

A BRENIER THEOREM FOR MEASURES ON $(\mathcal{P}_2(\dots \mathcal{P}_2(H) \dots), \mathcal{W}_2)$ AND APPLICATIONS TO ADAPTED TRANSPORT

MATHIAS BEIGLBÖCK, GUDMUND PAMMER, AND STEFAN SCHROTT

ABSTRACT. We develop Brenier theorems on iterated Wasserstein spaces. For a separable Hilbert space H and $N \geq 1$, we construct a full-support probability Λ on $\mathcal{P}_2^N(H) = \mathcal{P}_2(\dots \mathcal{P}_2(H) \dots)$ that is transport regular: for every Q with finite second moment, transporting Λ to Q with cost $\mathcal{W}_2^2$ admits a unique optimizer, and this optimizer is of Monge type. The analysis rests on a characterization of optimal couplings on $\mathcal{P}_2(H)$ and, more generally, on $\mathcal{P}_2^N(H)$ via convex potentials on the Lions lift; in the latter case we employ a new adapted version of the lift tailored to the N -step structure. A key idea is a new identification between optimal-transport c -conjugation (with c given by maximal covariance) and classical convex conjugation on the lift.

A primary motivation comes from the adapted Wasserstein distance $\mathcal{AW}_2$: our results yield a first Brenier theorem for $\mathcal{AW}_2$ and characterize $\mathcal{AW}_2^2$ -optimal couplings through convex functionals on the space of L_2 -processes.

keywords: optimal transport for random measures, Brenier's theorem, Lions lift, convex conjugate, Monge problem, adapted optimal transport.

1. INTRODUCTION

Let $(\mathcal{X}, d)$ be a Polish metric space and write $\mathcal{P}_2(\mathcal{X})$ for the probabilities with finite second moments. For $\mu, \nu \in \mathcal{P}_2(\mathcal{X})$ we write $\text{Cpl}(\mu, \nu)$ for the set of couplings or transport plans and

$$(1) \quad \mathcal{W}_2^2(\mu, \nu) = \inf_{\pi \in \text{Cpl}(\mu, \nu)} \int d^2 d\pi = \inf_{X, Y: (\Omega, \mathbb{P}) \rightarrow \mathcal{X}, X \sim \mu, Y \sim \nu} \mathbb{E}[d(X, Y)^2]$$

for the squared Wasserstein-2 distance. Brenier's theorem [26], see also [55, 72], asserts that if $\mathcal{X}$ is the Euclidean space and $\mu \ll \lambda$, the minimization problem (1) admits a unique optimizer π^* and that π^* is of Monge type, i.e., concentrated on the graph of a function. Moreover, the optimal Monge-transport is the gradient of a convex function. This result stands at the beginning of the breathtaking development of optimal transport over the last decades and has inspired a number of powerful generalizations and refinements, see [46, 61, 43, 62, 9, 4, 34, 39] among others.

We extend it to iterated Wasserstein spaces, i.e., we consider $(\mathcal{X}, d) = (\mathcal{P}_2(H), \mathcal{W}_2)$ and more generally $(\mathcal{P}_2(\dots \mathcal{P}_2(H) \dots), \mathcal{W}_2) = (\mathcal{P}_2^N(H), \mathcal{W}_2)$ where H denotes a separable Hilbert space.

Theorem 1.1 (Monge solutions). *There exists a full support probability Λ on $\mathcal{P}_2^N(H)$ which is transport regular. That is, for every probability Q on $\mathcal{P}_2^N(H)$ with finite second moment*

$$(2) \quad \inf_{\Pi \in \text{Cpl}(\Lambda, Q)} \int \mathcal{W}_2^2(\mu, \nu) d\Pi(\mu, \nu) = \inf_{X, Y: (\Omega, \mathbb{P}) \rightarrow \mathcal{P}_2^N(H), X \sim \Lambda, Y \sim Q} \mathbb{E}[\mathcal{W}_2^2(X, Y)]$$

admits a unique minimizer $\Pi^ \sim (X, Y)$ and Π^* is of Monge type.*

In fact we obtain here transport regularity of all orders, i.e., Λ is concentrated on probabilities μ which are again transport regular, etc. Since every $P \ll \Lambda$ is also transport regular, we have

We thank Francois Delarue, Lorenzo Dello Schiavo, Martin Huesmann and Daniel Lacker for insightful discussions.

Corollary 1.2. *The set of $\mathcal{W}_2$ -transport regular measures on $\mathcal{P}_2^N(H)$ is dense.*

Most closely related to our work are the concurrently written article of Savare–Pinzi [71] (see Section 1.4 below) and the work of Emami–Pass [39] which is concerned with $\mathcal{P}_2(M)$, where M is a smooth Riemannian manifold. As pointed out in [39] in the setting $X = \mathcal{P}_2(M)$, the choice of an appropriate reference measure is non-trivial. Recent deep work of Dello Schiavo [37] establishes a version of Rademacher’s theorem on Wasserstein spaces for a class of reference measures satisfying certain assumptions which allows Emami–Pass to conclude transport regularity. It can be shown that probabilities satisfying these assumptions exist in the case where M is the 1-dimensional torus but it seems difficult to extend the approach to higher dimensions and in particular the case $\mathcal{P}_2^N(M)$, for $N \geq 2$. Rather the approach put forward here and in [71] exploits a connection to the Lions lift which allows to use tools from convex analysis.

We also note that while our main interest lies in results for $\mathcal{P}_2^N(\mathbb{R}^d)$ and their applications to the adapted Wasserstein distance, we will work with a separable Hilbert space H throughout. The reason is that the proofs for both cases are identical, in fact our argument for the existence of transport regular measures on $\mathcal{P}_2^2(\mathbb{R}^d)$ already requires the construction of a transport regular measure on $\mathcal{P}_2(H)$ for infinite dimensional H .

Next we describe the construction of the measure Λ and the characterization of optimizers in terms convex functions on the Lions’ lift.

1.1. Characterization of optimal transport maps via convex Lions lifts. To set the stage we recall the formulation of Brenier’s theorem for measures on H in terms of couplings of random variables. We use Λ to denote Lebesgue measure in the classical case $H = \mathbb{R}^d$ (see [26]) or a regular Gaussian measure in the case of infinite-dimensional H (see [8]). If $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow H$ is a random variable with $\mathcal{L}_{\mathbb{P}}X \ll \Lambda$ and $\varphi : H \rightarrow (-\infty, \infty]$ is convex with $\partial\varphi(X) \neq \emptyset$ a.s., then $\partial\varphi(X)$ consists of exactly one element which we denote by $\nabla\varphi(X)$.

Theorem 1.3 (Brenier’s Theorem [26, 8]). *Let $\pi \in \text{Cpl}(\mu, \nu)$ where μ, ν are probability measures on H with finite second moment, $\mu \ll \Lambda$, and let $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow H$ be a random variable with $X \sim \mu$. Then π is an optimal coupling for squared-distance cost if and only if*

$$(3) \quad \pi \sim (X, \nabla\varphi(X))$$

for some convex lsc $\varphi : H \rightarrow (-\infty, \infty]$ with $\partial\varphi(X) \neq \emptyset$ a.s.

Moreover we have uniqueness, i.e., there is exactly one such coupling.

Following classical transport theory, optimal plans are characterized in terms of dual potentials. A key finding (see Theorem 1.8 below) is that for optimal transport of measures on $\mathcal{P}_2(H)$, optimal dual potentials $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, \infty]$ are best understood through their *Lions lift* given by

$$\bar{\varphi} : L_2([0, 1]; H) \rightarrow (-\infty, \infty], \quad \bar{\varphi}(X) := \varphi(\mathcal{L}_\lambda(X)).$$

Here we use $L_2([0, 1]; H)$ to denote the L_2 -space of H -valued square-integrable functions on $([0, 1], \lambda)$, where λ is the Lebesgue measure on $[0, 1]$. Evidently a functional $\bar{\varphi}$ on $L_2([0, 1]; H)$ appears as the Lions lift of some function φ on $\mathcal{P}_2(H)$ if and only if it is *law invariant*, i.e., $\bar{\varphi}(X) = \bar{\varphi}(Y)$ provided that $\mathcal{L}_\lambda(X) = \mathcal{L}_\lambda(Y)$.

In the case $H = \mathbb{R}^d$ we take Λ to be the *occupation measure* of d -dimensional Brownian motion $B = (B_t)_{t \in [0, 1]}$. Recall that for a path $b : [0, 1] \rightarrow \mathbb{R}^d$ of Brownian motion, the occupation measure of b is given by $\mathcal{L}_\lambda b = b_\#(\lambda) \in \mathcal{P}_2(\mathbb{R}^d)$ and the occupation measure of B has law

$$(4) \quad \Lambda := (\omega \mapsto \mathcal{L}_\lambda(B(\omega)))_\#(\mathbb{P}) = \mathcal{L}_{\mathbb{P}}(\mathcal{L}_\lambda B) \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d)).$$

In the case where H has infinite dimension, we take Λ to be the law of the occupation measure of a standard Wiener process, see Section 3.2 below.

The desired Brenier theorem for measures on $\mathcal{P}_2(H)$ is then obtained from Theorem 1.3 by replacing H with $L_2([0, 1]; H)$ in the domain of the convex potential and the state space of random variables, resp.

Theorem 1.4 (Brenier theorem on $\mathcal{P}_2(H)$). *Let $\Pi \in \text{Cpl}(P, Q)$, where P and Q are probability measures on $\mathcal{P}_2(H)$ with finite second moments, $P \ll \Lambda$, and let $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow L_2([0, 1]; H)$ be a random variable such that $X \sim P$. Then Π is a $\mathcal{W}_2^2$ -optimal coupling if and only if*

$$(5) \quad \Pi \sim (\mathcal{L}_\lambda X, \mathcal{L}_\lambda \nabla \varphi(X))$$

for a convex, lsc, law-invariant $\varphi : L_2([0, 1]; H) \rightarrow (-\infty, \infty]$ with $\partial \varphi(X) \neq \emptyset$ a.s.

Moreover we have uniqueness, i.e., there is exactly one such coupling.

To tackle the case of general $\mathcal{P}_2^N(H)$, $N \geq 1$, we introduce an adapted version of the Lions' lift. We consider the filtered probability space $([0, 1]^N, (\mathcal{F}_t)_{t=1}^N, \lambda)$, where λ denotes the uniform distribution on $[0, 1]^N$ and $(\mathcal{F}_t)_{t=1}^N$ the filtration generated by the coordinate projections. For $X, Y \in L_2([0, 1]^N; H)$ we define the *adapted law*

$$\mathcal{L}^{\text{ad}}(X) := \mathcal{L}(\mathcal{L}(\dots \mathcal{L}(X | \mathcal{F}_{N-1}) \dots | \mathcal{F}_1)) \in \mathcal{P}_2^N(H)$$

and write $X \sim_{\text{ad}} Y$ if $\mathcal{L}^{\text{ad}}(X) = \mathcal{L}^{\text{ad}}(Y)$. A basic but crucial fact which renders this representation useful is that

$$(6) \quad \mathcal{W}_2(P, Q) = \min_{X \sim_{\text{ad}} P, Y \sim_{\text{ad}} Q} \|X - Y\|_2,$$

see Proposition 5.7 below.¹ The *adapted Lions lift* of $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, \infty]$ is

$$\bar{\varphi} : L_2([0, 1]^N; H) \rightarrow (-\infty, \infty], \quad \bar{\varphi}(X) := \varphi(\mathcal{L}^{\text{ad}} X).$$

Evidently $\bar{\varphi}$ is of this form if and only if it is adapted-law invariant in the sense that $\bar{\varphi}(X) = \bar{\varphi}(Y)$ for X, Y with $X \sim_{\text{ad}} Y$.

Here Λ is from an H -valued Wiener sheet which is the N parameter version of the standard Wiener process, see Section 7 below. Using these notions we obtain:

Theorem 1.5 (N -iterated Brenier). *Let $\Pi \in \text{Cpl}(P, Q)$, where P and Q are probability measures on $\mathcal{P}_2^N(H)$ with finite second moments, $P \ll \Lambda$, and let $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow L_2([0, 1]^N; H)$ be a random variable such that $\mathcal{L}^{\text{ad}}(X) \sim P$. Then Π is a $\mathcal{W}_2^2$ -optimal coupling if and only if*

$$(7) \quad (\mathcal{L}^{\text{ad}}(X), \mathcal{L}^{\text{ad}}(\nabla \bar{\varphi}(X))) \sim \Pi,$$

for some convex, lsc, adapted-law invariant $\bar{\varphi} : L_2([0, 1]^N; H) \rightarrow (-\infty, \infty]$ with $\partial \bar{\varphi}(X) \neq \emptyset$ a.s.

Moreover we have uniqueness, i.e., there is exactly one such coupling.

A first ingredient to Theorems 1.4 and 1.5 is the fundamental theorem of optimal transport together with a new connection between c -conjugates in the sense of optimal transport and convex analysis on the Lions lift; we describe this in Section 1.2 below.

A further important role is played by the ability to switch back and forth between a stochastic process viewpoint and the Wasserstein on Wasserstein viewpoint. Specifically the map $\mathcal{L}^{\text{ad}}$ turns out to be an isometry between martingales equipped with the adapted Wasserstein distance and $(\mathcal{P}_2^N, \mathcal{W}_2)$. We discuss this connection as well as Brenier type theorems for the adapted Wasserstein distance in Section 1.3.

The construction of the transport regular measure Λ uses that convex functions on Hilbert spaces are almost surely differentiable with respect to Gaussian measures as well as path properties of Wiener processes. E.g. already for the case $N = 1$ and $H = \mathbb{R}^d$ we use the fact the for a

¹By Lisini's work [59], any $P_0, P_1 \in \mathcal{P}_2^N(H)$ are joined by a geodesic. We note that (6) yields this explicitly: Pick $X_0 \sim_{\text{ad}} P_0, X_1 \sim_{\text{ad}} P_1$ with $\mathcal{W}_2(P_0, P_1) = \|X_0 - X_1\|$ and set $P_t := \mathcal{L}^{\text{ad}}((1-t)X_0 + tX_1)$, $t \in [0, 1]$.

typical Brownian path $b : [0, 1] \rightarrow \mathbb{R}^d$, the occupation measure $\mathcal{L}_\lambda b$ is a transport regular probability on $\mathbb{R}^d$. This appears to be somewhat remarkable since the support of $\mathcal{L}_\lambda b$ has Hausdorff dimension 2 for every $d \geq 2$.

We conclude this section with a result that does not appeal to a particular reference measure. It provides a characterization of primal optimizers in terms convex potentials on $L_2([0, 1]^N; H)$ and asserts that *typically* (in the Baire category sense) couplings are of Monge type.

Theorem 1.6. *Let $\Pi \in \text{Cpl}(P, Q)$ where P, Q are probability measures on $\mathcal{P}_2^N(H)$ with finite second moment. Then Π is a $\mathcal{W}_2^2$ -optimal coupling if and only if for some $X, Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow L_2([0, 1]^N; H)$ and $\bar{\varphi} : L_2([0, 1]^N; H) \rightarrow (-\infty, \infty]$ convex lsc adapted-law invariant*

$$(8) \quad \Pi \sim (\mathcal{L}^{\text{ad}} X, \mathcal{L}^{\text{ad}} Y), \quad Y \in \partial \bar{\varphi}(X) \text{ a.s.}$$

Moreover, the set of (P, Q) for which there exists a unique $\mathcal{W}_2^2$ -optimal coupling π^ and π^* is concentrated on the graph of a bijection, is comeager in $\mathcal{P}_2^{N+1}(H) \times \mathcal{P}_2^{N+1}(H)$.*

1.2. Fundamental theorem of optimal transport, MC-convexity and L-convexity. As we are interested in the connection to convex analysis on $L_2([0, 1]^N; H)$, it is convenient to switch from the minimization of the squared-distance cost to the equivalent problem of finding the ‘maximal covariance’

$$(9) \quad \text{MC}(\mu, \nu) = \sup_{\pi \in \text{Cpl}(\mu, \nu)} \int x \cdot y \, d\pi(x, y) = \sup_{X \sim \mu, Y \sim \nu} \mathbb{E}[X \cdot Y]$$

of probabilities μ, ν on H with finite second moment. Likewise, for probabilities P, Q on $\mathcal{P}_2(H)$ with finite second moment we replace (1) by

$$(10) \quad \text{MC}(P, Q) = \sup_{\pi \in \text{Cpl}(P, Q)} \int \text{MC}(\mu, \nu) \, d\pi(\mu, \nu).$$

For $P, Q \in \mathcal{P}_2^N(H)$, $N > 2$ we define $\text{MC}(P, Q)$ through the natural iteration of (10).

Optimal transport plans can be characterized in terms of dual potentials. This is made precise in the ‘fundamental theorem of optimal transport’ (see e.g. [8] or [75]) which we recall in a form convenient for MC costs. For continuous symmetric $c : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ the c -convex conjugate of $\varphi : \mathcal{X} \rightarrow [-\infty, +\infty]$ is

$$\varphi^c(y) := \sup_{x \in \mathcal{X}} c(x, y) - \varphi(x).$$

Here, φ is called c -convex if $\varphi^{cc} = \varphi$ and each such φ is lsc. We also note that $\varphi^{ccc} = \varphi^c$. The subdifferential of a c -convex φ is

$\partial_c \varphi := \{(x, y) : \varphi^c(x) + \varphi(y) = c(x, y)\} = \{(x, y) : c(x, y) - \varphi(x) \geq c(z, y) - \varphi(z) \text{ for all } z \in X\}$, and $\partial_c \varphi(x) = \{y : (x, y) \in \partial_c \varphi\}$. The primal and dual transport problem w.r.t. c are given by

$$(11) \quad \text{OT}_{\text{primal}} := \sup_{\pi \in \text{Cpl}(\mu, \nu)} \int c \, d\pi, \quad \text{OT}_{\text{dual}} := \inf_{\varphi \in L^1(\mu), \varphi \text{ lsc}} \int \varphi \, d\mu + \int \varphi^c \, d\nu.$$

Theorem 1.7 (Fundamental theorem of optimal transport). *Assume that $c : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is symmetric, continuous and satisfies $|c| \leq a \oplus b$, $a \in L^1(\mu), b \in L^1(\nu)$. The primal problem is attained and the dual problem is attained by a c -convex function φ . Moreover for any such φ we have that $\pi \in \text{Cpl}(\mu, \nu)$ is optimal for the primal problem if and only if $\text{supp}(\pi) \subseteq \partial_c \varphi$.*

In the setting of Brenier’s theorem, $c(x, y) = x \cdot y$, c -convexity is just ordinary convexity and Theorem 1.7 yields that transport plans are optimal if and only if they are concentrated on the subgradient of a convex function.

In the case $\mathcal{X} = \mathcal{P}_2(H)$ and $c(\mu, \nu) = \text{MC}(\mu, \nu)$, Theorem 1.7 asserts that transport plans are optimal if and only if they are concentrated on the MC-subgradients of MC-convex functions.

The link to convex law invariant functions as in Theorem 1.4 is given in the following result which also appears to be of independent interest:

Theorem 1.8. *Let $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$, $N \geq 1$ be proper. Then*

$$\overline{\varphi}^* = \overline{\varphi^{\text{MC}}}.$$

In particular, the convex conjugate of an adapted-law invariant function is adapted-law invariant.

Theorem 1.8 provides a bridge between the abstract conjugates on $\mathcal{P}_2^N(H)$ and classical convex analysis on the Hilbert space $L_2([0, 1]^N; H)$. In particular, convex functions on $L_2([0, 1]^N; H)$ are significantly more tractable and amenable to the standard tools of convex analysis, making this identification conceptually and technically valuable.

In the case $N = 1$, convexity of the lift $\overline{\varphi}$ is tantamount to convexity of φ along all curves of the form

$$(12) \quad (\mu_t)_{t \in [0, 1]}, \quad \mu_t := ((1 - t) \text{pr}_0 + t \text{pr}_1)_\#(\pi), \pi \in \text{Cpl}(\mu, \nu).$$

Functionals with this property are called totally convex [32] or acceleration-free convex [66] or simply convex (in the context of risk measures). Total convexity is equivalent to displacement convexity if $\varphi : \mathcal{P}_2(H) \rightarrow \mathbb{R}$ is continuous [32], but not in general ([66]). By Theorem 1.8 we have

Corollary 1.9. *For a lsc functional $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, \infty]$, the following are equivalent:*

- (1) φ is totally convex.
- (2) $\overline{\varphi}$ is convex.
- (3) φ is MC-convex.

1.3. Stochastic processes and Brenier theorems for the adapted Wasserstein distance.

Authors from several fields have independently refined the weak topology on the laws of stochastic processes in order to take the temporal flow of information into account, see [5, 48, 49, 24, 67, 68, 64] among others. In finite discrete time, all these approaches define the same *adapted weak topology* on $\mathcal{P}_2(H^N)$ which is metrized by the adapted Wasserstein distance $\mathcal{AW}_2$, see [11, 65, 19].

The Monge-formulation of the (squared) adapted Wasserstein distance of $\mu, \nu \in \mathcal{P}_2(H)$ is

$$(13) \quad \mathcal{AW}_{2, \text{Monge}}^2(\mu, \nu) = \inf_{T: H^N \rightarrow H^N, T_\#(\mu) = \nu, T \text{ bi-triangular}} \int |T(x) - x|^2 d\mu(x).$$

Here $T : H^N \rightarrow H^N$ is called triangular (or adapted) if each T_k does depend on $(x_1, \dots, x_n)$ only through the first k coordinates. It is bi-triangular if it is invertible and T, T^{-1} are both triangular. Restricting to triangular / adapted mappings in the definition ensures that the adapted Wasserstein distance adequately respects the information structure inherent in stochastic processes.

The Kantorovich type relaxation is given by allowing for bi-causal couplings, i.e.,

$$(14) \quad \mathcal{AW}_2^2(\mu, \nu) = \inf_{\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)} \int |x - y|^2 d\pi(x, y).$$

A coupling π is called causal if for $t \leq N$ and Borel $B \subseteq H$

$$(15) \quad \pi(Y_t \in B | X_1, \dots, X_N) = \pi(Y_t \in B | X_1, \dots, X_t),$$

where we use $(X_1, \dots, X_N), (Y_1, \dots, Y_N)$ to denote the projections onto the first and second (respectively) coordinate of $H \times H$. The coupling π is called bi-causal, if this also holds when switching the roles of X and Y .

Already in the first non-trivial instance $H = \mathbb{R}$, $N = 2$ it is easy to find absolutely continuous probability measures which are not transport regular, see Example 8.26 below. The deeper reason for this is that via a specific isometry, adapted transport for processes with N time intervals can

be seen as a $\mathcal{W}_2$ -transport problem on $\mathcal{P}_2^N(H^N)$ and the corresponding push-forward of the Lebesgue measure to $\mathcal{P}_2^N(H^N)$ plays no particular role.

However based on our results for optimal transport on $\mathcal{P}_2^N(H^N)$ we are able to show that *typically* optimizers of (14) are unique and bi-triangular Monge.

Theorem 1.10 (Baire–Brenier theorem for $\mathcal{AW}_2$). *The set of (μ, ν) for which $\mathcal{AW}_2(\mu, \nu)$ admits a unique optimal coupling π^* and π^* is given by a bi-triangular map $T : H^N \rightarrow H^N$ is comeager in $\mathcal{P}_2(H^N) \times \mathcal{P}_2(H^N)$.*

As mentioned above, $\mathcal{AW}_2$ is considered as a natural metric for the adapted weak topology on $\mathcal{P}_2(H^N)$ but (like other compatible metrics) it is not complete. It is shown in [14] that the completion of $(\mathcal{P}_2(H^N), \mathcal{AW}_2)$ consists precisely in the space of stochastic processes with *filtration*, where two stochastic processes are identified, in signs $X \sim_{\text{ad}} Y$ if they have the same probabilistic properties. This can be made precise, e.g. by asserting that X and Y have Markovian lifts with the same laws, see [23].

To present the completion in terms of $L_2([0, 1]^N; H)$, we recall that $(\mathcal{F}_t)_{t=1}^N$ denotes the coordinate filtration on $([0, 1]^N, \lambda)$ and set

$$\begin{aligned} L_{2, \text{ad}}^N(H) &:= \{X = (X_t)_{t=1}^N \in L_2([0, 1]^N; H^N) : X_t \text{ is } \mathcal{F}_t\text{-measurable}\}, \\ \mathcal{AW}_2(X, Y) &:= \inf\{\|X' - Y'\|_2 : X' \sim_{\text{ad}} X, Y' \sim_{\text{ad}} Y\}. \end{aligned}$$

Elements of $\mathcal{P}_2(H^N)$ correspond to (equivalence classes of) naturally filtered processes $X \in L_{2, \text{ad}}^N(H)$. X is naturally filtered if $\mathbb{E}[f(X)|\mathcal{F}_t] = \mathbb{E}[f(X)|X_1, \dots, X_t]$ for all $t \leq N$ and $f : H^N \rightarrow \mathbb{R}$ continuous bounded. Intuitively this means that the coordinate filtration $(\mathcal{F}_t)_t$ contains no extra information on the process beyond what can be inferred from the past of the process. Using this embedding, the completion² of $(\mathcal{P}_2(H^N), \mathcal{AW}_2)$ is given by

$$(L_{2, \text{ad}}^N(H) / \sim_{\text{ad}}, \mathcal{AW}_2) =: (\text{filtered processes}, \mathcal{AW}_2).$$

Our main result for the adapted Wasserstein distance is the following characterization of optimal couplings in terms of convex functions on the set of adapted processes. Notice that the case $N = 1, H = \mathbb{R}^d$ corresponds once more to the classical Brenier theorem.

Theorem 1.11 (Brenier Theorem for $\mathcal{AW}_2$). *Let $X, Y \in L_{2, \text{ad}}^N(H)$. Then (X, Y) is an optimal coupling (i.e., $\|X - Y\|_2 = \mathcal{AW}_2(X, Y)$) if and only if*

$$(16) \quad Y(u_1, \dots) \in \partial \bar{\varphi}(X(u_1, \dots)), \quad u_1 \in [0, 1]$$

for some convex adapted-law invariant $\bar{\varphi} : H \times L_{2, \text{ad}}^{N-1}(H) \rightarrow (-\infty, \infty]$.

Moreover, a coupling $\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)$, $\mu, \nu \in \mathcal{P}_2(H^N)$ is optimal if and only if $\pi \sim \mathcal{L}(X, Y)$ for some naturally filtered processes X, Y satisfying (16).

Importantly $(\mathcal{P}_2(H^N), \mathcal{AW}_2)$ is a dense G_δ subset of its completion (filtered processes, $\mathcal{AW}_2$), see [38, 12]. This is crucial to deduce Theorem 1.10 from Theorem 1.11 and ultimately goes back to the fact that Monge couplings are a G_δ subset of the set of couplings with fixed first marginal, see [63] and [36, IV, 43].

²The space of filtered processes equipped with $\mathcal{AW}_2$ admits a number of convenient properties. It is a Polish geodesic space, the set of martingales is closed and geodesically convex, there is a Prohorov-type compactness criterion, finite state Markov chains are dense and Doob decomposition, optimal stopping, Snell-envelope, pricing, hedging, utility maximization, etc. are Lipschitz continuous w.r.t. $\mathcal{AW}_2$, cf. [14] and the references therein.

1.4. Related literature. Starting from the classical Monge and Kantorovich formulations, the modern theory of optimal transport originated around Brenier’s discovery that, for quadratic costs on Euclidean space, optimal plans are gradients of convex potentials [26], see also [55, 72]. In particular it paved the way for McCann’s displacement convexity [60] and the interpretation of many PDEs as gradient flows on the Wasserstein space. These developments were systematized in the metric-measure framework and gradient-flow theory of Ambrosio–Gigli–Savaré [8, 7] and in Villani’s monographs [74, 75], with more recent expositions by Santambrogio [73] and Figalli–Glaudo [42]. Beyond the Euclidean Brenier setting, there is a rich line of extensions and refinements — covering structure/uniqueness of optimal maps, displacement convexity and convex-order tools, and regularity/duality on general spaces — see, among others [43] (general convex costs), [61] (Brenier/polar factorization on manifolds), [44, 4] (multi-marginal transport and barycenters), [40] (Monge transport on abstract Wiener spaces via Cameron–Martin convexity, existence/uniqueness of maps) and [41] (infinite-dimensional Monge–Ampère on Wiener space, regularity/structure of the transport) [9] (Monge maps for strictly convex costs), [8] (Brenier theorem for Gaussian on Hilbert spaces), [33, 34] (existence of Monge optimizers under low regularity), [46] (convex potential beyond smooth Euclidean settings), [17] (Skorokhod embedding), [18, 50, 45] (martingale transport), [29] (connection to Knothe–Rosenblatt coupling), [47] (weak optimal transport).

In optimal transport and mean field games, the “Lions’ lift” refers to the idea of studying maps defined on the Wasserstein space of probability measures by lifting them to functions on a Hilbert space of square-integrable random variables. This was introduced by Lions in his Collège de France lectures [58] and became widely known through Cardaliaguet’s lecture notes [27] and Carmona–Delarue’s monographs [30, 31] which employ the Lions derivative as an essential tool in the theory of mean field games.

Emami and Pass [39] were the first to establish existence and uniqueness of Monge solutions for measures on $\mathcal{P}_2(M)$ and $\mathcal{W}_2^2$ -costs. Emami and Pass consider a smooth Riemannian manifold M and use structural assumptions on a reference measure on $\mathcal{P}_2(M)$. A central ingredient in their analysis is Dello Schiavo’s Rademacher theorem on Wasserstein spaces [37], that allows a classical optimal-map strategy to be carried out on the space of measures. Notably it is challenging to construct measures satisfying Dello Schiavo’s hypothesis together with the absolute continuity hypothesis necessary for the main Monge result of [39] and examples are only known in the case where M is the one dimensional torus.

Independently and in parallel with an earlier version of this manuscript, Pinzi–Savaré [71], see also [70, 69], analyzed the case $N = 1$ of measures on $\mathcal{P}_2(H)$ for separable Hilbert spaces H , developing a Brenier-type theory via totally convex functionals and their Lagrangian/Lions lift, and identifying natural classes of full-support, transport-regular laws for which the Monge formulation is uniquely solved. Our initial preprint [21], posted the same day their preprint appeared, likewise treated $N = 1$ but focused on $\mathbb{R}^d$; the restriction to $\mathbb{R}^d$ was due to the then-open problem of constructing a transport-regular reference measure on infinite-dimensional Hilbert spaces, while all other arguments were dimension-agnostic. The present version advances in two directions: it covers arbitrary N and it works on general separable Hilbert spaces. The extension relies on an adapted Lions lift tailored to the N -step structure. In particular, we construct a transport-regular measure Λ on $\mathcal{P}_2(H)$ and, by induction, on $\mathcal{P}_2^N(H)$; here it is essential to include the case of infinite dimensional H even to establish the N -level result in $\mathbb{R}^d$ for $N > 1$, since the induction from lower levels requires the Hilbert-space setting. We note that the construction of Λ as the occupation measure of a Wiener process (resp. based on the Wiener sheet for $N > 1$) is original to this paper to the best of our knowledge.

A main motivation for this article was to establish a Brenier theorem for the adapted Wasserstein distance, which requires to understand the $\mathcal{W}_2$ -optimizers on $\mathcal{P}_2^N(H)$ for general N . As

noted above, the adapted variant of Wasserstein distance is useful when measuring the distance between stochastic processes since it accounts for the inherent information structure of the processes. It has applications from stochastic optimization, stochastic control, and mathematical finance to the theory of geometric inequalities and machine learning; see [68, 57, 77, 14, 1, 35, 52, 15, 3] and the references therein.

1.5. Organization of the paper. Section 2 recalls the Lions lift and basic continuity properties. We identify c -conjugation for $c = \text{MC}$ with classical convex conjugation on the lift (Theorem 2.2) and relate total/Lions convexity, MC-convexity, and subdifferentials.

Section 3 treats the case of optimal transport between measures on $\mathcal{P}_2(H)$ ($N = 1$). We characterize $\mathcal{W}_2^2$ -optimizers on $\mathcal{P}_2(H)$ via convex, law-invariant potentials on the lift (Theorem 3.1) and construct transport-regular reference laws: on $\mathbb{R}^d$ using Brownian sheets and on separable Hilbert spaces using Q -Wiener processes (Theorem 3.5).

Appendix A develops a Baire-category viewpoint: we quantify non-regularity via τ^c , prove upper semicontinuity, and deduce that c -transport-regular laws form a G_δ subset of $\mathcal{P}_p(\mathcal{X})$ (Theorem A.5).

Section 4 introduces the iterated max-covariance MC on $\mathcal{P}_2^N(H)$ and the characterization of MC-optimal couplings in terms of MC-convex potentials. We also derive a DPP which will be necessary to establish the characterization of MC-convex potentials in terms of convex conjugation in Section 6.

In Section 5 we provide a Hilbert-space representation of laws on $\mathcal{P}_2^N(H)$, an adapted transfer principle, and a Skorokhod-type representation (Proposition 5.2, Corollary 5.6, Proposition 5.7).

Section 6 identifies MC-conjugation with the convex conjugate of the adapted lift (Theorem 6.1), establishes the MC-order which is used to establish the existence of the Lions derivative in the adapted setting, and links MC-subdifferentials with subdifferentials on the lift.

