, dt$$

$$\nabla F(\underline{\lambda}_t)\, d\underline{\lambda}_t \overset{\mathcal{M}}{=} \sum_{i=1}^n \sum_{k=1}^d \partial_k \Phi(\langle \underline{f}, \mu(\underline{\lambda})\rangle) \partial_i \langle f_k, \mu(\underline{\lambda})\rangle b^i(\lambda_t) dt$$

$$\nabla^2 F(\underline{\lambda}_t)\, d\underline{\lambda}_t d\underline{\lambda}_t = \sum_{i,j=1}^{n}\sum_{k,l=1}^{d} \partial_k\partial_l\Phi(\langle \underline{f}, \mu(\underline{\lambda})\rangle)\partial_i\langle f^l, \mu(\underline{\lambda}_t)\rangle\partial_j\langle f^k, \mu(\underline{\lambda}_t)\rangle a^{ij}(\underline{\lambda}_t)\, dt + \\ + \sum_{i,j=1}^{n}\sum_{k=1}^{d} \partial_k\Phi(\langle \underline{f}, \mu(\underline{\lambda})\rangle)\partial_j\partial_i\langle f^k, \mu(\underline{\lambda}_t)\rangle a^{ij}(\underline{\lambda}_t)\, dt$$

and on $\partial_r\Sigma_{n-1}$

$$d\underline{\lambda}_t^i d\underline{\lambda}_t^j = \sum_{k=1}^{n-1} \tau_k^i(r)(\underline{\lambda}_t)\tau_k^j(r)(\underline{\lambda}_t)d\phi(t) = \alpha^{ij}(\underline{\lambda}_t)\, d\phi(t)$$

$$\nabla F(\underline{\lambda}_t)\, d\underline{\lambda}_t \overset{\mathcal{M}}{=} \sum_{i\neq r}\sum_{k=1}^{d} \partial_k\Phi(\langle \underline{f}, \mu(\underline{\lambda}_t)\rangle)\partial_i\langle f^k, \mu(\underline{\lambda}_t)\rangle\beta^i(\underline{\lambda}_t)\rho_r(\underline{\lambda}_t)d\phi(t) + \\ + \sum_{k=1}^{d} \partial_k\Phi(\langle \underline{f}, \mu(\underline{\lambda})\rangle)\partial_r\langle f^k, \mu(\underline{\lambda})\rangle\delta_r(\lambda_t)d\phi(t)$$

$$\nabla^2 F(\underline{\lambda}_t)\, d\underline{\lambda}_t d\underline{\lambda}_t = \sum_{i,j\neq r}\sum_{k,l=1}^{d} \partial_k\partial_l\Phi(\langle \underline{f}, \mu(\underline{\lambda})\rangle)\partial_i\langle f^l, \mu(\underline{\lambda}_t)\rangle\partial_j\langle f^k, \mu(\underline{\lambda}_t)\rangle \alpha^{ij}(\underline{\lambda}_t)\, d\phi(t) + \\ + \sum_{i,j\neq r}\sum_{k=1}^{d} \partial_k\Phi(\langle \underline{f}, \mu(\underline{\lambda})\rangle)\partial_j\partial_i\langle f^k, \mu(\underline{\lambda}_t)\rangle \alpha^{ij}(\underline{\lambda}_t)\, d\phi(t)$$

□

Lemma 4.3.3. *As long as* $\mu_t \in \overset{\circ}{G}_n$ *the process* $t \mapsto Y_t := F(\mu_t)$ *satisfies the following stochastic differential equation:*

$$\begin{aligned} dF(\mu_t) &= \partial_l\Phi(\langle \underline{f}, \mu_t\rangle)\partial_i\langle f^l, \mu_t\rangle\sigma_k^i(m^{-1}(\mu_t))dB_t^k + \\ &+\partial_k\Phi(\langle \underline{f}, \mu_t\rangle)\{\partial_i\langle f^k, \mu_t\rangle b^i(m^{-1}(\mu_t)) + \partial_i\partial_j\langle f^k, \mu_t\rangle a^{ij}(m^{-1}(\mu_t))\}dt \\ &+\partial_k\partial_l\Phi(\langle \underline{f}, \mu_t\rangle)\partial_i\langle f^k, \mu_t\rangle\partial_j\langle f^l, \mu_t\rangle a^{ij}(m^{-1}(\mu_t))dt+ \end{aligned}$$

and the quadratic variation process of the semi-martingale Y_t *reads*

$$\langle Y, Y\rangle_t = \int_0^t (\partial_l\Phi(\langle \underline{f}, \mu_s\rangle)\partial_i\langle f^l, \mu_s\rangle)\partial_k\Phi(\langle \underline{f}, \mu_s\rangle)\partial_j\langle f^k, \mu_s\rangle)a^{ij}(m^{-1}(\mu_s))ds$$

If $\mu_t \in m(\partial_r\Sigma_{n-1})$ *then* Y_t *satisfies the following Skorohod stochastic differential equation:*

$$\begin{aligned} dF(\mu_t) &= \partial_l\Phi(\langle \underline{f}, \mu_t\rangle)\partial_i\langle f^l, \mu_t\rangle\tau_k^i(m^{-1}(\mu_t))d\tilde{B}_t^k + \\ &+\partial_k\Phi(\langle \underline{f}, \mu_t\rangle)\{\partial_i\langle f^k, \mu_t\rangle[(1-\delta_{jr})\beta^j(m^{-1}(\mu_t))\rho_r(m^{-1}(\mu_t)) + \delta_{jr}\delta_r(m^{-1}(\mu_t))] + \\ &+\partial_i\partial_j\langle f^k, \mu_t\rangle\alpha^{ij}(m^{-1}(\mu_t))\}d\phi_r(t) + \\ &+\partial_k\partial_l\Phi(\langle \underline{f}, \mu_t\rangle)\partial_i\langle f^k, \mu_t\rangle\partial_j\langle f^l, \mu_t\rangle\alpha^{ij}(m^{-1}(\mu_t))d\phi_r(t) \end{aligned}$$

Note that

$$\begin{aligned}\partial_i\langle f^l,\mu_t\rangle\tau_k^i(m^{-1}(\mu_t)) &= \frac{1}{n}\sum_{i=1}^{r-1}\{(f^l)'(\lambda_1(t)+\cdots+\lambda_{r-1}(t))\}\tau_k^i(m^{-1}(\mu_t)) + \\ &\quad + \sum_{i\neq r}^{n}\partial_i\{\int f^l\mu_t(dx)\}\tau_k^i(m^{-1}(\mu_t))\end{aligned}$$

and

$$\begin{aligned}\partial_i\partial_j\langle f^k,\mu_t\rangle\alpha^{ij}(m^{-1}(\mu_t)) &= \frac{1}{n}\sum_{i,j=1}^{r-1}\{(f^l)''(\lambda_1(t)+\cdots+\lambda_{r-1}(t))\}\alpha^{ij}(m^{-1}(\mu_t)) + \\ &\quad + \sum_{i,j\neq r}^{n}\partial_i\partial_j\{\int f^l\mu_t(dx)\}\alpha^{ij}(m^{-1}(\mu_t))\end{aligned}$$

Proof. Use Ito's formula. □

By virtue of a Σ_{n-1}-valued process $\underline{\lambda}_t$ we obtain for each $n\in\mathbb{N}$ a probability measure on the Skorohod space $D_P := D([0,\infty),P)$ of càdlàg functions (which is in turn a Polish space when equipped with the Skorohod topology, see for instance section 3.6. in [Daw93]):

$$P_n(A) := \mathbb{P}_{\underline{\lambda}_0}(\omega : m(\underline{\lambda}_t)(\omega)\in A; t\geq 0)$$

for all Borel sets A in D_P. The set of functions $\mathfrak{Z}$ (when restricted to G_n) is a family of real continuous functions on P which is closed under addition and separates points. Given $F\in\mathfrak{Z}$ we obtain a mapping $\tilde{F} : D_P \to D_{\mathbb{R}}$ by $(\tilde{F}m(\underline{\lambda}))(t) := F(m(\underline{\lambda}_t))$.

Lemma 4.3.4. *The family of probability measures $\{P_n\}_{n\in\mathbb{N}}$ on D_P is tight, i.e. it satisfies the following tightness criterion (see Theorem 3.7.1 [Daw93]): For each $F\in\mathfrak{Z}$ such that $F(\mu)=\int f\mu$ the sequence*

$$Q_n := \{P_n\circ\tilde{F}^{-1}\}$$

of probability measures on $D_{\mathbb{R}}$ is tight.

