

Wasserstein metric and subordination

by

PHILIPPE CLÉMENT (Leiden) and WOLFGANG DESCH (Graz)

Abstract. Let (X, d_X) , (Ω, d_Ω) be complete separable metric spaces. Denote by $\mathcal{P}(X)$ the space of probability measures on X , by W_p the p -Wasserstein metric with some $p \in [1, \infty)$, and by $\mathcal{P}_p(X)$ the space of probability measures on X with finite Wasserstein distance from any point measure. Let $f : \Omega \rightarrow \mathcal{P}_p(X)$, $\omega \mapsto f_\omega$, be a Borel map such that f is a contraction from (Ω, d_Ω) into $(\mathcal{P}_p(X), W_p)$. Let ν_1, ν_2 be probability measures on Ω with $W_p(\nu_1, \nu_2)$ finite. On X we consider the subordinated measures $\mu_i = \int_\Omega f_\omega d\nu_i(\omega)$. Then $W_p(\mu_1, \mu_2) \leq W_p(\nu_1, \nu_2)$. As an application we show that the solution measures $\varrho_\alpha(t)$ to the partial differential equation

$$\frac{\partial}{\partial t} \varrho_\alpha(t) = -(-\Delta)^{\alpha/2} \varrho_\alpha(t), \quad \varrho_\alpha(0) = \delta_0 \quad (\text{the Dirac measure at } 0),$$

depend absolutely continuously on t with respect to the Wasserstein metric W_p whenever $1 \leq p < \alpha < 2$.

1. Introduction. Consider a family $\{\sigma_t \mid t \geq 0\}$ of probability Borel measures on \mathbb{R}^N . Several topologies are available to characterize the continuous dependence of σ_t on t . Continuity in the sense of the variation norm is too restrictive for many applications. At the other end, continuity with respect to the narrow (weak) topology is characterized by the Prokhorov metric or equivalent metrics. For laws σ_t with finite p th moments ($1 \leq p < \infty$), the p -Wasserstein metric $W_p(\sigma_t, \sigma_s)$ is defined by (2.1) below. If σ_t is moreover absolutely continuous with respect to the p -Wasserstein metric with $1 < p < \infty$, i.e., there exists $m \in \mathbf{L}_{\text{loc}}^1([0, \infty))$ such that

$$W_p(\sigma_t, \sigma_s) \leq \int_s^t m(\tau) d\tau \quad \text{for } 0 \leq s < t,$$

then ([1, Theorem 8.3.1], see also [2]) there exists a vector field v_t such that (in an appropriate sense) the following continuity equation holds:

$$\partial_t \mu_t + \nabla \cdot (v_t \sigma_t) = 0.$$

2000 *Mathematics Subject Classification*: Primary 60B05; Secondary 28A30, 47D06.

Key words and phrases: Wasserstein metric, probability measures on metric spaces, subordination.

In this paper, we consider a method to investigate absolute continuity of σ_t with respect to the p -Wasserstein metric, if σ_t is obtained by subordination.

To introduce subordination, consider first an infinitely divisible family μ_t ($t \geq 0$) of probability measures on \mathbb{R}^N , i.e. a family of measures satisfying

- $\mu_0 = \delta_0$,
- $\mu_{t+s} = \mu_t * \mu_s$ for all $t, s \geq 0$,
- μ_t converges narrowly (weakly) to δ_0 as $t \rightarrow 0+$.

It is well known (see e.g. [13, Section 6.1]) that

$$[P_t(f)](x) := \int_{\mathbb{R}^N} f(x+y) d\mu_t(y)$$

defines a Markovian C_0 -semigroup on $C_{\text{ub}}(\mathbb{R}^N)$, the space of bounded, uniformly continuous functions on \mathbb{R}^N . Let ϱ be a Borel probability measure on \mathbb{R}^N . For shorthand we write $\varrho \in \mathcal{P}(\mathbb{R}^N)$. By abuse of notation we define

$$P_t\varrho = \mu_t * \varrho.$$

Then the function

$$[0, \infty) \rightarrow \mathcal{P}(\mathbb{R}^N), \quad t \mapsto P_t\varrho,$$

is a narrowly continuous curve.