Section 7 constructs full-support, transport-regular measures on $\mathcal{P}_2^N(H)$; this yields the N -level Brenier theorems stated in the introduction (Theorems 1.4, 1.1 and 1.5 and Corollary 1.2).

Section 8 applies the theory to the adapted Wasserstein distance $\mathcal{AW}_2$: via the adapted law isometry we obtain a Brenier theorem for $\mathcal{AW}_2$ and a Baire-Brenier uniqueness/Monge statement (Theorem 1.11, Theorem 1.10).

Technical measurability, selection, and auxiliary results are collected in the appendix.

CONTENTS

1. Introduction	1
1.1. Characterization of optimal transport maps via convex Lions lifts	2
1.2. Fundamental theorem of optimal transport, MC-convexity and L -convexity	4
1.3. Stochastic processes and Brenier theorems for the adapted Wasserstein distance	5
1.4. Related literature	7
1.5. Organization of the paper	8
2. Lions lift and convexity of law invariant functions	9
2.1. From MC-convexity to convexity on H	9
2.2. Subdifferentials	10
2.3. Examples and basic MC-convex functions	13
3. Brenier's theorem for measures on $\mathcal{P}_2(H)$	15
3.1. Characterization of MC-optimal couplings and criterion for transport regularity	15
3.2. Regularization on H by Brownian motion/standard Wiener process	16
4. Optimal transport on $\mathcal{P}_2^N(H)$	20
5. Probabilistic representation of iterated probability measures	23
6. The adapted Lions lift	26

6.1.	Connection between MC-transform and convex conjugate of the lift	26
6.2.	MC-order	27
6.3.	N -Monge Couplings	27
6.4.	MC-subdifferentials	29
6.5.	Characterization of strict Monge pairs by differentiability of MC	31
7.	Construction of transport-regular measures in $\mathcal{P}_2^N(H)$	34
7.1.	Criterion for transport regularity	34
7.2.	Existence of transport-regular measures	35
7.3.	Proofs of Section 1.1	36
8.	Application to adapted transport	36
8.1.	Outline of the framework	36
8.2.	Embedding into $\mathcal{P}_2^N(H^N)$	38
8.3.	Lions lift for filtered processes	40
8.4.	Naturally filtered Processes	43
8.5.	Lift of functions on $\mathcal{Z}_1$ and proof of Theorem 1.11	45
8.6.	Failure of an adapted Brenier theorem for absolutely continuous measures	46
Appendix A.	On transport-regularity and Baire category	46
Appendix B.	Auxiliary results	49
References		52

2. LIONS LIFT AND CONVEXITY OF LAW INVARIANT FUNCTIONS

Notation 1.1 $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a standard probability space. We use μ, ν for elements of $\mathcal{P}_2(H)$ and X, Y for their representatives in $L_2(H) = L_2(\Omega, \mathcal{F}, \mathbb{P}; H)$, i.e., random variables satisfying $X \sim \mu, Y \sim \nu$. Given a function $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$, the function $\bar{\varphi} : L_2(H) \rightarrow (-\infty, +\infty]$ always denotes its Lions lift $\bar{\varphi}(X) = \varphi(\mathcal{L}(X))$.

In this section we develop the connection of MC-convexity and MC-subdifferentials with their Hilbert space counterparts. We start by giving the representations of $W_2(\mu, \nu)$ on $L_2(H)$ in Lemma 2.1. The important Theorem 2.2 asserts that lifts of MC-conjugates are precisely convex conjugates of lifts. Likewise, Theorem 2.4 links lifts of MC-subdifferentials to classical Hilbertian subdifferentials of lifts. We call φ MC-differentiable at $\mu \in \mathcal{P}_2(H)$ if the MC-subdifferential is a singleton. Equivalently, the subdifferential of $\bar{\varphi}$ is a singleton at every random variable $X \sim \mu$. In this case there is a unique $\xi : H \rightarrow H$ such that $\partial_{\text{MC}}\varphi(\mu) = \{\xi_{\#}\mu\}$, or equivalently $\partial\bar{\varphi}(X) = \{\xi(X)\}$ for every $X \sim \mu$. That is, depending on the context, ξ is either the Brenier map or the Lions derivative at μ , see Theorem 2.7. The proof of this differs somewhat from the usual arguments, since we need to work with a weaker notion of derivative in view of our main results on the existence of transport-regular measures.

2.1. From MC-convexity to convexity on H . A technical but relevant detail is that not all representations $X \in L_2(H)$ of μ are ‘regular’ in the following sense: We say that X allows for *independent randomization* if it satisfies any of the following equivalent conditions:

- (1) there exists a uniformly distributed $U \in L_2(\mathbb{R})$ which is independent of X ;
- (2) the disintegration $(\mathbb{P}^x)_{x \in H}$ of $\mathbb{P}$ w.r.t. X satisfies that $\mathbb{P}^x$ is continuous μ -a.s.;
- (3) for any $\pi \in \mathcal{P}_2(H \times H)$ with first marginal μ there exists $Y \in L_2(H)$ such that $(X, Y) \sim \pi$.

The last characterization underlines the significance of independent randomization in our context.

The following lemma can be interpreted as an instance of the (L^2) -Skorokhod representation theorem. The non-standard bit is that the representation X of μ can be taken to be fixed.

Lemma 2.1 (Skorokhod representation). *Let $\mu, \nu \in \mathcal{P}_2(H)$ and $X \sim \mu$. Then*

$$(17) \quad \text{MC}(\mu, \nu) = \sup_{Y \sim \nu} \mathbb{E}[X \cdot Y], \quad \mathcal{W}_2(\mu, \nu) = \inf_{Y \sim \nu} \|X - Y\|_2.$$

If X allows independent randomization or μ is transport regular, then an optimal Y exists.

Proof. If X allows independent randomization or μ is transport regular, the result is clear. In the general case, pick $X' \sim \mu$ which allows for independent randomization. By [30, Lemma 5.23], there exists for each $n \in \mathbb{N}$ a measure preserving bijection $T_n : \Omega \rightarrow \Omega$ such that $\|X' - X \circ T_n\|_2 < 1/n$. Pick $Y' \sim \nu$ such that $\mathbb{E}[X' \cdot Y'] = \text{MC}(\mu, \nu)$. Then $Y_n := Y' \circ T_n^{-1}$ satisfies $Y_n \sim \nu$ and $\mathbb{E}[X \cdot Y_n] = \mathbb{E}[X \circ T_n \cdot Y'] \rightarrow \mathbb{E}[X' \cdot Y'] = \text{MC}(\mu, \nu)$. $\square$

Note that Lemma 2.1 implies that W_2 is the quotient metric of $\|\cdot\|_2$ w.r.t. the law map. In particular, the lift $\bar{\varphi}$ of a function φ is continuous / lsc if and only if φ has this property.

Recall from Section 1.2 that φ^{MC} denotes the c -transform of φ with cost $c = \text{MC}$, i.e. $\varphi^{\text{MC}}(\nu) = \sup_{\mu \in \mathcal{P}_2(H)} (\text{MC}(\mu, \nu) - \varphi(\mu))$. The following result plays a fundamental role for this paper.

Theorem 2.2 (Correspondence of conjugacy). *Let $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$ be proper. Then*

$$\bar{\varphi}^* = \overline{\varphi^{\text{MC}}}.$$

In particular, the convex conjugate of a law-invariant function is law-invariant.

Proof. For $Y \in L_2(H)$ we have by Lemma 2.1

$$\begin{aligned} \bar{\varphi}^*(Y) &= \sup_{X \in L_2(H)} \mathbb{E}[X \cdot Y] - \bar{\varphi}(X) = \sup_{\mu \in \mathcal{P}_2(H)} \sup_{X \sim \mu} \mathbb{E}[X \cdot Y] - \varphi(\mu) \\ &= \sup_{\mu \in \mathcal{P}_2(H)} \text{MC}(\mu, \mathcal{L}(Y)) - \varphi(\mu) = \varphi^{\text{MC}}(\mathcal{L}(Y)) = \overline{\varphi^{\text{MC}}}(Y). \end{aligned} \quad \square$$

Recall from Section 1.2 that a proper function $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$ is MC-convex if there exists a proper function $\psi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$ such that $\varphi = \psi^{\text{MC}}$. It follows immediately that a proper function φ is MC-convex if and only if it is of the form $\sup_{i \in I} \text{MC}(\cdot, \nu_i) + c_i$ for some $\nu_i \in \mathcal{P}_2(H)$, $c_i \in \mathbb{R}$, $i \in I$.

Further recall from (12) that $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, \infty]$ is totally convex if it is convex on all curves $(\mu_t)_{t \in [0,1]}$, $\mu_t := ((1-t)\text{pr}_0 + t\text{pr}_1)_{\#}\pi$, $\pi \in \text{Cpl}(\mu, \nu)$. Theorem 2.2 has the immediate

Corollary 2.3. *For proper, lsc $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$, the following are equivalent:*

- (1) φ is totally convex,
- (2) $\bar{\varphi}$ is convex,
- (3) φ is MC-convex.

Proof. It is straightforward that (1) $\Leftrightarrow$ (2). Further, (2) $\Leftrightarrow$ (3) follows from Theorem 2.2. $\square$

In particular, a function $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$ is MC-convex if and only if it is proper, lsc, and totally convex.

A further application of Corollary 2.3 arises by considering functions of the form $\varphi_\nu = \text{MC}(\cdot, \nu)$ for $\nu \in \mathcal{P}_2(H)$. Since $\bar{\varphi}_\nu(X) = \sup_{Y \sim \nu} \mathbb{E}[X \cdot Y]$ is convex, the function φ_ν is MC-convex. Expressing this observation in terms of $W_2^2(\cdot, \nu)$ recovers [8, Theorem 7.3.2].

2.2. Subdifferentials. Recall that the MC-subdifferential of $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$ is

$$(18) \quad \partial_{\text{MC}}\varphi(\mu) = \{\nu \in \mathcal{P}_2(H) : \varphi(\rho) \geq \varphi(\mu) + \text{MC}(\rho, \nu) - \text{MC}(\mu, \nu) \text{ for every } \rho \in \mathcal{P}_2(H)\}.$$

The following result shows how to express the subdifferential of the lift in terms of the MC-subdifferential and vice versa. A noteworthy caveat is that the subdifferential may be ‘too small’ if X does not admit independent randomization and $\mathcal{L}(X)$ is not transport-regular.

Theorem 2.4. *Let $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$ be MC-convex, $\mu \in \mathcal{P}_2(H)$, $X \sim \mu$. Then we have*

$$(19) \quad \partial\bar{\varphi}(X) = \{Y : \mathcal{L}(Y) \in \partial_{\text{MC}}\varphi(\mu), (X, Y) \text{ is optimal}\},$$

$$(20) \quad \{\mathcal{L}(Y) : Y \in \partial\bar{\varphi}(X)\} \subseteq \partial_{\text{MC}}\varphi(\mu).$$

If X allows independent randomization or μ is transport regular, then we have equality in (20).

Proof. We first prove ‘ $\subseteq$ ’ in (19). Let $Y \in \partial\bar{\varphi}(X)$ and $\nu = \mathcal{L}(Y)$. By definition we have

$$(21) \quad \bar{\varphi}(Z) \geq \bar{\varphi}(X) + \mathbb{E}[(Z - X) \cdot Y],$$

for every $Z \in L_2(H)$. If $Z \sim \mu$, we have $\bar{\varphi}(X) = \bar{\varphi}(Z)$ and (21) yields $\mathbb{E}[X \cdot Y] \geq \mathbb{E}[Z \cdot Y]$. Hence, $\mathbb{E}[X \cdot Y] = \text{MC}(\mu, \nu)$ by Lemma 2.1. Therefore $\mathcal{L}(X, Y)$ is an optimal coupling. Next, fix $\rho \in \mathcal{P}_2(H)$ and $\varepsilon > 0$. By Lemma 2.1 there is $Z \sim \rho$ such that $\mathbb{E}[Z \cdot Y] \geq \text{MC}(\rho, \nu) - \varepsilon$. Applying (21) then yields $\varphi(\rho) \geq \varphi(\mu) + \text{MC}(\rho, \nu) - \text{MC}(\mu, \nu) - \varepsilon$. Thus $\nu \in \partial_{\text{MC}}\varphi(\mu)$.

Conversely, to show ‘ $\supseteq$ ’ in (19), assume that $\nu \in \partial_{\text{MC}}\varphi(\mu)$ and that $Y \sim \nu$ satisfies $\mathbb{E}[X \cdot Y] = \text{MC}(\mu, \nu)$. Then, for every $Z \in L_2(H)$, we have writing $\rho = \mathcal{L}(Z)$

$$\bar{\varphi}(Z) = \varphi(\rho) \geq \varphi(\mu) + \text{MC}(\rho, \nu) - \text{MC}(\mu, \nu) \geq \bar{\varphi}(X) + \mathbb{E}[Z \cdot Y] - \mathbb{E}[X \cdot Y].$$

Hence, $Y \in \partial\bar{\varphi}(X)$.

Evidently, (19) implies (20). Finally, assume that X allows independent randomization or that μ is transport regular. Pick any $\nu \in \partial_{\text{MC}}\varphi(\mu)$. By Lemma 2.1 there is $Y \sim \nu$ such that (X, Y) is optimal. By (19), $Y \in \partial\bar{\varphi}(X)$. Thus we obtain equality in (20). $\square$

Note that (19) implies that the subdifferential of a law-invariant convex function $\bar{\varphi}$ consists of optimal couplings, i.e., if $(X, Y) \in \partial\bar{\varphi}$ then $\mathcal{L}(X, Y)$ is optimal between its marginals. In this sense the Lions lift contains more information than the MC-subdifferential of φ which carries only information on the marginals $\mathcal{L}(X)$ and $\mathcal{L}(Y)$.

The inclusion in (20) may be strict, see Example 2.17 below. In our definition of differentiability we exclude this:

Definition 2.5. Let $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, \infty]$ be MC-convex and $\mu \in \text{dom}(\varphi)$. We say that φ is MC-differentiable at μ if $\partial_{\text{MC}}\varphi(\mu)$ is a singleton.

Lemma 2.6. *Let $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, \infty]$ be MC-convex and $\mu \in \text{dom}(\varphi)$. Then the following are equivalent:*

- (1) φ is MC-differentiable in μ ,
- (2) for all X with $X \sim \mu$, $\partial\bar{\varphi}(X)$ is a singleton,
- (3) for some $X \sim \mu$ admitting independent randomization, $\partial\bar{\varphi}(X)$ is a singleton.

Proof. (1) $\Rightarrow$ (2): Suppose that there are $Z_1 \neq Z_2 \in \partial\bar{\varphi}(X)$ for some $X \sim \mu$. By Theorem 2.4, we have $\mathcal{L}(Z_1), \mathcal{L}(Z_2) \in \partial_{\text{MC}}\varphi(\mu)$. Hence, if $\mathcal{L}(Z_1) \neq \mathcal{L}(Z_2)$, we conclude that $\partial_{\text{MC}}\varphi(\mu)$ is not a singleton. Otherwise, observe that $Z := \frac{1}{2}(Z_1 + Z_2) \in \partial\bar{\varphi}(X)$ and hence $\mathcal{L}(Z) \in \partial_{\text{MC}}\varphi(\mu)$. If $f : H \rightarrow \mathbb{R}$ is any strictly convex function of quadratic growth, we have $\mathbb{E}[f(Z)] < \frac{1}{2}(\mathbb{E}[f(Z_1)] + \mathbb{E}[f(Z_2)]) = \mathbb{E}[f(Z_1)]$. Hence, $\mathcal{L}(Z) \neq \mathcal{L}(Z_1)$ showing that $\partial_{\text{MC}}\varphi(\mu)$ is no singleton.

(2) $\Rightarrow$ (3) is trivial.

(3) $\Rightarrow$ (1): Suppose there are distinct $\nu_1, \nu_2 \in \partial_{\text{MC}}\varphi(\mu)$. Since $X \sim \mu$ allows independent randomization, there are $Z_1, Z_2 \in \partial\bar{\varphi}(X)$ with $Z_1 \sim \nu_1$, $Z_2 \sim \nu_2$. In particular, $Z_1 \neq Z_2$. $\square$

Next we establish the existence of the Lions derivative in the case of MC-differentiable, MC-convex functions.

Theorem 2.7. *Let $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, \infty]$ be MC-convex and MC-differentiable at μ . Then there exists a measurable function $\xi : H \rightarrow H$ such that $\partial\bar{\varphi}(X) = \{\xi(X)\}$ for all $X \sim \mu$.*

We need some preparations for the proof. Recall that $\varphi_\nu = \text{MC}(\cdot, \nu)$, $\nu \in \mathcal{P}_2(H)$ is MC-convex. The convex order, denoted by $\preceq_c$, is the partial order on $\mathcal{P}_2(H)$ defined by $\mu \preceq_c \nu$ whenever $\int f d\mu \leq \int f d\nu$ for all convex functions $f : H \rightarrow (-\infty, +\infty]$. The following characterization of the convex order is due to Carlier [28] (see also Wiesel–Zhang [76] and Proposition 6.3 below for the case that H is infinite dimensional).

Proposition 2.8. *For $\mu, \nu \in \mathcal{P}_2(H)$, the following are equivalent:*

- (1) $\mu \preceq_c \nu$;
- (2) $\varphi_\rho(\mu) \leq \varphi_\rho(\nu)$ for every $\rho \in \mathcal{P}_2(H)$.

Note that Proposition 2.8 implies in particular that the functions φ_ν are increasing in the convex order. As every MC-convex function is the supremum of functions φ_{ν_i} plus constants and monotonicity is preserved under suprema, we obtain

Lemma 2.9. *Any MC-convex $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$ is increasing in the convex order.*

As a consequence we have

Lemma 2.10. *Let $\bar{\varphi} : L_2(H) \rightarrow (-\infty, +\infty]$ be law invariant convex. If $Z \in \partial\bar{\varphi}(X)$, then $\mathbb{E}[Z|X] \in \partial\bar{\varphi}(X)$.*

Proof. Using that $\bar{\varphi}$ is monotone in the convex order (see Lemma 2.9) we find for every $Y \in L_2(H)$

$$\bar{\varphi}(Y) \geq \bar{\varphi}(\mathbb{E}[Y|X]) \geq \bar{\varphi}(X) + \mathbb{E}[(\mathbb{E}[Y|X] - X) \cdot Z] = \bar{\varphi}(X) + \mathbb{E}[(Y - X) \cdot \mathbb{E}[Z|X]]. \quad \square$$

Proof of Theorem 2.7. Let $X \sim \mu$ and $Z \in \partial\bar{\varphi}(X)$. By Lemma 2.10 we have $\mathbb{E}[Z|X] \in \partial\bar{\varphi}(X)$. As $\partial\bar{\varphi}(X)$ is a singleton, we have $Z = \mathbb{E}[Z|X]$, hence there exists a measurable ξ such that $Z = \xi(X)$. Next, we need to argue that the function ξ does not depend on the choice of the random variable $X \sim \mu$. To that end, let $X_1, X_2 \sim \mu$ with $\partial\bar{\varphi}(X^i) = \{\xi^i(X^i)\}$ for $i \in \{1, 2\}$. Pick $X \sim \mu$ that allows for independent randomization. We then have $\xi^i(X) \in \partial\bar{\varphi}(X)$ for $i \in \{1, 2\}$ by Theorem 2.4. As $\partial\bar{\varphi}(X)$ is a singleton, we have $\xi^1 = \xi^2$ μ -a.s. $\square$

Corollary 2.11. *Let $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$ be MC-convex and $\mu \in \mathcal{P}_2(H)$. If φ is MC-differentiable at μ with derivative ν , then the optimal coupling of μ, ν is unique and of Monge type.*

Proof. Let $X \sim \mu$ allow for randomization. By Theorem 2.4, we have $\text{Cpl}_{\text{opt}}(\mu, \nu) = \{\mathcal{L}(X, Y) : Y \in \partial\bar{\varphi}(X)\}$. By Lemma 2.6 this set is a singleton and by Theorem 2.7 it consists of a Monge coupling. $\square$

Remark 2.12. There are different possible notions of differentiability for MC-convex functions $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$ at $\mu \in \text{dom}(\varphi)$:

- (1) $\bar{\varphi}$ is Frechet differentiable at some $X \sim \mu$. This is the common notion of L-differentiability, see [30, Definition 5.22]. By [30, Proposition 5.24] this is equivalent to $\bar{\varphi}$ being Frechet differentiable at any $X \sim \mu$.
- (2) φ is MC-differentiable according to Definition 2.5, i.e., $\partial_{\text{MC}}\varphi(\mu)$ is a singleton. This is the notion adopted in the present article and it is equivalent to $\partial\bar{\varphi}(X)$ being a singleton at any $X \sim \mu$ (see Lemma 2.6). Provided that φ is continuous at μ , this equivalent to $\bar{\varphi}$ being Gateaux differentiable at any $X \sim \mu$.³
- (3) $\partial\bar{\varphi}(X)$ is a singleton at some $X \sim \mu$.

³Indeed, $\bar{\varphi}$ is continuous at any $X \sim \mu$ by Lemma 2.1. Moreover, a convex function is Gateaux differentiable at a point if it is continuous there and its subdifferential is a singleton; see, e.g., [16, Proposition 17.26].

It is clear that (1) is stronger than (2) which is again stronger than (3). We will see in Example 2.16 and Example 2.17 below that the converse implications are in general not true, even if H is finite dimensional.

For the purposes of this article (2) is the correction notation as it allows to characterize and obtain existence of transport regular measures on $\mathcal{P}_2(H)$. This is because this weaker notion, compared to the classical L-differentiability outlined in (1), yields almost sure differentiability results for the dual potentials.

2.3. Examples and basic MC-convex functions. The results of this section elucidate the previous results but will not be required for our main theorems.

In usual convex analysis, linear functions serve as fundamental building blocks. In the present setting of MC-convexity, the role of linear functions is played by

$$\varphi_\nu = \text{MC}(\cdot, \nu), \quad \nu \in \mathcal{P}_2(H).$$

We summarize some important properties:

Proposition 2.13. *Let $\nu \in \mathcal{P}_2(H)$ and set $C_\nu = \overline{\text{conv}}\{Y \in L_2(H) : Y \sim \nu\}$. Then:*

- (1) φ_ν is the support function of C_ν , i.e., $\overline{\varphi}_\nu(X) = \sup_{Y \sim \nu} \mathbb{E}[X \cdot Y] = \sup_{Y \in C_\nu} \mathbb{E}[X \cdot Y]$;
- (2) The MC-conjugates satisfy $\chi_{\{\nu\}}^{\text{MC}} = \varphi_\nu$ and $\varphi_\nu^{\text{MC}} = \chi_{\{\mu : \mu \leq_c \nu\}}$;
- (3) $\{\mathcal{L}(Y) : Y \in C_\nu\} = \{\mu \in \mathcal{P}_2(H) : \mu \leq_c \nu\}$.

Proof. (1) For $X \in L_2(H)$, Lemma 2.1 gives $\overline{\varphi}_\nu(X) = \sup_{Y \sim \nu} \mathbb{E}[X \cdot Y]$. The right-hand side is unchanged if we take the supremum over the closed convex hull of $\{Y : Y \sim \nu\}$, hence $\overline{\varphi}_\nu(X) = \sup_{Y \in C_\nu} \mathbb{E}[X \cdot Y]$.

(2) It is immediate from the definition of the MC-transform that $\chi_{\{\nu\}}^{\text{MC}} = \varphi_\nu$. To see the second claim note that

$$\varphi_\nu^{\text{MC}}(\rho) = \sup_{\mu \in \mathcal{P}_2(H)} (\text{MC}(\mu, \rho) - \text{MC}(\mu, \nu)).$$

First consider the case $\rho \leq_c \nu$. By monotonicity of MC in the convex order we have $\text{MC}(\mu, \rho) - \text{MC}(\mu, \nu) \leq 0$ and choosing $\mu = \nu$ yields $\varphi_\nu^{\text{MC}}(\rho) = 0$. Now, let $\rho \not\leq_c \nu$. Then Proposition 2.8 provides some μ with $\text{MC}(\mu, \rho) > \text{MC}(\mu, \nu)$; scaling μ shows the supremum is $+\infty$.

(3) By (2) we have $\chi_{\{\nu\}}^{\text{MCMC}} = \chi_{\{\mu : \mu \leq_c \nu\}}$. On the other hand, by Theorem 2.2, the Lions lift of $\chi_{\{\nu\}}^{\text{MCMC}}$ is the lsc convex hull of in convex indicator of $\{Y \in L_2(H) : Y \sim \nu\}$, which is precisely the convex indicator of C_ν . $\square$

Call $B \subseteq \mathcal{P}_2(H)$ MC-convex if it is a sublevel set of a MC-convex function or equivalently if it contains all curves $(\mu_t)_{t \in [0,1]}$, $\mu_t := ((1-t)\text{pr}_0 + t\text{pr}_1)_\# \pi$, $\pi \in \text{Cpl}(\mu_0, \mu_1)$ where $\mu_0, \mu_1 \in B$.

Example 2.14. By Proposition 2.13, the closed MC-convex hull of $\{\nu\}$ is given by $\{\rho \in \mathcal{P}_2(H) : \rho \preceq_c \nu\}$. In particular, singletons are MC-convex sets if and only if they consist of a Dirac measure.

Below we will see that MC-subdifferentials of MC-convex functions are singletons in most points (cf. Example 2.14) and is usually not MC-convex by Example 2.14. However, it is convex along generalized geodesics with base point μ , as a consequence of Theorem 2.4.

It is an easy consequence of Theorem 2.4 that the subdifferential of a law invariant convex function $\overline{\varphi} : L_2(H) \rightarrow (-\infty, +\infty]$ can be decomposed into subdifferentials of basis functions, i.e. for $X \sim \mu$ we have

$$\partial \overline{\varphi}(X) = \bigcup \{Y \in \partial \overline{\varphi}_\nu(X) : \nu \in \partial_{\text{MC}} \varphi(\mu)\}.$$

This highlights the importance of understanding the subdifferential of these basis functions, which is addressed in the following proposition. This proposition can also be seen as an analog of the results of Alfonsi–Jourdain [6] in the language of MC-differentiability.

Proposition 2.15. *The following are equivalent for $\mu, \nu \in \mathcal{P}_2(H)$*

- (1) φ_μ is strictly $\preceq_c$ -increasing from below at ν , i.e., $\eta \preceq_c \nu, \eta \neq \nu \Rightarrow \varphi_\mu(\eta) < \varphi_\mu(\nu)$;
- (2) For $\mu \in \mathcal{P}_2(H)$, the map φ_ν is MC-differentiable at μ ;
- (3) The optimal coupling of μ, ν is unique and of Monge type.

In this case, $\partial_{\text{MC}}\varphi_\nu(\mu) = \{\nu\}$, and the Lions derivative is given by the optimal map from μ to ν .

Proof. (1) $\Leftrightarrow$ (2): Recall that $\eta \in \partial_{\text{MC}}\varphi_\nu(\mu)$ if and only if $\varphi_\nu(\mu) + \varphi_\nu^{\text{MC}}(\eta) = \text{MC}(\mu, \eta)$. By Proposition 2.13 (2) this is equivalent to $\text{MC}(\mu, \nu) = \text{MC}(\mu, \eta)$ and $\eta \preceq_c \nu$.

(2) $\Rightarrow$ (3) follows from Corollary 2.11.

(3) $\Rightarrow$ (2): $\eta \in \partial_{\text{MC}}\varphi_\nu(\mu)$. By the argument above, $\eta \preceq_c \nu$. Hence, there exist random variables (X, Y', Y) with (X, Y') is optimal and $\mathbb{E}[Y|Y', X] = Y'$ a.s. Then, $\text{MC}(\mu, \nu) = \text{MC}(\mu, \eta) = \mathbb{E}[X \cdot Y'] = \mathbb{E}[X \cdot Y]$ and thus (X, Y) is optimal. By our hypothesis, $Y = T(X)$ for some map T . Hence, $Y = \mathbb{E}[Y|Y', X] = Y'$ a.s. and $\eta = \nu$. $\square$

We conclude this section with two examples that show that the notions of differentiability discussed in Remark 2.12 are indeed different.

Example 2.16. For every $P \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ there exists a continuous MC-convex function $\varphi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ such that for P -a.e. μ the Lions lift $\bar{\varphi}$ is not Frechet differentiable at any $X \sim \mu$. This function φ can be constructed using a modification of the construction in [25, Example 4.6.10]: As P is a Borel measure, there exists an increasing sequence $(K'_n)_n$ of compact subsets of $\mathcal{P}_2(\mathbb{R}^d)$ such that $P(K'_n) \rightarrow 1$. We write K_n for the MC-convex hull of K'_n and note that K_n is still compact⁴. We set $\varphi_n(\mu) = \text{dist}_{\mathcal{W}_2}(\mu, K_n)$ and $\varphi := \sum_n 2^{-n} \varphi_n$.

We write $C_n := \{X : \mathcal{L}(X) \in K_n\}$ and note that as C_n is law-invariant by Lemma 2.1

$$\bar{\varphi}_n(Y) = \text{dist}_{\mathcal{W}_2}(\mathcal{L}(Y), K_n) = \inf_{X \in C_n} \|X - Y\|_2 = \text{dist}_{\|\cdot\|_2}(Y, C_n).$$

This shows in particular that φ_n and hence φ is MC-convex. Moreover, it is shown in [25, Exercise 4.2.6] that if C_n is a closed convex set with empty interior, then $\text{dist}_{\|\cdot\|_2}(\cdot, C_n)$ is in every $X \in C_n$ not Frechet differentiable. Hence, $\bar{\varphi} = \sum_n 2^{-n} \bar{\varphi}_n$ is not Frechet differentiable in every point of $\bigcup_n C_n$.

Below we will show that for certain $P \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ every MC-convex $\varphi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, \infty]$ is P -a.s. MC-differentiable. In particular that φ is MC-differentiable in μ does not imply that $\bar{\varphi}$ is Frechet differentiable in some $X \sim \mu$.

Example 2.17. There exist a MC-convex $\varphi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ and X, X' with $X \sim X'$ such that $|\partial\bar{\varphi}(X)| = 1, |\partial\bar{\varphi}(X')| > 1$. Specifically let $d = 2$, assume wlog that $(\Omega, \mathcal{F}, \mathbb{P})$ is the unit square $[0, 1]^2$ equipped with two dimensional Lebesgue measure $\lambda^{(2)}$. We set $\mu := \lambda \otimes \delta_0$, $\nu := \lambda \otimes \frac{\delta_{-1} + \delta_{+1}}{2}$ (where λ denotes one dimensional Lebesgue measure), $\varphi = \varphi_\nu = \text{MC}(\cdot, \nu)$, take X to be an isomorphism of $([0, 1]^2, \mathfrak{B}, \lambda^{(2)})$ and $(\mathbb{R}^2, \mathfrak{B}, \mu)$ and set $X'(u_1, u_2) := (u_1, 0)$.

Then $X \sim \mu \sim X'$ and it is straightforward to see that $\#\partial\bar{\varphi}(X) = 1$, while $\#\partial\bar{\varphi}(X') > 1$. In particular $\bar{\varphi}$ is Gateaux differentiable in X but not in X' .

⁴Recall that $A \subseteq \mathcal{P}_2(\mathbb{R}^d)$ is $\mathcal{W}_2$ -relatively compact if and only if there is a convex increasing $\varphi : [0, +\infty) \rightarrow [0, +\infty)$ with $\lim_{r \rightarrow \infty} \varphi(r)/r^2 = +\infty$ such that $\sup_{\mu \in A} \int \varphi(|x|) d\mu(x) < \infty$. As $\mu \mapsto \int \varphi(|x|) d\mu(x)$ is MC-convex, the closed MC-convex hull of a compact set is compact.

3. BRENIER'S THEOREM FOR MEASURES ON $\mathcal{P}_2(H)$

This section is concerned with proving Theorem 1.4 announced in the introduction. To establish this theorem, we will work with random variables X, Y taking values in $L_2(H)$. To avoid ambiguity regarding the involved probability spaces, we reserve from here on $(\Omega, \mathcal{F}, \mathbb{P})$ for an abstract probability space, and choose $L_2(H) = L_2([0, 1]^k, \lambda; H)$ where λ denotes here the k -dimensional Lebesgue measure on $[0, 1]^k$. This choice is solely for notational clarity, and analogous results hold when $([0, 1]^k, \lambda)$ is replaced by any standard probability space. In particular, we write $\mathcal{L}_\lambda(X)$ for the $\mathcal{P}_2(H)$ -valued random measure obtained by applying pointwise the pushforward of λ by $X(\omega) \in L_2(H)$.

3.1. Characterization of MC-optimal couplings and criterion for transport regularity.

Theorem 3.1. *Let $P, Q \in \mathcal{P}_2(\mathcal{P}_2(H))$ and $\Pi \in \text{Cpl}(P, Q)$. Then, the following are equivalent:*

- (1) Π is optimal;
- (2) there exists an MC-convex $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$ with $\Pi(\partial_{\text{MC}}\varphi) = 1$;
- (3) there exist proper, lsc, convex, and law-invariant $\bar{\varphi} : L_2(H) \rightarrow (-\infty, +\infty]$ and random variables $X, Y : \Omega \rightarrow L_2(H)$ such that

$$(\mathcal{L}_\lambda(X), \mathcal{L}_\lambda(Y)) \sim \Pi \quad \text{and} \quad (X, Y) \in \partial \bar{\varphi} \text{ almost surely.}$$

Proof. The equivalence of (1) and (2) is due to Theorem 1.7.