Proof. We consider the Doob-Meyer decomposition of the semimartingale $Y_t = M_t + A_t$ as defined in Lemma 4.3.3 with $\Phi = \mathrm{id}$. Then

$$\langle Y\rangle_t = \langle M\rangle_t = \int_0^t \partial_i\langle f,\mu_s\rangle\partial_j\langle f,\mu_s\rangle a^{ij}(m^{-1}(\mu_s))ds.$$

In order to verify in a first step that there exist Lipschitz estimates of the quadratic variation that are uniform in n we assume the existence of a scaling function $\kappa(n)$ and define $\mu_t^\kappa := m(\underline{\lambda}_{t\kappa(n)})$ resp. $Y_t^\kappa := F(\mu_t^\kappa)$. We suppose $\kappa(n)=1$ whenever the superscript κ is omitted in μ_t.
Set $f=g'$. Let us look closer at

$$\frac{\partial}{\partial\lambda_i}\int f\mu(\underline{\lambda}) = \frac{\partial}{\partial\lambda_i}\sum_{j=1}^{n}\int_0^1 f(x)\frac{1}{n\lambda_j}\mathbf{1}_{[\sum_{k=1}^{j-1}\lambda_k,\sum_{k=1}^{j}\lambda_k)}(x)dx, \tag{4.1}$$

this equals

$$-\frac{1}{n\lambda_i^2}\left\{g(\sum_{k=1}^{i}\lambda_k)-g(\sum_{k=1}^{i-1}\lambda_k)\right\}+\sum_{j=1}^{n}\frac{1}{n\lambda_j}\left\{g'(\sum_{k=1}^{j}\lambda_k)\frac{\partial}{\partial_i}\sum_{k=1}^{j}\lambda_k-g'(\sum_{k=1}^{j-1}\lambda_k)\frac{\partial}{\partial_i}\sum_{k=1}^{j-1}\lambda_k\right\}$$

which is

$$-\frac{1}{n\lambda_i^2}\int_{\sum_{k=1}^{i-1}\lambda_k}^{\sum_{k=1}^{i}\lambda_k}f(x)dx+\sum_{j=1}^{n}\frac{1}{n\lambda_j}\left\{g'(\sum_{k=1}^{j}\lambda_k)\mathbf{1}_{i\le j}-g'(\sum_{k=1}^{j-1}\lambda_k)\mathbf{1}_{i\le j-1}\right\}$$

i.e.

$$|\frac{\partial}{\partial\lambda_i}\int f\mu(\underline{\lambda})|\le\frac{1}{n\lambda_i}||f||_\infty+\frac{1}{n\min\{\lambda_j;j=1,\dots,n\}}||f'||_\infty. \tag{4.2}$$

Thus

$$\begin{aligned}|\langle Y^\kappa\rangle_t-\langle Y^\kappa\rangle_s| &\le \kappa(n)\int_s^t|\partial_i\langle f,\mu_u\rangle\partial_j\langle f,\mu_u\rangle a^{ij}(m^{-1}(\mu_u))|du\\ &\le \kappa(n)\frac{2n-1}{n^2}(t-s)C(||f||_\infty+||f'||_\infty)^2\end{aligned}$$

for some positive constant C.

Choosing $\kappa(n)=\frac{n^2}{2n-1}$ we obtain Lipschitz constants that are uniform in n. Since Y_t^κ is conservative we obtain for some fixed time $T>0$ the Lyons-Zheng decomposition (see [FOT94] Theorem 5.7.1)

$$Y_t^\kappa-Y_0^\kappa=\frac{1}{2}(X_t-(\tilde{X}_T-\tilde{X}_{T-t})),$$

here X is a $\mathcal{F}_t=\sigma(\underline{\lambda}_{s\kappa};0\le s\le t)$-martingale and $\tilde{X}$ is a $\tilde{\mathcal{F}}_t=\sigma(\underline{\lambda}_{(T-s)\kappa};0\le s\le t)$-martingale for $0\le t\le T$. Then for the quadratic variation of X we have

$$\langle X\rangle_t-\langle X\rangle_s=\kappa(n)\int_s^t\partial_i\langle f,\mu_s\rangle)\partial_j\langle f,\mu_s\rangle)a^{ij}(m^{-1}(\mu_s))ds\le C|t-s|$$

For the quadratic variation of $\tilde{X}$ a similar estimate holds by symmetry.

Then

$$\begin{aligned}\mathbb{E}|(F(\mu_{\kappa t})-F(\mu_{\kappa s})| &= \frac{1}{2}\mathbb{E}|X_t-X_s|+\frac{1}{2}\mathbb{E}|\tilde{X}_{T-t}-\tilde{X}_{T-s}|\\ &\le \frac{1}{2}(\mathbb{E}|X_t-X_s|^2)^{\frac{1}{2}}+\frac{1}{2}(\mathbb{E}|\tilde{X}_{T-t}-\tilde{X}_{T-s}|^2)^{\frac{1}{2}}\\ &\le \frac{1}{2}(\mathbb{E}|\langle X\rangle_t-\langle X\rangle_s|)^{\frac{1}{2}}+\frac{1}{2}(\mathbb{E}|\langle\tilde{X}\rangle_{T-t}-\langle\tilde{X}\rangle_{T-s}|)^{\frac{1}{2}}\\ &\le C|t-s|^{\frac{1}{2}}\end{aligned}$$

Tightness follows by Theorem 7.2 in chapter 3 of [EK05]. On the boundary we use the same reasoning: replace a by α and remark that for formula (4.1) the same estimate holds (the function f is integrated with respect to a box-type measure where at least one indicator function is replaced by a Dirac measure but since f was supposed to have bounded derivatives, an estimate as in formula (4.2) holds.) □

Chapter 5

Finite-dimensional diffusion processes via projections

5.1 Riemannian geometry of the space of histograms endowed with Wasserstein distance

Definition 5.1.1 (Histograms with respect to a fixed partition). *Let* $\mathbb{A} = \bigcup_{i=1}^{n} A_i$ *denote any finite partition of a compact Riemannian manifold* M *where each* A_i *has non-empty interior and is convex.*
To each $\underline{\lambda} \in \Sigma_{n-1}$ *we associate a probability measure on* M *which has the following density function with respect to the volume measure*

$$\iota : \underline{\lambda} \mapsto \mu^{\underline{\lambda}}(x) := \sum_{i=1}^{n} \frac{\lambda_i}{vol(A_i)} 1_{A_i}(x); \;\; x \in M$$

We denote $P_2^{\mathbb{A}}(M) := \iota(\Sigma_{n-1}) \subset P_2(M)$ *dense with respect to the quadratic Wasserstein distance. The map* ι *gives by construction an isometry between the spaces* $(P_2^{\mathbb{A}}(M), d_W)$ *and* (Σ_{n-1}, d) *where* $d(\underline{\lambda}, \underline{\lambda}') := d_W(\iota(\underline{\lambda}), \iota(\underline{\lambda}'))$*.*
To each partition $\mathbb{A}$ *we associate a projection operator*

$$Pr : P(M) \ni \mu \mapsto \sum_{i=1}^{n} \frac{\mu(A_i)}{vol(A_i)} 1_{A_i}(x) \in P_2^{\mathbb{A}}(M)$$

Lemma 5.1.1 (Wasserstein distance of histograms on the unit interval via quantile functions). *Henceforth we confine ourselves to* $M = [0,1]$ *and the following*

$$\iota : \underline{\lambda} \mapsto \mu^{\underline{\lambda}}(x) := \sum_{i=1}^{n} n\lambda_i 1_{[\frac{i-1}{n}, \frac{i}{n})}(x); \;\; x \in [0,1].$$