As in [13, Section 6.3] we now consider another infinitely divisible family ν_t with values in $[0, \infty)$, and define

$$(1.1) \quad \sigma_t = \int_0^\infty \mu_s d\nu_t(s).$$

This is again an infinitely divisible family on \mathbb{R}^N which is called *subordinated to μ_t by ν_t* . This situation occurs, for instance, if σ_t is associated with a C_0 -semigroup generated by $-(-A)^\beta$ ($\beta \in (0, 1)$), where $-A$ is the generator of a C_0 contraction semigroup associated with μ_t (cf. Lemma 4.3).

Checking the absolute continuity leads to the natural question of estimating the Wasserstein distances of subordinated measures σ_t in terms of the subordinator ν_t . In this paper we prove the following result:

Let ν_1, ν_2 be Borel probability measures on $[0, \infty)$, and for $t \in [0, \infty)$ let μ_t be a Borel probability measure in $\mathcal{P}(\mathbb{R}^N)$, depending measurably on t in an appropriate sense. We define subordinated measures

$$\sigma_i = \int_0^\infty \mu_t d\nu_i(t) \quad (i = 1, 2).$$

The distance $d(t, s) := W_p(\mu_t, \mu_s)$ is then a metric on $[0, \infty)$. If the Wasserstein distance $W_p(\nu_1, \nu_2)$ of the subordinators is defined with respect to this metric on $[0, \infty)$, then the Wasserstein distance between the subordinated

measures is estimated by

$$W_p(\sigma_1, \sigma_2) \leq W_p(\nu_1, \nu_2).$$

A precise and somewhat more general formulation of this result is given in Theorems 3.2 and 3.3.

The proof is quite straightforward, with the exception of some fine points concerning the precise notion of integration of measure-valued functions. For technical reasons, we need to select a single-valued measurable function with values in \mathcal{P} from a multi-valued function. It turns out that we cannot achieve Borel measurability, but only measurability with respect to the sigma-algebra of universally measurable sets, using a sophisticated selection theorem [10, 5.5.3].

The paper is organized as follows: In Section 2 we introduce the basic definitions of tools used in this paper. Some definitions and results which are slightly more general than the standard results in the literature are justified in the appendix. In Section 3 we establish the main results. Finally, in Section 4 we consider the semigroup generated by $-(-\Delta)^\beta$ which can be obtained by subordination from the heat semigroup, using Gaussian kernels (cf. [13, Section 6.3.4]).

Acknowledgements. This work was done during mutual visits of the authors to Karl-Franzens University of Graz and to the University of Leiden. It is our pleasure to thank both universities for their kind hospitality. We are grateful to Onno van Gaans for helpful and stimulating discussions on Lévy processes. We also thank an anonymous referee for his careful reading of our manuscript.

2. Preliminaries. In this section we give the basic definitions and properties of the objects considered in this paper. For more details and proofs see the appendix.

Let (X, d_X) be a complete, separable metric space. We denote by $\mathcal{B}(X)$ the Borel sigma-algebra on X ; moreover, $\mathcal{P}(X)$ denotes the set of probability measures on $(X, \mathcal{B}(X))$. On $\mathcal{P}(X)$ we consider the topology of narrow convergence and the corresponding Borel sigma-algebra. It is known that this topology is induced by several metrics, among them the Prokhorov metric and the dual bounded Lipschitz metric β . (For definition of β see [6, p. 395, Proposition 11.3.2].) With this metric, $\mathcal{P}(X)$ is a complete separable metric space ([6, 11.5.5]; for the separability see [6, Section 11, Problem 4]).

If $\gamma \in \mathcal{P}(X_1 \times X_2)$ is a probability measure on a product space, then the marginals will be denoted by $\pi_{\#}^1 \gamma(A_1) = \gamma(A_1 \times X_2)$ and $\pi_{\#}^2 \gamma(A_2) = \gamma(X_1 \times A_2)$ for $A_i \in \mathcal{B}(X_i)$.

DEFINITION 2.1. Let $\mu_1, \mu_2 \in \mathcal{P}(X)$. We define the *Wasserstein distance*

$$(2.1) \quad W_{p,X}(\mu_1, \mu_2) = \inf \left[\int_{X \times X} d_X(x_1, x_2)^p d\gamma(x_1, x_2) \right]^{1/p},$$

where the infimum is taken over all probability measures $\gamma \in \mathcal{P}(X \times X)$ such that $\pi_{\#}^i \gamma = \mu_i$ for $i = 1, 2$. If $\gamma \in \mathcal{P}(X \times X)$ is such that $\pi_{\#}^i \gamma = \mu_i$ for $i = 1, 2$ and $W_{p,X}^p(\mu_1, \mu_2) = \int_{X \times X} d_X(x_1, x_2)^p d\gamma(x_1, x_2)$, then γ is called an *optimal plan* for (μ_1, μ_2) .