As MC-convex functions are proper and lsc, we find by Corollary 2.3 that φ is MC-convex if and only if $\bar{\varphi}$ is proper, lsc, convex and law-invariant. Hence, (2) and (3) are equivalent. $\square$

Proposition 3.2. *Let $P \in \mathcal{P}_2(\mathcal{P}_2(H))$. Then, the following are equivalent:*

- (1) P is transport-regular;
- (2) For every MC-convex $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$, we have

$$(22) \quad P(\{\mu : \#\partial_{\text{MC}}\varphi(\mu) > 1\}) = 0;$$

- (3) For every Lipschitz MC-convex $\varphi : \mathcal{P}_2(H) \rightarrow \mathbb{R}$, we have (22).

In particular, for every $Q \in \mathcal{P}_2(\mathcal{P}_2(H))$ the optimal coupling $\Pi \in \text{Cpl}(P, Q)$ is induced by a Monge map $T : \mathcal{P}_2(H) \rightarrow \mathcal{P}_2(H)$ and, in addition, for P -almost all μ , $\text{Cpl}_{\text{opt}}(\mu, T(\mu))$ consists of a single element which is again of Monge type.

Proof. The characterization of transport-regularity of P simply follows from applying Lemma A.1 to the current setting. Specifically, we let $(\mathcal{X}, d) = (\mathcal{P}_2(\mathcal{P}_2(H)), \mathcal{W}_2)$ and $c = -\text{MC}$, and note that for every bounded set $K \subseteq \mathcal{P}_2(H)$ we have that $\varphi_K := (\varphi^{\text{MC}} + \chi_K)^{\text{MC}}$ is Lipschitz continuous. Indeed, for $\mu, \mu' \in \mathcal{P}_2(H)$ we have

$$\varphi_K(\mu) - \varphi_K(\mu') \leq \sup_{\nu \in K} \text{MC}(\mu, \nu) - \text{MC}(\mu', \nu) \leq \mathcal{W}_2(\mu, \mu') \sup_{\nu \in K} \left(\int |y|^2 d\nu \right)^{\frac{1}{2}}.$$

To see the remaining claim, we recall that by Corollary 2.11

$$\partial_{\text{MC}}\varphi(\mu) = \{\nu\} \implies \exists! \pi \in \text{Cpl}_{\text{opt}}(\mu, \nu) \text{ and } \pi \text{ is Monge.}$$

We conclude that Π has the claimed properties. $\square$

Theorem 3.3. *Let $P \in \mathcal{P}_2(\mathcal{P}_2(H))$ and X with $\mathcal{L}_\lambda(X) \sim P$. Assume that*

$$(23) \quad \text{almost surely } \mathcal{L}_\lambda(X) \in \mathcal{P}_2(H) \text{ is transport regular,}$$

$$(24) \quad \mathcal{L}(X) \in \mathcal{P}_2(L_2(H)) \text{ is transport regular.}$$

Then, P is transport regular.

Proof. By Proposition 3.2, we have to show that for every MC-convex $\varphi : \mathcal{P}_2(H) \rightarrow (-\infty, +\infty]$, we have $\#\partial_{\text{MC}}\varphi(\mu) \leq 1$ for P -almost every μ . Recall that the lift $\bar{\varphi} : L_2(H) \rightarrow (-\infty, +\infty]$ is proper, convex lsc and law-invariant by Theorem 2.2. Hence, we have by (24) that

$$\#\partial\bar{\varphi}(X) \leq 1 \quad \text{almost surely.}$$

Because of (23), we can use Theorem 2.4 to obtain

$$\{\mathcal{L}_\lambda(Y) : Y \in \partial\bar{\varphi}(X)\} = \partial_{\text{MC}}\varphi(\mathcal{L}_\lambda(X)) \quad \text{almost surely.}$$

We deduce that $\#\partial_{\text{MC}}\varphi(\mu) \leq 1$ for P -almost every μ . $\square$

Example 3.4 (Regularization on $\mathbb{R}^d$ by Brownian sheet). We construct transport-regular measures based on Brownian sheet. To this end, we consider the law of a (d, d) -Brownian sheet (d -parameters and d -dimensions), denoted by $\mathbb{W}$ and viewed as a probability measure on $L_2(\mathbb{R}^d) = L_2([0, 1]^d, \lambda; \mathbb{R}^d)$. That is, $\mathbb{W}$ is the centered, non-degenerate Gaussian measure on $L_2(\mathbb{R}^d)$ induced by the (d, d) -Brownian sheet. Naturally, $\mathbb{W}$ induces a measure $P \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ via

$$P(A) = \mathbb{W}(\{X \in L_2(\mathbb{R}^d) : \mathcal{L}_\lambda(X) \in A\}) = (\mathcal{L}_\lambda)_\# \mathbb{W}(A),$$

for measurable $A \subseteq \mathcal{P}_2(\mathbb{R}^d)$.

To see that P is transport-regular, we check the assumptions of Theorem 3.3. Since $\mathbb{W}$ is a non-degenerate Gaussian measure on $L_2(\mathbb{R}^d)$, we have by [8, Theorem 6.2.10] that $\mathbb{W}$ is transport-regular. Furthermore, by [10, Theorem 3.1 and the preceding discussion] the Brownian sheet admits almost surely a local time. Thus, $\mathcal{L}_\lambda(X)$ is absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}^d$, for $\mathbb{W}$ -a.e. X . Hence, the assumptions of Theorem 3.3 are satisfied which shows that P is transport regular.

Clearly, measures that are absolutely continuous w.r.t. a transport-regular measure are also transport-regular. Since $\mathbb{W}$ is non-degenerate, P has full support $\mathcal{P}_2(\mathbb{R}^d)$, and it follows immediately that transport-regular measures are dense in $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$.

3.2. Regularization on H by Brownian motion/standard Wiener process. Throughout this section, let $B = (B_t)_{t \in [0, 1]}$ be a one-dimensional Brownian motion and let $W = (W_t)_{t \in [0, 1]}$ be a Q -Wiener process, where $Q : H \rightarrow H$ is a symmetric, positive semi-definite linear operator with finite trace. Recall that a Q -Wiener process $(W_t)_{t \in [0, 1]}$ on H is an H -valued stochastic process such that each W_t is a centered Gaussian random variable with covariance operator tQ , the trajectories are continuous, $W_0 = 0$, and for $t > s$, the increment $W_t - W_s$ is independent of the past and Gaussian with covariance $(t - s)Q$.

The main result of this section establishes the transport regularity of the law of the occupation measure associated with a Q -Wiener process, where the covariance operator Q is non-degenerate. In particular, it applies to the law of the occupation measure of Brownian motion when $H = \mathbb{R}^d$.

Theorem 3.5. *Let Q be a positive-definite, symmetric linear operator on H with finite trace. Let W be a Q -Wiener process. Then, $\mathcal{L}_\lambda(W)$ is almost surely transport-regular.*

Since $\mathcal{L}(W)$ is a non-degenerate Gaussian measure on $L_2(H)$, we have that it is transport-regular by [8, Theorem 6.2.10]. Therefore, using Theorem 3.5, we deduce by Theorem 3.3 the following

Corollary 3.6. *Let Q be a positive-definite, symmetric linear operator on H with finite trace. Let W be a Q -Wiener process. Then, $\mathcal{L}(\mathcal{L}_\lambda(W))$ is a transport-regular measure on $\mathcal{P}_2(H)$.*

To establish Theorem 3.5, following the classical approach, it suffices to show that continuous, convex functions on H are almost surely Gateaux differentiable. By Zajíček [78], the set of non-differentiability points of a continuous convex function on a separable Hilbert space is contained

in a countable union of (c-c)-hypersurfaces. Here, a (c-c)-hypersurface S is a subset of H such that there are $v \in H$ and Lipschitz, convex $f, g : \{v\}^\perp \rightarrow \text{span}(v)$ with

$$S = \{x + f(x) - g(x) : x \in H \text{ with } \langle x, v \rangle = 0\}$$

Consequently, it suffices to prove that, almost surely, the occupation measure of a Q -Wiener process does not charge any (c-c)-hypersurface in H .

In fact, we will even show that, almost surely, Lipschitz hypersurfaces are not charged. To prove this, the crucial observation is that the occupation measure $b_\# \lambda$ associated with a path $b : [0, 1] \rightarrow H$ charges no Lipschitz hypersurface if for all unit vectors $v \in H$ we have

$$\forall A \subseteq [0, 1], \lambda(A) > 0 : \sup_{s, t \in A} \frac{|\langle b(s) - b(t), v \rangle|}{\|b^\perp(s) - b^\perp(t)\|} = \infty,$$

where $b = \langle b, v \rangle v + b^\perp$ and using the convention that $\frac{0}{0} = 0$.

Lemma 3.7. *Let B be a 1-dimensional Brownian motion and let W be an independent Q -Wiener process. Assume that*

$$(25) \quad \mathbb{P}\left(\exists A \subseteq [0, 1], \lambda(A) > 0, \sup_{s, t \in A} \frac{|B_s - B_t|}{\|W_s - W_t\|} < \infty\right) > 0.$$

Then there exists $L > 0$ such that for all $\varepsilon \in (0, 1)$

$$\mathbb{P}\left(\exists A \subseteq [0, 1], \lambda(A) \geq 1 - \varepsilon, \sup_{s, t \in A} \frac{|B_s - B_t|}{\|W_s - W_t\|} \leq L\right) > 0.$$

Proof. Suppose that (25) holds. Then there exist $\delta \in (0, 1)$ and $L > 0$ such that the event $\mathcal{E}(\delta, L, 0, 1)$, defined for $0 \leq a \leq b \leq 1$ by

$$\mathcal{E}(\delta, L, a, b) := \left\{ \exists A \subseteq [a, b], \lambda(A) \geq (b - a)\delta, \sup_{s, t \in A} \frac{|B_s - B_t|}{\|W_s - W_t\|} \leq L \right\},$$

has positive probability.

Fix $\varepsilon \in (0, 1)$. Let $A \subseteq [0, 1]$ be a measurable set with $\lambda(A) > 0$. By the Lebesgue density theorem, there exists $t \in (0, 1)$ and $r > 0$ such that for all $\varepsilon' \in (0, r)$,

$$\frac{\lambda(A \cap (t - \varepsilon', t + \varepsilon'))}{2\varepsilon'} \geq 1 - \varepsilon/2.$$

Let $(q_k)_{k \in \mathbb{N}}$ be a sequence of rational numbers converging to t . Then for sufficiently large n , we have

$$\lim_{k \rightarrow \infty} \frac{\lambda(A \cap (q_k - 1/n, q_k + 1/n))}{2/n} = \frac{\lambda(A \cap (t - 1/n, t + 1/n))}{2/n} \geq 1 - \varepsilon/2.$$

Hence, there exists $(q, n) \in \mathbb{Q} \times \mathbb{N}$ such that

$$\frac{\lambda(A \cap [q, q + 1/n])}{1/n} \geq 1 - \varepsilon.$$

Since $A \subseteq [0, 1]$ with $\lambda(A) > 0$ was arbitrary, we deduce

$$\mathcal{E}(\delta, L, 0, 1) \subseteq \bigcup_{(q, n) \in \mathbb{Q} \times \mathbb{N}} \mathcal{E}(1 - \varepsilon, L, q, q + 1/n).$$

Therefore, there exists some $(q, n) \in \mathbb{Q} \times \mathbb{N}$ such that $\mathbb{P}(\mathcal{E}(1 - \varepsilon, L, q, q + 1/n)) > 0$.

Since (B, W) is a Wiener process, we may assume w.l.o.g. that $q = 0$. Then

$$\mathcal{E}(1 - \varepsilon, L, 0, 1/n) = \left\{ \exists A \subseteq [0, 1/n], \lambda(A) \geq \frac{1 - \varepsilon}{n} : \sup_{s, t \in A} \frac{|B_{s/n} - B_{t/n}|}{\|W_{s/n} - W_{t/n}\|} \leq L \right\}.$$

Note that by the scaling property

$$\sqrt{n} \left((B_{t/n}, W_{t/n}) \right)_{t \in [0,1]} \sim (B, W).$$

Hence,

$$0 < \mathbb{P}(\mathcal{E}(1 - \varepsilon, L, 0, 1/n)) = \mathbb{P}(\mathcal{E}(1 - \varepsilon, L, 0, 1)),$$

which proves the claim. $\square$

Theorem 3.8. *Let B be a 1-dimension Brownian motion and let W be an independent Q -Wiener process. Then, we have*

$$(26) \quad \mathbb{P}\left(\exists A \subseteq [0, 1], \lambda(A) > 0, \sup_{s, t \in A} \frac{|B_s - B_t|}{\|W_s - W_t\|} < \infty\right) = 0.$$

Proof. Suppose, for contradiction, that the probability in (26) is positive. Then, by Lemma 3.7, there exists $L > 0$ such that for all $\varepsilon \in (0, 1)$,

$$\mathbb{P}\left(\exists A \subseteq [0, 1], \lambda(A) \geq 1 - \varepsilon, \sup_{s, t \in A} \frac{|B_s - B_t|}{\|W_s - W_t\|} \leq L\right) > 0.$$

Consider the event

$$E := \left\{ \sup_{s, t \in [0, 1]} \|W_s - W_t\| \leq 1, \inf_{s \in [0, 1/3], t \in [2/3, 1]} |B_s - B_t| \geq 2L \right\},$$

and set $p := \mathbb{P}(E) > 0$. Fix $\varepsilon > 0$ with $4\varepsilon < p$. Let

$$\mathcal{A} := \{A \subseteq [0, 1] : \lambda(A) \geq 1 - \varepsilon\}.$$

For $n \in \mathbb{N}$ and $k = 0, \dots, 2^n - 1$, define $h_n := 2^{-n}$, $t_{n,k} := kh_n$, $t_{n,2^n} := 1$ and

$$E(n, k) := \left\{ \sup_{s, t \in [t_{n,k}, t_{n,k+1}]} \|W_s - W_t\| \leq \sqrt{h_n}, \inf_{s \in [t_{n,k}, t_{n,k} + h_n/3], t \in [t_{n,k+1} - h_n/3, t_{n,k+1}]} |B_s - B_t| \geq 2L\sqrt{h_n} \right\}.$$

Observe that $\{E(n, k) : k = 0, \dots, 2^n - 1\}$ are independent. By scale invariance of Brownian motion and the Wiener process,

$$\mathbb{P}(E(n, k)) = \mathbb{P}(E(n, 0)) = \mathbb{P}(E) = p > 0.$$

For each $n \in \mathbb{N}$, define

$$S_n := \left\{ \omega \in \Omega : h_n \# \{k : \omega \in E(n, k)\} \geq 4\varepsilon \right\}, \quad S := \bigcup_{n \in \mathbb{N}} S_n.$$

Thus, for $\omega \in S$, there exist $n \in \mathbb{N}$ and pairwise distinct indices $k_1, \dots, k_J$ with $4\varepsilon 2^n \leq J \in \mathbb{N}$ such that $\omega \in \bigcap_{j=1}^J E(n, k_j)$.

Now, fix $A \in \mathcal{A}$ and $\omega \in S_n$. We claim that there exists at least on $j_0 \in \{1, \dots, J\}$ such that

$$[t_{n,k_j}, t_{n,k_j} + h_n/3] \cap A \neq \emptyset \quad \text{and} \quad [t_{n,k_j+1} - h_n/3, t_{n,k_j+1}] \cap A \neq \emptyset.$$

To see this, write $B := \bigcup_{j=1}^J [t_{n,k_j}, t_{n,k_j+1}]$. Since $\lambda(A) \geq 1 - \varepsilon$ and $\lambda(B) = Jh_n \geq 4\varepsilon$, we have $\lambda(A \cap B) \geq 3\varepsilon$ and $\lambda(A^C \cap B) \leq \varepsilon$. Therefore,

$$\lambda(A|B) = \frac{1}{1 + \frac{\lambda(A^C \cap B)}{\lambda(A \cap B)}} \geq \frac{1}{1 + \frac{\varepsilon}{3\varepsilon}} = \frac{3}{4}.$$

If $\lambda(A \cap [t_{n,k_j}, t_{n,k_j+1}]) \leq 2/3h_n$ for all $j \in \{1, \dots, J\}$. Then,

$$\lambda(A|B) = \sum_{j=1}^J \frac{\lambda(A \cap [t_{n,k_j}, t_{n,k_j+1}])}{\lambda(B)} \leq \frac{2}{3},$$

which is a contradiction. Hence, we have shown the claim

As consequence of the claim, we find for every $\omega \in S$ and every $A \in \mathcal{A}$ that

$$\sup_{s,t \in A} \frac{|B_s(\omega) - B_t(\omega)|}{\|W_s(\omega) - W_t(\omega)\|} \geq 2L.$$

Finally, write $X_{n,k} := \mathbb{1}_{E(n,k+1)}$ and $Y_n := h_n \sum_{k=1}^{2^n} X_{n,k}$, and note that $(X_{n,k})_{k=1}^{2^n}$ are iid and Bernoulli-distributed with parameter p . We have Y_n has mean p and variance $h_n(1-p)p$. Since $p > 4\varepsilon$ we conclude that $\lim_{n \rightarrow \infty} P(Y_n > 4\varepsilon) = 1$, and therefore $\lim_{n \rightarrow \infty} \mathbb{P}(S_n) = 1$. Consequently,

$$\left\{ \exists A \in \mathcal{A}, \sup_{s,t \in A} \frac{|B_s - B_t|}{\|W_s - W_t\|} \leq L \right\} \subseteq S^C \subseteq S_n^C,$$

and therefore

$$\mathbb{P}\left(\exists A \in \mathcal{A}, \sup_{s,t \in A} \frac{|B_s - B_t|}{\|W_s - W_t\|} \leq L\right) = 0,$$

contradicting our assumption. This proves the theorem. $\square$

When v is some vector in H and $f : \{v\}^\perp \rightarrow \text{span}(v)$, we write

$$\Gamma(f) := \{x + f(x) : x \in \{v\}^\perp\}.$$

Lemma 3.9. *Let $v, w \in H$ with $\|v\| = \|w\| = 1$, and let $f : \{v\}^\perp \rightarrow \text{span}(v)$ be L -Lipschitz. If*

$$\langle v, w \rangle > \frac{1}{2} \left(1 - \frac{L}{\sqrt{1+L^2}}\right)^2,$$

then there exists a Lipschitz function $g : \{w\}^\perp \rightarrow \mathbb{R}$ such that $\Gamma(g) \supseteq \Gamma(f)$.

Proof. We call $\Gamma \subseteq H$ L -Lipschitz in direction v if

$$\forall x, y \in \Gamma, \quad |\langle x - y, v \rangle| \leq L \|\text{pr}_{\{v\}^\perp}(x - y)\|.$$

By the Pythagorean theorem this is equivalent to

$$\forall x, y \in \Gamma, \quad |\langle x - y, v \rangle|^2 \leq \frac{L^2}{1+L^2} \|x - y\|^2.$$

We apply this observation to $\Gamma(f)$ and find by Cauchy-Schwarz, for $x, y \in \Gamma(f)$,

$$|\langle x - y, w \rangle|^2 \leq \left(|\langle x - y, w - v \rangle| + |\langle x - y, v \rangle| \right)^2 \leq \|x - y\|^2 \left(\|v - w\| + \frac{L}{\sqrt{1+L^2}} \right)^2.$$

Hence, to have that $\Gamma(f)$ is Lipschitz in direction w we want $\|v - w\| + \frac{L}{\sqrt{1+L^2}} < 1$. We compute

$$\begin{aligned} \|v - w\|^2 - \left(1 - \frac{L}{\sqrt{1+L^2}}\right)^2 &= 1 - 2\langle v, w \rangle + 2\frac{L}{\sqrt{1+L^2}} - \frac{L^2}{1+L^2} \\ &= \left(1 - \frac{L}{\sqrt{1+L^2}}\right)^2 - 2\langle v, w \rangle < 0, \end{aligned}$$

which yields $\|v - w\| + \frac{L}{\sqrt{1+L^2}} < 1$.

We have shown that $\Gamma(f)$ is also Lipschitz in direction w , and $\{(\text{pr}_{\{w\}^\perp} x, \langle x, w \rangle) : x \in \Gamma(f)\}$ defines the graph of a (real-valued) Lipschitz function. By extending this function, we get Lipschitz $\tilde{g} : \{w\}^\perp \rightarrow \mathbb{R}$ such that $g : \{w\}^\perp \rightarrow \text{span}(w)$ given by $g(x) = x + \tilde{g}(x)w$ is Lipschitz with $\Gamma(g) \supseteq \Gamma(f)$. $\square$

Proof of Theorem 3.5. Fix $v \in H$, $v \neq 0$. We decompose W into a one-dimensional process B^v along $\text{span}(v)$ and an orthogonal component $B^{v,\perp}$ in $\{v\}^\perp$:

$$B_t^v := \langle W_t, v \rangle v, \quad B_t^{v,\perp} := W_t - B_t^v, \quad t \in [0, 1].$$

For $w, w' \in \{v\}^\perp$, we compute

$$\begin{aligned} \text{Cov}(B_t^v, B_s^v) &= \mathbb{E}[\langle W_t, v \rangle \langle W_s, v \rangle] = \min(t, s) \langle Qv, v \rangle, \\ \mathbb{E}[\langle B_t^{v,\perp}, w \rangle \langle B_s^{v,\perp}, w' \rangle] &= \min(t, s) \langle Qw, w' \rangle. \end{aligned}$$

Since Q is non-degenerate and non-negative, we have $\langle Qv, v \rangle > 0$, and $B^{v,\perp}$ is a centered $Q|_{\{v\}^\perp}$ -Wiener process. Thus, by Theorem 3.8,

$$(27) \quad \mathbb{P}\left(\exists A \subseteq [0, 1], \lambda(A) > 0 : \sup_{s, t \in A} \frac{\|B_s^v - B_t^v\|}{\|B_s^{v,\perp} - B_t^{v,\perp}\|} < \infty\right) = 0.$$

Let $\text{Lip}(v)$ denote the set of Lipschitz functions $f : \{v\}^\perp \rightarrow \text{span}(v)$. As a direct consequence of (27),

$$\mathbb{P}\left(\forall f \in \text{Lip}(v) : \lambda(\{t \in [0, 1] : W_t \in \Gamma(f)\}) = 0\right) = 1.$$

Denote by Ω^v the corresponding full-probability event.

Since H is separable, there exists a countable dense family of unit vectors $(v_n)_{n \in \mathbb{N}}$ in the unit ball of H . Define

$$\tilde{\Omega} := \bigcap_{n \in \mathbb{N}} \Omega^{v_n}.$$

By Lemma 3.9, for every $v \in H$ and $f \in \text{Lip}(v)$ there exist $n \in \mathbb{N}$ and $g \in \text{Lip}(v_n)$ such that $\Gamma(f) = \Gamma(g)$. Consequently, for every $\omega \in \tilde{\Omega}$,

$$\forall v \in H, \forall f \in \text{Lip}(v) : \lambda(\{t \in [0, 1] : W_t(\omega) \in \Gamma(f)\}) = 0.$$

Hence, on $\tilde{\Omega}$, the occupation measure $(t \mapsto W_t(\omega))_{\#} \lambda$ does not charge Lipschitz hypersurfaces of H . By Zajíček [78, Theorem 2], this implies

$$\mathbb{P}\left(\forall \varphi : H \rightarrow \mathbb{R} \text{ continuous, convex, } \lambda(\{t \in [0, 1], \#\partial\varphi(W_t) > 1\}) = 0\right) = 1.$$

Therefore, $\mathcal{L}_\lambda(W)$ is almost surely transport-regular by Lemma A.1. $\square$

4. OPTIMAL TRANSPORT ON $\mathcal{P}_2^N(H)$

The aim of the section is apply the fundamental theorem of optimal transport to transport on $\mathcal{P}_2^N(H)$ to establish a duality in terms of MC-convex functions. Properties of MC-convex functions, in particular, the connection to convex analysis via an appropriate version of the Lions lift will be the subject of the subsequent section.

Starting with the max-covariance of two probabilities $\mu, \nu \in \mathcal{P}_2(H)$, which is given by

$$\text{MC}(\mu, \nu) = \sup_{\pi \in \text{Cpl}(\mu, \nu)} \int \langle x, y \rangle d\pi(x, y),$$

we now define its iterated counterpart for probabilities in $\mathcal{P}_2^N(H)$.

Definition 4.1. Let $N > 1$ and $P, Q \in \mathcal{P}_2^N(H)$. We inductively define

$$\text{MC}(P, Q) := \sup_{\Pi \in \text{Cpl}(P, Q)} \int \text{MC}(p, q) d\Pi(p, q),$$

where $\text{MC}(p, q)$ is the max-covariance of $p, q \in \mathcal{P}_2^{N-1}(H)$.

Note that we always use the same symbol MC for all max-covariance functionals $\text{MC} : \mathcal{P}_2^N(H) \times \mathcal{P}_2^N(H) \rightarrow \mathbb{R}$, $N \geq 1$ and trust that this causes no confusion as we always specify the level of iterations we work with. In particular, for fixed N , we denote with capital letters probabilities $P, Q \in \mathcal{P}_2^N(H)$ and with lower case letters probabilities $p, q \in \mathcal{P}_2^{N-1}(H)$ on the underlying space.

We briefly state the connection between the iterated MC-functional and the iterated Wasserstein distance. Note that the iterated Wasserstein distance on $\mathcal{P}_2^N(H)$, $N > 1$, is inductively defined as

$$\mathcal{W}_2(P, Q) := \inf_{\Pi \in \text{Cpl}(P, Q)} \int \mathcal{W}_2(p, q) d\Pi(p, q).$$

Given a Polish space $\mathcal{X}$, we write I for the intensity operator $I : \mathcal{P}_2^2(\mathcal{X}) = \mathcal{P}_2(\mathcal{P}_2(\mathcal{X})) \rightarrow \mathcal{P}_2(\mathcal{X})$ satisfying the characteristic property, for all measurable bounded $f : \mathcal{X} \rightarrow \mathbb{R}$,

$$\int_{\mathcal{X}} f dIP := \int_{\mathcal{P}_2(\mathcal{X})} \int_{\mathcal{X}} f(x) dp(x) dP(p).$$

As this operator is well-defined on any Polish space $\mathcal{X}$, it is in particular possible if $\mathcal{X}$ is itself a space of measures. This enables us to define the iterated intensity operator

$$I^{N-1} : \mathcal{P}_2^N(H) \rightarrow \mathcal{P}_2(H).$$

The next lemma is then a straightforward consequence of the definition of $\mathcal{W}_2$, MC, and I^{N-1} .

Lemma 4.2. *Let $P, Q \in \mathcal{P}_2(H)$. We then have*

- (1) $\mathcal{W}_2^2(P, Q) = \text{MC}(P, P) + \text{MC}(Q, Q) - 2\text{MC}(P, Q)$,
- (2) $\text{MC}(P, P) = \int |x|^2 (I^{N-1}P)(dx)$,
- (3) $\Pi \in \text{Cpl}(P, Q)$ is $\mathcal{W}_2$ -optimal iff it is MC-optimal.

This simple observation motivates the study of the max-covariance functional on iterated spaces of probability measures. In order to obtain a characterization of its optimizers, we follow in the footsteps of classical transport theory and specialize the corresponding definition of c -transform and c -subdifferential to the case of $c = \text{MC}$.

Definition 4.3. Let $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$ be proper. We define its MC-transform as

$$\varphi^{\text{MC}}(Q) = \sup_{P \in \mathcal{P}_2^N(H)} \text{MC}(P, Q) - \varphi(P).$$

A function $\psi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$ is called MC-convex if there exists a proper function $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$ such that $\psi = \varphi^{\text{MC}}$. Moreover, the MC-subdifferential of φ at $P \in \mathcal{P}_2^N(H)$ is defined as

$$\partial_{\text{MC}}\varphi(P) = \{Q \in \mathcal{P}_2^N(H) : \varphi(R) \geq \varphi(P) + \text{MC}(R, Q) - \text{MC}(P, Q) \text{ for every } R \in \mathcal{P}_2^N(H)\},$$

and we write

$$\partial_{\text{MC}}\varphi := \{(P, Q) : Q \in \partial_{\text{MC}}\varphi(P), P \in \mathcal{P}_2^N(H)\}.$$

We remark that for an MC-convex function φ , the MC-subdifferential $\partial_{\text{MC}}\varphi$ consists precisely of those $(P, Q) \in \mathcal{P}_2^N(H) \times \mathcal{P}_2^N(H)$ with $\varphi(P) + \varphi^{\text{MC}}(Q) = \text{MC}(P, Q)$, see also [75, Definition 5.2]. Applying the fundamental theorem of optimal transport to this setting yields

Theorem 4.4. *Let $P, Q \in \mathcal{P}_2^N(H)$. Then we have the duality relation*

$$\text{MC}(P, Q) = \inf_{\substack{\varphi : \mathcal{P}_2^{N-1}(H) \rightarrow (-\infty, +\infty] \\ \text{MC-convex}}} \int \varphi dP + \int \varphi^{\text{MC}} dQ,$$

and there exist both a primal optimizer $\Pi \in \text{Cpl}(P, Q)$ and a dual optimizer $\varphi : \mathcal{P}_2^{N-1}(H) \rightarrow (-\infty, +\infty]$.

Moreover, the following complementary slackness condition holds: Candidates Π and φ are optimal for primal and dual problem, respectively, if and only

$$\Pi(\partial_{\text{MC}}\varphi) = 1,$$

that is, Π is concentrated on the MC-subdifferential of φ .

Remark 4.5. Any MC-convex function $\varphi : \mathcal{P}_2^{N-1}(H) \rightarrow (-\infty, +\infty]$ is integrable with respect to every measure $P \in \mathcal{P}_2^N(H)$. In particular, the integrals $\int \varphi dP$ and $\int \varphi^{\text{MC}} dQ$ in the dual formulation are always well defined (possibly taking the value $+\infty$).

Indeed, since φ is MC-convex, its MC-transform is proper. Hence, there is $q \in \mathcal{P}_2^{N-1}(H)$ such that $\varphi^{\text{MC}}(q) \in \mathbb{R}$. Recalling that $\mathcal{W}_2^2(p, q) = \text{MC}(p, p) - 2\text{MC}(p, q) + \text{MC}(q, q)$ by Lemma 4.2 and using the inequality $\varphi(p) \geq \text{MC}(p, q) - \varphi^{\text{MC}}(q)$, we conclude that 2φ is dominated from below by the P -integrable function $p \mapsto \text{MC}(p, p) + \text{MC}(q, q) - \mathcal{W}_2^2(p, q) - 2\varphi^{\text{MC}}(q)$.

Above we discussed how to solve the optimal transport problem on the outer most layer. However, this procedure can be iterated according to the dynamic programming principle. More specifically, if $P, Q \in \mathcal{P}_2^N(H)$ and $\Pi \in \text{Cpl}_{\text{opt}}(P, Q)$, then one can again solve the optimal transport problems between Π -almost every pair $(p, q) \in \mathcal{P}_2^{N-1}(H) \times \mathcal{P}_2^{N-1}(H)$ and gets a selection of optimizers $\pi_{p,q} \in \text{Cpl}_{\text{opt}}(p, q)$. By gluing Π and the kernel $(\pi_{p,q})_{p,q}$, one arrives at a measure that contains the information of the optimal transport on the first two layers. Continuing this procedure leads to the following concept of N -couplings.

To give the definition of N -couplings, we need to introduce a notation for iterated pushforwards. Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces and $f : \mathcal{X} \rightarrow \mathcal{Y}$ be Borel. We write $\mathcal{P}[f]$ for the pushforward map

$$\mathcal{P}[f] : \mathcal{P}(\mathcal{X}) \rightarrow \mathcal{P}(\mathcal{Y}) : \mu \mapsto f_{\#}\mu$$

and inductively

$$\mathcal{P}^N[f] : \mathcal{P}^N(\mathcal{X}) \rightarrow \mathcal{P}^N(\mathcal{Y}) : P \mapsto (\mathcal{P}^{N-1}[f])_{\#}P.$$

Definition 4.6. For $P, Q \in \mathcal{P}_2^N(H)$, we set

$$\text{Cpl}^N(P, Q) = \{\Pi \in \mathcal{P}_2^N(H \times H) : \mathcal{P}^N[\text{pr}_1](\Pi) = P, \mathcal{P}^N[\text{pr}_2](\Pi) = Q\},$$

and call elements of $\text{Cpl}^N(P, Q)$ N -couplings of P and Q .

The following results make precise that elements of $\text{Cpl}^N(P, Q)$ encode the full information of the transport across all layers.

Proposition 4.7. For $P, Q \in \mathcal{P}_2^N(H)$ we have

$$(28) \quad \text{MC}(P, Q) = \max_{\Pi \in \text{Cpl}^N(P, Q)} \int x \cdot y dI^{N-1}\Pi(x, y).$$

Proof. For $N = 1$, the assertion follows directly from the definition of MC.