To each probability density $\mu^{\underline{\lambda}}$ *we associate its cumulative distribution function*

$$c^{\mu^{\underline{\lambda}}}(t) := \mu^{\underline{\lambda}}([0,t))$$

and its quantile function

$$q^{\mu^{\underline{\lambda}}}(t) := \inf \left\{ s \in [0,1] : \mu^{\underline{\lambda}}([0,s]) > t \right\}, \inf \emptyset := 1$$

which is a nondecreasing, piecewise linear function from $[0,1]$ *to* $[0,1]$. *For* $\underline{\lambda} \in \overset{\circ}{\Sigma}_{n-1}$

$$q^{\mu^{\underline{\lambda}}}(t) = \sum_{k=1}^{n} \left\{ \frac{t}{n\lambda_k} + \frac{k\lambda_k - \sum_{j=1}^{k} \lambda_j}{n\lambda_k} \right\} 1_{[\sum_{j=0}^{k-1} \lambda_j, \sum_{j=0}^{k} \lambda_j)}(t),$$

with the convention $\lambda_0 = 0$ *and* $q^{\mu^{\underline{\lambda}}}(1) = 1$. *Whenever there exists a* $\lambda_l = 0$ *then the quantile function reads*

$$\begin{aligned} q^{\mu^{\underline{\lambda}}}(t) &= \sum_{k=1}^{l} \left\{ \frac{t}{n\lambda_k} + \frac{k\lambda_k - \sum_{j=1}^{k} \lambda_j}{n\lambda_k} \right\} 1_{[\sum_{j=0}^{k-1} \lambda_j, \sum_{j=0}^{k} \lambda_j)}(t) \\ &\quad + \sum_{k=l+1}^{n} \left\{ \frac{t}{n\lambda_k} + \frac{(k-1)\lambda_k - \sum_{j=1}^{k-1} \lambda_j}{n\lambda_k} \right\} 1_{[\sum_{j=0}^{k-1} \lambda_j, \sum_{j=0}^{k} \lambda_j)}(t) \\ &= \sum_{k=1}^{l} \left\{ \frac{t}{n\lambda_k} + \frac{k\lambda_k - \sum_{j=1}^{k} \lambda_j}{n\lambda_k} \right\} 1_{[\sum_{j=0}^{k-1} \lambda_j, \sum_{j=0}^{k} \lambda_j)}(t) \\ &\quad + \sum_{k=l+1}^{n} \left\{ \frac{t}{n\lambda_k} + \frac{k\lambda_k - \sum_{j=1}^{k} \lambda_j}{n\lambda_k} \right\} 1_{[\sum_{j=0}^{k-1} \lambda_j, \sum_{j=0}^{k} \lambda_j)}(t) \end{aligned}$$

This generalizes to higher dimensional boundary parts of simplices: We assume that $\underline{\lambda} \in \partial\Sigma_{n-1}$, $\lambda_{\alpha_l} = 0$ *and* $\lambda_j > 0$ *for* $j \neq \alpha_l$ *with* $l = 1, \ldots, L$. *This means that the point* $\underline{\lambda}$ *is in a* $(n-1-L)$*-dimensional part of* $\partial\Sigma_{n-1}$. *The quantile function of* $\mu^{\underline{\lambda}}$ *reads*

$$q^{\mu^{\underline{\lambda}}}(t) = \sum_{k \in \{1,\ldots,n\} \setminus \{\alpha_1,\ldots,\alpha_L\}} \left\{ \frac{t}{n\lambda_k} + \frac{k\lambda_k - \sum_{j=1}^{k} \lambda_j}{n\lambda_k} \right\} 1_{[\sum_{j=0}^{k-1} \lambda_j, \sum_{j=0}^{k} \lambda_j)}(t).$$

Henceforth we denote $I_k = [\sum_{j=0}^{k-1} \lambda_j, \sum_{j=0}^{k} \lambda_j)$ *resp.* $\tilde{I}_k = [\sum_{j=0}^{k-1} \tilde{\lambda}_j, \sum_{j=0}^{k} \tilde{\lambda}_j)$. *Given* $\underline{\lambda}, \underline{\tilde{\lambda}} \in \Sigma_{n-1}$ *one calculates the quadratic Wasserstein distance via quantile functions:*

$$d_W(\mu^{\underline{\lambda}}, \mu^{\underline{\tilde{\lambda}}})^2 = \int_0^1 |q^{\mu^{\underline{\lambda}}}(t) - q^{\mu^{\underline{\tilde{\lambda}}}}(t)|^2 \, dt$$

In a first step we calculate

$$\left(q^{\mu^{\underline{\lambda}}}(t) - q^{\mu^{\underline{\tilde{\lambda}}}}(t)\right)^2 = \left[\sum_{k=1}^{n} \left\{ \frac{t}{n\lambda_k} + \frac{k\lambda_k - \sum_{j=1}^{k} \lambda_j}{n\lambda_k} \right\} 1_{I_k}(t) - \left\{ \frac{t}{n\tilde{\lambda}_k} + \frac{k\tilde{\lambda}_k - \sum_{j=1}^{k} \tilde{\lambda}_j}{n\tilde{\lambda}_k} \right\} 1_{\tilde{I}_k}(t) \right]^2$$

and observe that we will have to integrate terms of the type

$$\mathcal{I}_k = \left\{ \frac{t}{n\lambda_k} + \frac{k\lambda_k - \sum_{j=1}^{k} \lambda_j}{n\lambda_k} \right\}^2 1_{I_k}(t), \qquad \tilde{\mathcal{I}}_k = \left\{ \frac{t}{n\tilde{\lambda}_k} + \frac{k\tilde{\lambda}_k - \sum_{j=1}^{k} \tilde{\lambda}_j}{n\tilde{\lambda}_k} \right\}^2 1_{\tilde{I}_k}(t)$$

and

$$\mathcal{I}_i\tilde{\mathcal{I}}_k = \left\{ \frac{t}{n\lambda_i} + \frac{i\lambda_i - \sum_{j=1}^{i} \lambda_j}{n\lambda_i} \right\} \left\{ \frac{t}{n\tilde{\lambda}_k} + \frac{k\tilde{\lambda}_k - \sum_{j=1}^{k} \tilde{\lambda}_j}{n\tilde{\lambda}_k} \right\} 1_{I_i \cap \tilde{I}_k}(t)$$

For the terms of type $\mathcal{I}\mathcal{I}$ we distinguish two different cases. Either

$$(\underline{\lambda}, \underline{\tilde{\lambda}}) \in B_{ik} := \left\{ \sum_{j=0}^{k-1} \tilde{\lambda}_j < \sum_{j=0}^{i} \lambda_j \quad \wedge \quad \sum_{j=0}^{i-1} \lambda_j < \sum_{j=0}^{k} \tilde{\lambda}_j \right\} \subset \Sigma_{n-1} \times \Sigma_{n-1}$$

then the intersection

$$I_i \cap \tilde{I}_k = A_{ik} := [\sum_{j=0}^{i-1} \lambda_j, \sum_{j=0}^{i} \lambda_j) \cap [\sum_{j=0}^{k-1} \tilde{\lambda}_j, \sum_{j=0}^{k} \tilde{\lambda}_j) = [\sum_{j=0}^{i-1} \lambda_j \vee \sum_{j=0}^{k-1} \tilde{\lambda}_j, \sum_{j=0}^{i} \lambda_j \wedge \sum_{j=0}^{k} \tilde{\lambda}_j) \neq \emptyset,$$

or

$$(\underline{\lambda}, \underline{\tilde{\lambda}}) \notin B_{ik}$$

then $A_{ik} = \emptyset$ for all $i \neq k$. Note that in general $B_{ik} \neq B_{ki}$ and that $\mathcal{I}_i \tilde{\mathcal{I}}_k \neq \mathcal{I}_k \tilde{\mathcal{I}}_i$.