Detailed information about the Wasserstein distance can be found, e.g., in [1] and [6]. Notice that, according to the definition above, $W_{p,X}^p(\mu_1, \mu_2)$ may be infinite. In this case, any $\gamma \in \mathcal{P}(X \times X)$ such that $\pi_{\#}^i \gamma = \mu_i$ is an optimal plan. Notice also that in some references $W_{p,X}^p(\mu_1, \mu_2)$ is only considered for measures μ_i which have finite Wasserstein distance to some point measure. The space of these measures is usually denoted by $\mathcal{P}_p(X)$. The generalization to the case $\mu_i \in \mathcal{P}(X)$ with the possibility that the Wasserstein distance is infinite causes only minor technical changes; some of them will be indicated in the appendix.

In particular, in this case the infimum in (2.1) is also in fact a minimum ([1, Chapter 6, p. 133], [6, Theorem 11.8.2]). (More details are given in Lemma 5.2 and Corollary 5.3 in the appendix.)

Although the Wasserstein metric is not necessarily finite, it satisfies the triangle inequality, so the space $\mathcal{P}(X)$ can be topologized by $W_{p,X}$, for instance using the metric $\tilde{d}(\mu_1, \mu_2) = \min(1, W_{p,X}(\mu_1, \mu_2))$. This topology is stronger than the topology of narrow convergence. $W_{p,X}^p$ is lower semicontinuous with respect to narrow convergence (Corollary 5.3 in the appendix; for the case of $\mu \in \mathcal{P}_p(X)$, see [1, 7.1.3]).

Equivalently, we may characterize the Wasserstein metric by

$$(2.2) \quad W_{p,X}(\mu_1, \mu_2) = \inf [\mathbb{E}(d(X_1, X_2)^p)]^{1/p},$$

where the infimum is taken over all pairs of X -valued random variables X_1, X_2 such that the law of X_i is μ_i . (This approach is taken, e.g., in [7, Definition 1.1], where also infinite Wasserstein distances are allowed.)

We can reiterate the machinery and define a Wasserstein distance again on the space $\mathcal{P}(\mathcal{P}(X))$: If $\tilde{\mu}_1, \tilde{\mu}_2$ are probability measures on $\mathcal{P}(X)$ (with respect to the Borel sigma-algebra induced by β), then

$$(2.3) \quad W_{p,\mathcal{P}(X)}(\tilde{\mu}_1, \tilde{\mu}_2) = \inf \left[\int_{\mathcal{P}(X) \times \mathcal{P}(X)} W_{p,X}^p(\varrho_1, \varrho_2) d\tilde{\gamma}_{\varrho_1, \varrho_2} \right]^{1/p}$$

with the infimum taken over all $\tilde{\gamma}$ whose marginals are $\tilde{\mu}_1, \tilde{\mu}_2$.

We also need some basic facts about integration of measure-valued functions. Consider a probability space $(\Omega, \mathcal{F}, \theta)$ and a function $f : \Omega \rightarrow \mathcal{P}(X)$, $\omega \mapsto f_\omega$. Below, X will again be a separable complete metric space.

DEFINITION 2.2. The map f is called *measurable* iff for each Borel set $A \subset X$ the map

$$f_A : \Omega \rightarrow [0, 1], \quad \omega \mapsto f_\omega(A),$$

is measurable from (Ω, \mathcal{F}) to $([0, 1], \mathcal{B}([0, 1]))$.

By standard arguments it is sufficient to check measurability of f_A for any closed (or any open) subset $A \subset X$. If Ω is a metric space, and \mathcal{F} is the Borel sigma-algebra on Ω , then Definition 2.2 is just the definition of a Borel map according to, e.g., [1, Section 5.3], where we can also find the following concept of integration.

DEFINITION 2.3. If $f : \Omega \rightarrow \mathcal{P}(X)$ is measurable, then we define a probability measure

$$\nu = \int_{\Omega} f_\omega d\theta$$

by the following formula for each $A \in \mathcal{B}(X)$:

$$(2.4) \quad \nu(A) = \int_{\Omega} f_\omega(A) d\theta(\omega).$$

In fact, $\int_{\Omega} f_\omega d\theta(\omega)$ is a Borel probability measure on X . For each non-negative Borel measurable function $g : X \rightarrow [0, \infty]$ we have ([1])

$$(2.5) \quad \int_X g(x) \nu(dx) = \int_{\Omega} \int_X g(x) df_\omega(x) d\theta(\omega).$$

A more detailed discussion of this concept of integration will be given in Lemma 5.8 in the appendix.