Next, assume that the statement holds for $N - 1$. By definition we have

$$\text{MC}(P, Q) = \max_{\pi \in \text{Cpl}(P, Q)} \int \text{MC}(p, q) d\pi(p, q).$$

Fix an optimizer $\pi \in \text{Cpl}(P, Q) \subseteq \mathcal{P}_2(\mathcal{P}_2^{N-1}(H) \times \mathcal{P}_2^{N-1}(H))$. By the induction hypothesis,

$$(29) \quad \text{MC}(p, q) = \inf_{\Pi_{p,q} \in \text{Cpl}^{N-1}(p, q)} \int x \cdot y dI^{N-2}\Pi_{p,q}(x, y),$$

for $p, q \in \mathcal{P}_2^{N-1}(H)$. By a standard measurable selection argument (e.g. the Jankov-von Neumann uniformization theorem [54, Theorem 18.1]), there is a universally measurable optimal selector

$$\Phi : \mathcal{P}_2^{N-1}(H) \times \mathcal{P}_2^{N-1}(H) \rightarrow \mathcal{P}_2^{N-1}(H \times H) : (p, q) \mapsto \Pi_{p,q}$$

for the right-hand side of (29). Define $\Pi := \Phi_{\#}\pi \in \mathcal{P}_2^N(H \times H)$.

The marginal constraint $\Pi_{p,q} \in \text{Cpl}^{N-1}(p, q)$ amounts to $\mathcal{P}^{N-1}[\text{pr}_i] \circ \Phi = \text{pr}_i$ for $i \in \{1, 2\}$. Therefore, we obtain that

$$\mathcal{P}^N[\text{pr}_1](\Pi) = \mathcal{P}^N[\text{pr}_1](\mathcal{P}[\Phi](\pi)) = \mathcal{P}[\mathcal{P}^{N-1}[\text{pr}_1] \circ \Phi](\pi) = \mathcal{P}[\text{pr}_1](\pi) = P,$$

and with the same argument $\mathcal{P}^N[\text{pr}_2](\Pi) = Q$, that is, $\Pi \in \text{Cpl}^N(P, Q)$. Note that $\int \Pi_{p,q} \pi(dp, dq) = I(\Phi_{\#}\pi) = I\Pi$. Hence, by optimality of π ,

$$\begin{aligned} \text{MC}(P, Q) &= \int \text{MC}(p, q) \pi(dp, dq) \\ &= \iint x \cdot y I^{N-2} \Pi_{p,q}(dx, dy) \pi(dp, dq) \\ &= \int x \cdot y I^{N-1} \Pi(dx, dy). \end{aligned}$$

This completes the induction step and proves (28). $\square$

5. PROBABILISTIC REPRESENTATION OF ITERATED PROBABILITY MEASURES

The adapted Lions lift is designed to provide a Hilbert space representation of functionals on $\mathcal{P}_2^N(H)$ and will play a central role in analyzing the structure of $\mathcal{W}_2^2$ -optimal couplings of measures on $\mathcal{P}_2^N(H)$. In this section, we introduce the necessary notation and establish basic properties of the representation of iterated probability measures in terms of filtrations. These results form the foundation for the study of the adapted Lions lift in the subsequent Section 6.

Recall from Section 1.1 that $(\mathcal{F}_t)_{t=1}^N$ denotes the coordinate filtration⁵ on $([0, 1]^N, \lambda)$ and that

$$(30) \quad \mathcal{L}^{\text{ad}}(X) = \mathcal{L}(\mathcal{L}(\dots \mathcal{L}(X|\mathcal{F}_{N-1}) \dots |\mathcal{F}_1)) \in \mathcal{P}_2^N(H), \quad X \sim_{\text{ad}} Y \text{ if } \mathcal{L}^{\text{ad}}(X) = \mathcal{L}^{\text{ad}}(Y)$$

for $X, Y \in L_2([0, 1]^N; H)$. In this section we show that every $P \in \mathcal{P}_2^N(H)$ can be represented in this way, establish an adapted transfer principle and derive a Skorokhod representation theorem that links convergence on $L_2([0, 1]^N; H)$ and $\mathcal{P}_2^N(H^N)$.

To this end, it will be necessary to disentangle towers of measures / iterated conditional laws which will require considerable notation which we now start to set up. In this section $\mathcal{X}$ will always denote a Polish metric space. We will use the notation (30) also for $\mathcal{X}$ instead of the separable Hilbert space H and we employ the shorthand

$$L_2^N(\mathcal{X}) := L_2([0, 1]^N; \mathcal{X}).$$

For $t \in \{1, \dots, N\}$ we write $\mathcal{L}^{\mathcal{F}_t}(X) = \mathcal{L}(X|\mathcal{F}_t)$. Writing multiple σ -algebras in the superscript indicates iterated conditional expectations, e.g. $\mathcal{L}^{\mathcal{F}_{2:1}}(X) = \mathcal{L}^{\mathcal{F}_2, \mathcal{F}_1}(X) = \mathcal{L}(\mathcal{L}(X|\mathcal{F}_2)|\mathcal{F}_1)$. Moreover, we use the abbreviations

$$(31) \quad \begin{aligned} \text{ip}_t(X) &= \mathcal{L}^{\mathcal{F}_{(N-1):t}}(X) \quad t = 1, \dots, N-1 \\ \text{ip}_N(X) &= X. \end{aligned}$$

Note that $\text{ip}_t(X)$ takes values in the space $A_t := \mathcal{P}_2^{N-t}(H)$, where we use the convention $\mathcal{P}_2^0(H) = H$. We write

$$\text{ip}_{1:t}(X) = (\text{ip}_1(X), \dots, \text{ip}_t(X)), \quad \text{ip}(X) = \text{ip}_{1:N}(X), \quad A_{1:t} = A_1 \times \dots \times A_t$$

and note that $\mathcal{L}^{\text{ad}}(X) = \mathcal{L}(\text{ip}_1(X))$.

We also recall from the introduction that $T : [0, 1]^N \rightarrow [0, 1]^N$ being bi-adapted means that T is a bijection for which T, T^{-1} are $(\mathcal{F}_t, \mathcal{F}_t)$ -measurable for every $t \leq N$.

⁵Note that instead of $([0, 1]^N, \lambda, (\mathcal{F}_t)_{t=1}^N)$ we could also work with an abstract filtered space $(\Omega, \mathcal{G}, \mathbb{P}, (\mathcal{G}_t)_{t=1}^N)$ where $(\Omega, \mathcal{G}, \mathbb{P})$ is standard Borel and for each $t < N$ there exists a continuous $\mathcal{G}_{t+1}$ -measurable random variable which is independent of $\mathcal{G}_t$.

Lemma 5.1. *Let $T : [0, 1]^N \rightarrow [0, 1]^N$ be bi-adapted with $T_{\#}\lambda = \lambda$. Then for $s \leq t$, $X \in L_2^N(H)$*

$$\mathcal{L}^{\mathcal{F}_t, \dots, \mathcal{F}_s}(X \circ T^{-1}) = \mathcal{L}^{\mathcal{F}_t, \dots, \mathcal{F}_s}(X) \circ T^{-1}.$$

In particular, $\text{ip}_t(X \circ T^{-1}) = \text{ip}_t(X) \circ T^{-1}$ and $\mathcal{L}^{\text{ad}}(X \circ T^{-1}) = \mathcal{L}^{\text{ad}}(X)$.

Proof. We first show that for every random variable $X \in L_2^N(\mathcal{X})$ and every t , we have $\mathcal{L}(X \circ T^{-1} | \mathcal{F}_t) = \mathcal{L}(X | \mathcal{F}_t) \circ T^{-1}$. To that end, let $f : \mathcal{X} \rightarrow \mathbb{R}$ be bounded and Borel, and write

$$Y := \mathbb{E}[f(X \circ T^{-1}) | \mathcal{F}_t] \circ T$$

and note that Y is $\mathcal{F}_t$ -measurable. Moreover, we find for every bounded $\mathcal{F}_t$ -measurable random variable Z using that $Z \circ T^{-1}$ is $\mathcal{F}_t$ -measurable

$$\mathbb{E}[YZ] = \mathbb{E}[\mathbb{E}[f(X \circ T^{-1}) | \mathcal{F}_t] Z \circ T^{-1}] = \mathbb{E}[f(X \circ T^{-1}) Z \circ T^{-1}] = \mathbb{E}[f(X) Z].$$

Hence, $Y = \mathbb{E}[f(X) | \mathcal{F}_t]$. As this is true for every bounded Borel $f : \mathcal{X} \rightarrow \mathbb{R}$, we conclude $\mathcal{L}(X \circ T^{-1} | \mathcal{F}_t) = \mathcal{L}(X | \mathcal{F}_t) \circ T^{-1}$.

Since this claim is proven for random variables with values in an Polish space, we can iteratively apply it and obtain

$$\mathcal{L}^{\mathcal{F}_t, \dots, \mathcal{F}_s}(X \circ T^{-1}) = \mathcal{L}^{\mathcal{F}_t, \dots, \mathcal{F}_{s+1}}(\mathcal{L}^{\mathcal{F}_s}(X) \circ T^{-1}) = \dots = \mathcal{L}^{\mathcal{F}_t, \dots, \mathcal{F}_s}(X) \circ T^{-1}. \quad \square$$

Proposition 5.2. *Let $P \in \mathcal{P}_2^N(\mathcal{X})$. Then there is $X \in L_2^N(\mathcal{X})$ such that $\mathcal{L}^{\text{ad}}(X) = P$.*

Proof. We assume w.l.o.g. that $\mathcal{X}$ is uncountable and we denote with $(U_t)_{t=1}^N$ the coordinate process on $[0, 1]^N$. We show by induction on N that there is a Borel map

$$\Phi_N : \mathcal{P}_2^N(\mathcal{X}) \times [0, 1]^N \rightarrow \mathcal{X}$$

such that $\mathcal{L}^{\text{ad}}(\Phi_N(P, U_{1:N})) = P$ for every $P \in \mathcal{P}_2^N(\mathbb{R}^d)$.

To show the claim for $N = 1$, let $\Psi : \mathbb{R} \rightarrow \mathcal{X}$ be a Borel isomorphism. If $\mu \in \mathcal{P}_2(\mathbb{R})$, we write Q_μ for its quantile function. We set

$$\Phi_1(P, U_1) = \Psi(Q_{\Psi_{\#}^{-1}P}(U_1))$$

and observe that

$$\mathcal{L}(\Phi_1(P, U_1)) = \Psi_{\#} \mathcal{L}((Q_{\Psi_{\#}^{-1}P}(U_1))) = \Psi_{\#} \Psi_{\#}^{-1} P = P.$$

Now suppose that the claim is true for $N - 1$. By the claim for $N = 1$ applied with the metric space $\mathcal{P}_2^{N-1}(\mathcal{X})$ there is

$$\Phi_1 : \mathcal{P}_2^N(\mathcal{X}) \times [0, 1] \rightarrow \mathcal{P}_2^{N-1}(\mathcal{X})$$

such that $\mathcal{L}(\Phi_1(P, U_1)) = P$. By the claim for $N - 1$ there is

$$\Phi_{N-1} : \mathcal{P}_2^{N-1}(\mathcal{X}) \times [0, 1]^{N-1} \rightarrow \mathcal{X}$$

such that $\mathcal{L}(\mathcal{L}^{\mathcal{F}_{N-1:2}}(\Phi_{N-1}(p, U_{2:N}))) = p$ for every $p \in \mathcal{P}_2^{N-1}(\mathcal{X})$. We then set

$$\Phi_N(P, U_{1:N}) := \Phi_{N-1}(\Phi_1(P, U_1), U_{2:N}).$$

As $\mathcal{F}_1 = \sigma(U_1)$ is independent of $U_{2:N}$ and $\Phi_1(P, U_1)$ is $\mathcal{F}_1$ -measurable, $\mathcal{L}^{\mathcal{F}_{N-1:1}}(\Phi_N(P, U_{1:N})) = \Phi_1(P, U_1)$. Hence, $\mathcal{L}^{\text{ad}}(\Phi_N(P, U_{1:N})) = \mathcal{L}(\Phi_1(P, U_1)) = P$. $\square$

Note that $\text{Cpl}(\mathcal{L}^{\text{ad}}(X), \mathcal{L}^{\text{ad}}(Y))$ does not capture the entire information of the joint adapted distribution of X and Y . For this reason we introduce the following notion.

Lemma 5.3. *For $P, Q \in \mathcal{P}_2^N(H)$ we have*

$$\text{Cpl}^N(P, Q) = \{\mathcal{L}^{\text{ad}}(X, Y) : X \sim_{\text{ad}} P, Y \sim_{\text{ad}} Q\}.$$

Proof. If $\Pi \in \text{Cpl}^N(P, Q)$, then by Proposition 5.2 there is an $H \times H$ -valued random variable (X, Y) with $\mathcal{L}^{\text{ad}}(X, Y) = \Pi$. We then have $\mathcal{L}^{\text{ad}}(X) = \mathcal{L}^{\text{ad}}(\text{pr}_1(X, Y)) = \mathcal{P}^N[\text{pr}_1](\Pi) = P$ and with the same argument $\mathcal{L}^{\text{ad}}(Y) = Q$. Conversely, if $X \sim_{\text{ad}} P$ and $Y \sim_{\text{ad}} Q$, then $\mathcal{P}^N[\text{pr}_1](\mathcal{L}^{\text{ad}}(X, Y)) = \mathcal{L}^{\text{ad}}(X) = P$ and $\mathcal{P}^N[\text{pr}_2](\mathcal{L}^{\text{ad}}(X, Y)) = \mathcal{L}^{\text{ad}}(Y) = Q$. $\square$

Corollary 5.4. For $P, Q \in \mathcal{P}_2^N(H)$ we have

$$\text{MC}(P, Q) = \max_{X \sim_{\text{ad}} P, Y \sim_{\text{ad}} Q} \mathbb{E}[X \cdot Y].$$

Proof. By Proposition 4.7, there is $\Pi \in \text{Cpl}^N(P, Q) \subseteq \mathcal{P}_2^N(H \times H)$ optimal for $\text{MC}(P, Q)$. By Proposition 5.2, there exists an $H \times H$ -valued random variable (X, Y) such that $\mathcal{L}^{\text{ad}}(X, Y) = \Pi$. In particular, we have $\mathcal{L}^{\text{ad}}(X) = \mathcal{L}^{\text{ad}}(\text{pr}_1(X, Y)) = \mathcal{P}^N[\text{pr}_1](\mathcal{L}^{\text{ad}}(X, Y)) = \mathcal{P}^N[\text{pr}_1](\Pi) = P$ and with the same argument $\mathcal{L}^{\text{ad}}(Y) = Q$. $\square$

Lemma 5.5. Let $P, P' \in \mathcal{P}^N(\mathcal{X})$ and X, X' be random variables with $X \sim_{\text{ad}} P$, $X' \sim_{\text{ad}} P'$. Let $c^{(0)} := c : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be continuous and $|c(x, x')| \leq f(x) + g(x')$ for $f(X), g(X') \in L^1(\mathbb{P})$. Define recursively for $t = 1, \dots, N$

$$\begin{aligned} c^{(t)} : \mathcal{P}^t(\mathcal{X}) \times \mathcal{P}^t(\mathcal{X}) &\rightarrow \mathbb{R}, \\ (Q, Q') &\mapsto \inf_{\pi \in \text{Cpl}(Q, Q')} \int c^{(t-1)}(q, q') d\pi(q, q'). \end{aligned}$$

Then, for $X \sim_{\text{ad}} P$

$$c^{(N)}(P, P') = \inf_{\substack{T: [0,1]^N \rightarrow [0,1]^N \\ \text{bi-adapted, } T_{\#}\lambda = \lambda}} \mathbb{E}[c(X \circ T, X')].$$

The proof of Lemma 5.5 is postponed to Appendix B.

Corollary 5.6. Let $X, X' \in L_2^N(\mathcal{X})$ with $\mathcal{L}^{\text{ad}}(X) = \mathcal{L}^{\text{ad}}(X')$. Then, for every $\varepsilon > 0$, there exists a measure-preserving bi-adapted bijection $T : [0, 1]^N \rightarrow [0, 1]^N$ such that $\lambda(|X' - X \circ T| \geq \varepsilon) < \varepsilon$.

Proof. Using that $\mathcal{L}^{\text{ad}}(X) = \mathcal{L}^{\text{ad}}(X')$, the claim follows directly from Lemma 5.5 applied with $\mathcal{X} := (0, 1) \times H$ and $c(x, x') := \min(d(x, x'), 1)$. $\square$

Proposition 5.7. For $P, Q \in \mathcal{P}_2^N(H)$ we have

$$\mathcal{W}_2(P, Q) = \min_{X \sim_{\text{ad}} P, Y \sim_{\text{ad}} Q} \|X - Y\|_2.$$

Let $P, P_1, P_2, \dots \in \mathcal{P}_2^N(H)$ and $X \sim_{\text{ad}} P$. Then the following are equivalent:

- (1) $\mathcal{W}_2(P_n, P) \rightarrow 0$,
- (2) there are $X_n \sim_{\text{ad}} P_n$ such that $\|X_n - X\|_2 \rightarrow 0$.

In particular, the map $\mathcal{L}^{\text{ad}} : (L_2^N(H), \|\cdot\|_2) \rightarrow (\mathcal{P}_2^N(H), \mathcal{W}_2)$ is continuous.

Proof. First note that Corollary 5.4 and Lemma 4.2 imply that for $P, Q \in \mathcal{P}_2^N(H)$ we have

$$\mathcal{W}_2(P, Q) = \min_{X \sim_{\text{ad}} P, Y \sim_{\text{ad}} Q} \|X - Y\|_2.$$

To see that (1) implies (2) first observe that applying this fact to P and P_n yields $Y_n, Z_n \in L_2^N(H)$ with $Y_n \sim_{\text{ad}} P$ and $Z_n \sim_{\text{ad}} P_n$ such that $\mathcal{W}_2(P_n, P) = \|Y_n - Z_n\|_2$. By Corollary 5.6 there are isomorphisms T_n such that $\|Y_n \circ T_n - X\|_{L_2} \rightarrow 0$. We set $X_n := Z_n \circ T_n$ and not that by Lemma 5.1, we have $\mathcal{L}^{\text{ad}}(X_n) = \mathcal{L}^{\text{ad}}(Z_n) = P_n$. Moreover, we have

$$\|X_n - X\|_2 = \|Z_n \circ T_n - Y_n \circ T_n + Y_n \circ T_n - X\|_2 \leq \|Z_n - Y_n\|_2 + \|Y_n \circ T_n - X\|_2 \rightarrow 0.$$

Next, we show that (2) implies (1). To that end, we show by backward induction on t that $\text{ip}_t(X_n) \rightarrow \text{ip}_t(X)$ in $L_2^N(\mathcal{P}_2^{N-t}(H))$. For $t = N$ this is trivial. Suppose it is true for $t + 1$. Then,

$$\begin{aligned} \mathcal{W}_2^2(\text{ip}_t(X_n), \text{ip}_t(X)) &= \mathcal{W}_2^2(\mathcal{L}^{\mathcal{F}_t}(\text{ip}_{t+1}(X_n)), \mathcal{L}^{\mathcal{F}_t}(\text{ip}_{t+1}(X))) \\ &\leq \mathbb{E}[\mathcal{W}_2^2(\text{ip}_{t+1}(X_n), \text{ip}_{t+1}(X)) | \mathcal{F}_t] \quad \text{a.s.}, \end{aligned}$$

where the inequality follows from the observation that $\mathcal{L}^{\mathcal{F}_t}(\text{ip}_{t+1}(X_n), \text{ip}_{t+1}(X))$ is a.s. a coupling in $\text{Cpl}(\text{ip}_t(X_n), \text{ip}_t(X))$. Hence,

$$\mathbb{E}[\mathcal{W}_2^2(\text{ip}_t(X_n), \text{ip}_t(X))] \leq \mathbb{E}[\mathcal{W}_2^2(\text{ip}_{t+1}(X_n), \text{ip}_{t+1}(X))] \rightarrow 0.$$

This finishes the induction and in particular we have $\text{ip}_1(X_n) \rightarrow \text{ip}_1(X)$ in $L_2^N(\mathcal{P}_2^{N-1}(H))$. Hence, $\mathcal{L}^{\text{ad}}(X_n) = \mathcal{L}(\text{ip}_1(X_n)) \rightarrow \mathcal{L}(\text{ip}_1(X)) = \mathcal{L}^{\text{ad}}(X)$. $\square$

6. THE ADAPTED LIONS LIFT

In this section we develop the basic theory of adapted Lions lift and its applications to MC-convex functions.

In the following we write $L_2^N(H) := L_2([0, 1]^N; H)$ and for $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$, the function $\bar{\varphi} : L_2^N(H) \rightarrow (-\infty, +\infty]$ always denotes the adapted Lions lift $\bar{\varphi}(X) = \varphi(\mathcal{L}^{\text{ad}}(X))$ of φ .

We will first establish the crucial result Theorem 6.1 on the identification of MC-conjugacy and convex conjugacy on $L_2^N(H)$. In particular this allows us to prove Theorem 1.3 on the characterization of optimizers in terms of adapted-law invariant convex functions. We also introduce the notion of MC-differentiability of MC-functions which yields a characterization of transport regularity.

6.1. Connection between MC-transform and convex conjugate of the lift. Note that for $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$, we have $\text{dom}(\bar{\varphi}) = \{X \in L_2^N(H) : \mathcal{L}^{\text{ad}}(X) \in \text{dom}(\varphi)\}$ and we will see in Lemma 6.10 below that the corresponding statement holds also for the continuity points.

Theorem 6.1. *Let $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$ be proper. Then*

$$(32) \quad \bar{\varphi}^* = \overline{\varphi^{\text{MC}}}.$$

In particular, the convex conjugate of an adapted-law invariant function is adapted-law invariant. Moreover, the following are equivalent:

- (1) $\bar{\varphi}$ is lsc convex;
- (2) φ is MC-convex.

Proof. For $Y \in L_2^N(H)$ we write $Q = \mathcal{L}^{\text{ad}}(Y)$. By Lemma 5.5 we find

$$\begin{aligned} \bar{\varphi}^*(Y) &= \sup_{X \in L_2^N(H)} \mathbb{E}[X \cdot Y] - \bar{\varphi}(X) = \sup_{P \in \mathcal{P}_2^N(H)} \sup_{X \sim_{\text{ad}} P} \mathbb{E}[X \cdot Y] - \varphi(P) \\ &= \sup_{P \in \mathcal{P}_2^N(H)} \text{MC}(P, Q) - \varphi(P) = \varphi^{\text{MC}}(Q) = \overline{\varphi^{\text{MC}}}(Y). \end{aligned}$$

Next, we prove the equivalence of (1) and (2).

Suppose that $\bar{\varphi}$ is lsc convex. By first applying the Fenchel–Moreau theorem and then (32) twice we find

$$\bar{\varphi} = \bar{\varphi}^{**} = \overline{\varphi^{\text{MC}}}^* = \overline{\varphi^{\text{MCMC}}}.$$

Hence, $\varphi = \varphi^{\text{MCMC}}$, which shows that φ is MC-convex.

Conversely assume that φ is MC-convex. Then $\varphi = \psi^{\text{MC}}$ for some proper function $\psi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$. By (32), we have $\bar{\varphi} = \psi^{\text{MC}} = \bar{\psi}^*$, which shows that $\bar{\varphi}$ is lsc convex. $\square$

6.2. MC-order. In this section we introduce an order on $\mathcal{P}_2^N(H)$ which will subsequently be used to establish the existence of the Lions derivative in the adapted setting.

Definition 6.2. Let $P, Q \in \mathcal{P}_2^N(H)$. We say that P is smaller than Q in the MC-order, written as $P \preceq_{\text{MC}} Q$ if $\int \varphi dP \leq \int \varphi dQ$ for every MC-convex function $\varphi : \mathcal{P}_2^{N-1}(H) \rightarrow (-\infty, +\infty]$.

We recall the convention $\mathcal{P}_2^0(H) = H$ and note that the MC-order extends the classical convex order on $\mathcal{P}_2(H)$. The main result of this subsection is the following characterization of the MC-order, which is an analogue of Proposition 2.8.

Proposition 6.3. Let $P, Q \in \mathcal{P}_2^N(H)$. Then the following are equivalent:

- (1) $P \preceq_{\text{MC}} Q$,
- (2) $\text{MC}(P, R) \leq \text{MC}(Q, R)$ for every $R \in \mathcal{P}_2^N(H)$.

We note that implies Proposition 6.3 that for every $R \in \mathcal{P}_2^N(H)$ the function $\text{MC}(\cdot, R)$ is increasing in the $\preceq_{\text{MC}}$ -order.

Proof. We first show that (1) implies (2). To that end, let $P \preceq_{\text{MC}} Q \in \mathcal{P}_2^N(H)$ and fix $R \in \mathcal{P}_2^N(H)$. Let φ be a dual optimizer for $\text{MC}(Q, R)$. We then have

$$\text{MC}(P, R) \leq \int \varphi dP + \int \varphi^{\text{MC}} dR \leq \int \varphi dQ + \int \varphi^{\text{MC}} dR = \text{MC}(Q, R).$$

To establish the reverse implication suppose that (1) is not satisfied and assume first that $N \geq 2$. Then there exists a MC-convex $\varphi : \mathcal{P}_2^{N-1}(H) \rightarrow (-\infty, +\infty]$ such that $\int \varphi dP > \int \varphi dQ$. By approximating φ with MC-convex functions of the form $(\varphi^{\text{MC}} + \chi_A)^{\text{MC}}$ for bounded $A \subseteq \mathcal{P}_2^{N-1}(H)$, we can additionally assume that φ is Lipschitz and, in particular, finitely valued. Let $f^\varepsilon : \mathcal{P}_2^{N-1}(H) \rightarrow \mathcal{P}_2^{N-1}(H)$ be an ε -selection of φ^{MCMC} , i.e.,

$$\varphi(p) + \varphi^{\text{MC}}(f^\varepsilon(p)) \leq \text{MC}(p, f^\varepsilon(p)) + \varepsilon, \quad p \in \mathcal{P}_2^{N-1}(H).$$

Let $2\varepsilon = \int \varphi dP - \int \varphi dQ$. Define the measure

$$R_\varepsilon := (f^\varepsilon)_\# P.$$

Then, we have

$$\text{MC}(P, R_\varepsilon) \geq P(\varphi) + R_\varepsilon(\varphi^{\text{MC}}) - \varepsilon > Q(\varphi) + R_\varepsilon(\varphi^{\text{MC}}) \geq \text{MC}(Q, R_\varepsilon),$$

which yields a contradiction.

In the case $N = 1$, the same argument applies, replacing $\mathcal{P}_2^{N-1}(H)$ by H , MC-conjugation with convex conjugation, and MC by the inner product on H . $\square$

Proposition 6.4. Let $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, \infty]$ be MC-convex. Then φ is increasing in the $\preceq_{\text{MC}}$ -order.

Proof. This follows because $\text{MC}(\cdot, R)$ is increasing in the $\preceq_{\text{MC}}$ -order (see Proposition 6.3) and every MC-convex function is by definition supremum of such functionals (plus constants). $\square$

6.3. N-Monge Couplings. In this section we use $\mathcal{X}, \mathcal{Y}$ to denote Polish metric spaces.

Recall from (31) that $\text{ip}_t = \mathcal{L}^{\mathcal{F}^{(N-1):t}}(X) \in A_t = \mathcal{P}_2^{N-t}(H)$.

Definition 6.5. Let $P, Q \in \mathcal{P}_2^N(\mathcal{X})$. We say that $\Pi \in \text{Cpl}^N(P, Q)$ is N -Monge if there exists a map $\xi : A_{1:N} \rightarrow \mathcal{X}$ such that, whenever $(X, Y) \sim_{\text{ad}} \Pi$, we have $Y = \xi(\text{ip}(X))$.

Crucially, being N -Monge is equivalent to the fact that $\text{ip}(Y) = T(\text{ip}(X))$ as it is possible to ‘unfold’ $\xi : A_{1:N} \rightarrow \mathcal{X}$ to a map $T : A_{1:N} \rightarrow A_{1:N}$ that satisfies $\text{ip}(Y) = T(\text{ip}(X))$.

Proposition 6.6. *Let $P \in \mathcal{P}_2^N(\mathcal{X})$ and let $\xi : \prod_{t=1}^N \mathcal{P}^{N-t}(\mathcal{X}) \rightarrow \mathcal{Y}$ be measurable. Write $Y = \xi(\text{ip}(X))$ and $Q := \mathcal{L}^{\text{ad}}(Y)$ and inductively define mappings $T_N := \xi$ and, for $t = N-1, \dots, 1$,*

$$T_t : A_{1:t} \rightarrow \mathcal{P}_2^{N-t}(\mathcal{X}), \quad T_t(p^1, \dots, p^t) := T_{t+1}(p^1, \dots, p^t, \cdot) \# p^t.$$

Then, for every $t = 1, \dots, N$, we have $\text{ip}_t(Y) = T_t(\text{ip}_{1:t}(X))$ almost surely.

Proof. We show the claim by backward induction. For $t = N$, we trivially have

$$\text{ip}_N(Y) = Y = \xi(\text{ip}(X)) = T_N(\text{ip}_{1:N}(X)).$$

Next, suppose the claim is true for $t+1$. We have almost surely that

$$\begin{aligned} \text{ip}_t(Y) &= \mathcal{L}(\text{ip}_{t+1}(Y) | \mathcal{F}_t) = \mathcal{L}(T_{t+1}(\text{ip}_{1:t+1}(X)) | \mathcal{F}_t) \\ &= (T_{t+1}(\text{ip}_{1:t}(X), \cdot) \# \mathcal{L}(\text{ip}_{t+1}(X) | \mathcal{F}_t)) \\ &= (T_{t+1}(\text{ip}_{1:t}(X), \cdot) \# \text{ip}_t(X)) = T_t(\text{ip}_{1:t}(X)), \end{aligned}$$

where the first, fourth and fifth equality hold by definition, the second by inductive assumption and the third by $\mathcal{F}_t$ -measurability of $\text{ip}_{1:t}(X)$. This yields the claim. $\square$

Corollary 6.7. *Let $P, Q \in \mathcal{P}_2^N(H)$. Let $\Pi^1, \Pi^2 \in \text{Cpl}^N(P, Q)$ be N -Monge. Then,*

$$(\Pi^1 + \Pi^2)/2 \text{ is } N\text{-Monge} \iff \Pi^1 = \Pi^2.$$

Proof. Note that the $\Leftarrow$ -implication is trivial.

To see the $\Rightarrow$ -implication, apply Proposition 6.6 with $\mathcal{X} = H$, $\mathcal{Y} = H \times H$ and $\xi(\text{ip}(X)) = (X, \xi^i(\text{ip}(X)))$, where ξ^i is the corresponding map associated with the N -Monge coupling Π^i , $i = 1, 2$. Using the assertion of Proposition 6.6 with $t = 1$, we obtain Borel maps T^i with $T^i_\# P = \Pi^i$. As $(\Pi^1 + \Pi^2)/2$ is assumed to be N -Monge, we deduce (by analogous reasoning as above) that $(\Pi^1 + \Pi^2)/2$ is given by pushforward of some measurable map. This enforces that $T^1 = T^2$ P -almost surely, and hence $\Pi^1 = \Pi^2$. $\square$

Definition 6.8. We call $(\mu, \nu) \in \mathcal{P}_2(H) \times \mathcal{P}_2(H)$ a strict Monge pair if the optimal coupling between μ and ν is unique and of Monge type.

For $N \geq 2$, we call $(P, Q) \in \mathcal{P}_2^N(H) \times \mathcal{P}_2^N(H)$ a strict Monge pair if

- the optimal coupling between P and Q is unique and of Monge type, i.e., $\text{Cpl}_{\text{opt}}(P, Q) = \{(\text{id}, T)_\# P\}$ for some $T : \mathcal{P}_2^{N-1}(H) \rightarrow \mathcal{P}_2^{N-1}(H)$;
- For P -almost every p , the pair $(p, T(p))$ is a strict Monge pair.

Proposition 6.9. *Let $P, Q \in \mathcal{P}_2^N(H)$. Then, (P, Q) is a strict Monge pair if and only if $\text{Cpl}_{\text{opt}}^N(P, Q) = \{\Pi\}$ and Π is N -Monge.*

Proof. First, suppose that (P, Q) is a strict Monge pair. We need to show that there exists a unique optimizer $\Pi \in \text{Cpl}^N(P, Q)$ and that it is N -Monge. To this end, we show that there is $\xi : A_{1:N} \rightarrow H$ such that for every optimizer $\Pi \in \text{Cpl}^N(P, Q)$ and random variables X, Y with $(X, Y) \sim_{\text{ad}} \Pi$ it holds $Y = \xi(\text{ip}(X))$ almost surely.

We show the claim by induction on N . For $N = 1$ it is clear. Suppose that the claim holds for $N-1$. That is, if $(p, q) \in \mathcal{P}_2^{N-1}(H) \times \mathcal{P}_2^{N-1}(H)$ is a strict Monge pair then there is $\xi^{p,q} : A_{2:N} \rightarrow H$ such that for all $X', Y' \in L_2^{N-1}(H)$ with $\text{ip}_1(X', Y') \in \text{Cpl}_{\text{opt}}^{N-1}(p, q)$ we have $Y' = \xi^{p,q}(\text{ip}(X'))$ almost surely.