$$\begin{aligned}
\int_0^1 \mathcal{I}_k dt &= \int_0^1 \left\{ \frac{t}{n\lambda_k} + \frac{k\lambda_k - \sum_{j=1}^k \lambda_j}{n\lambda_k} \right\}^2 1_{I_k}(t) dt \\
&= \frac{(\sum_{j=0}^k \lambda_j)^3 - (\sum_{j=0}^{k-1} \lambda_j)^3}{3n^2\lambda_k^2} + \\
&\quad + \frac{k\lambda_k - \sum_{j=1}^k \lambda_j}{n\lambda_k} \frac{(\sum_{j=0}^k \lambda_j)^2 - (\sum_{j=0}^{k-1} \lambda_j)^2}{n\lambda_k} \\
&\quad + \lambda_k \frac{(k\lambda_k - \sum_{j=1}^k \lambda_j)^2}{(n\lambda_k)^2} \\
&= O(\lambda_1, \ldots, \lambda_{n-1})
\end{aligned}$$

and

$$\begin{aligned}
\int_0^1 \mathcal{I}_i \tilde{\mathcal{I}}_k dt &= \frac{(\sum_{j=0}^i \lambda_j \wedge \sum_{j=0}^k \tilde{\lambda}_j)^3 - (\sum_{j=0}^{i-1} \lambda_j \vee \sum_{j=0}^{k-1} \tilde{\lambda}_j)^3}{3n^2\lambda_i\tilde{\lambda}_k} + \\
&\quad + \frac{(\sum_{j=0}^i \lambda_j \wedge \sum_{j=0}^k \tilde{\lambda})^2 - (\sum_{j=0}^{i-1} \lambda_j \vee \sum_{j=0}^{k-1} \tilde{\lambda}_j)^2}{2n^2\lambda_i\tilde{\lambda}_k} \left[i\lambda_i - \sum_{j=1}^{i} \lambda_j + k\tilde{\lambda}_k - \sum_{j=1}^{k} \tilde{\lambda}_j \right] \\
&\quad + \left((\sum_{j=0}^{i} \lambda_j \wedge \sum_{j=0}^{k} \tilde{\lambda}) - (\sum_{j=0}^{i-1} \lambda_j \vee \sum_{j=0}^{k-1} \tilde{\lambda}_j) \right) \frac{i\lambda_i - \sum_{j=1}^i \lambda_j}{n\lambda_i} \frac{k\tilde{\lambda}_k - \sum_{j=1}^k \tilde{\lambda}_j}{n\tilde{\lambda}_k} \\
&= O(\lambda_1^2, \ldots, \lambda_{n-1}^2)
\end{aligned}$$

for $\underline{\tilde{\lambda}}$ fixed. Finally

$$d_W(\mu^{\underline{\lambda}}, \mu^{\underline{\tilde{\lambda}}})^2 = \int_0^1 \sum_{k=1}^{n} (\mathcal{I}_k + \tilde{\mathcal{I}}_k)\, dt - \int_0^1 \sum_{i,k=1}^{n} (\mathcal{I}_i \tilde{\mathcal{I}}_k + \mathcal{I}_k \tilde{\mathcal{I}}_i)\, dt$$

Example 5.1.1 (Σ_1)**.** *Given $\underline{\lambda}, \underline{\tilde{\lambda}} \in \mathring{\Sigma}_1$:*

$$d_W(\mu^{\underline{\lambda}}, \mu^{\underline{\tilde{\lambda}}})^2 = \frac{|\lambda_1 - \tilde{\lambda}_1|^2 (2|\lambda_1 - \tilde{\lambda}_1| + 1)}{12(1 - \lambda_1)\tilde{\lambda}_1} \tag{5.1}$$

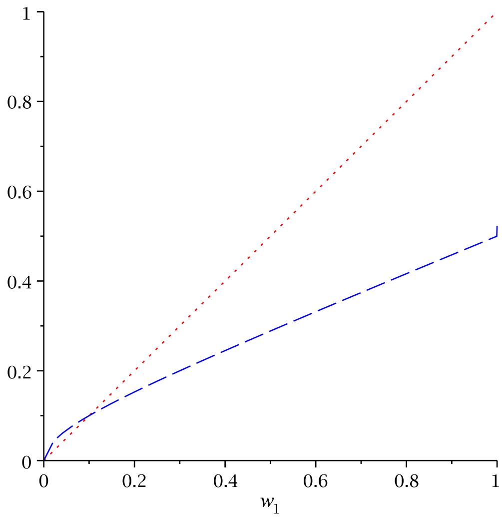
Assuming $\lambda_1 = 0$ *and* $0 < \tilde{\lambda}_1 < 1$*:*

$$d_W(\mu^{\underline{\lambda}}, \mu^{\underline{\tilde{\lambda}}})^2 = \frac{2\tilde{\lambda}_1^2 + \tilde{\lambda}_1}{12}$$

and

$$d_W(\mu^{(1,0)}, \mu^{(0,1)})^2 = \frac{1}{4}$$

In the following plot the dashed line shows the graph of $w_1 \mapsto d_W\left(\mu^{(0,1)}, \mu^{(w_1,1-w_1)}\right)$ *for* $w_1 \in [0,1]$ *compared to the plot of* $w_1 \mapsto w_1$ *represented by the doted line.*



Remark 5.1.1 (Wasserstein geodesics and their projection to the space of histograms)**.** *Given* $\mu, \nu \in P_{ac}([0,1])$ *by a general theorem of Brenier and McCann ([Bre91], [McC01]) we know that there exists a* μ*-a.s. unique (optimal) map* $T : [0,1] \to [0,1]$ *such that* $d_W(\mu,\nu)^2 = \int_0^1 (x - T(x))^2 \mu(dx)$*; in dimension one we know even better:* $T = q^\nu \circ c^\mu$*, which is a monotonous mapping from* $[0,1]$ *to itself (see [Vil03]). By ([AGS08] Theorem 7.2.2) the curve*

$$s \mapsto \gamma_s := ((1-s)id + sT)\#\mu$$

is a constant-speed geodesic in $P_{ac}([0,1])$*. If we had started directly with* $\mu, \nu \in P^{\mathbb{A}}$ *then the constant-speed geodesic* $\gamma_s = T_s\#\mu := ((1-s)id + sT)\#\mu$ *has as quantile function* $q^{\gamma_s} = (1-s)q^\mu + sq^\nu$ *(see [AGS08] 7.2.8) which proves that* γ_s *does not stay in* $P^{\mathbb{A}}$ *for* $s \in (0,1)$ *but is contained in a bigger space* $P^{\mathbb{B}}$ *where* $\mathbb{B}$ *is a refinement of* $\mathbb{A}$*. Consider a geodesic* $\gamma \in P_{ac}$ *linking the histograms* μ *and* ν*, then*

$$s \mapsto \Gamma_s := \iota^{-1}(Pr(\gamma_s)) = (\gamma_s(A_i))_{i=1}^n \in \Sigma_{n-1}$$

which is the projection of the Wasserstein geodesic to the space of histograms.

Lemma 5.1.2. *By the first order variation of the distance along projected geodesics we obtain a Riemannian metric on* $\overset{\circ}{\Sigma}_1 \subset \mathbb{R}^2$ *by*

$$g_p(V,V) := \frac{|V|^2}{12(1-p)p}$$

Proof. Denote $p := \iota^{-1}(\Pr(\mu))^1$ and consider at first the case of Σ_1 with $A_i = [(i-1)/2, i/2)$ for $i = 1, 2$. By (5.1) the first order variation of the distance d along the curve Γ is

$$\lim_{t\to 0}\frac{d(p,\Gamma_t^1)}{t} = \lim_{t\to 0} 1/t \begin{cases} \frac{-\sqrt{3}(p-\Gamma_t^1)\sqrt{-2p+2\Gamma_t^1+1}}{6\sqrt{1-p}\sqrt{\Gamma_t^1}} & 0 < p < \Gamma_t^1 < 1 \\ \frac{\sqrt{3}(p-\Gamma_t^1)\sqrt{2p-2\Gamma_t^1+1}}{6\sqrt{1-p}\sqrt{\Gamma_t^1}} & 0 < \Gamma_t^1 < p < 1 \end{cases}$$

which equals

$$\frac{\sqrt{3}\lim_{t\to 0}\frac{1}{t}|p-\Gamma_t^1|}{6\sqrt{1-p}\sqrt{p}} = \frac{\sqrt{3}|\dot\Gamma_0^1|}{6\sqrt{1-p}\sqrt{p}}$$

$|p|$ means $\frac{1}{\sqrt{2}}||(0,1)-(p,1-p)||_{\mathbb{R}^2}$ and

$$\begin{aligned}
\dot\Gamma_t^i &= \frac{d}{dt}\gamma_t(A_i) = \frac{d}{dt}\int_0^1 \mathbf{1}_{A_i}((1-t)x + t(q^\nu\circ c^\mu)(x))\mu(dx) \\
&= \int_0^1 (\nabla \mathbf{1}_{A_i})(T_t(x))\frac{d}{dt}T_t(x)\mu(dx) \\
&= \int_0^1 (\nabla \mathbf{1}_{A_i})(T_t(x))((q^\nu\circ c^\mu)(x) - x)\mu(dx) \\
&= \int_{T_t([0,1])} (\nabla \mathbf{1}_{A_i})(y)((q^\nu\circ c^\mu)(T_t^{-1}(y)) - T_t^{-1}(y))\gamma_t(dy)) \\
&= \int_0^1 (d\delta_{(i-1)/2} - d\delta_{i/2})(y)((q^\nu\circ c^\mu)(T_t^{-1}(y)) - T_t^{-1}(y))\gamma_t(dy)) \\
&= q^\nu\circ c^\mu(T_t^{-1}((i-1)/2)) - T_t^{-1}((i-1)/2)) - q^\nu\circ c^\mu)(T_t^{-1}((i)/2)) + T_t^{-1}((i)/2))
\end{aligned}$$