3. The main theorem. Before we state the main result we prove a technical lemma: By Lemma 5.2, each pair $\mu_i \in \mathcal{P}(X)$ ($i = 1, 2$) admits at least one optimal plan $\gamma \in \mathcal{P}(X \times X)$. For later use we require a measurable function mapping each pair μ_1, μ_2 into an optimal plan γ_{μ_1, μ_2} . However, for technical reasons, we do not achieve Borel measurability of this function. The following lemma states that the optimal plan can be chosen to depend measurably on μ_i if we endow $\mathcal{P}(X) \times \mathcal{P}(X)$ with the sigma-algebra of universally measurable sets, which contains the Borel sigma-algebra.

LEMMA 3.1. *Let (X, d_X) be a complete separable metric space. Let $p \in [1, \infty)$ and let W_p denote the Wasserstein distance on $\mathcal{P}(X)$. Let $\mathcal{B}_U(\mathcal{P}(X) \times \mathcal{P}(X))$ be the sigma-algebra of universally measurable sets on $\mathcal{P}(X) \times \mathcal{P}(X)$*

(see [5, II.32]). Then there exists a map

$$\gamma : \mathcal{P}(X) \times \mathcal{P}(X) \rightarrow \mathcal{P}(X \times X), \quad (\mu_1, \mu_2) \mapsto \gamma_{\mu_1, \mu_2},$$

such that γ_{μ_1, μ_2} is an optimal plan for each μ_1, μ_2 . Moreover, γ is measurable if $\mathcal{P}(X) \times \mathcal{P}(X)$ is endowed with the sigma-algebra $\mathcal{B}_U(\mathcal{P}(X) \times \mathcal{P}(X))$, and $\mathcal{P}(X \times X)$ with the Borel sigma-algebra.

Proof. Set

$$H = \{(\mu_1, \mu_2, \gamma) \in \mathcal{P}(X) \times \mathcal{P}(X) \times \mathcal{P}(X \times X) \mid \gamma \text{ is an optimal plan for } (\mu_1, \mu_2)\}.$$

Define $h : \mathcal{P}(X \times X) \rightarrow [0, \infty]$ by

$$h(\gamma) := \int_{X \times X} d_X(x_1, x_2)^p d\gamma(x_1, x_2).$$

Then $H = H_1 \cap H_2$ where

$$H_1 = \{(\mu_1, \mu_2, \gamma) \in \mathcal{P}(X) \times \mathcal{P}(X) \times \mathcal{P}(X \times X) \mid h(\gamma) = W_p^p(\mu_1, \mu_2)\},$$

$$H_2 = \{(\mu_1, \mu_2, \gamma) \in \mathcal{P}(X) \times \mathcal{P}(X) \times \mathcal{P}(X \times X) \mid \pi_{\sharp}^i \gamma = \mu_i \ (i = 1, 2)\}.$$

According to [1, (5.1.15)], h is lower semicontinuous with respect to the topology of narrow convergence. By Corollary 5.3, the same holds for W_p^p . Therefore, according to Lemma 5.10, H_1 is a Borel set. The set H_2 is closed with respect to narrow convergence. Thus H is a Borel set.

Now we use [10, Theorem 5.5.3] (cf. also [6, Theorem 13.2.7]) to see that there exists a map $\gamma : \mathcal{P}(X) \times \mathcal{P}(X) \rightarrow \mathcal{P}(X \times X)$ such that each triple $(\mu_1, \mu_2, \gamma_{\mu_1, \mu_2})$ is contained in H and γ is measurable from $(\mathcal{P}(X) \times \mathcal{P}(X), \mathcal{B}_U(\mathcal{P}(X) \times \mathcal{P}(X)))$ into $(\mathcal{P}(X \times X), \mathcal{B}(\mathcal{P}(X \times X)))$. ■

Now we are ready to state our main theorem:

THEOREM 3.2. *Let (X, d_X) be a complete separable metric space. Let $p \in [1, \infty)$. On X and $\mathcal{P}(X)$ we consider the p -Wasserstein distances $W_{p, X}$ and $W_{p, \mathcal{P}(X)}$, respectively. Let $\tilde{\mu}_1, \tilde{\mu}_2 \in \mathcal{P}(\mathcal{P}(X))$, and define $\mu_i \in \mathcal{P}(X)$ ($i = 1, 2$) by*

$$\mu_i(A) := \int_{\mathcal{P}(X)} \varrho(A) d\tilde{\mu}_i(\varrho)$$

for each $A \in \mathcal{B}(X)$. Then $W_{p, X}(\mu_1, \mu_2) \leq W_{p, \mathcal{P}(X)}(\tilde{\mu}_1, \tilde{\mu}_2)$.