Fix an optimizer $\Pi \in \text{Cpl}_{\text{opt}}^N(P, Q)$ and random variables $(X, Y) \sim_{\text{ad}} \Pi$. We then have a.s.

$$\mathcal{P}^{N-1}[\text{pr}_1](\text{ip}_1(X, Y)) = \text{ip}_1(X) \sim P \text{ and } \mathcal{P}^{N-1}[\text{pr}_2](\text{ip}_1(X, Y)) = \text{ip}_1(Y) \sim Q.$$

Therefore, we get

$$\begin{aligned} \text{MC}(P, Q) &= \int x \cdot y \, dI^{N-1}\Pi(x, y) = \mathbb{E} \left[\int x \cdot y \, I^{N-2} \text{ip}_1(X, Y)(dx, dy) \right] \\ &\leq \mathbb{E}[\text{MC}(\text{ip}_1(X), \text{ip}_1(Y))] \leq \text{MC}(P, Q). \end{aligned}$$

Thus, all inequalities are equalities. In particular, $\mathcal{L}(\text{ip}_1(X), \text{ip}_1(Y)) \in \text{Cpl}_{\text{opt}}(P, Q)$ and almost surely $\text{ip}_1(X, Y) \in \text{Cpl}_{\text{opt}}^{N-1}(\text{ip}_1(X), \text{ip}_1(Y))$. As (P, Q) is a strict Monge pair, this yields $\text{ip}_1(Y) = T(\text{ip}_1(X))$ for some map $T : \mathcal{P}_2^{N-1}(H) \rightarrow \mathcal{P}_2^{N-1}(H)$ and there is λ -full set $\Omega' \in \mathcal{F}_1$ such that for every $\omega \in \Omega'$ we have that

$$\begin{aligned} (\text{ip}_1(X)(\omega), \text{ip}_1(Y)(\omega)) &\in \mathcal{P}_2^{N-1}(H) \times \mathcal{P}_2^{N-1}(H) \text{ is a strict Monge pair,} \\ \text{ip}_1(X, Y)(\omega) &\in \text{Cpl}_{\text{opt}}^{N-1}(\text{ip}_1(X)(\omega), T(\text{ip}_1(X)(\omega))). \end{aligned}$$

By the induction hypothesis, we obtain that for all $\omega \in \Omega'$

$$Y(\omega_1, \cdot) = \xi^{\text{ip}_1(X)(\omega), \text{ip}_1(Y)(\omega)}(\text{ip}_{2:N}(X)(\omega_1, \cdot)).$$

Using this we can define (on a $\mathcal{L}(\text{ip}(X))$ -full set) the desired map $\xi : A_{1:N} \rightarrow H$ via

$$\xi(\text{ip}(X)) := \xi^{\text{ip}_1(X), T(\text{ip}_1(X))}(\text{ip}_{2:N}(X)),$$

which satisfies $\xi(\text{ip}(X)) = Y$ almost surely.

Conversely, we need to show that if $\text{Cpl}_{\text{opt}}^N(P, Q) = \{\Pi\}$ and Π is N -Monge that then (P, Q) is a strict Monge pair. Again, the claim is trivially satisfied when $N = 1$. Assume that the claim holds for $N - 1$. That is, if $(p, q) \in \mathcal{P}_2^{N-1}(H) \times \mathcal{P}_2^{N-1}(H)$ satisfies $\text{Cpl}_{\text{opt}}^{N-1}(p, q) = \{\Pi_{p,q}\}$ and $\Pi_{p,q}$ is $(N-1)$ -Monge, then (p, q) is strict Monge. If there are $\pi_1, \pi_2 \in \text{Cpl}_{\text{opt}}^N(P, Q)$ with $\pi_1 \neq \pi_2$ and $\tilde{\Pi}_{p,q} \in \text{Cpl}_{\text{opt}}^{N-1}(p, q)$ is a measurable selection, then $\Pi_i := ((p, q) \mapsto \tilde{\Pi}_{p,q} \# \pi_i) \in \text{Cpl}_{\text{opt}}^N(P, Q)$. Thus, $\Pi = \Pi_1 = \Pi_2$ which can only be the case if $\pi_1 = \pi_2$. Since Π is N -Monge there is a measurable map $\xi : A_{1:N} \rightarrow \mathcal{X}$ such that $(X, \xi(\text{ip}(X))) \sim_{\text{ad}} \Pi$ for $X \sim_{\text{ad}} P$. As in the first part, we have that $\mathcal{L}(\text{ip}_1(X), \text{ip}_1(\xi(\text{ip}(X)))) \in \text{Cpl}_{\text{opt}}^N(P, Q)$, from where it follows that $\pi_1 = \mathcal{L}(\text{ip}_1(X), \text{ip}_1(\xi(\text{ip}(X))))$ is also of Monge type. By the inductive assumption, we have that for almost surely $(\text{ip}_1(X), \text{ip}_1(Y))$ is a strict Monge pair. This completes the induction step and concludes the proof. $\square$

6.4. MC-subdifferentials. Before considering subdifferentials, we briefly compare continuity on $\mathcal{P}_2^N(H)$ and $L_2^N(H)$.

Lemma 6.10. *Let $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$ and $P \in \mathcal{P}_2^N(H)$. Then the following are equivalent:*

- (1) φ is continuous at P .
- (2) For all $X \sim_{\text{ad}} P$, $\bar{\varphi}$ is continuous at X .
- (3) There exists $X \sim_{\text{ad}} P$ such that $\bar{\varphi}$ is continuous at X .

In particular,

$$\text{cont}(\bar{\varphi}) = \{X \in L_2^N(H) : \mathcal{L}^{\text{ad}}(X) \in \text{cont}(\varphi)\},$$

and hence $\text{cont}(\bar{\varphi})$ is adapted-law invariant.

With the same arguments as in the proof of Lemma 6.10 it follows that a function $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$ is lsc if and only if its adapted Lions lift $\bar{\varphi}$ is lsc.

Proof of Lemma 6.10. To see that (1) implies (2) let $X \sim_{\text{ad}} P$ and $\|X_n - X\|_2 \rightarrow 0$. Then, $\mathcal{W}_2(\mathcal{L}^{\text{ad}}(X_n), \mathcal{L}^{\text{ad}}(X)) \rightarrow 0$ and hence $\bar{\varphi}(X_n) = \varphi(\mathcal{L}^{\text{ad}}(X_n)) \rightarrow \varphi(\mathcal{L}^{\text{ad}}(X)) = \bar{\varphi}(X)$.

The implication from (2) to (3) is trivial.

Suppose that (3) holds true. Let $P_n \rightarrow P$ in $\mathcal{W}_2$. By Proposition 5.7 there are $X_n \sim_{\text{ad}} P_n$ such that $\|X_n - X\|_2 \rightarrow 0$. Hence, $\varphi(P_n) = \bar{\varphi}(X_n) \rightarrow \bar{\varphi}(X) = \varphi(P)$. $\square$

We recall the notion of MC-subdifferentials from Definition 4.3. For a function $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, \infty]$ and $P \in \mathcal{P}_2^N(H)$ its MC-subdifferential at P is defined as

$$\partial_{\text{MC}}\varphi(P) = \{Q \in \mathcal{P}_2^N(H) : \varphi(R) \geq \varphi(P) + \text{MC}(R, Q) - \text{MC}(P, Q) \text{ for every } R \in \mathcal{P}_2^N(H)\}.$$

Proposition 6.11. *Let $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$ be proper and let $X, Y \in L_2^N(H)$. If $X \sim_{\text{ad}} P$ and $Z \sim_{\text{ad}} Q$ we have*

$$Z \in \partial\bar{\varphi}(X) \iff Q \in \partial_{\text{MC}}\varphi(P) \text{ and } \mathbb{E}[X \cdot Z] = \text{MC}(P, Q)$$

Proof. The proof is line by line as the proof of (19). $\square$

A crucial observation about subdifferentials of law invariant convex functions is that if $Z \in \partial\bar{\varphi}(X)$, then $\mathbb{E}[Z|X] \in \partial\bar{\varphi}(X)$. This allowed us to conclude that if φ is MC-differentiable at $\mathcal{L}(X)$ then $\partial\bar{\varphi}(X) = \{\xi(X)\}$. Our next goal is to establish an analogue of this for adapted-law invariant functionals. Specifically, we show that if $Z \in \partial\bar{\varphi}(X)$ then $\mathbb{E}[Z|\text{ip}(X)] \in \partial\bar{\varphi}(X)$ and therefore conclude that MC-differentiability of φ at $\mathcal{L}^{\text{ad}}(X)$ implies that $\partial\bar{\varphi}(X) = \{\xi(\text{ip}(X))\}$. To this end, we need some auxiliary results.

Lemma 6.12. *Let $X, Y \in L_2^N(H)$ such that $\mathbb{E}[Y|\text{ip}(X)] = 0$. Then, we have*

$$\mathcal{L}^{\text{ad}}(X + Y) \succ_{\text{MC}} \mathcal{L}^{\text{ad}}(X).$$

Proof. By Lemma B.2 we can assume w.l.o.g. that X has a $\sigma(Z_t, U_{t+1:N})$ -measurable version, where $Z_t := \mathcal{L}^{\mathcal{F}_{N-1:t}}(X)$, for all $t = 1, \dots, N-1$. Consider the family of random variables

$$Y_0 := Y, \quad Y_t := \mathbb{E}[Y|Z_{1:t}, U_{t+1:N}], \quad Y_N := \mathbb{E}[Y|X, Z_{1:N-1}] = 0.$$

We claim that $\mathcal{L}^{\text{ad}}(X + Y_t) \succ_{\text{MC}} \mathcal{L}^{\text{ad}}(X + Y_{t+1})$ for $t = 0, \dots, N-1$, which we proceed to show by induction.

Let φ be an MC-convex function, then we have as $\bar{\varphi}$ is convex and lsc

$$\mathbb{E}[\bar{\varphi}(X + Y_0)|Z_1, U_{2:N}] \geq \bar{\varphi}(\mathbb{E}[X + Y|Z_1, U_{2:N}]) = \bar{\varphi}(X + \mathbb{E}[Y|Z_1, U_{2:N}]) = \bar{\varphi}(X + Y_1).$$

By taking the expectation on both sides, the resulting inequality shows the claim for $t = 0$.

Next, assume that the claim holds true for $t < N$. Again, let φ be MC-convex, then

$$\begin{aligned} \mathbb{E}[\bar{\varphi}(X + Y_t)|Z_{1:t}, U_{t+1:N}] &\geq \bar{\varphi}(\mathbb{E}[X + Y_t|Z_{1:t}, U_{t+1:N}]) \\ &= \bar{\varphi}(X + \mathbb{E}[Y_t|Z_{1:t}, U_{t+1:N}]) = \bar{\varphi}(X + Y_{t+1}). \end{aligned}$$

As above we conclude that $\mathcal{L}^{\text{ad}}(X + Y_t) \succ_{\text{MC}} \mathcal{L}^{\text{ad}}(X + Y_{t+1})$ and the assertion follows by transitivity of $\prec_{\text{MC}}$ and noting that $X + Y_1 = X + Y$ as well as $X = X + Y_N$. $\square$

Proposition 6.13. *Let $\bar{\varphi} : L_2^N(H) \rightarrow (-\infty, +\infty]$ be adapted-law invariant and convex. If $Z \in \partial\bar{\varphi}(X)$, then*

$$\mathbb{E}[Z|\text{ip}(X)] \in \partial\bar{\varphi}(X).$$

Proof. Let $Z \in \partial\bar{\varphi}(X)$. For every $Y \in L_2^N(H)$, we have $\mathbb{E}[Y - \mathbb{E}[Y|\text{ip}(X)]|\text{ip}(X)] = 0$. Hence, by Lemma 6.12 we find

$$\begin{aligned} \bar{\varphi}(Y) &\geq \bar{\varphi}(\mathbb{E}[Y|\text{ip}(X)]) \\ &\geq \bar{\varphi}(X) + \mathbb{E}[(\mathbb{E}[Y|\text{ip}(X)] - X) \cdot Z] \\ &= \bar{\varphi}(X) + \mathbb{E}[\mathbb{E}[Y - X|\text{ip}(X)] \cdot Z] \\ &= \bar{\varphi}(X) + \mathbb{E}[(Y - X) \cdot \mathbb{E}[Z|\text{ip}(X)]]. \end{aligned}$$

Hence, $\mathbb{E}[Z|\text{ip}(X)] \in \partial\bar{\varphi}(X)$. $\square$

Definition 6.14. Let $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, \infty]$ be MC-convex and $P \in \text{dom}(\varphi)$. We say that φ is MC-differentiable at P if $|\partial_{\text{MC}}\varphi(P)| = 1$.

Lemma 6.15. *Let $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, \infty]$ be MC-convex and $P \in \text{dom}(\varphi)$. Then the following are equivalent:*

- (1) φ is MC-differentiable in P .
- (2) For all X with $X \sim_{\text{ad}} P$ we have $\#\partial\overline{\varphi}(X) = 1$.

Proof. (1) implies (2): Suppose that there are $Z_1 \neq Z_2 \in \partial\overline{\varphi}(X)$ for some $X \sim_{\text{ad}} P$. By Proposition 6.11, we have $\mathcal{L}^{\text{ad}}(Z_1), \mathcal{L}^{\text{ad}}(Z_2) \in \partial_{\text{MC}}\varphi(P)$. Hence, if not $Z_1 \sim_{\text{ad}} Z_2$, we conclude that $\partial_{\text{MC}}\varphi(P)$ is no singleton. Otherwise, observe that $Z := \frac{1}{2}(Z_1 + Z_2) \in \partial\overline{\varphi}(X)$ and hence $\mathcal{L}(Z) \in \partial_{\text{MC}}\varphi(P)$. As $\|\cdot\|_2^2$ is strictly convex, $\|Z\|_2^2 < \frac{1}{2}(\|Z_1\|_2^2 + \|Z_2\|_2^2) = \|Z_1\|_2^2$. Hence, $\mathcal{L}(Z) \neq \mathcal{L}(Z_1)$ and in particular the adapted laws are different. This shows that $\partial_{\text{MC}}\varphi(P)$ is no singleton.

(2) implies (1): Let $Q_1, Q_2 \in \partial_{\text{MC}}\varphi(P)$. By Proposition 5.2 and Corollary 5.4 there are random variables $X_i, Y_i \in L_2^N(H)$ with $X_i \sim_{\text{ad}} P$ and $Y_i \sim_{\text{ad}} Q_i$ such that $\text{MC}(P, Q_i) = \mathbb{E}[X_i \cdot Y_i]$, for $i = 1, 2$. By Proposition 6.11 we have that $\{Y_i\} = \partial\overline{\varphi}(X_i)$ and due to Proposition 6.13 there exist measurable functions such that $\xi_i(\text{ip}(X_i)) = Y_i$. Since

$$\text{MC}(P, Q_1) = \mathbb{E}[X_1 \cdot Y_1] = \mathbb{E}[X_1 \cdot \xi_1(\text{ip}(X_1))] = \mathbb{E}[X_2 \cdot \xi_1(\text{ip}(X_2))]$$

Hence, $\xi_1(\text{ip}(X_2)) \in \partial\overline{\varphi}(X_2)$. By assumption $\#\partial\overline{\varphi}(X_2) = 1$ and it follows that $\xi_1(\text{ip}(X_2)) = Y_2$. Thus $Q_1 = \mathcal{L}^{\text{ad}}(Y_1) = \mathcal{L}^{\text{ad}}(Y_2) = Q_2$. Hence φ is MC-differentiable at P . $\square$

Proposition 6.16. *Let $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, \infty]$ be MC-differentiable in P . Then there exists a measurable function $\xi : A_{1:N} \rightarrow H$ such that for all $X \sim_{\text{ad}} P$*

$$\partial\overline{\varphi}(X) = \{\xi(\text{ip}(X))\}.$$

Proof. Let $X \sim_{\text{ad}} P$. Then Proposition 6.13 implies that $\partial\overline{\varphi}(X) = \{\xi(\text{ip}(X))\}$. It remains to show that ξ does not depend on the choice of $X \sim_{\text{ad}} P$. To that end, let $X' \sim_{\text{ad}} P$ and ξ' such that $\partial\overline{\varphi}(X') = \{\xi'(\text{ip}(X'))\}$. By Proposition 6.11, we have $\mathcal{L}^{\text{ad}}(\xi'(\text{ip}(X'))) \in \partial_{\text{MC}}\varphi(P)$ and $\mathbb{E}[X' \cdot \xi'(\text{ip}(X'))] = \text{MC}(P, Q)$. As $\mathbb{E}[X \cdot \xi'(\text{ip}(X))] = \mathbb{E}[X' \cdot \xi'(\text{ip}(X'))] = \text{MC}(P, Q)$ and $\mathcal{L}^{\text{ad}}(\xi'(\text{ip}(X))) = \mathcal{L}^{\text{ad}}(\xi'(\text{ip}(X'))) \in \partial_{\text{MC}}\varphi(P)$, Proposition 6.11 implies that $\xi'(\text{ip}(X)) \in \partial\overline{\varphi}(X)$. As $\partial\overline{\varphi}(X)$ is a singleton, we have $\xi = \xi'$ almost surely. $\square$

6.5. Characterization of strict Monge pairs by differentiability of MC. In the following we generalize the result of Alfonsi–Jourdain [6] that characterizes the existence of a unique optimal transport that is induced by a map in terms of Lions differentiability of the Wasserstein distance to the case of iterated probability measures.

Proposition 6.17. *For $P, Q \in \mathcal{P}_2^N(H)$ the following are equivalent:*

- (1) $\text{MC}(\cdot, Q)$ is MC-differentiable at P .
- (2) There is $\xi : A_{1:N} \rightarrow H$ such that for some $X \sim_{\text{ad}} P$

$$\{Y \sim_{\text{ad}} Q : \mathbb{E}[X \cdot Y] = \text{MC}(P, Q)\} = \{\xi(\text{ip}(X))\} = \overline{\partial\text{MC}(\cdot, Q)}(X).$$

- (3) For some $X \sim_{\text{ad}} P$, the function $\overline{\text{MC}(\cdot, Q)}$ is Frechet differentiable at X .
- (4) The pair (P, Q) is a strict Monge pair.

Moreover, whenever one (and hence all) of the above conditions holds, statements (2) and (3) are valid for every $X \sim_{\text{ad}} P$, and $\partial_{\text{MC}}\text{MC}(\cdot, Q)(P) = \{Q\}$.

Proof. We start with proving the equivalence of (1), (2), (3), and (4).

(4) $\implies$ (2): Assume that (4) holds and let $X \sim_{\text{ad}} P$. By Proposition 6.9 there is $\xi : A_{1:N} \rightarrow H$ such that $\text{Cpl}_{\text{opt}}^N(P, Q) = \{\mathcal{L}^{\text{ad}}(X, \xi(\text{ip}(X)))\}$. In particular, if $Y \sim_{\text{ad}} Q$ with $\mathbb{E}[X \cdot Y]$

$\mathbb{E}[X \cdot \xi(\text{ip}(X))] = \text{MC}(P, Q)$. Next, we show that $\overline{\text{MC}(\cdot, Q)}$ is Frechet differentiable at X with derivative $\xi(\text{ip}(X))$. To this end, let $X_n \rightarrow X$ in L_2 . We then have

$$(33) \quad \text{MC}(\mathcal{L}^{\text{ad}}(X_n), Q) \geq \mathbb{E}[X_n \cdot \xi(\text{ip}(X))] = \text{MC}(P, Q) + \mathbb{E}[(X_n - X) \cdot \xi(\text{ip}(X))].$$

Next, let $Y_n \sim_{\text{ad}} Q$ with $\mathbb{E}[X_n \cdot Y_n] \geq \text{MC}(\mathcal{L}^{\text{ad}}(X_n), Q) - \|X_n - X\|_2/n$. We then find

$$\begin{aligned}
 \text{MC}(P, Q) &\geq \mathbb{E}[X \cdot Y_n] = \mathbb{E}[X_n \cdot Y_n] + \mathbb{E}[(X - X_n) \cdot Y_n] \\
 (34) \quad &\geq \text{MC}(\mathcal{L}^{\text{ad}}(X_n), Q) + \mathbb{E}[(X - X_n) \cdot Y_n] - \|X_n - X\|/n \\
 &= \text{MC}(\mathcal{L}^{\text{ad}}(X_n), Q) + \mathbb{E}[(X - X_n) \cdot \xi(\text{ip}(X))] + \mathbb{E}[(X - X_n) \cdot (Y_n - \xi(\text{ip}(X)))]
 \end{aligned}$$

Rearranging the inequalities (33) and (34) yields

$$\begin{aligned}
 |\text{MC}(\mathcal{L}^{\text{ad}}(X_n), Q) - \text{MC}(P, Q) - \mathbb{E}[(X_n - X) \cdot \xi(\text{ip}(X))]| \\
 \leq |\mathbb{E}[(X - X_n) \cdot (Y_n - \xi(\text{ip}(X)))]| \leq \|X_n - X\|_2 \|Y_n - \xi(\text{ip}(X))\|_2.
 \end{aligned}$$

Since $\lim_{n \rightarrow \infty} \|Y_n - \xi(\text{ip}(X))\|_2 = 0$ by Lemma 6.18, we find that $\overline{\text{MC}(\cdot, Q)}$ is Frechet differentiable at X with derivative $\xi(\text{ip}(X))$.

(3) $\implies$ (1): This implication is clear because Frechet differentiability of the Lions lift at some X implies, by Theorem 6.19, its Frechet differentiability at all $X' \sim_{\text{ad}} P$. Hence, $|\{\partial \overline{\text{MC}(\cdot, Q)}(X')\}| = 1$ for all $X' \sim_{\text{ad}} P$. Thus, it follows from Lemma 6.15 that $\text{MC}(\cdot, Q)$ is MC-differentiability at P .

(1) $\implies$ (4): Assume that $\partial_{\text{MC}} \text{MC}(\cdot, Q)(P) = \{Q\}$. For any $\Pi \in \text{Cpl}_{\text{opt}}^N(P, Q)$ and $(X, Y) \sim_{\text{ad}} \Pi$, we have that $\text{MC}(P, Q) = \mathbb{E}[X \cdot Y]$ and, by Proposition 6.11 we have also $Y \in \partial \overline{\text{MC}(\cdot, Q)}(X)$. Consequently, by Proposition 6.13 we obtain that $\mathbb{E}[Y|\text{ip}(X)] \in \partial \overline{\text{MC}(\cdot, Q)}(X)$ and, again by Proposition 6.11, $\mathcal{L}^{\text{ad}}(\mathbb{E}[Y|\text{ip}(X)]) \in \partial_{\text{MC}} \text{MC}(\cdot, Q)(P)$. Hence, $Q = \mathcal{L}^{\text{ad}}(\mathbb{E}[Y|\text{ip}(X)])$ and $\mathbb{E}[|Y|^2] = \mathbb{E}[\mathbb{E}[Y|\text{ip}(X)]^2]$, from where we deduce from the equality case of Jensen's inequality that $Y = \mathbb{E}[Y|\text{ip}(X)]$ almost surely, i.e., Y is $\text{ip}(X)$ -measurable. We have shown that every $\Pi \in \text{Cpl}_{\text{opt}}^N(P, Q)$ is N -Monge, which, by Corollary 6.7, yields that $\text{Cpl}_{\text{opt}}^N(P, Q)$ contains a single element. We conclude that (P, Q) is a strict Monge pair by Proposition 6.9. $\square$

Lemma 6.18. *Let $P, Q, P_1, Q_1, P_2, Q_2, \dots \in \mathcal{P}_2^N(H)$ and let $X, Y, X_1, Y_1, X_2, Y_2, \dots$ be random variables with $X \sim_{\text{ad}} P$, $Y \sim_{\text{ad}} Q$, $X_n \sim_{\text{ad}} P_n$, $Y_n \sim_{\text{ad}} Q_n$ such that*

$$\mathbb{E}[X_n \cdot Y_n] \geq \text{MC}(P_n, Q_n) - \varepsilon_n, \quad Q_n \rightarrow Q \text{ in } \mathcal{W}_2, \text{ and } \varepsilon_n \rightarrow 0.$$

If (P, Q) is a strict Monge pair and $X_n \rightarrow X$ in L_2 , then

$$Y_n \rightarrow \xi(\text{ip}(X)) \text{ in } L_2,$$

where ξ is the unique map with $\mathcal{L}(\text{ip}(X), \xi(\text{ip}(X))) \in \text{Cpl}_{\text{opt}}^N(P, Q)$.

Proof. Note that since (P, Q) is a strict Monge pair, there exists a unique map $\xi : \mathcal{P}_2^{N-1}(H) \times \dots \times H \rightarrow H$ with $Y := \xi(\text{ip}(X)) \sim_{\text{ad}} Q$ and $\mathcal{L}^{\text{ad}}(X, Y) =: \hat{\Pi}$ is the unique element in $\text{Cpl}_{\text{opt}}^N(P, Q)$. By $\mathcal{W}_2$ -relative compactness, up to extracting a subsequence, we can assume that

$$\lim_{n \rightarrow \infty} \hat{\Pi}_n := \mathcal{L}^{\text{ad}}(X_n, Y_n) = \tilde{\Pi} \text{ in } \mathcal{W}_2.$$

Write $c(x, y) = x \cdot y$, by $\mathcal{W}_2$ -continuity of $c^{(N)}$ we get

$$\begin{aligned}
 \int x \cdot y dI^{N-1} \hat{\Pi}(x, y) &= \text{MC}(P, Q) = \lim_{n \rightarrow \infty} \text{MC}(P_n, Q_n) \\
 &= \lim_{n \rightarrow \infty} \int x \cdot y dI^{N-1} \hat{\Pi}_n(x, y) = \int x \cdot y dI^{N-1} \tilde{\Pi}(x, y).
 \end{aligned}$$

Hence, $\tilde{\Pi} \in \text{Cpl}_{\text{opt}}^N(P, Q)$ and by uniqueness $\tilde{\Pi} = \hat{\Pi}$. We have shown that

$$(X_n, Y_n) \rightarrow (X, Y) = (X, \xi(\text{ip}(X))) \text{ in adapted distribution.}$$

In particular, we have $(\text{ip}(X_n), \text{ip}(Y_n)) \rightarrow (\text{ip}(X), \text{ip}(Y))$ in distribution and as $\text{ip}_N(Y_n) = Y_n$ and $\text{ip}_N(Y) = Y$, $(\text{ip}(X_n), Y_n) \rightarrow (\text{ip}(X), Y) = (\text{ip}(X), \xi(\text{ip}(X)))$. Additionally, $(U, \text{ip}(X_n)) \rightarrow (U, \text{ip}(X))$ in probability and therefore in distribution. We can invoke [22, Lemma 6.4] to obtain

$$(35) \quad (U, \text{ip}(X_n), Y_n) \rightarrow (U, \text{ip}(X), Y) \text{ in distribution,}$$

and thus $(U, Y_n) \rightarrow (U, Y)$ in distribution. As $Y = \xi(\text{ip}(X))$ and Y_n are functions of U , (35) entails by [53, Lemma 3.14] that $Y_n \rightarrow Y$ in probability. Note that $\mathbb{E}[|Y_n|^2] = \int |y|^2 dI^N(Q_n) \rightarrow \int |y|^2 dI^N(Q) = \mathbb{E}[|Y|^2]$ and therefore $Y_n \rightarrow Y$ in L_2 , which was the claim. $\square$

The Ekeland–Lebourg theorem (see e.g. [16, Theorem 18.3]) guarantees that a convex function on an infinite dimensional Hilbert space that is continuous in at least one point is Frechet differentiable on a dense G_δ subset of the closure of its domain. For our purposes, we need a version of this result that is tailored to adapted-law invariant functions.

Theorem 6.19. *Let $\bar{\varphi} : L_2^N(H) \rightarrow (-\infty, +\infty]$ be adapted-law invariant, convex and continuous at some $\mu \in \text{dom}(\bar{\varphi})$. Then the set of Frechet-differentiability points of $\bar{\varphi}$ is dense in the closure of $\text{dom}(\bar{\varphi})$ and it is the intersection of countable many open adapted-law invariant sets.*

Sketch of the proof of Theorem 6.19. We write $A \subseteq L_2^N(H)$ for the set of Frechet differentiability points of $\bar{\varphi}$. The Ekeland–Lebourg theorem [16, Theorem 18.3] asserts that A is a dense G_δ subset of the closure of $\text{dom}(\bar{\varphi})$. It remains to observe that the dense open set O_n such that $A = \bigcap_n O_n$ can be chosen to be adapted-law invariant.

By [16, Proposition 18.1] we have that $A = \bigcap_{n \in \mathbb{N}} O_n$ with

$$O_n = \bigcup_{\eta > 0} \left\{ X \in \text{cont}(\bar{\varphi}) : \sup_{\|Y\|_2=1} \bar{\varphi}(X + \eta Y) + \bar{\varphi}(X - \eta Y) - 2\bar{\varphi}(X) < \frac{\eta}{n} \right\}.$$

Note that $\text{cont}(\bar{\varphi})$ is adapted-law invariant by Lemma 6.10. Hence, it is clear that O_n is adapted-law invariant provided that $\bar{\varphi}$ was. $\square$

Remark 6.20. Lemma 6.10 implies that a map $f : \mathcal{P}_2^N(H) \rightarrow \mathbb{R}$ is continuous if and only if $f \circ \mathcal{L}^{\text{ad}} : L_2^N(H) \rightarrow \mathbb{R}$ is continuous. This is precisely the universal property of quotient topologies, i.e., $\mathcal{P}_2^N(H) = L_2^N(H) / \sim_{\text{ad}}$. Clearly, adapted-law invariant sets are precisely the saturated sets w.r.t. $\sim_{\text{ad}}$ and there is a one-to-one correspondence between open sets in $\mathcal{P}_2^N(H)$ and open adapted-law invariant sets in $L_2^N(H)$. Hence, dense sets in $\mathcal{P}_2^N(H)$ correspond to adapted-law invariant dense sets in $L_2^N(H)$ and G_δ sets in $\mathcal{P}_2^N(H)$ correspond to subsets of $L_2^N(H)$ that are a countable intersection of open adapted-law invariant sets.

Corollary 6.21. *Let $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$ and suppose that $\text{cont}(\varphi) \neq \emptyset$. Then φ is MC-differentiable on a dense G_δ subset of the closure of $\text{dom}(\varphi)$.*

Proof. We write $A \subseteq L_2^N(H)$ for the set of Frechet differentiability points of $\bar{\varphi}$. By Theorem 6.19, the set $A = \bigcap_n O_n$ with O_n adapted-law invariant, open and dense in the closure of $\text{dom}(\bar{\varphi})$. By Remark 6.20, $U_n := \{\mathcal{L}^{\text{ad}}(X) : X \in O_n\}$ is open and dense in the closure of $\text{dom}(\varphi)$. As Frechet-differentiability implies that the subdifferential is a singleton, Lemma 6.15 yields that φ is MC-differentiable on $\bigcap_n U_n$. $\square$

Theorem 6.22. *Let $Q \in \mathcal{P}_2^N(H)$. Then the set of $P \in \mathcal{P}_2^N(H)$ such that (P, Q) is a strict Monge pair is a dense G_δ set.*

Proof. This follows from Proposition 6.17 and Corollary 6.21. $\square$

7. CONSTRUCTION OF TRANSPORT-REGULAR MEASURES IN $\mathcal{P}_2^N(H)$

7.1. Criterion for transport regularity. The aim of this section is to provide a more accessible criterion for transport regularity of measures on iterated spaces of probability measures. To this end, we leverage the adapted Lion's lift which permits us to formulate the following criterion:

Theorem 7.1. *Let $P \in \mathcal{P}_2^N(H)$ and $X \in L_2^N(H)$ with $\mathcal{L}^{\text{ad}}(X) = P$. Assume that*

$$(36) \quad \text{for almost every } u, \quad (\tilde{u}_t \mapsto X(u_{1:t-1}, \tilde{u}_t, \cdot))_{\#} \lambda \in \mathcal{P}_2(L_2^{N-t}(H)) \text{ is transport regular,}$$

for all $t = 1, \dots, N$. Then, P is transport regular.

An important role in the proof of Theorem 7.1 plays the following

Proposition 7.2. *Let $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$ be MC-convex and let $P \in \mathcal{P}_2^N(H)$. Then, for all $X \in L_2^N(H)$ with $X \sim_{\text{ad}} P$ we have*

$$\{\mathcal{L}^{\text{ad}}(Y) : Y \in \partial \bar{\varphi}(X)\} \subseteq \partial_{\text{MC}} \varphi(P).$$

Assume further that P is transport regular, then we have

$$\{\mathcal{L}^{\text{ad}}(Y) : Y \in \partial \bar{\varphi}(X)\} = \partial_{\text{MC}} \varphi(P).$$

Proof. The first inclusion follows immediately from Proposition 6.11.

For the second part, since P is transport regular, we have by Proposition 6.9 that for every $Q \in \partial_{\text{MC}} \varphi(P)$, $\text{Cpl}_{\text{opt}}^N(P, Q) = \{\mathcal{L}(\text{ip}(X), \xi(\text{ip}(X)))\}$ for some measurable map ξ . This readily yields the desired equality. $\square$

Corollary 7.3. *Let $P \in \mathcal{P}_2^N(H)$ and $X \in L_2^N(H)$ with $\mathcal{L}^{\text{ad}}(X) = P$. Assume that*

- (1) $(u_1 \mapsto X(u_1, \cdot))_{\#} \lambda$ is a transport regular measure in $\mathcal{P}_2(L_2^{N-1}(H))$,
- (2) for λ -almost every u_1 , $\mathcal{L}^{\text{ad}}(X(u_1, \cdot))$ is a transport regular measure in $\mathcal{P}_2^{N-1}(H)$.

Then, P is transport regular.