$d\delta_a$ denotes the Dirac measure at a. It is important that γ_t neither charges points nor does it verify $\int_{B\supset\partial A_i}\gamma_t = 0$ for any measurable B which is guaranteed by the fact that if μ and ν do have full support, then their displacement interpolation γ_t does so. (Just look at the interpolating quantile functions $q^{\gamma_t} = (1-t)q^\mu + tq^\nu$!) Finally

$$\dot\Gamma_0^i = q^\nu\circ c^\mu((i-1)/2) - (i-1)/2 - q^\nu\circ c^\mu(i/2) + i/2 \qquad \text{in particular } \dot\Gamma_0^1 = -\dot\Gamma_0^2$$

Note that for a fixed partition the velocity vector of the projected geodesic depends on the quantile functions of the starting point $\mu \in P^{\mathbb{A}}$ and on the quantile function of $\nu \in P^{\mathbb{A}}$ which determines the direction of the unit speed geodesic. Within the set $\Pr^{-1}(\nu) \subset P_{ac}$ we are free to choose a representative which determines the velocity vector of the projected curve: Taking for instance a smooth density $\tilde\nu$ with $\Pr(\tilde\nu) = \nu$ gives rise to a projected geodesic $\tilde\Gamma_s$ which is different from Γ_s for $s \in (0,1)$ but still

$$\dot{\tilde\Gamma}_0^1 = -q^{\Pr(\tilde\nu)}\circ c^\mu(1/2) + 1/2 = \dot\Gamma_0^1.$$

By the above considerations the first order variation does not depend on the representative of the velocity of the curve. We obtain a Riemannian metric on $\mathring{\Sigma}_1 \subset \mathbb{R}^2$ by

$$g_p(\dot\Gamma_0^1, \dot\Gamma_0^1) := \frac{|\dot\Gamma_0^1|^2}{12(1-p)p} = \frac{|-q^\nu\circ c^\mu(1/2) + 1/2|^2}{12(1-p)p},$$

□

By the same reasoning we obtain a Riemannian metric on the 2-simplex:

Lemma 5.1.3. *Denote $p := \iota^{-1}(Pr(\mu))^1$ and consider the case of Σ_2 with $A_i = [(i-1)/3, i/3)$ for $i = 1, 2, 3$. We denote the vector $X = \dot{\Gamma}_0$. Then by the first order variation of the Wasserstein distance along projected geodesic we obtain a Riemannian metric on the 2-simplex:*

$$h(X, X) \quad = \quad X^i h_{ij} X^j$$

with

$$h_{ij}(p) = f(p) \begin{pmatrix} 3\,p_1 - 3\,{p_1}^2 - {p_2}^2 - 3\,p_2 p_1 + p_2 & \frac{1}{2}(3\,p_1 - p_1 p_2 - 3{p_1}^2) \\ \frac{1}{2}(3\,p_1 - p_1 p_2 - 3{p_1}^2) & p_1 - {p_1}^2 \end{pmatrix},$$

with $f(p) = \frac{1}{27 p_1 p_2 (1 - p_1 - p_2)}$

Proof. Note that by executing the Maple script (see chapter 6) we obtain that $\lim_{t\to 0} \frac{1}{t} d_W(\Gamma_t, \mu^{\underline{w}})$ equals

$$\frac{\sqrt{3}}{9}\sqrt{\frac{(3\,p_1 - 3\,{p_1}^2 - {p_2}^2 - 3\,p_2 p_1 + p_2){X_1}^2 + (3\,p_1 - p_1 p_2 - 3{p_1}^2)X_1 X_2 + (p_1 - {p_1}^2){X_2}^2}{p_1 p_2\,(1 - p_1 - p_2)}}.$$

I.e. in a global chart we obtain a Riemannian metric

$$h_{ij}(p) := f(p) \begin{pmatrix} 3\,p_1 - 3\,{p_1}^2 - {p_2}^2 - 3\,p_2 p_1 + p_2 & \frac{1}{2}(3\,p_1 - p_1 p_2 - 3{p_1}^2) \\ \frac{1}{2}(3\,p_1 - p_1 p_2 - 3{p_1}^2) & p_1 - {p_1}^2 \end{pmatrix},$$

with $f(p) = \frac{1}{27 p_1 p_2 (1 - p_1 - p_2)}$ and it can be (numerically) verified that this matrix is positive definite in the interior of the simplex. □

5.2 (Non)-explosion of Brownian motion on the simplex with respect to projected Wasserstein metrics - Case study

The 1-simplex. Obviously the Riemannian metric g is conformally equivalent to the Euclidian metric, we denote the conformal factor by $\varphi(p) = \frac{1}{12p(1-p)}$, furthermore we write Δ_g for the Laplacian with respect to the metric g, then([GKM68] p.90)

$$\Delta_g f = \frac{1}{\varphi}\left\{\Delta f + \left(\frac{n}{2} - 1\right) g\left(\nabla f, \nabla \log \varphi\right)\right\}$$

i.e.

$$\begin{aligned} \Delta_g f(x) &= 12x(1-x)\left\{f''(x) - \frac{1}{2} g_x\left(f'(x), -\frac{12 - 24x}{(12x(1-x))^2} 12x(1-x)\right)\right\} \\ &= 12x(1-x)f''(x) + \frac{1 - 2x}{2x(1-x)} f'(x) \end{aligned}$$

Let us consider the solution of the following SDE on Σ_1

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t \qquad X_0 = x \in \overset{\circ}{\Sigma}_1 \tag{5.2}$$

we write $b(x) = \frac{1-2x}{4x(1-x)}$ and $\sigma(x) = \sqrt{12x(1-x)}$, which are Lipschitz on any compact $K \subset (0,1)$. Hence the equation (5.2) has a strong solution X_t with generator $\frac{1}{2}\Delta_g$. Let us fix $c \in (0,1)$ and define the function i.e. $a(x) = \sigma^2/2 = 12x(1-x)$

$$H : \Sigma_1 \to \mathbb{R}\cup\{\pm\infty\} \qquad H(r) = \exp\left\{-\int_c^r \frac{b(\rho)}{a(\rho)}\,d\rho\right\} \qquad s(r) = \int_c^r H(\rho)\,d\rho$$

According to([HT94] p.343) the process X_t is recurrent if and only if

$$s(0) = -\infty \qquad s(1) = \infty,$$

additionally this condition implies that X_t has infinite lifetime in $(0,1)$. Let us verify this for a constant $c = \frac{1}{10}$ and $C > 0$:

$$\begin{aligned} s(0) &= \int_c^0 \exp\left\{-\left\{\int_c^y \frac{1-2x}{48x^2(1-x)^2}\right\}\right\} dy = \int_c^0 \exp\left\{\frac{-1}{48(y-1)} + \frac{1}{48y} - \frac{25}{108}\right\} dy \\ &\leq -C \lim_{\epsilon\to 0} \int_\epsilon^c \exp\left\{\frac{1}{48y} - \frac{25}{108}\right\} = -\infty \text{ resp. } s(1) = +\infty \end{aligned}$$

We can state

Proposition 5.2.1. *g-Brownian motion is recurrent and has infinite lifetime in the interior of the 1-simplex.*

The 2-simplex. The Laplacian with respect to the metric h reads

$$\Delta_h = F(u,v)\{A(u,v)\partial_u + B(u,v)\partial_v + C(u,v)\partial_{uu} + D(u,v)\partial_{vv} + E(u,v)\partial_{uv}\}$$