Proof. By Lemma A.1, we have to show that for every MC-convex $\varphi : \mathcal{P}_2^{N-1}(H) \rightarrow (-\infty, +\infty]$, we have $\#\partial_{\text{MC}} \varphi(p) \leq 1$ for P -almost every p . Using assumption (1), we get

$$\#\partial \bar{\varphi}(X(u_1, \cdot)) \leq 1 \quad \text{for } \lambda\text{-a.e. } u_1.$$

Next, assumption (2) allows us to invoke Proposition 7.2 and we obtain

$$\{\mathcal{L}^{\text{ad}}(Y) : Y \in \partial \bar{\varphi}(X(u_1, \cdot))\} = \partial_{\text{MC}} \varphi(\mathcal{L}^{\text{ad}}(X(u_1, \cdot))) \quad \text{for } \lambda\text{-a.e. } u_1.$$

Hence, $\#\partial_{\text{MC}} \varphi(\mathcal{L}^{\text{ad}}(X(u_1, \cdot))) \leq 1$ for λ -a.e. u_1 or equivalently $\#\partial_{\text{MC}} \varphi(p) \leq 1$ for P -a.e. p . As φ is an arbitrary MC-convex function, this shows transport regularity of P . $\square$

Proof of Theorem 7.1. We prove the claim by backward induction over t . More precisely, we show for $t = N, \dots, 1$ that

$$\text{for almost every } u, \quad \mathcal{L}^{\text{ad}}(X(u_{1:t-1}, \cdot)) \in \mathcal{P}_2^{N+1-t}(H) \text{ is transport regular.}$$

Base case ($t = N$): Using the assumption (36) with $t = N$, we find that for almost every $u_{1:N-1}$

$$\mathcal{L}^{\text{ad}}(X(u_{1:N-1}, \cdot)) = (u_N \mapsto X(u_{1:N-1}, u_N))_{\#} \lambda \in \mathcal{P}_2(H) \text{ is transport regular.}$$

Inductive step: Assume the claim holds for $t + 1$. Using (36) with t , we have

$$\text{for a.e. } u, \quad P^{u_{1:t-1}} := (u_t \mapsto X(u_{1:t-1}, u_t, \cdot))_{\#} \lambda \in \mathcal{P}_2(L_2^{N-t}(H)) \text{ is transport regular.}$$

Further, we derive from the induction hypothesis that, for a.e. u and $P^{u_{1:t-1}}$ -a.e. Y , $\mathcal{L}^{\text{ad}}(Y) \in \mathcal{P}_2^{N-t}(H)$ is transport regular. Thus, we can invoke Corollary 7.3 and conclude that, for almost every u , $\mathcal{L}^{\text{ad}}(X(u_{1:t-1}, \cdot))$ is transport regular.

Taking $t = 1$, we yields that $\mathcal{L}^{\text{ad}}(X)$ is a transport regular measure in $\mathcal{P}_2^N(H)$. $\square$

7.2. Existence of transport-regular measures. The criterion in Theorem 7.1 permits us to build transport-regular measures on $\mathcal{P}_2^N(H)$ using non-degenerate Gaussian random fields. More precisely, we will consider the H -valued analogue to a Brownian sheet.

Recall that an H -valued N -parameter Brownian sheet W with covariance operator Q is an H -valued, continuous, centered Gaussian process indexed by $[0, 1]^N$ such that

$$(37) \quad \mathbb{E}[\langle W(u), x \rangle \langle W(v), y \rangle] = \prod_{t=1}^N \min(u_t, v_t) \langle Qx, y \rangle,$$

for $u, v \in [0, 1]^N$ and $x, y \in H$. Here, $Q : H \rightarrow H$ is a non-negative, symmetric linear operator with finite trace.

Theorem 7.4. *Let $Q \in L(H)$ be a positive-definite, symmetric linear operator with finite trace. Let W be an H -valued N -parameter Brownian sheet. Then $\mathcal{L}_{\mathbb{P}}(\mathcal{L}^{\text{ad}}(W))$ is a transport-regular measure on $\mathcal{P}_2^N(\mathbb{R}^d)$ with full support.*

Proof. For $t = 1, \dots, N$, we write $H_t := L_2^{N-t}(H)$ using the convention that $L_2^0(H) = H$. For fixed $u_{1:t}, v_{1:t} \in [0, 1]^t$ we have

$$\begin{aligned} & \mathbb{E}[\langle W(u_{1:t}, \cdot), f \rangle_{H_t} \langle W(v_{1:t}, \cdot), g \rangle_{H_t}] \\ &= \int_{[0,1]^{N-t} \times [0,1]^{N-t}} \prod_{s=1}^N \min(u_s, v_s) \langle Qf(u_{t+1:N}), g(v_{t+1:N}) \rangle d(u_{t+1:N}, v_{t+1:N}). \end{aligned}$$

Therefore, we can interpret W as an H_t -valued, t -parameter Brownian sheet whose covariance operator Q_t is given by

$$(Q_t f)(u_{t+1:N}) = \int_{[0,1]^{N-t}} \prod_{s=t+1}^N \min(u_s, v_s) Qf(v_{t+1:N}) dv_{t+1:N},$$

where $u_{t+1:N} \in [0, 1]^{N-t}$ and $f \in H_t$. Since H_t can be identified with $L_2^{N-t}(\mathbb{R}) \otimes H$, Q_t can be identified with the tensor product of the positive-definite, symmetric trace class operators $\tilde{Q}$ and Q on the separable Hilbert spaces $L_2^{N-t}(\mathbb{R})$ and H , respectively. Therefore, Q_t is a positive-definite, symmetric trace class operator on H_t . Indeed, Q_t is symmetric, positivity follows from

$$\langle Q_t f, f \rangle_{H_t} = \int_{[0,1]^{N-t}} \left\langle Q \int_{[w,1]} f(u_{t+1,N}) du_{t+1,N}, \int_{[w,1]} f(u_{t+1,N}) du_{t+1,N} \right\rangle dw,$$

where $[w, 1] := [w_{N-t}, 1] \times \dots \times [w_N, 1]$, for all $f \in H_t$, and it is trace class since

$$\begin{aligned} \text{tr}(Q_t) &= \sum_{n,m \in \mathbb{N}} \langle Q_t e_n \otimes f_m, e_n \otimes f_m \rangle_{H_t} \\ &= \left(\sum_{n \in \mathbb{N}} \int_{[0,1]^{N-t} \times [0,1]^{N-t}} \prod_{s=t+1}^N \min(u_s, v_s) e_n(u_{t+1:N}) e_n(v_{t+1:N}) d(u_{t+1:N}, v_{t+1:N}) \right) \\ &\quad \cdot \left(\sum_{m \in \mathbb{N}} \langle Q f_m, f_m \rangle \right) \\ &= \text{tr}(\tilde{Q}) \text{tr}(Q) < \infty, \end{aligned}$$

for orthonormal bases $(e_n)_{n \in \mathbb{N}}$ of $L_2^{N-t}(\mathbb{R})$ and $(f_m)_{m \in \mathbb{N}}$ of H .

For fixed $u_{1:t-1} \in (0, 1]^{t-1}$ we have that

$$u_t \mapsto \mathbb{E}[\langle W(u_{1:t}, \cdot), f \rangle_{H_t} \langle W(u_{1:t}, \cdot), g \rangle_{H_t}] = \prod_{s=1}^t u_s \langle Q_t f, g \rangle_{H_t},$$

and thus $u_t \mapsto W(u_{1:t-1}, u_t, \cdot)$ is a Wiener process with positive-definite covariance operator. By Theorem 3.5, we deduce that

$$(38) \quad \text{a.s.} \quad (u_t \mapsto W(u_{1:t}, \cdot))_{\#} \lambda \text{ is a transport-regular measure on } H_t = L_2^{N-t}(H).$$

By Theorem 7.1, $\mathcal{L}^{\text{ad}}(W)$ is almost surely transport regular in $\mathcal{P}_2^N(H)$. Since $\mathcal{L}(W)$ is a non-degenerate Gaussian measure on $L_2^N(H)$ it is transport regular, see e.g. [8, Theorem 6.2.10]. Another application of Proposition 7.2 yields that $\mathcal{L}_{\mathbb{P}}(\mathcal{L}^{\text{ad}}(W)) \in \mathcal{P}_2^{N+1}(H)$ is transport regular, as claimed.

Finally, we show that the support of P is the whole space. Since $\mathcal{L}(W)$ as Gaussian measure on $L_2^N(H)$ is non-degenerate, we have that the support of $\mathcal{L}(W)$ is the whole space. As $\mathcal{L}^{\text{ad}} : L_2^N(H) \rightarrow \mathcal{P}_2^N(H)$ is continuous by Proposition 5.7 and surjective by Proposition 5.2, we conclude that the support of $\mathcal{L}_{\mathbb{P}}(\mathcal{L}^{\text{ad}}(B))$ coincides with $\mathcal{P}_2^N(H)$. $\square$

7.3. Proofs of Section 1.1.

Proof of Theorem 1.1. Theorem 7.4 provides a transport-regular measure Λ on $\mathcal{P}_2^N(H)$ with full support, and yields the claim. $\square$

Proof of Corollary 1.2. By Theorem 1.1 there exists a transport-regular measure Λ on $\mathcal{P}_2^N(H)$ with full support. Therefore, $\{Q \in \mathcal{P}_2^N(H) : Q \ll P\}$ is dense and consists of transport-regular measures as consequence of Lemma A.1. $\square$

Proof of Theorem 1.5. First, let φ be any MC-convex function. Since $P \ll \Lambda$ and the latter is transport regular, we have by Lemma A.1 that $\#\partial_{\text{MC}}\varphi(p) \leq 1$ for P -a.e. $p \in \mathcal{P}_2^N(H)$. Because of Proposition 6.11 and Lemma 6.15 the adapted Lions lift satisfies $\#\partial\bar{\varphi}(X) \leq 1$ a.s.

Next, recall that by Lemma 4.2 and Theorem 4.4, $\Pi \in \text{Cpl}(P, Q)$ is $\mathcal{W}_2^2$ -optimal if and only if there exists an MC-convex φ with $\Pi(\partial_{\text{MC}}\varphi) = 1$. By the first part we find that, for P -a.e. p , $\partial_{\text{MC}}\varphi(p) = \{\mathcal{L}^{\text{ad}}(\nabla\bar{\varphi}(Y))\}$ with $Y \sim_{\text{ad}} p$. Hence, $\mathcal{W}_2^2$ -optimality of Π is equivalent to

$$\Pi \sim (\mathcal{L}^{\text{ad}}(X), \mathcal{L}^{\text{ad}}(\nabla\bar{\varphi}(X)))$$

for some MC-convex $\varphi : \mathcal{P}_2^N(H) \rightarrow (-\infty, +\infty]$ with $\partial_{\text{MC}}\varphi(p) \neq \emptyset$ for P -a.e. p . The claim then follows by Theorem 6.1. $\square$

Proof of Theorem 1.6. The first claim follows from Theorem 1.5. To see the second claim, note that $(P, Q) \in \mathcal{P}_2^N(H) \times \mathcal{P}_2^N(H)$ has a unique $\mathcal{W}_2$ -optimal coupling and this coupling is concentrated on the graph of a bijection if and only if (P, Q) is a strict Monge pair. For each fixed $Q \in \mathcal{P}_2^N(H)$, the set $\{P \in \mathcal{P}_2^N(H) : (P, Q) \text{ is strict Monge}\}$ is a dense G_δ -set. It follows from the Kuratowski–Ulam theorem (see e.g. [54, Theorem 8.41]) that the sets

$$\{(P, Q) \in \mathcal{P}_2^N(H) \times \mathcal{P}_2^N(H) : (P, Q) \text{ is strict Monge}\},$$

$$\{(P, Q) \in \mathcal{P}_2^N(H) \times \mathcal{P}_2^N(H) : (Q, P) \text{ is strict Monge}\},$$

are both comeager. Hence, the same is true for their intersection, which yields the claim. $\square$

8. APPLICATION TO ADAPTED TRANSPORT

8.1. Outline of the framework. The aim of this section is to apply the results obtained above to H -valued stochastic processes, equipped with the adapted Wasserstein distance.

As noted in the introduction, the set $\mathcal{P}_2(H^N)$ of laws of N -step processes is not complete w.r.t. $\mathcal{AW}_2$. It is shown in [14] that the completion of $(\mathcal{P}_2(H^N), \mathcal{AW}_2)$ consists in the stochastic processes with *filtration*. In detail, we use $\mathcal{FP}_2$ to denote the class of all 5-tuples

$$\mathbb{X} = (\Omega^X, \mathcal{F}^X, \mathbb{P}^X, (\mathcal{F}_t^X)_{t=1}^N, (X_t)_{t=1}^N),$$

where X is an adapted square integrable process. In analogy to (14), the (squared) adapted Wasserstein distance of two filtered processes is defined as

$$\mathcal{AW}_2(\mathbb{X}, \mathbb{Y}) := \inf_{\pi \in \text{Cpl}_{\text{bc}}(\mathbb{X}, \mathbb{Y})} \int |X - Y|^2 d\pi.$$

Here π is a causal coupling of $\mathbb{X}, \mathbb{Y}$ if $\pi \in \text{Cpl}(\mathbb{P}^X, \mathbb{P}^Y)$ and

$$\pi(\Omega^X \times B | \mathcal{F}_t^X \otimes \mathcal{F}_N^Y) = \pi(\Omega^X \times B | \bar{\mathcal{F}}_N^X \otimes \mathcal{F}_N^Y), \quad B \in \mathcal{F}^Y, t \leq N.$$

Bi-causality of π , that is $\pi \in \text{Cpl}_{\text{bc}}(\mathbb{X}, \mathbb{Y})$ means that this holds also when the roles of $\mathbb{X}, \mathbb{Y}$ are exchanged.

We say that $\mathbb{X}, \mathbb{Y}$ are $\mathcal{AW}$ -equivalent, in signs $\mathbb{X} \sim_{\mathcal{AW}} \mathbb{Y}$ if $\mathcal{AW}_2(\mathbb{X}, \mathbb{Y}) = 0$ and write FP_2 for $\mathcal{FP}_2 / \sim_{\mathcal{AW}}$. Using these notations $(\text{FP}_2, \mathcal{AW}_2)$ is the completion of $(\mathcal{P}_2(H^N), \mathcal{AW}_2)$. Moreover, $(\text{FP}_2, \mathcal{AW}_2)$ is a Polish geodesic space. We note that $\mathcal{AW}_2(\mathbb{X}, \mathbb{Y}) = 0$ can be expressed in a number of equivalent ways: it is tantamount to $\mathbb{X}, \mathbb{Y}$ having the same probabilistic properties in the sense of [51], to $\mathbb{X}, \mathbb{Y}$ having Markov lifts with the same laws (see [23]) and to $\mathbb{X}, \mathbb{Y}$ having the same adapted law in the sense defined in (41) below.

Every filtered process $\mathbb{X}$ has a representative in

$$(39) \quad L_{2,ad}^N(H) := \{X \in L_2([0, 1]^N, \lambda; H^N) : X_t \text{ is } \mathcal{F}_t\text{-measurable}\},$$

where $(\mathcal{F}_t)$ denotes the coordinate filtration as above, see e.g. [23]. Moreover we have for $X, Y \in L_{2,ad}^N(H)$

$$(40) \quad \mathcal{AW}_2(X, Y) = \inf_{X \sim_{\mathcal{AW}} X', Y \sim_{\mathcal{AW}} Y'} \|X'_1 - Y'_1\|_2$$

and the infimum is attained. It is therefore without loss of generality to work with filtered stochastic processes $X \in L_{2,ad}^N$ and we shall do so from now on.

The stochastic interpretation of the framework developed in the previous sections is that we consider an H -valued random variable X together with a filtration $(\mathcal{F}_t)_{t=1}^N$. We assume that X is $\mathcal{F}_N$ -measurable and interpret $(\mathcal{F}_t)_{t=1}^N$ as a model of how information about X is gradually revealed. (Alternatively we could identify X with its Doob martingale.)

In order to apply our results for $\mathcal{P}_2^N(H)$ to the adapted Wasserstein setting we need to consider the relationship of $L_2^N(H)$ and $L_{2,ad}^N(H)$ in some detail.

The aim of this section is to translate these results to the classical setting of H -valued stochastic processes $X = (X_t)_{t=1}^N$ that are adapted to the filtration $(\mathcal{F}_t)_{t=1}^N$.

Following [14], we define the information process of an adapted process $X \in L_{2,ad}^N(H)$ by backward induction on t as

$$\begin{aligned} \mathbf{ip}_N(X) &= X_N, \\ \mathbf{ip}_t(X) &= (X_t, \mathcal{L}^{\mathcal{F}_t}(\mathbf{ip}_{t+1}(X))), \quad t = N-1, \dots, 1. \end{aligned}$$

The information process takes values in the following nested space: Set $(\mathcal{Z}_N, d_{\mathcal{Z}_N}) := (H, |\cdot|)$ and

$$\mathcal{Z}_t := H \times \mathcal{P}_2(\mathcal{Z}_{t+1}) \quad d_{\mathcal{Z}_t}^2 := |\cdot|^2 + W_{2, \mathcal{Z}_{t+1}}^2,$$

where $W_{2, \mathcal{Z}_{t+1}}$ is the 2-Wasserstein distance w.r.t. the underlying metric $d_{\mathcal{Z}_{t+1}}$. The adapted distribution of X is then defined as

$$(41) \quad \mathcal{L}^{\text{ad}}(X) := \mathcal{L}(\mathbf{ip}_1(X)) \in \mathcal{P}_2(\mathcal{Z}_1).$$

Theorem 8.1. *The space of filtered processes is isometrically isomorphic to $\mathcal{P}_2(\mathcal{Z}_1)$. That is*

$$\mathcal{AW}_2(X, Y) = \mathcal{W}_2(\mathcal{L}^{\text{ad}}(X), \mathcal{L}^{\text{ad}}(Y)).$$

In particular $X \sim_{\mathcal{AW}} Y$ if and only if $X \sim_{\text{ad}} Y$. Moreover, for every $P \in \mathcal{P}_2(\mathcal{Z}_1)$, there exists $X \in L_{2,ad}^N(H)$ such that $\mathcal{L}^{\text{ad}}(X) = P$.

Proof. The first assertion is due to [14, Theorem 1.3]. The second claim follows analogously as in Proposition 5.2. $\square$

8.2. Embedding into $\mathcal{P}_2^N(H^N)$. In order to apply the results from the previous sections, we need to embed $\mathcal{P}_2(\mathcal{Z}_1)$ into $\mathcal{P}_2^N(H^N)$. We start to outline this procedure at the level of processes / random variables.

Every process $X = (X_t)_{t=1}^N \in L_{2,ad}^N(H)$ is (by forgetting the adaptedness constraint) in particular an element of $L_2^N(H^N)$, i.e., there is a natural embedding

$$(42) \quad \iota_{ad} : L_{2,ad}^N(H) \rightarrow L_2^N(H^N).$$

If it is notionally convenient, we suppress ι_{ad} and in particular consider $L_{2,ad}^N(H)$ as a subspace of $L_2^N(H^N)$.

On the other hand we can interpret $X \in L_2^N(H^N)$ as a process $X = (X_t)_{t=1}^N$ that is not necessarily adapted. We can naturally assign to it the adapted process $(\mathbb{E}[X_t|\mathcal{F}_t])_{t=1}^N$. This defines the orthogonal projection

$$(43) \quad \text{pr}_{ad} : L_2^N(H^N) \rightarrow L_{2,ad}^N(H) : X \mapsto (\mathbb{E}[X_t|\mathcal{F}_t])_{t=1}^N.$$

Next, we outline this procedure at the level of adapted distributions. To that end, we introduce the auxiliary spaces

$$E_t := \mathcal{P}_2^{N-t}(H^{N+1-t}) \quad t = 1, \dots, N.$$

Note that $\mathcal{P}_2(E_1) = \mathcal{P}_2^N(H^N)$. We recall from Section 5 that the ip of a random variable is given by $\text{ip}_t(X) = \mathcal{L}^{\mathcal{F}_{t+1}}(X)$ and further note that $\text{ip}_t(X_{t:N})$ takes values in E_t .

It is crucial that $\mathbf{ip}_t(X)$ and $\text{ip}_t(X_{t:N})$ contain precisely the same information. In particular, $\mathbf{ip}_1(X)$ and $\text{ip}_1(X)$ contain the same information and hence the difference of $\mathcal{L}^{\text{ad}}$ and $\mathcal{L}^{\mathbf{ad}}$ is only of notational nature. This is made rigorous in the next proposition:

Proposition 8.2. *For $t \in \{1, \dots, N\}$, there exist mappings*

$$\begin{aligned} \iota_t : \mathcal{Z}_t &\rightarrow E_t \\ j_t : E_t &\rightarrow \mathcal{Z}_t \end{aligned}$$

with the following properties:

- (1) j_t is continuous and the left-inverse of ι_t , i.e., $j_t \circ \iota_t = \text{id}_{\mathcal{Z}_t}$,
- (2) ι_t is an isometric embedding,
- (3) $\iota_t(\mathbf{ip}_t(X)) = \text{ip}_t(X_{t:N})$ for every $X \in L_{2,ad}^N$,
- (4) $j_t(\text{ip}(X_{t:N})) = \mathbf{ip}_t(\text{pr}_{ad}(X))$ for every $X \in L_2^N(H^N)$.

Moreover, we have:

- (5) $j_{1\#}\mathcal{L}^{\text{ad}}(X) = \mathcal{L}^{\mathbf{ad}}(\text{pr}_{ad}(X))$ for every $X \in L_2^N(H^N)$,
- (6) $\iota_{1\#}\mathcal{L}^{\mathbf{ad}}(X) = \mathcal{L}^{\text{ad}}(X)$ for every $X \in L_{2,ad}^N(H)$.

Proof. We start with the construction of the mappings. To this end, we write δ for the map

$$\delta : \mathcal{X} \rightarrow \mathcal{P}_2(\mathcal{X}) : x \mapsto \delta_x$$

which assigns to a point x the Dirac measure at x and we also consider its t -fold iteration $\delta^t : \mathcal{X} \rightarrow \mathcal{P}_2^t(\mathcal{X})$.

Next, we define the mappings $\iota_t : \mathcal{Z}_t \rightarrow B_t$ by backward induction on t . For $t = N$ we set $\iota_N := \text{id} : H \rightarrow H$ and inductively, for $t < N$,

$$(44) \quad \iota_t : \mathcal{Z}_t \rightarrow E_t, \quad \iota_t(x, p) := \delta_x^{N-t} \otimes \iota_{t+1\#}p.$$

The mappings j_t are then defined by $j_N := \text{id} : H \rightarrow H$ and by induction for $t < N$ (writing $\text{pr}_t, \dots, \text{pr}_N$ for the projections from H^{N+1-t} onto H) defined by

$$(45) \quad j_t : E_t \rightarrow \mathcal{Z}_t, \quad j_t(P) = (\text{mean}(\text{pr}_{t\#}(I^{N-t-1}(P))), j_{t+1\#}(\mathcal{P}^{N-t}[\text{pr}_{t+1:N}](P))),$$

where $\text{mean}(\mu) := \int x \mu(dx)$ for $\mu \in \mathcal{P}_2(H)$.

Next, we show the claims (1) to (4) simultaneously by induction on t . For $t = N$, they are trivial as ι_N and j_N are both the identity.

Suppose that the claim is true for $t + 1$. We start with claim (1). As the intensity, mean and pushforwards with continuous maps, are all continuous operations, j_t is continuous. For $(x, p) \in \mathcal{Z}_t = H \times \mathcal{P}_2(\mathcal{Z}_{t+1})$ we have using that $j_{t+1} \circ \iota_{t+1} = \text{id}_{\mathcal{Z}_{t+1}}$

$$j_t(\iota_t(x, p)) = j_t(\delta_x^{N-t} \otimes \iota_{t+1\#} p) = (\text{mean}(I^{N-t-1}(\delta_x^{N-t})), j_{t+1\#} \iota_{t+1\#} p) = (x, p).$$

Concerning (2), it is easy to check from (44) and from the definition of the Wasserstein distance that ι_t is an isometry. It is an embedding because j_t is its continuous left-inverse.

In order to show (3) observe that

$$\iota_t(\mathbf{ip}_t(X)) = \iota_t(X_t, \mathcal{L}^{\mathcal{F}_t}(\mathbf{ip}_{t+1}(X))) = \delta_{X_t}^{N-t} \otimes \mathcal{L}^{\mathcal{F}_t}(\iota_{t+1}(\mathbf{ip}_{t+1}(X)))$$

and using the inductive claim this is further equal to

$$\delta_{X_t}^{N-t} \otimes \mathcal{L}^{\mathcal{F}_t}(\mathbf{ip}_{t+1}(X)) = \delta_{X_t}^{N-t} \otimes \mathcal{L}^{\mathcal{F}_{N-1:t}}(X_{t+1:N}) = \mathcal{L}^{\mathcal{F}_{N-1:t}}(X_{t:N}),$$

where the last equality is true because X_t is $\mathcal{F}_t$ -measurable.

Claim (4) follows from (3) and the fact that j_t is the left inverse of ι_t . This finishes the induction.

The claims (5) and (6) follow from (3) and (4) for $t = 1$. □

Corollary 8.3. *We have $\mathcal{P}_2(\mathcal{Z}_1) = \{\mathcal{L}^{\text{ad}}(X) : X \in L_{2,\text{ad}}^N(H)\}$. In particular, $L_{2,\text{ad}}^N(H)$ is an adapted-law invariant subspace and the projection pr_{ad} is adapted-law invariant, i.e., if $X \sim_{\text{ad}} X'$, then $\text{pr}_{\text{ad}}(X) \sim_{\text{ad}} \text{pr}_{\text{ad}}(X')$.*

Proof. The first assertion follows from Proposition 8.2 and the fact that the space of filtered processes and $\mathcal{P}_2(\mathcal{Z}_1)$ are isomorphic (see Theorem 8.1). The second assertion follows from Proposition 8.2 (5). □

We note that it is more generally true that the projection onto an adapted-law invariant subspace is adapted-law invariant.

Corollary 8.4. *For $t \leq N$, there exists a measurable map $g_t : \mathcal{Z}_{1:t} \rightarrow \mathcal{P}_2^{N-t}(H^N)$ such that $\mathbf{ip}_t(X) = g_t(\mathbf{ip}_{1:t}(X))$ for every $X \in L_{2,\text{ad}}^N(H)$.*

Proof. For every $s \leq t$, X_s is the first component of $\mathbf{ip}_s(X)$. Hence, there is a projection $\text{pr} : \mathcal{Z}_{1:t-1} \rightarrow H^{t-1}$ such that $\text{pr}(\mathbf{ip}_{1:t-1}(X)) = X_{1:t-1}$ for every $X \in L_{2,\text{ad}}^N(H)$. The map $\iota_t : \mathcal{Z}_t \rightarrow E_t$ defined in Proposition 8.2 satisfies $\iota_t(\mathbf{ip}_t(X)) = \mathbf{ip}_t(X_{t:N})$ for every $X \in L_{2,\text{ad}}^N(H)$. Thus,

$$g_t : \mathcal{Z}_{1:t} \rightarrow \mathcal{P}_2^{N-t}(H^N), \quad g(z_{1:t}) := \delta_{\text{pr}(z_{1:t-1})}^{N-t} \otimes \iota_t(z_t)$$

satisfies $\mathbf{ip}_t(X) = g_t(\mathbf{ip}_{1:t}(X))$ for every $X \in L_{2,\text{ad}}^N(H)$. □

It was central in the previous sections (in particular in the context of MC-subdifferential and N -Monge couplings) to establish that a random variable Y is of the form $Y = \xi(\mathbf{ip}(X))$. The next assertion shows that in the present setting the mapping ξ translates to an adapted mapping.

Proposition 8.5. *Let $X, Y \in L_{2,ad}^N(H)$ and let $\tilde{\xi} : A_{1:N} \rightarrow H^N$ be measurable such that $Y = \tilde{\xi}(\text{ip}(X))$. Then there is an adapted map $\xi : \mathcal{Z} \rightarrow H^N$ such that $Y = \xi(\text{ip}(X))$, i.e., $Y_t = \xi_t(\text{ip}_t(X))$ for every $t \leq N$. Moreover, there is an adapted map $S : \mathcal{Z} \rightarrow \mathcal{Z}$ such that $\text{ip}(Y) = S(\text{ip}(X))$.*

Proof. Fix $t \leq N$. By Proposition 6.6, the map $\tilde{\xi}$ induces a map $T_t : \prod_{s=1}^t \mathcal{P}_2^{N-s}(H^N) \rightarrow \mathcal{P}_2^{N-t}(H^N)$ such that $\text{ip}_t(Y) = T_t(\text{ip}_{1:t}(X))$. The map g_t defined in Corollary 8.4 satisfies $\text{ip}_t(X) = g_t(\text{ip}_{1:t}(X))$. We set $h_t(P) := \text{mean}(I^{N-t-1}(\mathcal{P}^{N-t}[\text{pr}_t](P)))$ where $\text{pr}_t : H^N \rightarrow H$ is the projection onto the t -th component. Note that

$$h_t(\text{ip}_t(Y)) = \text{mean}(I^{N-t-1}(\mathcal{L}^{\mathcal{F}_{N-1:t}}(Y_t))) = \text{mean}(I^{N-t-1}(\delta_{Y_t}^{N-t})) = Y_t.$$

We then define the desired map ξ_t as $\xi_t := h_t \circ T_t \circ g_t$ and note that

$$\xi_t(\text{ip}_t(X)) = h_t(T_t(g_t(\text{ip}_{1:t}(X)))) = h_t(T_t(\text{ip}_{1:t}(X))) = h_t(\text{ip}_t(Y)) = Y_t.$$

For the second claim, we set $S_t = j_t \circ T_t \circ g_t$ and note that

$$S_t(\text{ip}_{1:t}(X)) = j_t(T_t(g_t(\text{ip}_{1:t}(X)))) = j_t(T_t(\text{ip}_{1:t}(X))) = j_t(\text{ip}_t(Y)) = \text{ip}_t(Y). \quad \square$$

8.3. Lions lift for filtered processes. From now on, we occasionally write $\mathcal{Z}_t^{(N)}$ instead of $\mathcal{Z}_t$ to make the dependence on the number of time steps explicit when needed. As the space of filtered processes is isometrically isomorphic to $\mathcal{P}_2(\mathcal{Z}_1^{(N)})$, defining a lift for functions $\psi : \mathcal{P}_2(\mathcal{Z}_1) \rightarrow (-\infty, +\infty]$ is equivalent to defining a lift for functions on the space of filtered processes with N time periods.

Definition 8.6. Let $\psi : \mathcal{P}_2(\mathcal{Z}_1^{(N)}) \rightarrow (-\infty, +\infty]$. Then its adapted Lions lift is defined as the function

$$\bar{\psi} : L_{2,ad}^N(H) \rightarrow (-\infty, +\infty], \quad \bar{\psi}(X) = \psi(\mathcal{L}^{\text{ad}}(X)).$$

In order to establish a connection to convex analysis and Brenier-type results for the adapted Wasserstein distance, we consider the adapted maximal covariance

$$\text{AMC}(X, Y) := \sup_{X' \sim_{\text{ad}} X, Y' \sim_{\text{ad}} Y} \mathbb{E}[X' \cdot Y'].$$

This adapted max-covariance functional is connected to the iterated max-covariance on $\mathcal{P}_2(\mathcal{Z}_1^{(N)})$. To define this functional we first set

$$\begin{aligned} \langle x, y \rangle_{\mathcal{Z}_N} &:= \langle x, y \rangle_H \\ \langle (x, p), (y, q) \rangle_{\mathcal{Z}_t} &:= \langle x, y \rangle_H + \langle p, q \rangle_{\mathcal{Z}_{t+1}} \end{aligned}$$

and then define the max-covariance functional on $\mathcal{P}_2(\mathcal{Z}_1)$ as

$$\text{MC}(P, Q) = \sup_{\Pi \in \text{Cpl}(P, Q)} \int \langle (x, p), (y, q) \rangle_{\mathcal{Z}_1} d\Pi(x, p, y, q).$$

Proposition 8.7. *For $X, Y \in L_{2,ad}^N(H)$ we have*

- (1) $\mathcal{AW}_2^2(X, Y) = \|X\|_2^2 + \|Y\|_2^2 - 2\text{AMC}(X, Y)$
- (2) $\mathcal{W}_2^2(\mathcal{L}^{\text{ad}}(X), \mathcal{L}^{\text{ad}}(Y)) = \mathbb{E}|X|^2 + \mathbb{E}|Y|^2 - 2\text{MC}(\mathcal{L}^{\text{ad}}(X), \mathcal{L}^{\text{ad}}(Y))$
- (3) $\text{AMC}(X, Y) = \text{MC}(\mathcal{L}^{\text{ad}}(X), \mathcal{L}^{\text{ad}}(Y))$

Proof. The claims (1) and (2) are straightforward by completing squares. (3) follows then from Theorem 8.1 by invoking (1) and (2). $\square$

Transport theory on $\mathcal{P}_2(\mathcal{Z}_1)$ with MC costs then yields the notions of MC-transform and MC-subgradient in the present setting:

Definition 8.8. Let $\psi : \mathcal{P}_2(\mathcal{Z}_1) \rightarrow (-\infty, +\infty]$ be proper. We define its MC-transform as

$$\psi^{\text{MC}}(Q) = \sup_{P \in \mathcal{P}_2(\mathcal{Z}_1)} \text{MC}(P, Q) - \psi(P).$$

A function $\psi : \mathcal{P}_2(\mathcal{Z}_1) \rightarrow (-\infty, +\infty]$ is called MC-convex if there exists a proper function $\vartheta : \mathcal{P}_2(\mathcal{Z}_1) \rightarrow (-\infty, +\infty]$ such that $\psi = \vartheta^{\text{MC}}$. Moreover, the MC-subdifferential of ψ at $P \in \mathcal{P}_2(\mathcal{Z})$ is defined as

$$\partial_{\text{MC}}\psi(P) = \{Q \in \mathcal{P}_2(\mathcal{Z}_1) : \psi(R) \geq \psi(P) + \text{MC}(R, Q) - \text{MC}(P, Q) \text{ for every } R \in \mathcal{P}_2(\mathcal{Z}_1)\},$$

and we write

$$\partial_{\text{MC}}\varphi := \{(P, Q) : Q \in \partial_{\text{MC}}\varphi(P), P \in \mathcal{P}_2(\mathcal{Z}_1)\}.$$

Analogous to the theory on $\mathcal{P}_2^N(H)$, we have the following crucial connection between the MC-transform and the convex conjugate on $L_{2,ad}^N(H)$.