Where we denote $F = \frac{-2v\sqrt{2916}}{(3u^2+3u+3uv-4v)^2}$ and

$$A(u,v) = -6u^2 + 3u + 3u^3 + 3u^2v + 2v - 6uv$$
$$B(u,v) = -3u - v - 3v^2 + 3u^2 + 3u^2v + 3uv^2$$
$$C(u,v) = -6u^3 + 3u^4 + 3u^2 + 3u^3 - 7u^2v + 4uv$$
$$D(u,v) = -18u^3 - 4v^3 + 9u^4 + 4v^2 + 9u^2 + 3uv^2 - 33u^2v - 18uv^2 + 18u^3v + 12u^2v^2 + 15uv$$
$$E(u,v) = 18u^3 - 9u^2 - 9u^4 + 24u^2v + 4uv^2 - 12u^3v - 3u^2v^2 - 12uv$$

By X_t we denote the h-Brownian motion in $\overset{\circ}{\Sigma}_2$, i.e. for all $f \in \mathcal{C}^\infty(\overset{\circ}{\Sigma}_2)$

$$f \circ X - \frac{1}{2}\int \Delta_h f \circ X dt$$

is a local martingale. In other words given a two-dimensional real Brownian motion B then X is solution of the SDE

$$dX = \beta(X)dt + \sigma(X)dB_t, \qquad X_0 = x \in \overset{\circ}{\Sigma}_2,$$

for

$$\beta = F\begin{pmatrix} A \\ B \end{pmatrix}, \qquad \sigma\sigma^* = K := F\begin{pmatrix} C & \frac{1}{2}E \\ \frac{1}{2}E & D \end{pmatrix}$$

Proposition 5.2.2. *The h-Brownian motion in $\overset{\circ}{\Sigma}_2$ explodes, i.e. it has a.s. finite lifetime.*

Proof. We will apply an application of the comparison theorem (Theorem 3.1 in [IW77]) by Ikeda-Watanabe to explosion tests for non-degenerate diffusions on open sets of $\mathbb{R}^d$. Denote by $\pi_k : \Sigma_2 \ni \underline{p} \to p_k \in \Sigma_1$ for $k = 1, 2$ the projection to the k-th face of the simplex. Some more notations are needed:

$$a_k(x) := \sum_{i,j} K_{ij}(x) \frac{\partial \pi_k}{\partial x_i} \frac{\partial \pi_k}{\partial x_j} = \sum_{i,j} K_{ij}(x) \delta_{ik} \delta_{jk} = K_{kk};$$

$$b_k(x) := a_k(x)^{-1} \sum_i \beta_i(x) \frac{\partial \pi_k}{\partial x_i} = K_{kk}^{-1} \beta_k$$

$$a_k^+(\xi) = \sup_{x \in \overset{\circ}{\Sigma}_2 : \pi_k(x) = \xi} a_k(x); \qquad a_k^-(\xi) = \inf_{x \in \overset{\circ}{\Sigma}_2 : \pi_k(x) = \xi} a_k(x)$$

$$b_k^+(\xi) = \sup_{x \in \overset{\circ}{\Sigma}_2 : \pi_k(x) = \xi} b_k(x); \qquad b_k^-(\xi) = \inf_{x \in \overset{\circ}{\Sigma}_2 : \pi_k(x) = \xi} b_k(x)$$

We denote ξ_k^{++} resp. ξ_k^{--} for diffusion processes on $(0, 1)$ with generators

$$L_k^{++} = a_k^+ \left(\frac{1}{2} \frac{d^2}{d\xi^2} + b_k^+ \frac{d}{d\xi} \right); \qquad L_k^{--} = a_k^- \left(\frac{1}{2} \frac{d^2}{d\xi^2} + b_k^- \frac{d}{d\xi} \right)$$

and with explosion times e_k^+ resp. e_k^-. By Ikeda-Watanabe we know $e_k^+ < e_{\pi_k(X)} < e_k^-$. But for the diffusion processes ξ_k^{++} resp. ξ_k^{--} we can apply the Feller test. We will show that for $k = 1, 2$ the explosion times e_k^- are finite: Define

$$H_k(r) := \exp \left\{ - \int_c^r \frac{a_k^-(\rho) b_k^-(\rho)}{\frac{1}{2} a_k^-(\rho)} d\rho \right\} = \exp \left\{ - \int_c^r 2 b_k^-(\rho) d\rho \right\}$$

We have to show that

$$\int_c^{c_i} H_k(r) \left\{ \int_c^r \frac{1}{a_k^-(\rho) H_k(\rho)} d\rho \right\} dr < \left\{ \begin{array}{ll} \infty & c_1 = 0 \\ \infty & c_2 = 1 \end{array} \right.$$

Let us treat

$$\int_c^r b_1^-(\rho) d\rho = \int_c^r \inf_{x_2 \in (0,1)} \frac{A(\rho, x_2)}{C(\rho, x_2)} d\rho$$

$$b_1^-(r) \approx -O(\frac{1}{r})$$

Asymptotically this means that $H_1(r) \approx O(r)$ and so $\int_c^{c_i} H_1(r) \left\{ \int_c^r \frac{1}{a_1^-(\rho) H_1(\rho)} d\rho \right\} dr \approx \int O(r \log r) \approx c_i^2 \log c_i + c_i^2$ which is finite when evaluated at the end points 0 and 1.

In the same manner

$$\int_c^r b_2^-(\rho) d\rho = \int_c^r \inf_{x_1 \in (0,1)} \frac{B(x_1, \rho)}{D(x_1, \rho)} d\rho$$

$$b_2^-(r) \approx O(\frac{1}{r})$$

Asymptotically this means that $H_2(r) \approx O(\frac{1}{r})$ and so $\int_c^{c_i} H_2(r) \left\{ \int_c^r \frac{1}{a_2^-(\rho)H_2(\rho)} d\rho \right\} dr \approx \int O(\frac{1}{r^3}) \approx \frac{1}{c_i^2}$ which is finite when evaluated at the end point 1 but infinite at 0. To prove finite lifetime of ξ_2^{--} it is sufficient to know that $\int_c^0 H_2(\rho)d\rho = -\infty$ which is the case since $H_2(r) \approx O(\frac{1}{r})$.

□

Chapter 6

Maple worksheets

Riemannian metrics on the 2-simplex via symbolic computation of Wasserstein distance

In part 1) we perform symbolic computions of the Wasserstein distance between two histograms on the unit interval [0, 1] with partition points {0, 1/3, 2/3, 1}.
In part 2) we perform symbolic computions of a Riemannian metric obtained by projecting a Wasserstein geodesic between two histograms on the unit interval [0, 1] with partition points {0, 1/3, 2/3, 1}.

$N := 4 :; \ with(LinearAlgebra) :; with(DifferentialGeometry) :; with(Tensor) :;$

1) Symbolic computation of Wasserstein distances

$V := [0, v_1, v_2, 1 - v_1 - v_2] :; W := [0, w_1, w_2, 1 - w_1 - w_2] :;$

(1)

$$g(i, k, V, W) := \begin{cases} 1 & \sum_{j=1}^{k} W[j] < \sum_{j=1}^{i+1} V[j] \textbf{ and } \sum_{j=1}^{i} V[j] < \sum_{j=1}^{k+1} W[j] \\ 0 & \textbf{otherwise} \end{cases} :$$

$$J1(k, V) := \left(\frac{\left(\sum_{j=1}^{k+1} V[j]\right)^3 - \left(\sum_{j=1}^{k} V[j]\right)^3}{3 \cdot (V[k+1])^2} + \frac{\left(k \cdot V[k+1] - \sum_{j=1}^{k+1} V[j]\right) \cdot \left(\left(\sum_{j=1}^{k+1} V[j]\right)^2 - \left(\sum_{j=1}^{k} V[j]\right)^2\right)}{(V[k+1])^2} + \frac{\left(k \cdot V[k+1] - \sum_{j=1}^{k+1} V[j]\right)^2}{V[k+1]} \right) :$$

$$J2(k,W) := \left(\frac{\left(\sum_{j=1}^{k+1} W[j]\right)^3 - \left(\sum_{j=1}^{k} W[j]\right)^3}{3\cdot(W[k+1])^2} + \frac{\left(k\cdot W[k+1] - \sum_{j=1}^{k+1} W[j]\right)\cdot\left(\left(\sum_{j=1}^{k+1} W[j]\right)^2 - \left(\sum_{j=1}^{k} W[j]\right)^2\right)}{(W[k+1])^2} + \frac{\left(k\cdot W[k+1] - \sum_{j=1}^{k+1} W[j]\right)^2}{W[k+1]} \right):$$

$$H(i,k,V,W) := \left(\frac{\left(\min\left(\sum_{j=1}^{i+1} V[j], \sum_{j=1}^{k+1} W[j]\right)\right)^3 - \left(\max\left(\sum_{j=1}^{i} V[j], \sum_{j=1}^{k} W[j]\right)\right)^3}{3\cdot V[i+1]\cdot W[k+1]} + \frac{\left(\min\left(\sum_{j=1}^{i+1} V[j], \sum_{j=1}^{k+1} W[j]\right)\right)^2 - \left(\max\left(\sum_{j=1}^{i} V[j], \sum_{j=1}^{k} W[j]\right)\right)^2}{2\cdot V[i+1]\cdot W[k+1]} \cdot \left(\mathrm{i}\cdot V[i+1] - \sum_{j=1}^{i+1} V[j] + k\cdot W[k+1] - \sum_{j=1}^{k+1} W[j]\right) + \frac{\left(\min\left(\sum_{j=1}^{i+1} V[j], \sum_{j=1}^{k+1} W[j]\right) - \max\left(\sum_{j=1}^{i} V[j], \sum_{j=1}^{k} W[j]\right)\right)}{V[i+1]\cdot W[k+1]} \cdot \left(\mathrm{i}\cdot V[i+1] - \sum_{j=1}^{i+1} V[j]\right)\cdot\left(k\cdot W[k+1] - \sum_{j=1}^{k+1} W[j]\right) \right):$$

$$JJ1(k,V) := \begin{cases} \mathrm{J1}(\mathrm{k},\mathrm{V}) & \mathrm{V}[\mathrm{k}+1] > 0 \\ 0 & V[k+1] = 0 \end{cases}:$$

$$JJ2(k,W) := \begin{cases} \mathrm{J2}(\mathrm{k},\mathrm{W}) & \mathrm{W}[\mathrm{k}+1] > 0 \\ 0 & \mathrm{W}[\mathrm{k}+1] = 0 \end{cases}:$$

$$HH(i,k,V,W) := \begin{cases} H(i,k,V,W) & W[k+1] > 0 \ \mathbf{and}\ \mathrm{V}[\mathrm{i}+1] > 0 \\ 0 & \mathit{otherwise} \end{cases}:$$

$$dist(n,V,W) := \frac{1}{n^2}\sum_{k=1}^{n} JJ1(k,V) + \frac{1}{n^2}\sum_{k=1}^{n} JJ2(k,W) - 2\cdot\frac{1}{n^2}\cdot\sum_{k=1}^{n}\left(\sum_{i=1}^{n}(HH(i,k,V,W)\cdot g(i,k,V,W))\right):;$$

$sqd := unapply(dist(3, V, W), v_1, v_2, w_1, w_2)$ assuming $0 < v_1 < 1$ **and** $0 < v_2 < 1$ **and** $0 < w_1 < 1$ **and** $0 < w_2 < 1$:;

Distances between different points in the simplex are studied and compared:

$A1 := simplify(\mathrm{sqd}(v_1, v_2, w_1, w_2))$ assuming $0 < v_1 \le v_2 \le w_1 \le w_2 \le 1$ **and** $1 - v_1 - v_2 > 0$ **and** $1 - w_1 - w_2 > 0$ **and** $w_1 < v_1 + v_2$:; $AA1 := simplify(\mathrm{sqd}(v_1, v_2, w_1, w_2))$ assuming $0 < v_1 < v_2 < w_1 < w_2 < 1$ **and** $1 - v_1 - v_2 > 0$ **and** $1 - w_1 - w_2 > 0$ **and** $w_1 < v_1 + v_2$:;

$$\frac{1}{27}\,\frac{1}{w_2 v_2 w_1 (v_1 + v_2 - 1)}\big(w_1^4 + 6 w_1^2 v_1 v_2 - v_1^2 w_2 v_2 - 3 v_1^2 w_1 v_2 - 3 v_1 v_2^2 w_1 - 4 v_2 w_1^2 w_2 - v_2 w_1 w_2^2 + 9 v_1 v_2 w_1^3 - 12 v_1^2 w_1^2 v_2 - 15 v_1 v_2^2 w_1^2 + 3 w_1^2 v_1 w_2 + w_1^3 v_1 w_2 - 3 v_2 w_1^3 w_2 - 5 v_2 w_1^2 w_2^2 - 2 v_2 w_2^3 w_1 - 3 v_1^2 w_2 w_1^2 - 3 v_1^2 w_2 w_1 + 12 v_2^2 w_2 w_1^2 + 6 v_2^2 w_2^2 w_1 + 2 v_2^2 w_2 w_1 + 3 v_1^3 w_2 w_1 - 6 v_2^3 w_2 w_1 + v_1^2 w_2 v_2^2 + 5 v_1^3 w_1 v_2 + 9 v_1^2 w_1 v_2^2 + 7 v_2^3 w_1 v_1 + 9 v_1 v_2 w_2 w_1^2 + 6 v_1 v_2 w_2^2 w_1 + 5 v_1 v_2 w_2 w_1 - 6 v_1^2 w_2 v_2 w_1 - 15 v_1 v_2^2 w_2 w_1 - 3 v_2 w_1^3 - v_1^4 w_2 + v_1^4 w_1 + 2 v_2^4 w_1 + 3 w_1^2 v_1^2 + 3 v_2^2 w_1^2 + v_1^3 w_2 - v_1^3 w_1 - v_2^3 w_1 - w_1^3 w_2 - 3 w_1^3 v_1 - w_1^4 v_1 - 2 v_2 w_1^4 + 3 v_1^2 w_1^3 + 6 v_2^2 w_1^3 - 3 v_1^3 w_1^2 - 6 v_2^3 w_1^2\big) \qquad \textbf{(2)}$$

$A2 := simplify(\mathrm{sqd}(v_1, v_2, w_1, w_2))$ assuming $0 < v_1 < w_1 < v_2 < w_2 < 1$ **and** $1 - v_1 - v_2 > 0$ **and** $1 - w_1 - w_2 > 0$ **and** $w_1 < v_1 + v_2$:; $A2 - A1$:;

$A3 := simplify(\mathrm{sqd}(v_1, v_2, w_1, w_2))$ assuming $0 < v_1 < w_1 < v_2 < w_2 < 1$ **and** $1 - v_1 - v_2 > 0$ **and** $1 - w_1 - w_2 > 0$ **and** $w_2 < v_1 + v_2$:;

$A4 := simplify(\mathrm{sqd}(v_1, v_2, w_1, w_2))$ assuming $0 < v_1 < v_2 < w_1 < w_2 < 1$ **and** $1 - v_1 - v_2 > 0$ **and** $1 - w_1 - w_2 > 0$ **and** $w_1 > v_1 + v_2$:;
$A5 := simplify(\mathrm{sqd}(v_1, v_2, w_1, w_2))$ assuming $0 < v_1 < v_2 < w_1 < w_2 < 1$ **and** $1 - v_1 - v_2 > 0$ **and** $1 - w_1 - w_2 > 0$ **and** $w_1 < v_1 + v_2$ **and** $w_2 > w_1 + v_2$:;
$A6 := simplify(\mathrm{sqd}(v_1, v_2, w_1, w_2))$ assuming $0 < v_1 < w_1 < v_2 < w_2 < 1$ **and** $1 - v_1 - v_2 > 0$ **and** $1 - w_1 - w_2 > 0$ **and** $w_2 > v_1 + v_2$:;

$A7 := simplify(\mathrm{sqd}(v_1, v_2, w_1, w_2))$ assuming $0 < w_1 < v_1 < v_2 < w_2 < 1$ **and** $1 - v_1 - v_2 > 0$ **and** $1 - w_1 - w_2 > 0$ **and** $w_2 < v_1 + v_2$ **and** $w_1 + w_2 < v_1 + v_2$:; $A8$

$:= simplify(\mathrm{sqd}(v_1, v_2, w_1, w_2))$ assuming $0 < w_1 < v_1 < v_2 < w_2 < 1$ **and** $1 - v_1 - v_2 > 0$ **and** $1 - w_1 - w_2 > 0$ **and** $w_1 + w_2 > v_1 + v_2$:;