Lemma 8.9. Let $P, Q \in \mathcal{P}_2(\mathcal{Z}_1)$ and $Y \in L_{2,ad}^N(H)$ with $\mathcal{L}^{\text{ad}}(Y) = Q$ be given. Then,

$$\text{MC}(P, Q) = \sup_{\substack{X \in L_{2,ad}^N(H) \\ \mathcal{L}^{\text{ad}}(X) = P}} \mathbb{E}[X \cdot Y].$$

Proof. By using that ι_1 is an isometry (see Proposition 8.2), the corresponding claim in the setup of $\mathcal{P}_2^N(H^N)$ (see Lemma 5.5) and the self-adjointness of pr_{ad} we find

$$\text{MC}(P, Q) = \text{MC}(\iota_{1\#}P, \iota_{1\#}Q) = \sup_{\substack{X \in L_2^N(H^N) \\ \mathcal{L}^{\text{ad}}(X) = P}} \mathbb{E}[X \cdot Y] = \sup_{\substack{X \in L_{2,ad}^N(H) \\ \mathcal{L}^{\text{ad}}(X) = P}} \mathbb{E}[X \cdot Y].$$

□

Proposition 8.10. Let $\psi : \mathcal{P}_2(\mathcal{Z}_1) \rightarrow (-\infty, +\infty]$ be proper. Then we have

$$(46) \quad \overline{\psi^{\text{MC}}} = \overline{\psi}^*.$$

In particular, the convex conjugate of an adapted-law invariant function on $L_{2,ad}^N(H)$ is adapted-law invariant.

Moreover, the following are equivalent:

- (1) $\overline{\psi}$ is lsc convex,
- (2) ψ is MC-convex.

Proof. For $Y \in L_{2,ad}^N(H)$ we write $Q = \mathcal{L}^{\text{ad}}(Y)$. By Lemma 8.9 we find

$$\begin{aligned} \overline{\psi}^*(Y) &= \sup_{X \in L_{2,ad}^N(H)} \mathbb{E}[X \cdot Y] - \overline{\psi}(X) = \sup_{P \in \mathcal{P}_2(\mathcal{Z}_1)} \sup_{X \sim_{ad} P} \mathbb{E}[X \cdot Y] - \psi(P) \\ &= \sup_{P \in \mathcal{P}_2(\mathcal{Z}_1)} \text{MC}(P, Q) - \psi(P) = \psi^{\text{MC}}(Q) = \overline{\psi^{\text{MC}}}(Y). \end{aligned}$$

Next, suppose that $\overline{\psi}$ is lsc convex. Then by first applying the Fenchel–Moreau theorem and then (46) twice we find

$$\overline{\psi} = \overline{\psi}^{**} = \overline{\psi^{\text{MC}}}^* = \overline{\psi^{\text{MCMC}}}.$$

Hence, $\psi = \psi^{\text{MCMC}}$, which shows that ψ is MC-convex.

Conversely assume that ψ is MC-convex. Then $\psi = \vartheta^{\text{MC}}$ for some proper function $\vartheta : \mathcal{P}_2(\mathcal{Z}_1) \rightarrow (-\infty, +\infty]$. By (46), we have $\overline{\psi} = \overline{\vartheta^{\text{MC}}} = \overline{\vartheta}^*$, which shows that $\overline{\psi}$ is lsc convex. □

Remark 8.11. Next, we discuss how to extend adapted-law invariant functions on $L_{2,ad}^N(H)$ to adapted-law invariant functions on $L_2^N(H^N)$. This allows us to also translate results on the structure of the subdifferentials of MC-convex functions from Section 6 to the present setting. There are two ways to extend a given adapted-law invariant function $\bar{\psi} : L_{2,ad}^N(H) \rightarrow (-\infty, +\infty]$.

- (1) We set it $+\infty$ on non-adapted processes, i.e., writing χ for convex indicators

$$\bar{\varphi}(X) = \bar{\psi}(X) + \chi_{L_{2,ad}^N(H)}(X).$$

- (2) We use the projection onto adapted processes, i.e.

$$\bar{\varphi}(X) := \bar{\psi}(\text{pr}_{ad}(X)).$$

As $L_{2,ad}^N(H)$ and pr_{ad} are adapted-law invariant (see Corollary 8.3), this defines indeed adapted-law invariant functions. Moreover, these two ways are dual to each other via the convex conjugate on $L_2^N(H^N)$, i.e.

$$(\bar{\psi} \circ \text{pr}_{ad})^* = \bar{\psi}^* + \chi_{L_{2,ad}^N(H)}.$$

Remark 8.12. If we are in the situation that $\bar{\varphi}(X) := \bar{\psi}(\text{pr}_{ad}(X))$, we have (when suppressing the embedding ι_{ad} and considering $L_{2,ad}^N(H)$ as subspace of $L_2^N(H^N)$)

$$\partial \bar{\varphi}(X) = \partial \bar{\psi}(X).$$

Here, the subdifferential on the left hand side is meant in $L_2^N(H^N)$, whereas the one on the right hand side is meant in the subspace $L_{2,ad}^N(H)$. In particular, this asserts that $\partial \bar{\varphi}(X) \subseteq L_{2,ad}^N(H)$. Moreover, for $X \in L_{2,ad}^N(H)$, $\bar{\varphi}$ is Gateaux (Frechet) differentiable at X if and only if $\bar{\psi}$ is Gateaux (Frechet) differentiable at X .

Proposition 8.13. Let $\psi : \mathcal{P}_2(\mathcal{Z}_1) \rightarrow (-\infty, +\infty]$ be proper and let $X, Y \in L_{2,ad}^N(H)$. If $X \sim_{ad} P$ and $Z \sim_{ad} Q$ we have

$$Z \in \partial \bar{\psi}(X) \iff Q \in \partial_{MC} \psi(P) \text{ and } \mathbb{E}[X \cdot Z] = MC(P, Q).$$

Proof. The proof is line by line as the proof of (19). $\square$

Lemma 8.14. Let $\psi : \mathcal{P}_2(\mathcal{Z}_1) \rightarrow (-\infty, \infty]$ be MC-convex and $P \in \text{dom}(\psi)$. Then the following are equivalent:

- (1) ψ is MC-differentiable at P .
- (2) For all X with $X \sim_{ad} P$ we have $\# \partial \bar{\psi}(X) = 1$.

Proof. This follows line by line as in the proof of Lemma 6.15. $\square$

Corollary 8.15. Let $\psi : \mathcal{P}_2(\mathcal{Z}_1) \rightarrow (-\infty, \infty]$ be MC-convex and $P \in \text{dom}(\varphi)$. Set $\varphi := \psi \circ \mathcal{P}(\iota_1)$. Then the MC-subdifferentials coincide in the sense that

$$\partial_{MC} \varphi(\iota_{1\#} P) = \{\iota_{1\#} Q : Q \in \partial_{MC} \psi(P)\}.$$

In particular, ψ is MC-differentiable at P if and only if φ is MC-differentiable at $\iota_{1\#} P$.

Proof. Let $R \in \partial_{MC} \varphi(\iota_{1\#} P)$. Let $X \sim_{ad} \iota_{1\#} P$ (and hence $X \in L_{2,ad}^N(H)$) and $Y \sim_{ad} R$ such that $\mathbb{E}[X \cdot Y] = MC(\iota_{1\#} P, R)$. By Proposition 6.11, we have $Y \in \partial \bar{\varphi}(X)$. By Remark 8.12, we have $Y \in \partial \bar{\psi}(X)$. In particular, $Y \in L_{2,ad}^N(H)$ and hence $R = \mathcal{L}^{ad}(Y)$ is of the form $\iota_{1\#} Q$ for some $Q \in \mathcal{P}_2(\mathcal{Z}_1)$. Then Proposition 8.13 yields $Q \in \partial_{MC} \psi(P)$.

Conversely, let $Q \in \partial_{MC} \psi(P)$. Then there are $X, Y \in L_{2,ad}^N(H)$ such that $MC(P, Q) = \mathbb{E}[X \cdot Y]$. By Proposition 8.13, we have $Y \in \partial \bar{\psi}(X)$, by Remark 8.12 we have $Y \in \partial \bar{\varphi}(X)$. Then, $\iota_{1\#} Q \in \partial_{MC} \varphi(\iota_{1\#} P)$ by Proposition 6.11. $\square$

Proposition 8.16. *Let $\psi : \mathcal{P}_2(\mathcal{Z}_1) \rightarrow (-\infty, \infty]$ be MC-differentiable at P . Then there exists an adapted map $\xi : \mathcal{Z} \rightarrow H^N$ such that for all $X \sim_{\text{ad}} P$*

$$\partial \bar{\psi}(X) = \{\xi(\text{ip}(X))\}.$$

Proof. We set $\varphi := \psi \circ \mathcal{P}(\iota_1)$ and note that $\bar{\varphi}(X) = \bar{\psi}(\text{pr}_{\text{ad}}(X))$ and that φ is MC-differentiable at $\iota_1 \# P$ by Corollary 8.15. By Proposition 6.16 there exists a measurable function $\tilde{\xi} : A_{1:N} \rightarrow H^N$ such that $\partial \bar{\varphi}(X) = \{\tilde{\xi}(\text{ip}(X))\}$ for all $X \sim_{\text{ad}} P$. By Proposition 8.5, there is an adapted map $\xi : \mathcal{Z} \rightarrow H^N$ such that $\tilde{\xi}(\text{ip}(X)) = \xi(\text{ip}(X))$. Using Remark 8.12 we find

$$\partial \bar{\psi}(X) = \partial \bar{\varphi}(X) = \{\xi(\text{ip}(X))\}.$$

□

Proposition 8.17. *For $P, Q \in \mathcal{P}_2(\mathcal{Z}_1)$ the following are equivalent:*

- (1) $\text{MC}(\cdot, Q)$ is MC-differentiable at P .
- (2) *There is an adapted map $\xi : \mathcal{Z} \rightarrow H^N$ such that for some $X \sim_{\text{ad}} P$*

$$\{Y \sim_{\text{ad}} Q : \mathbb{E}[X \cdot Y] = \text{MC}(P, Q)\} = \{\xi(\text{ip}(X))\}.$$

- (3) *For some $X \sim_{\text{ad}} P$, the function $\overline{\text{MC}(\cdot, Q)}$ is Frechet differentiable at X .*

Moreover, whenever one (and hence all) of the above conditions holds, statements (2) and (3) are valid for every $X \sim_{\text{ad}} P$.

Proof. We derive this result from Proposition 6.17, that is the corresponding result on $\mathcal{P}_2^N(H^N)$. For this note that the extension of $\text{MC}(\cdot, Q)$ to a function on $\mathcal{P}_2^N(H^N)$ in the way described in Remark 8.11(2) is precisely $\text{MC}(\cdot, \iota_{1\#} Q)$ (here MC denotes the max-covariance on $\mathcal{P}_2^N(H^N)$).

Assertion (1) in Proposition 6.17 and assertion (1) in Proposition 8.17 are equivalent by Corollary 8.15. Moreover, the assertion (2) in Proposition 6.17 and assertion (2) in Proposition 8.17 are equivalent due to Proposition 8.5. Finally, assertion (3) in Proposition 6.17 and assertion (3) in Proposition 8.17 are equivalent by Remark 8.12. □

Proposition 8.18. *Let $\psi : \mathcal{P}_2(\mathcal{Z}_1) \rightarrow (-\infty, +\infty]$ and suppose that $\text{cont}(\psi) \neq \emptyset$. Then ψ is MC-differentiable on a dense G_δ subset of the closure of $\text{dom}(\psi)$.*

Proof. The proof follows the proof of Corollary 6.21 line by line. □

8.4. Naturally filtered Processes. A process is called natural if its filtration contains no more information about its future evolution than is provided by its past trajectory. This is made precise in the following definition.

Definition 8.19. X is called naturally filtered if for every $t \leq N$

$$\mathcal{L}(X | \mathcal{F}_t) = \mathcal{L}(X | X_{1:t}).$$

Naturally filtered processes are already determined by the law of the processes itself and it is not necessary to consider the adapted law in this case. This is made precise in the next lemma.

Lemma 8.20. *Let $X, Y \in L_{2,\text{ad}}^N(H)$. Then, we have:*

- (1) *X is naturally filtered if and only if there is an adapted map f such that $\text{ip}(X) = f(X)$.*
- (2) *If X, Y are naturally filtered, then $\mathcal{L}^{\text{ad}}(X) = \mathcal{L}^{\text{ad}}(Y)$ if and only if $\mathcal{L}(X) = \mathcal{L}(Y)$.*

Proof. See e.g. [23, Section 3] and [65, Lemma 2.3]. □

As the law of the random variables determines a natural filtered processes, it is also meaningful to consider the adapted Wasserstein distance between these laws, defined in (14) in the introduction. The following lemma clarifies that these concepts are consistent.

Lemma 8.21. *We have*

- (1) *If $X, Y \in L_{2,\text{ad}}^N(H)$ are naturally filtered, $\mathcal{L}(X, Y) \in \text{Cpl}_{\text{bc}}(\mathcal{L}(X), \mathcal{L}(Y))$.*

- (2) If $\mu, \nu \in \mathcal{P}_2(H^N)$ and $\pi \in \text{Cpl}(\mu, \nu)$, there are naturally filtered $X, Y \in L_{2,ad}^N(H)$ such that $\mathcal{L}(X, Y) = \pi$.
- (3) If $X, Y \in L_{2,ad}^N(H)$ are naturally filtered, $\mathcal{AW}_2(X, Y) = \mathcal{AW}_2(\mathcal{L}(X), \mathcal{L}(Y))$.

Proof. For (1) and (2), see e.g. [23, Lemma 5.3]. For (3) see [14, Theorem 1.2]. $\square$

Next, we give an embedding J of the (adapted distributions of) natural filtered processes $(\mathcal{P}_2^N(H^N), \mathcal{AW}_2)$ into set of (adapted distributions of) general processes. For $N = 2$ periods, writing $d\mu(x_1, x_2) = d\mu_1(x_1)\mu_{x_1}(x_2)$, it reads as follows:

$$J(\mu) = (x_1 \mapsto (x_1, \mu_{x_1}))_{\#}\mu_1.$$

For $N > 2$ it is an iteration of such maps (see [11, 12]) for notational simplicity we define the map J in processes language as

$$(47) \quad J : \mathcal{P}_2(H^N) \rightarrow \mathcal{P}_2(\mathcal{Z}_1) : \mu \mapsto \mathcal{L}^{\text{ad}}(X) \text{ where } X \text{ naturally filtered with } \mathcal{L}(X) = \mu.$$

Proposition 8.22. *The map $J : (\mathcal{P}_2(H^N), \mathcal{AW}_2) \rightarrow (\mathcal{P}_2(\mathcal{Z}_1), \mathcal{W}_2)$ is an isometric embedding and its range is dense G_δ .*

Proof. See e.g. [12, Section 5]. $\square$

Corollary 8.23. *Let $X, Y \in L_{2,ad}^N(H)$. If X is naturally filtered and $Y = \xi(\mathbf{ip}(X))$, then there is an adapted map $\mathsf{T} : H^N \rightarrow H^N$ such that $Y = \mathsf{T}(X)$.*

Proof. This follows immediately from Proposition 8.5 and Lemma 8.20 (1). $\square$

Proposition 8.24. *For every $\nu \in \mathcal{P}_2(H^N)$, the set of $\mu \in \mathcal{P}_2(H^N)$ such that $\text{Cpl}_{\text{opt}}^{bc}(\mu, \nu) = \{\pi\}$ and π is induced by an adapted map is comeager in $\mathcal{P}_2(H^N)$.*

Proof. Fix $\nu \in \mathcal{P}_2(H^N)$ and write $Q = J(\nu) \in \mathcal{P}_2(\mathcal{Z}_1)$, where J is the map defined in (47). By Proposition 8.18, the set of $P \in \mathcal{P}_2(\mathcal{Z}_1)$ such that $\text{MC}(\cdot, Q)$ is MC-differentiable in P is dense G_δ in $\mathcal{P}_2(\mathcal{Z}_1)$. By Proposition 8.22, also the set

$$D := \{\mu \in \mathcal{P}_2(H^N) : \text{MC}(\cdot, Q) \text{ is MC-differentiable in } J(\mu)\}$$

is dense G_δ in $\mathcal{P}_2(H^N)$.

Let $\mu \in D$. By Proposition 8.17 there is an adapted map $\xi : \mathcal{Z} \rightarrow H^N$ such that for every $X \sim_{\text{ad}} J(\mu)$

$$(48) \quad \{Y \sim_{\text{ad}} Q : \mathbb{E}[X \cdot Y] = \text{MC}(P, Q)\} = \{\xi(\mathbf{ip}(X))\}.$$

By Lemma 8.20, there is an adapted map $f : H^N \rightarrow \mathcal{Z}$ such that $\mathbf{ip}(X) = f(X)$. We define

$$\mathsf{T} := \xi \circ f : H^N \rightarrow H^N.$$

Note that T is adapted as concatenation of adapted maps. Now, let $\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)$ be any optimizer. We observe that $\pi = (\text{id}, \mathsf{T})_{\#}\mu$. Indeed, by Lemma 8.21, there are $X, Y \in L_{2,ad}^N(H)$ natural such that $\mathcal{L}(X, Y) = \pi$. By (48), we have $Y = \xi(\mathbf{ip}(X)) = \xi(f(X)) = \mathsf{T}(X)$. Hence, the optimal π is unique and Monge. $\square$

Proposition 8.25. *The set of pairs $(\mu, \nu) \in \mathcal{P}_2(H^N) \times \mathcal{P}_2(H^N)$ such that $\text{Cpl}_{\text{opt}}^{bc}(\mu, \nu) = \{\pi\}$ and π is induced by a bi-adapted map is comeager in $\mathcal{P}_2(H^N) \times \mathcal{P}_2(H^N)$.*

Proof. We write $e : H^N \times H^N : (x, y) \mapsto (y, x)$ and

$$\begin{aligned} A &= \{(\mu, \nu) : \text{Cpl}_{\text{opt}}^{bc}(\mu, \nu) = \{\pi\} \text{ and } \pi \text{ is Monge}\}, \\ B &= \{(\mu, \nu) : \text{Cpl}_{\text{opt}}^{bc}(\mu, \nu) = \{\pi\} \text{ and } e_{\#}\pi \text{ is Monge}\}. \end{aligned}$$

By Proposition 8.25, for every $\nu \in \mathcal{P}_2(H^N)$, the set $A^\nu := \{\mu \in \mathcal{P}_2(H^N) : (\mu, \nu) \in A\}$ is comeager. For the same reason, for every $\mu \in \mathcal{P}_2(H^N)$, the set $B_\mu = \{\nu \in \mathcal{P}_2(H^N) : (\mu, \nu) \in B\}$ is comeager. It follows from the Kuratowski–Ulam theorem (see e.g. [54, Theorem 8.41]) that A and B are comeager in $\mathcal{P}_2(H^N) \times \mathcal{P}_2(H^N)$. Hence, $A \cap B$ is comeager in $\mathcal{P}_2(H^N) \times \mathcal{P}_2(H^N)$. If $(\mu, \nu) \in A \cap B$, then $\text{Cpl}_{\text{opt}}^{bc}(\mu, \nu) = \{\pi\}$ and both π and $e_\# \pi$ are induced by an adapted map. Hence, π is induced by a bi-adapted map (see e.g. [20, Lemma A.2]). $\square$

8.5. Lift of functions on $\mathcal{Z}_1$ and proof of Theorem 1.11. For transport of measures $P, Q \in \mathcal{P}_2(\mathcal{Z}_1^{(N)})$, the dual potentials are functions on $\mathcal{Z}_1^{(N)} = H \times \mathcal{P}_2(\mathcal{Z}_2^{(N)}) = H \times \mathcal{P}_2(\mathcal{Z}_1^{(N-1)})$. For this reason, we also need to consider adapted Lions lifts of such functions, that is

$$(49) \quad \bar{\psi} : H \times L_{2,ad}^{N-1}(H) \rightarrow (-\infty, +\infty], \quad \bar{\psi}(x_1, X_{2:N}) = \psi(x_1, \mathcal{L}^{\text{ad}}(X_{2:N})).$$

We write

$$V_N := H \times L_{2,ad}^{N-1}(H)$$

and note that if $X \in L_{2,ad}^N(H)$ and $u_1 \in [0, 1]$ is fixed, then

$$X(u_1, \dots) = (X_1(u_1), X_2(u_1, \cdot), \dots, X_N(u_1, \dots)) \in V_N.$$

Hence, we can consider $X \in L_{2,ad}^N(H)$ also as V_N -valued random variable on $([0, 1], \lambda)$ and apply a function $\bar{\psi} : H \times L_{2,ad}^{N-1}(H) \rightarrow (-\infty, +\infty]$ to it, in this case we write $\bar{\psi}(X(u_1, \cdot))$.

It is easy to see from the results in Section 8.3 that via this Lions lift defined in (49) the c -transform for the cost $c((x, p), (y, q)) = \langle (x, p), (y, q) \rangle_{\mathcal{Z}_1}$ coincides with the convex conjugate on V_N . Writing $\langle \cdot, \cdot \rangle_{V_N}$ for the scalar product on V_N we also have the analogue of Proposition 8.13, namely

$$(50) \quad (y_1, Y_{2:N}) \in \partial \bar{\varphi}((x_1, X_{2:N})) \iff \begin{cases} (Q, y_1) \in \partial_c \varphi((P, x_1)) \text{ and} \\ \langle (x_1, X_{2:N}), (y_1, Y_{2:N}) \rangle_{V_N} = c((x_1, P), (y_1, Q)). \end{cases}$$

Proof of Theorem 1.11. We write $P = \mathcal{L}^{\text{ad}}(X)$ and $Q = \mathcal{L}^{\text{ad}}(Y)$. Suppose that there is an adapted-law invariant function $\bar{\varphi} : V_N \rightarrow (-\infty, +\infty]$ such that $Y(u_1, \dots) \in \partial \bar{\varphi}(X(u_1, \dots))$ for a.e. $u_1 \in [0, 1]$. Writing φ for the corresponding function on $\mathcal{Z}_1$ we have

$$\varphi(\mathcal{L}^{\text{ad}}(X(u_1, \cdot))) + \varphi^c(\mathcal{L}^{\text{ad}}(Y(u_1, \cdot))) = \bar{\varphi}(X(u_1, \cdot)) + \bar{\varphi}^*(Y(u_1, \cdot)) = \langle X(u_1, \cdot), Y(u_1, \cdot) \rangle_{V_N}.$$

By integration over $\lambda(du_1)$ we find

$$\int \varphi dP + \int \varphi^c dQ = \mathbb{E}[X \cdot Y].$$

On the other hand Theorem 8.1 and the fundamental theorem of optimal transport yield that

$$\text{AMC}(X, Y) = \text{MC}(P, Q) = \inf_{\varphi} \int \varphi dP + \int \varphi^c dQ.$$

Hence, (X, Y) is optimal.

Conversely, assume that (X, Y) is optimal. Then on the probability space $([0, 1], \lambda)$ (denoting the elements of it with u_1) the pair of random variables

$$(\mathcal{L}^{\text{ad}}(X(u_1, \cdot)), \mathcal{L}^{\text{ad}}(Y(u_1, \cdot)))$$

is optimal for the transport problem between P and Q with cost c . By the fundamental theorem of optimal transport there exists a c -convex function $\varphi : \mathcal{Z}_1 \rightarrow (-\infty, +\infty]$ such that for λ -a.e. u_1 we have

$$\mathcal{L}^{\text{ad}}(Y(u_1, \cdot)) \in \partial_c \varphi(\mathcal{L}^{\text{ad}}(X(u_1, \cdot))).$$

Writing $\bar{\varphi}$ for the lift of φ as in (49), display (50) yields that $Y(u_1, \cdot) \in \partial \bar{\varphi}(X(u_1, \cdot))$ for almost every u_1 .

This finishes the proof of the first equivalence. The second claim concerning naturally filtered processes follows immediately as being naturally filtered is a property of the adapted law (i.e., if $X \sim_{\text{ad}} J(\mu)$ and $Y \sim_{\text{ad}} J(\nu)$ they are necessarily naturally filtered). $\square$

8.6. Failure of an adapted Brenier theorem for absolutely continuous measures. Finally we provide an example showing that a naive attempt to extend Brenier's theorem to the adapted Wasserstein fails already in the first non-trivial instance of one-dimensional state space and one time interval:

Example 8.26. There exist absolutely continuous, compactly supported probabilities on $\mathbb{R}^2$ such that the adapted transport problem for the quadratic costs admits no Monge type solution:

Let $\mu_1 = \nu_1$ be the uniform probability measure on $[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$, and define μ, ν on $\mathbb{R}^2$ by

$$\mu(dx_1, dx_2) := \mu(dx_1)q_{a(x_1)}(dx_2), \quad \nu(dy_1, dy_2) := \nu(dy_1)q_{b(y_1)}(dy_2)$$

where we write q_x for the uniform distribution on $[x, x+1]$ and

$$a(x) := \begin{cases} \sqrt{2}x - 1, & x \geq 0, \\ \sqrt{2}x + 1, & x < 0, \end{cases} \quad b(y) := \frac{y}{\sqrt{2}}.$$

By the dynamic programming principle the squared adapted Wasserstein distance of μ, ν is given by

$$(51) \quad \mathcal{AW}_2^2(\mu, \nu) = \inf_{\pi_1 \in \text{Cpl}(\mu_1, \nu_1)} \int (x_1 - y_1)^2 + \mathcal{W}_2^2(q_{a(x_1)}, q_{b(y_1)}) d\pi_1(x_1, y_1) =$$

$$(52) \quad = \inf_{\pi_1 \in \text{Cpl}(\mu_1, \nu_1)} \int c(x_1, y_1) d\pi_1(x_1, y_1),$$

where $c(x, y) = (x - y)^2 + (a(x) - b(y))^2$. The one dimensional transport problem (52) admits an elementary solution. Indeed, the optimal transport plan between μ_1 and ν_1 is of Monge type and is uniquely given μ_1 -a.e. by the transport map

$$T_1(x) = \begin{cases} 2x - \frac{1}{\sqrt{2}}, & x \geq 0, \\ 2x + \frac{1}{\sqrt{2}}, & x < 0. \end{cases}$$

and the dual maximizers are given by

$$\varphi_1(x) = \frac{1}{2} - x^2, \quad \psi_1(y) = \frac{1}{2}y^2, \quad x, y \in \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right].$$

It is straightforward to verify this using the complementary slackness conditions.

It follows that the unique optimal bi-adapted transport between μ and ν is induced by the map $T(x_1, x_2) = (T_1(x_1), x_2 + b(T_1(x_1)) - a(x_1))$.

Crucially, the map T_1 is not invertible on $[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$, hence there exists not optimal transport map from ν_1 to μ_1 and hence also no Monge solution for the adapted transport problem from ν to μ .

APPENDIX A. ON TRANSPORT-REGULARITY AND BAIRE CATEGORY

In this section, we let $(\mathcal{X}, d)$ be a Polish metric space and let $c : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be a continuous cost function satisfying the growth condition $|c(x, y)| \leq C(1 + d(x, x_0)^p + d(y, y_0)^p)$ for some $C > 0$, $p \geq 1$, and $x_0, y_0 \in \mathcal{X}$. We write $\text{Cpl}_{\text{opt}}^c(\mu, \nu)$ for the set of couplings that are optimal for the transport problem between μ and ν with cost c , that is, for $\pi' \in \text{Cpl}_{\text{opt}}^c(\mu, \nu)$ we have

$$\int c(x, y) d\pi'(x, y) = \inf_{\pi \in \text{Cpl}(\mu, \nu)} \int c(x, y) \pi(dx, dy).$$

A measure μ is then called c -transport-regular if for every $\nu \in \mathcal{P}_p(\mathcal{X})$ there is a unique optimal transport plan from μ to ν and this plan is of Monge type. Recall that the c -transform of a Borel measurable function $\varphi : \mathcal{X} \rightarrow \mathbb{R} \cup \{-\infty\}$, $\varphi \neq -\infty$, is given by

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

which is thus an upper semicontinuous function, and φ is called c -concave if $\varphi \neq -\infty$ and

$$\varphi(x) = \inf_{y \in \mathcal{X}} c(x, y) - \varphi^c(y).$$

In this case, the c -superdifferential of φ is given by

$$\partial^c \varphi = \{(x, y) \in \mathcal{X} \times \mathcal{X} : \varphi(x) + \varphi^c(y) = c(x, y)\},$$

where we also write $\partial^c \varphi(x) = \{y \in \mathcal{X} : (x, y) \in \partial^c \varphi\}$ for $x \in \mathcal{X}$.

Lemma A.1. *Let $\mu \in \mathcal{P}_p(\mathcal{X})$. The following are equivalent:*

- (1) μ is c -transport regular;
- (2) For every c -concave φ , we have $\mu(\{x \in \mathcal{X} : \#\partial^c \varphi(x) > 1\}) = 0$;
- (3) For every proper φ such that there is a compact $K \subseteq \mathcal{X}$ with $\varphi(x) = \inf_{y \in K} c(x, y) - \varphi^c(y)$, we have $\mu(\{x \in \mathcal{X} : \#\partial^c \varphi(x) > 1\}) = 0$.

Proof. Clearly, (2) implies (3).

Assume that (2) holds and our goal is to show (1). By the fundamental theorem of optimal transport there is for each $\nu \in \mathcal{P}_p(\mathcal{X})$ an optimal c -concave potential φ such that

$$\pi(\partial^c \varphi) = 1, \quad \pi \in \text{Cpl}_{\text{opt}}^c(\mu, \nu).$$

Since $\partial^c \varphi(x)$ contains at most one point for μ -a.e. x , it readily follows that π is of Monge type. Consequently, as $\text{Cpl}_{\text{opt}}^c(\mu, \nu)$ is a non-empty convex set, it has to be a singleton.

Now, we want to show that (1) and (3) each imply (2). To this end, note that

$$\{(x, y_1, y_2) \in \mathcal{X}^3 : y_1, y_2 \in \partial^c \varphi(x), y_1 \neq y_2\}$$

is Borel, whenever φ is c -concave. Hence, by the Jankov–von Neumann uniformization theorem there exists an analytically measurable selection $h = (h_1, h_2) : A \rightarrow \mathcal{X} \times \mathcal{X}$, and observe that $h_1(x) \neq h_2(x)$ for every $x \in A$ by construction. With these maps at hand, we define subprobability measures $\nu_i := (h_i)_\# \mu(\cdot \cap A)$ for $i = 1, 2$. By inner regularity, there exists a compact set $K \subseteq \mathcal{X}$ such that $\nu_i(K) > \mu(A)/2$. Define the potential $\varphi_K := (\varphi^c - \chi_K)^c$, which satisfies

$$\varphi_K \geq \varphi \text{ and } \varphi_K(x) + \varphi_K^c(y) = c(x, y) \text{ for all } (x, y) \in \partial^c \varphi \cap (\mathcal{X} \times K),$$

and, as K is compact, $\partial^c \varphi_K(x) \neq \emptyset$ for all $x \in \mathcal{X}$. Let g be an analytically measurable selector of $\partial^c \varphi_K$, and set

$$g_i(x) := \begin{cases} h_i(x) & x \in A \cap h_1^{-1}(K) \cap h_2^{-1}(K), \\ g(x) & \text{otherwise.} \end{cases}$$

Then we have that

$$\pi := \frac{1}{2}((x \mapsto (x, g_1(x)))_\# \mu + (x \mapsto (x, g_2(x)))_\# \mu)$$

has first marginal μ , its second marginal $\nu \in \mathcal{P}_p(\mathcal{X})$ is concentrated on K , and $\pi(\partial^c \varphi_K) = 1$. Since $\nu_i(K) > \mu(A)/2$, we have $\mu(A \cap h_1^{-1}(K) \cap h_2^{-1}(K)) \geq \mu(A) - \nu_1(K^c) - \nu_2(K^c) > 0$. Hence, we have constructed an MC-concave function φ_K with $\varphi_K(x) = \inf_{y \in \mathcal{X}} c(x, y) - \varphi_K^c(y)$ as well as a coupling $\pi \in \text{Cpl}_{\text{opt}}^c(\mu, \nu)$, which is not Monge. This leads to a contradiction to (1) as well as (3), and we conclude that (1) and (3) each imply (2). $\square$

We introduce functionals which quantify how far a pair of measures (μ, ν) is from admitting a unique optimal transport map between them. Based on this, we will be able to quantify how far a measure is from being c -transport-regular. We introduce the quantities

$$\tau^c(\mu, \nu) := \sup_{\pi \in \text{Cpl}_{\text{opt}}^c(\mu, \nu)} \int \text{Var}(\pi_x) \mu(dx), \quad \text{and} \quad \tau_R^c(\mu) := \sup_{\nu: \mathcal{W}_p(\nu, \delta_{x_0}) \leq R} \tau(\mu, \nu),$$

where the variance for probabilities in $\nu \in \mathcal{P}_p(\mathcal{X})$ is defined as

$$\widehat{\text{Var}}(\nu) := \sum_{n \in \mathbb{N}} 2^{-n} \text{Var}((f_n)_\# \nu),$$

where $(f_n)_{n \in \mathbb{N}}$ is a family of continuous, point-separating functions on $\mathcal{X}$ bounded by 1, and

$$\text{Var}(\rho) = \int |x - x'|^2 \rho \otimes \rho(dx, dx'),$$

for $\rho \in \mathcal{P}_2(\mathbb{R})$.