$A9 := simplify(\mathrm{sqd}(v_1, v_2, w_1, w_2))$ assuming $0 < v_1 < w_2 < w_1 < v_2 < 1$ **and** $1 - v_1 - v_2 > 0$ **and** $1 - w_1 - w_2 > 0$ **and** $w_2 < v_1 + v_2$ **and** $v_1 + v_2 > w_1 + w_2$:; $A9 - A4$:; $A7 - A1$:;

$AA := simplify(\mathrm{sqd}(v_1, v_2, w_1, w_2))$ assuming $0 < v_1 \le v_2 \le w_1 \le w_2 < 1$ **and** $1 - v_1 - v_2 > 0$ **and** $1 - w_1 - w_2 > 0$ **and** $w_1 \le v_1 + v_2$:;

$a4 := unapply(A8, v_1, v_2, w_1, w_2)$:; $a1 := unapply(A1, v_1, v_2, w_1, w_2)$:; $a2 := unapply(A4, v_1, v_2, w_1, w_2)$:; $a3 := unapply(A7, v_1, v_2, w_1, w_2)$:

2) Symbolic computation of the Riemannian metric

$$simplify\left(\lim_{t \to 0} \sqrt{\frac{1}{t^2} a1(u + t \cdot X_1, v + t \cdot X_2, u, v)}\right)$$

$$\frac{1}{9} \left(\frac{1}{(u+v-1)\, v\, u} \left(-v X_1^2 + X_1^2 v^2 - 3 X_1 u X_2 + 3 X_1 u^2 X_2 + u v X_1 X_2 + u^2 X_2^2 + 3 u^2 X_1^2 - 3 u X_1^2 + 3 v u X_1^2 - u X_2^2\right)\right)^{1/2} \sqrt{3} \tag{3}$$

$$simplify\left(\lim_{t \to 0} \sqrt{\frac{1}{t^2} a3(u + t \cdot X_1, v + t \cdot X_2, u, v)}\right)$$

$$\frac{1}{9} \left(\frac{1}{(u+v-1)\, v\, u} \left(-v X_1^2 + X_1^2 v^2 - 3 X_1 u X_2 + 3 X_1 u^2 X_2 + u v X_1 X_2 + u^2 X_2^2 + 3 u^2 X_1^2 - 3 u X_1^2 + 3 v u X_1^2 - u X_2^2\right)\right)^{1/2} \sqrt{3} \tag{4}$$

$$simplify\left(\lim_{t \to 0} \sqrt{\frac{1}{t^2} a4(u + t \cdot X_1, v + t \cdot X_2, u, v)}\right)$$

$$\frac{1}{9}\left(\frac{1}{(u+v-1)\,v\,u}\left(-v X_1^2 + X_1^2 v^2 - 3 X_1 u X_2 + 3 X_1 u^2 X_2 + u v X_1 X_2 + u^2 X_2^2 + 3 u^2 X_1^2 - 3 u X_1^2 + 3 v u X_1^2 - u X_2^2\right)\right)^{1/2} \sqrt{3} \qquad (5)$$

Bibliography

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows im metric spaces and in the space of probability measures.* Lectures in Mathematics, ETH Zürich. Birkhäuser Verlag, 2008.

[Apo76] T. Apostol. *Introduction to Analytic Number Theory.* Undergraduate texts in Mathematics. Springer, 1976.

[AS70] M. Abramowitz and I. Stegun. *Handbook of mathematical functions.* Dover, 1970.

[Bax84] P. Baxendale. Brownian motions in diffeomorphism group i. *Compositio Math.*, 53:19–50, 1984.

[BB00] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numer. Math.*, 84:375–393, 2000.

[BBS10] L. Brasco, G. Buttazzo, and F. Santambrogio. A Benamou-Brenier approach to branched transport. *preprint*, 2010.

[BE85] D. Bakry and M. Émery. Diffusions hypercontractives. *Séminaire de probabilités de Strasbourg*, 19:177–206, 1985.

[Bre91] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44:375–417, 1991.

[Daw93] D. Dawson. Measure-valued Markov processes. In *École d'Été de Probabilités de Saint-Flour XXI-1991*, Lecture Notes in Mathematics. Springer, 1993.

[EK05] S. Ethier and T. Kurtz. *Markov processes. Characterization and convergence.* Wiley-Interscience, 2005.

[Fan02] S. Fang. Canonical Brownian Motion on the Diffeomorphism Group of the Circle. *Journal of Functional Analysis*, 196:162–179, 2002.

[Fig07] A. Figalli. Existence, uniqueness and regularity of optimal transport maps. *SIAM J. Math. Anal.*, 39, 2007.

[FOT94] M. Fukushima, Y. Oshima, and M. Takeda. *Dirichlet forms and symmetric Markov processes.* de Gruyter, 1994.

[Gig09] N. Gigli. Second order analysis on $(P_2(M), W_2)$. *preprint*, 2009.

[GKM68] D. Gromoll, W. Klingenberg, and W. Meyer. *Riemannsche Geometrie im Großen*. Lecture notes in Mathematics. Springer, 1968.

[GKP10] W. Gangbo, H. K. Kim, and T. Pacini. Differential forms on Wasserstein space and infinite-dimensional Hamiltonian systems. *to appear in Memoirs AMS*, 2010.

[HT94] W. Hackenbrock and A. Thalmaier. *Stochastische Analysis. Eine Einführung in die Theorie der stetigen Semimartingale*. Mathematische Leitfäden. Teubner, 1994.

[Hux03] M. N. Huxley. Exponential sums and lattice points III. *Proc. London Math. Soc.*, 87:5910–609, 2003.

[IW77] N. Ikeda and S. Watanabe. A comparison theorem for solutions of stochastic differential equations and its applications. *Osaka J. Math.*, 14(3):619–633, 1977.

[IW89] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North Holland Mathematical Library. North Holland, 1989.

[KM97] A. Kriegl and P. Michor. *A convenient setting of global analysis*, volume 53 of *Mathematical Surveys and Monographs*. AMS, 1997.

[Kun90] H. Kunita. *Stochastic flows and stochastic differential equations*. Cambridge studies in advanced mathematics. Cambridge University Press, 1990.

[Lot08] J. Lott. Some geometric calculations on Wasserstein space. *Comm. Math. Phys.*, 277:423–437, 2008.

[LR04] Y. LeJan and O. Raimond. Flows, coalescence and noise. *Ann. Probab.*, 2:1247–1315, 2004.

[LR05] Y. LeJan and O. Raimond. Stochastic flows on the circle. In *Probability and partial differential equations in modern applied mathematics*, IMA Vol. Math. Appl., 140. Springer, 2005.

[LV09] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Math.*, 169:903–991, 2009.

[Mal99] P. Malliavin. The canonic diffusion above the diffeomorphism group of the circle. *C.R. Acad. Sci. Paris*, 329:325–329, 1999.

[Mal08] P. Malliavin. Invariant or quasi-invariant probability measures for infinite dimensional groups. *Japan. J. Math.*, 3:1–17, 2008.

[McC01] R. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11:589–608, 2001.

[MH08] J. Mattingly and M. Hairer. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Annals of Probability*, 6:993–1032, 2008.

[Omo78] H. Omori. On Banach Lie groups acting on finite dimensional manifolds. *Tohoku Math. J.*, 30:223–250, 1978.

[Ott01] F. Otto. The geometry of dissipative evolution equations: The porous medium equation. *Comm. Partial Differential Equations*, 26:101–174, 2001.

[RR98] S. Rachev and L. Rüschendorf. *Mass transportation problems. Volume I: Theory*. Springer, 1998.

[Sel06] C. Selinger. Gradient flows on the space of probability measures. On differential-geometric aspects of optimal transport. Master's thesis, Universität Wien, unpublished, 2006.

[Stu06a] K.-Th. Sturm. On the geometry of metric measure spaces. *Acta Math.*, 196:65–131, 2006.

[Stu06b] K.-Th. Sturm. On the geometry of metric measure spaces II. *Acta Math.*, 196:133–177, 2006.

[SvR09] K.-Th. Sturm and M.-K. von Renesse. Entropic measure and Wasserstein diffusion. *Ann. Prob.*, 37:1114–1191, 2009.

[Vil03] C. Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. AMS, 2003.

[Vil08] C. Villani. *Optimal transport, old and new*. Grundlehren der mathematischen Wissenschaften. Springer, 2008.