Lemma A.2. *For $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$ we have:*

- (1) *There is a unique $\pi \in \text{Cpl}_{\text{opt}}^c(\mu, \nu)$ and π is Monge iff $\tau^c(\mu, \nu) = 0$;*
- (2) *μ is c -transport-regular iff $\tau_R^c(\mu) = 0$ for all $R > 0$.*

Proof. To see (1), first note that $\widehat{\text{Var}}(\nu) > 0$ if and only if ν is not a Dirac measure. Indeed, if ν is not a Dirac measure then, as $(f_n)_{n \in \mathbb{N}}$ is point-separating, there exists $n \in \mathbb{N}$, such that $(f_n)_\# \nu$ is not a Dirac measure, which yields that

$$\widehat{\text{Var}}(\nu) \geq 2^{-n} \text{Var}((f_n)_\# \nu) > 0.$$

Hence, $\tau^c(\mu, \nu) = 0$ iff for every $\pi \in \text{Cpl}_{\text{opt}}^c(P, Q)$ we have that π_x is μ -almost surely a Dirac. Since $\text{Cpl}_{\text{opt}}^c(\mu, \nu)$ is a convex set, this yields that $\text{Cpl}_{\text{opt}}^c(\mu, \nu)$ contains a single transport plan and this plan is Monge.

To see (2), note that by the first part we have for every $\nu \in \mathcal{P}_p(\mathcal{X})$ a unique optimal transport plan $\pi \in \text{Cpl}_{\text{opt}}^c(\mu, \nu)$, which means that μ is c -transport-regular. $\square$

Lemma A.3. *The map $\mathcal{P}_p(\mathcal{X}) \times \mathcal{P}_p(\mathcal{X}) \rightarrow \mathbb{R} : (\mu, \nu) \mapsto \tau^c(\mu, \nu)$ is $\mathcal{W}_p$ -upper semi-continuous.*

Proof. We first show that

$$\widehat{\text{Var}} : \mathcal{P}_p(\mathcal{X}) \rightarrow \mathbb{R},$$

is bounded, continuous and concave. Clearly, Var is continuous and concave on $\mathcal{P}_2(\mathbb{R})$ as $\text{Var}(\rho) = \int x^2 \rho(dx) - (\int x \rho(dx))^2$. Since $(f_n)_{n \in \mathbb{N}}$ is a point-separating family of continuous functions on $\mathcal{X}$ bounded by 1, we have that $\widehat{\text{Var}}$ is the sum of concave and continuous functions on $\mathcal{P}_p(\mathcal{X})$ bounded by 1, from where we infer the claimed properties of $\widehat{\text{Var}}$.

Next, we invoke [13, Proposition 2.8] to find that the map

$$\mathcal{P}_p(\mathcal{X} \times \mathcal{X}) \rightarrow \mathbb{R} : \pi \mapsto \int \widehat{\text{Var}}(\pi_x) \pi(dx),$$

is upper semi-continuous, concave and bounded.

Finally, consider sequences $(\mu^n)_n$ and $(\nu^n)_n$ and measures $\mu^\infty, \nu^\infty \in \mathcal{P}_p(\mathcal{X})$ with $\mathcal{W}_p(\mu^n, \mu) \rightarrow 0$ and $\mathcal{W}_p(\nu^n, \nu) \rightarrow 0$. Let $\pi^n \in \text{Cpl}_{\text{opt}}^c(\mu^n, \nu^n)$ be such that

$$\int \text{Var}(\pi_x^n) \pi^n(dx) \geq \tau(\mu^n, \nu^n) - 1/n.$$

As $\{\mu_n : n \in \mathbb{N} \cup \{\infty\}\}$ and $\{\nu_n : n \in \mathbb{N} \cup \{\infty\}\}$ are $\mathcal{W}_p$ -compact, we have that

$$A := \bigcup_{n \in \mathbb{N} \cup \{\infty\}} \text{Cpl}(\mu^n, \nu^n)$$

is $\mathcal{W}_p$ -compact as immediate consequence of Prokhorov's theorem and the characterization of $\mathcal{W}_p$ -convergence, see [75, Definition 6.8]. Therefore, there is $\pi \in A$ such that, after possibly passing to a subsequence, $W_p(\pi^n, \pi) \rightarrow 0$. By continuity of the map which assigns a couplings its marginals, we have $\pi \in \text{Cpl}(\mu, \nu)$. As the coupling π^n is optimal between its marginals, the same holds true for the limiting π , i.e., $\pi \in \text{Cpl}_{\text{opt}}^c(\mu, \nu)$. This is because optimality is characterized in terms of c -cyclical monotonicity of the support and the latter is preserved under weak limits (see e.g. [75, Theorem 5.20]).

All in all, we have

$$\tau^c(\mu, \nu) \geq \int \widehat{\text{Var}}(\pi_x) \pi(dx) \geq \limsup_n \int \widehat{\text{Var}}(\pi_x^n) \pi^n(dx) = \limsup_n \tau^c(\mu^n, \nu^n). \quad \square$$

Lemma A.4. *Assume that $\mathcal{X}$ has the property that $\mathcal{W}_p$ -bounded sets in $\mathcal{P}_p(\mathcal{X})$ are tight. Then, for $R \geq 0$, the map $\mathcal{P}_p(\mathcal{X}) \rightarrow \mathbb{R} : \mu \mapsto \tau_R^c(\mu)$ is $\mathcal{W}_p$ -upper semi-continuous.*

Proof. The proof is line by line as for Lemma A.3 subject to replacing $\bigcup_{n \in \mathbb{N} \cup \{\infty\}} \text{Cpl}(\mu^n, \nu^n)$ with

$$A := \bigcup_{n \in \mathbb{N} \cup \{\infty\}} \bigcup_{\nu : \mathcal{W}_p(\nu, \delta_{x_0}) \leq R} \text{Cpl}(\mu_n, \nu).$$

While A is not $\mathcal{W}_p$ -compact, it is tight by our assumption, as the marginals form a $\mathcal{W}_p$ -bounded set. Furthermore, A is weakly closed since any limit point π of a sequence $(\pi^n)_n$ in A and p -th moment of the second marginals bounded by R , has second marginal with p -th moment bounded by R . This follows directly from weak lower semi-continuity of the p -th moment w.r.t. weak convergence.

We can proceed the proof as in Lemma A.3 and obtain that π is in $\text{Cpl}_{\text{opt}}^c(\mu, \cdot)$ by [75, Theorem 5.20], and thereby analogously conclude $\mathcal{W}_p$ -upper semi-continuity of τ_R^c . $\square$

Theorem A.5. *Assume that $\mathcal{X}$ has the property that $\mathcal{W}_p$ -bounded sets in $\mathcal{P}_p(\mathcal{X})$ are tight. Then the set of c -transport-regular measures is G_δ in $\mathcal{P}_p(\mathcal{X})$.*

Proof. By Lemma A.2, the set of c -transport-regular measures is precisely given by

$$\bigcap_{R \in \mathbb{N}} \bigcap_{k \in \mathbb{N}} \{\mu \in \mathcal{P}_p(\mathcal{X}) : \tau_R^c(\mu) < 1/k\}.$$

As τ_R^c is upper semicontinuous by Lemma A.4, the sets $\{\mu \in \mathcal{P}_p(\mathcal{X}) : \tau_R^c(\mu) < 1/k\}$ are open. Therefore, the set of c -transport-regular measures is a G_δ -subset of $\mathcal{P}_p(\mathcal{X})$. $\square$

APPENDIX B. AUXILIARY RESULTS

In this section $U : [0, 1] \rightarrow \mathbb{R}$ defines the identity and $U = (U_t)_{t=1}^N : [0, 1]^N \rightarrow \mathbb{R}^N$ defines the coordinate process. We equip $[0, 1]$ and $[0, 1]^N$ with the Lebesgue measure λ .

Lemma B.1. *Let $\mathcal{X}, \mathcal{Y}$ be Polish and $f : \mathcal{X} \rightarrow \mathcal{Y}$ measurable. Then, there is a Borel map*

$$T : \mathcal{P}(\mathcal{X}) \times (0, 1) \rightarrow \mathcal{X}$$

such that $T(\mu, U) \sim \mu$ and, for all $\mu, \nu \in \mathcal{P}(\mathcal{X})$ with $f_\# \mu = f_\# \nu$, $f(T(\mu, U)) = f(T(\nu, U))$ a.s.

Proof. Using a Borel isomorphism we assume that $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. By [56, Lemma B.1] there exists a jointly measurable version of disintegration of measure, i.e., there is a Borel map $S : \mathcal{P}(\mathcal{X}) \times \mathcal{Y} \rightarrow \mathcal{P}(\mathcal{X})$ such that for $\mu \in \mathcal{P}(\mathcal{X})$

$$\begin{aligned} \mu(dx) &= \int S(\mu, y; dx) f_\# \mu(dy), \\ S(\mu, y; f^{-1}(\{y\})) &= 1 \quad \text{for } f_\# \mu\text{-a.e. } y. \end{aligned}$$

Further, let $R = (R_1, R_2) : (0, 1) \rightarrow (0, 1)^2$ be Borel measurable with $R_{\#}\lambda = \lambda^2$. Writing Q_ν for the quantile function of $\nu \in \mathcal{P}(\mathbb{R})$, we define

$$T(\mu, u) := Q_{S(\mu, Q_{f_{\#}\mu}(R_1(u)))}(R_2(u)),$$

for $(\mu, u) \in \mathcal{P}(\mathcal{X}) \times (0, 1)$. It follows directly from inverse transform sampling that $T(\mu, U) \sim \mu$. Next, let $\mu, \nu \in \mathcal{P}(\mathcal{X})$ with $f_{\#}\mu = f_{\#}\nu =: \eta$, then we have

$$S(\mu, y; f^{-1}(\{y\})) = 1 = S(\nu, y; f^{-1}(\{y\})) \quad \text{for } \eta\text{-a.e. } y.$$

We deduce that for η -a.e. y

$$f(Q_{S(\mu, y)}(U)) = y = f(Q_{S(\nu, y)}(U)) \quad \text{a.s.}$$

We conclude that

$$f(T(\mu, U)) = Q_\eta(R_1(U)) = f(T(\nu, U)) \quad \text{a.s.} \quad \square$$

Lemma B.2. *Let X, Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t=1}^N)$. Then, there exist random variables $X', Y' : (0, 1)^N \rightarrow H$ on $((0, 1)^N, \mathcal{B}((0, 1)^N), \lambda^N, (\sigma(U_{1:t}))_{t=1}^N)$ such that*

- for $t = 1, \dots, N-1$, X' has a $(\mathcal{L}^{\mathcal{F}_{N-1:t}}(X'), U_{t+1:N})$ -measurable version,
- $(X', Y') \sim_{\text{ad}} (X, Y)$.

Proof. We show the claim by induction. More specifically, we show the following:

There exists a measurable map $S_N = (X_N, Y_N) : \mathcal{P}^N(H \times H) \times [0, 1]^N \rightarrow H \times H$ with

- for $P, Q \in \mathcal{P}^N(H \times H)$, $u, v \in (0, 1)^N$ and $t = 0, \dots, N-1$ with $X_N(P, u_{1:t}, U_{t+1:N}) \sim_{\text{ad}} X_N(Q, v_{1:t}, U_{t+1:N})$, then $X_N(P, u_{1:t}, U_{t+1:N}) = X_N(Q, v_{1:t}, U_{t+1:N})$ a.s.,
- for $P \in \mathcal{P}^N(H \times H)$, $(X_N(P, U_{1:N}), Y_N(P, U_{1:N})) \sim_{\text{ad}} P$.

First, let $N = 1$. Then the claim follows from Lemma B.1 with $\mathcal{X} = H \times H$, $\mathcal{Y} = H$ and $f(x_1, x_2) = x_1$.

Next, assume the claim holds true for $N-1$. We invoke Lemma B.1 with $\mathcal{X} = \mathcal{P}^{N-1}(H \times H)$, $\mathcal{Y} = \mathcal{P}^{N-1}(H)$ and $f = \mathcal{P}^{N-1}[\text{pr}_1]$ to find a Borel map $T : \mathcal{P}^N(H \times H) \times (0, 1) \rightarrow \mathcal{P}^{N-1}(H \times H)$ with $T(P, U_1) \sim P$ and, for $P, Q \in \mathcal{P}^N(H \times H)$ with $f_{\#}P = f_{\#}Q$, then $f(T(P, U_1)) = f(T(Q, U_1))$ a.s. We set

$$X_N(P, u_{1:N}) := X_{N-1}(T(P, u_1), u_{2:N}) \text{ and } Y_N(P, u_{1:N}) := Y_{N-1}(T(P, u_1), u_{2:N}),$$

which defines a measurable map $S_N = (X_N, Y_N) : \mathcal{P}^N(H \times H) \times (0, 1)^N \rightarrow H \times H$. By construction we have for fixed $(P, u_1) \in \mathcal{P}^N(H \times H) \times (0, 1)$

$$\begin{aligned} S_N(P, u_1, U_{2:N}) &= (X_N(P, u_1, U_{2:N}), Y_N(P, u_1, U_{2:N})) \\ &= (X_{N-1}(T(P, u_1), U_{2:N}), Y_{N-1}(T(P, u_1), U_{2:N})) \sim_{\text{ad}} T(P, u_1). \end{aligned}$$

Hence, $S_N(P, U_{1:N}) \sim_{\text{ad}} (T(P, \cdot))_{\#}\lambda = P$ which also yields $X_N(P, U_{1:N}) \sim_{\text{ad}} f_{\#}P$ since $f = \mathcal{P}^{N-1}[\text{pr}_1]$. Therefore, if $X_N(P, U_{1:N}) \sim_{\text{ad}} X_N(Q, U_{1:N})$ then $f_{\#}P = f_{\#}Q$ and thus $f(T(P, U_1)) = f(T(Q, U_1))$ almost surely. Consequently, we get

$$X_N(P, U_{1:N}) = X_{N-1}(T(P, U_1), U_{2:N}) = X_{N-1}(T(Q, U_1), U_{2:N}) = X_N(Q, U_{1:N}) \quad \text{a.s.}$$

If additionally $u, v \in (0, 1)^N$ are such that $X_N(P, u_{1:t}, U_{t+1:N}) \sim_{\text{ad}} X_N(Q, v_{1:t}, U_{t+1:N})$, then we also have that $X_{N-1}(T(P, u_1), u_{2:t}, U_{t+1:N}) \sim_{\text{ad}} X_{N-1}(T(Q, v_1), v_{2:t}, U_{t+1:N})$. Using the inductive assumption we obtain that a.s.

$$\begin{aligned} X_N(P, u_{1:t}, U_{t+1:N}) &= X_{N-1}(T(P, u_1), u_{2:t}, U_{t+1:N}) \\ &= X_{N-1}(T(Q, v_1), v_{2:t}, U_{t+1:N}) = X_N(Q, v_{1:t}, U_{1:N}). \end{aligned}$$

This completes the inductive step.

Finally, we let $(X', Y') := S_N(P, U_{1:N})$. We note that for $t = 1, \dots, N-1$

$$X' = \mathbb{E}[X' | \mathcal{L}^{\mathcal{F}_{N-1:t}}(X'), U_{t+1:N}] \quad \text{a.s.},$$

hence, (X', Y') have the desired properties. $\square$

Proof of Lemma 5.5. First note that

$$\begin{aligned} c^{(N)}(P, P') &= \inf_{\Pi \in \text{Cpl}^N(P, P')} \int c(x, x') dI^{N-1}\Pi(x, x') \\ &\leq \inf_{Y \sim_{\text{ad}} P} \mathbb{E}[c(Y, X')] \leq \inf_{\substack{T: \Omega \rightarrow \Omega \\ \text{isomorphism}}} \mathbb{E}[c(X \circ T, X')], \end{aligned}$$

the equality follows as in Proposition 4.7, the first inequality by Lemma 5.3 and last inequality follows from Lemma 5.1.

We write $U = (U_t)_{t=1}^N$ for the coordinate process on $[0, 1]^N$. The distributions

$$\mu := \mathcal{L}(U_1, \dots, U_{N-1}, (U_N, X)) \text{ and } \nu := (U_1, \dots, U_{N-1}, (U_N, X'))$$

have atomless disintegrations, that is, $\lambda = \mathcal{L}(U_1) = \mathcal{L}(U_{t+1}|\mathcal{F}_t)$ as well as

$$\begin{aligned} \mu_{u_{1:N-1}} &:= \mathcal{L}(U_N, X | U_{1:N-1} = u_{1:N-1}), \\ \nu_{u_{1:N-1}} &:= \mathcal{L}(U_N, X' | U_{1:N-1} = u_{1:N-1}), \end{aligned}$$

are λ^N -almost surely atomless and thus w.l.o.g. atomless. In what follows we view μ and ν as laws of the N -time step stochastic processes $(U_1, \dots, U_{N-1}, (U_N, X))$ and $(U_1, \dots, U_{N-1}, (U_N, X'))$, respectively, and use the denseness of biadapted Monge couplings in the set of bicausal couplings to show the remaining inequality. Therefore we have to relate the adapted optimal transport problem between μ and ν with $c^{(N)}(P, P')$. To this end, we define

$$\bar{\mu}_{U_{1:t}} := \mathcal{L}^{\mathcal{F}_{N-1:t}}(X) = \mathcal{L}^{\mathcal{F}_{N-2:t}}(\mu_{U_{1:N-1}}) \text{ and } \bar{\nu}_{U_{1:t}} := \mathcal{L}^{\mathcal{F}_{N-1:t}}(Y) = \mathcal{L}^{\mathcal{F}_{N-2:t}}(\nu_{U_{1:N-1}}).$$

Due to the dynamic programming principle of optimal transport under bicausality constraints

$$\inf_{\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)} \int c(x, x') d\pi((u_1, \dots, u_{N-1}, (u_N, x)), (v_1, \dots, v_{N-1}, (v_N, x'))) = V^{(0)},$$

see, for example, [2, Theorem 3.2], where $\text{Cpl}_{\text{bc}}(\mu, \nu)$ denotes the set of bicausal couplings, see [2, Definition 2.4], and $V^{(0)}$ is inductively given by

$$\begin{aligned} V^{(N-1)}(u_{1:N-1}, v_{1:N-1}) &:= \inf_{\pi \in \text{Cpl}(\mu_{u_{1:N-1}}, \nu_{v_{1:N-1}})} \int c(x, x') d\pi((u_N, x), (v_N, x')), \\ V^{(t)}(u_{1:t}, v_{1:t}) &:= \inf_{\pi \in \text{Cpl}(\lambda, \lambda)} \int V_{t+1} d\pi, \quad V^{(0)} := \inf_{\pi \in \text{Cpl}(\lambda, \lambda)} \int V_1 d\pi. \end{aligned}$$

By comparing the definition of $c^{(t)}$ and $V^{(t)}$, we have for all u, v in a λ -full set

$$V^{(N-1)}(u_{1:N-1}, v_{1:N-1}) = c^{(N-1)}(\bar{\mu}_{u_{1:N-1}}, \bar{\nu}_{v_{1:N-1}}), \quad V^{(t)}(u_{1:t}, u'_{1:t}) = c^{(t)}(\bar{\mu}_{u_{1:t}}, \bar{\nu}_{u'_{1:t}}),$$

and thus $V^{(0)} = c^{(N)}(P, P')$. By [20], bicausal transport plans given by biadapted Monge maps are dense in $\text{Cpl}_{\text{bc}}(\mu, \nu)$, since μ and ν both have atomless successive disintegration as argued above. Let $T = (T^1, \dots, T^N)$ be a biadapted Monge map from μ to ν , that is, $T_{\#}\mu = \nu$ and

$$T^t : [0, 1]^t \rightarrow [0, 1], \quad T^t(u_{1:t-1}, \cdot)_{\#}\lambda = \lambda,$$

$$T^N = (R^N, S^N) : [0, 1]^N \times \mathcal{X} \rightarrow [0, 1]^N \times \mathcal{X}, \quad T^N(u_{1:N-1}, \cdot)_{\#}\mu_{u_{1:N-1}} = \nu_{u_{1:N-1}},$$

and $(T^1(u_1), \dots, T^t(u_{1:t})) : [0, 1]^t \rightarrow [0, 1]^t$ is bijective, for $t = 1, \dots, N-1$. In particular, $R := (T^1, \dots, T^{N-1}, R^N)$ is a biadapted Monge map from λ to itself, and $S^N = Y \circ R$ λ -almost surely. Consequently, we obtain

$$c^{(N)}(P, P') = V^{(0)} \geq \inf_{\substack{R: [0, 1]^N \rightarrow [0, 1]^N, \\ \text{bi-adapted}, R_{\#}\lambda = \lambda}} \int c(X(u), X'(R(u))) d\lambda(u),$$

which provides the missing inequality and shows the first claim. $\square$

REFERENCES

- [1] B. Acciaio, A. Kratsios, and G. Pammer. Designing universal causal deep learning models: The geometric (hyper) transformer. *Mathematical Finance*, 34(2):671–735, 2024.
- [2] B. Acciaio, D. Kršek, and G. Pammer. Multicausal transport: barycenters and dynamic matching. *arXiv preprint arXiv:2401.12748*, 2024.
- [3] B. Acciaio, D. Kršek, G. Pammer, and M. Rodrigues. Absolutely continuous curves of stochastic processes, 2025.
- [4] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.*, 43(2):904–924, 2011.
- [5] D. J. Aldous. Weak convergence and general theory of processes. Unpublished monograph: Department of Statistics, University of California, Berkeley, 1981.
- [6] A. Alfonsi and B. Jourdain. Squared quadratic Wasserstein distance: optimal couplings and Lions differentiability. *ESAIM: Probability and Statistics*, 24:703–717, 2020.
- [7] L. Ambrosio and N. Gigli. A user’s guide to optimal transport. In *Modelling and optimisation of flows on networks*, volume 2062 of *Lecture Notes in Math.*, pages 1–155. Springer, Heidelberg, 2013.
- [8] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
- [9] L. Ambrosio, B. Kirchheim, and A. Pratelli. Existence of optimal transport maps for crystalline norms. *Duke Mathematical Journal*, 125(2):207–241, Nov. 2004.
- [10] A. Ayache, D. Wu, and Y. Xiao. Joint continuity of the local times of fractional brownian sheets. In *Annales de l’IHP Probabilités et statistiques*, volume 44, pages 727–748, 2008.
- [11] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. All adapted topologies are equal. *Probab. Theory Relat. Fields*, 178(3-4):1125–1172, 2020.
- [12] J. Backhoff-Veraguas, M. Beiglböck, M. Eder, and A. Pichler. Fundamental properties of process distances. *Stoch. Proc. Appl.*, 130(9):5575–5591, 2020.
- [13] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations*, 58(6):1–28, 2019.
- [14] D. Bartl, M. Beiglböck, and G. Pammer. The Wasserstein space of stochastic processes. *J. Eur. Math. Soc.*, 2024. To appear. [arXiv:2104.14245](https://arxiv.org/abs/2104.14245).
- [15] D. Bartl and J. Wiesel. Sensitivity of multiperiod optimization problems with respect to the adapted wasserstein distance. *SIAM J. Financ. Math.*, 14(2):704–720, 2023.
- [16] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, second edition, 2017.
- [17] M. Beiglböck, A. Cox, and M. Huesmann. Optimal transport and Skorokhod embedding. *Invent. Math.*, 208(2):327–400, 2017.
- [18] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, 44(1):42–106, 2016.
- [19] M. Beiglböck, G. Pammer, and A. Posch. The Knothe–Rosenblatt distance and its induced topology. *arXiv:2312.16515*, 2023.
- [20] M. Beiglböck, G. Pammer, and S. Schrott. Denseness of biadapted Monge mappings. *Ann. Inst. H. Poincaré Probab. Statist.*, Oct. 2024.
- [21] M. Beiglböck, G. Pammer, and S. Schrott. A brenier theorem on $(\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d)), w_2)$ and applications to adapted transport, 2025.
- [22] M. Beiglböck, G. Pammer, S. Schrott, and X. Zhang. Representing general stochastic processes as martingale laws. *ArXiv e-prints*, 2023.
- [23] M. Beiglböck, S. Pflügl, and S. Schrott. A probabilistic view on the adapted wasserstein distance. *arXiv e-prints*, 2024.
- [24] J. Bion-Nadal and D. Talay. On a Wasserstein-type distance between solutions to stochastic differential equations. *Ann. Appl. Probab.*, 29(3):1609–1639, 2019.
- [25] J. Borwein and J. Vanderwerff. *Convex Functions: Constructions, Characterizations and Counterexamples*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2010.
- [26] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.*, 44(4):375–417, 1991.
- [27] P. Cardaliaguet. Notes on mean field games (from p.-l. lions’ lectures at collège de france). <https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf>, 2013. Version of September 27, 2013.
- [28] G. Carlier. Remarks on Toland’s duality, convexity constraint and optimal transport. *Pac. J. Optim.*, 4(3):423–432, 2008.

- [29] G. Carlier, A. Galichon, and F. Santambrogio. From Knothe’s transport to Brenier’s map and a continuation method for optimal transport. *SIAM J. Math. Anal.*, 41(6):2554–2576, 2009/10.
- [30] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I*. Springer, 2018.
- [31] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications II: Mean Field Games with Common Noise and Master Equations*, volume 85 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2018.
- [32] G. Cavignari, G. Savaré, and G. E. Sodini. A lagrangian approach to totally dissipative evolutions in wasserstein spaces. 2023.
- [33] T. Champion and L. De Pascale. The monge problem for strictly convex norms in $\{R\}^d$. *Journal of the European Mathematical Society*, 012(6):1355–1369, 2010.
- [34] T. Champion and L. D. Pascale. The monge problem in $\mathbb{R}^d$. *Duke Mathematical Journal*, 157(3):551–572, Apr. 2011.
- [35] R. Cont and F. R. Lim. Causal transport on path space. *arxiv:2412.02948*, 2024.
- [36] C. Dellacherie and P.-A. Meyer. *Probabilities and potential. B*, volume 72 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1982. Theory of martingales, Translated from the French by J. P. Wilson.
- [37] L. Dello Schiavo. A rademacher-type theorem on l_2 -wasserstein spaces over closed riemannian manifolds. *Journal of Functional Analysis*, 278(6):108397, 2020.
- [38] M. Eder. Compactness in adapted weak topologies. *arXiv:1905.00856*, 2019.
- [39] P. Emami and B. Pass. Optimal transport with optimal transport cost: the monge–kantorovich problem on wasserstein space. *Calculus of Variations and Partial Differential Equations*, 64(2):11pp, 2025.
- [40] D. Feyel and A. Üstünel. Measure transport on Wiener space and the Girsanov theorem. *C. R. Math. Acad. Sci. Paris*, 334(11):1025–1028, 2002.
- [41] D. Feyel and A. Üstünel. Monge-Kantorovich measure transportation and Monge-Ampère equation on Wiener space. *Probab. Theory Related Fields*, 128(3):347–385, 2004.
- [42] A. Figalli and F. Glaudo. *An invitation to optimal transport, Wasserstein distances, and gradient flows*. EMS Textbooks in Mathematics. EMS Press, Berlin, 2021.
- [43] W. Gangbo and R. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.
- [44] W. Gangbo and A. Świąch. Optimal maps for the multidimensional monge–kantorovich problem. *Communications on Pure and Applied Mathematics*, 51(1):23–45, 1998.
- [45] N. Ghoussoub, Y.-H. Kim, and T. Lim. Structure of optimal martingale transport plans in general dimensions. *arXiv:1508.01806*, 2016.
- [46] N. Gigli. On the inverse implication of brenier–mccann theorems and the structure of $(p_2(m), w_2)$. *Methods and Applications of Analysis*, 18(2):127–158, June 2011.
- [47] N. Gozlan and N. Juillet. On a mixture of Brenier and Strassen theorems. *Proceedings of the London Mathematical Society*, 120(3):434–463, 2020.
- [48] M. F. Hellwig. Sequential decisions under uncertainty and the maximum theorem. *J. Math. Econom.*, 25(4):443–464, 1996.
- [49] M. F. Hellwig and K. M. Schmidt. Discrete–time approximations of the holmström–milgrom brownian–motion model of intertemporal incentive provision. *Econometrica*, 70(6):2225–2264, 2002.
- [50] P. Henry-Labordère and N. Touzi. An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.*, 20(3):635–668, 2016.
- [51] D. N. Hoover and H. J. Keisler. Adapted probability distributions. *Transactions of the American Mathematical Society*, 286(1):159–201, 1984.
- [52] Y. Jiang and J. Obloj. Sensitivity of causal distributionally robust optimization. 2024.
- [53] B. Jourdain and G. Pammer. An extension of martingale transport and stability in robust finance. 2023.
- [54] A. S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.
- [55] M. Knott and C. S. Smith. On the optimal mapping of distributions. *J. Optim. Theory Appl.*, 43(1):39–49, 1984.
- [56] D. Lacker. On the convergence of closed-loop nash equilibria to the mean field game limit. *Annals of Applied Probability*, 30(4):1693–1761, 2020. Preprint version *arXiv:1808.02745*.
- [57] R. Lassalle. Causal transference plans and their Monge–Kantorovich problems. *Stoch. Anal. Appl.*, 36(3):452–484, 2018.
- [58] P.-L. Lions. Cours au collège de france: Théorie des jeux à champ moyen (mean field games). <https://www.college-de-france.fr/fr/agenda/cours/jeux-champ-moyen-suite-1>, 2009–2013. Lecture series where the Lions derivative (“Lions lift”) was introduced; videos and summaries available on the Collège de France site.
- [59] S. Lisini. Characterization of absolutely continuous curves in Wasserstein spaces. *Calc. Var. Partial Differential Equations*, 28(1):85–120, 2007.

- [60] R. McCann. A convexity theory for interacting gases and equilibrium crystals. *PhD thesis, Princeton University*, 1994.
- [61] R. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.
- [62] R. J. McCann. Polar factorization of maps on riemannian manifolds. *Geometric & Functional Analysis GAFA*, 11(3):589–608, 2001.
- [63] P. A. Meyer and W. A. Zheng. Tightness criteria for laws of semimartingales. *Annales de l’I.H.P. Probabilités et statistiques*, 20(4):353–372, 1984.
- [64] F. Nielsen and K. Sun. *Chain Rule Optimal Transport*, pages 191–217. Springer International Publishing, Cham, 2021.
- [65] G. Pammer. A note on the adapted weak topology in discrete time. *Electron. Commun. Probab.*, 29:1 – 13, 2024.
- [66] G. Parker. Some convexity criteria for differentiable functions on the 2-wasserstein space. *Bulletin of the London Mathematical Society*, 56(5):1839–1858, 2024.
- [67] G. C. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM J. Optim.*, 22(1):1–23, 2012.
- [68] G. C. Pflug and A. Pichler. *Multistage Stochastic Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2014.
- [69] A. Pinzi. First order equation on random measures as superposition of weak solutions to the mckean-vlasov equation, 2025.
- [70] A. Pinzi and G. Savaré. Nested superposition principle for random measures and the geometry of the wasserstein on wasserstein space, 2025.
- [71] A. Pinzi and G. Savaré. Totally convex functions, l^2 -optimal transport for laws of random measures, and solution to the Monge problem. *arXiv preprint arXiv:2509.01768*, Sept. 2025. Available at *arXiv*. Version v1, submitted 1 Sep 2025.
- [72] S. T. Rachev and L. Rüschendorf. A characterization of random variables with minimum L^2 -distance. *J. Multivariate Anal.*, 32(1):48–54, 1990.
- [73] F. Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Cham, 2015. Calculus of variations, PDEs, and modeling.
- [74] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [75] C. Villani. *Optimal Transport, Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.
- [76] J. Wiesel and E. Zhang. A characterisation of convex order using the 2-Wasserstein distance. *arXiv preprint*, 2022.
- [77] T. Xu and B. Acciaio. Conditional COT-GAN for video prediction with kernel smoothing. In *NeurIPS 2022 Workshop on Robustness in Sequence Modeling*, 2022.
- [78] L. Zajíček. On the differentiation of convex functions in finite and infinite dimensional spaces. *Czechoslovak Math. J.*, 29(104)(3):340–348, 1979.