Proof. Let $\tilde{\theta} \in \mathcal{P}(\mathcal{P}(X) \times \mathcal{P}(X))$ be such that $\pi_{\sharp}^i \tilde{\theta} = \tilde{\mu}_i$ for $i = 1, 2$, and

$$(3.1) \quad \int_{\mathcal{P}(X) \times \mathcal{P}(X)} W_{p, X}^p(\varrho_1, \varrho_2) d\tilde{\theta}(\varrho_1, \varrho_2) = W_{p, \mathcal{P}(X)}^p(\tilde{\mu}_1, \tilde{\mu}_2).$$

Choose $\gamma : \mathcal{P}(X) \times \mathcal{P}(X) \rightarrow \mathcal{P}(X \times X)$ according to Lemma 3.1. We define a measure $\nu \in \mathcal{P}(X \times X)$ by

$$\nu = \int_{\mathcal{P}(X) \times \mathcal{P}(X)} \gamma_{\varrho_1, \varrho_2} d\tilde{\theta}(\varrho_1, \varrho_2).$$

Notice that, although γ is not necessarily a Borel function, this definition makes sense, since γ is measurable with respect to the sigma-algebra $\mathcal{B}_U(\mathcal{P}(X) \times \mathcal{P}(X))$ of universally measurable sets. By construction ([5, II.32]), this sigma-algebra is a subalgebra of the algebra of all $\tilde{\theta}$ -measurable sets, and $\tilde{\theta}$ can be uniquely extended to a measure on $\mathcal{B}_U(\mathcal{P}(X) \times \mathcal{P}(X))$. In particular, for $A \in \mathcal{B}(X)$ we obtain, using Lemma 5.8,

$$\begin{aligned} \pi_{\sharp}^1 \nu(A) &= \nu(A \times X) = \int_{\mathcal{P}(X) \times \mathcal{P}(X)} \gamma_{\varrho_1, \varrho_2}(A \times X) d\tilde{\theta}(\varrho_1, \varrho_2) \\ &= \int_{\mathcal{P}(X) \times \mathcal{P}(X)} \pi_{\sharp}^1 \gamma_{\varrho_1, \varrho_2}(A) d\tilde{\theta}(\varrho_1, \varrho_2) = \int_{\mathcal{P}(X) \times \mathcal{P}(X)} \varrho_1(A) d\tilde{\theta}(\varrho_1, \varrho_2) \\ &= \int_{\mathcal{P}(X)} \varrho_1(A) d\pi_{\sharp}^1 \tilde{\theta}(\varrho_1) = \int_{\mathcal{P}(X)} \varrho_1(A) d\tilde{\mu}_1(\varrho_1) = \mu_1(A). \end{aligned}$$

Similarly, $\pi_{\sharp}^2 \nu = \mu_2$. Therefore, using the definition of the Wasserstein metric, (2.5), and (3.1), we obtain

$$\begin{aligned} W_{p,X}^p(\mu_1, \mu_2) &\leq \int_{X \times X} d_X(x, y)^p d\nu(x, y) \\ &= \int_{\mathcal{P}(X) \times \mathcal{P}(X)} \int_{X \times X} d_X(x, y)^p d\gamma_{\varrho_1, \varrho_2}(x, y) d\tilde{\theta}(\varrho_1, \varrho_2) \\ &= \int_{\mathcal{P}(X) \times \mathcal{P}(X)} W_{p,X}^p(\varrho_1, \varrho_2) d\tilde{\theta}(\varrho_1, \varrho_2) = W_{p,\mathcal{P}(X)}^p(\tilde{\mu}_1, \tilde{\mu}_2). \blacksquare \end{aligned}$$

We will use this theorem in the following form:

THEOREM 3.3. *Let (X, d_X) and (Ω, d_Ω) be complete separable metric spaces. Let $p \in [1, \infty)$. On $\mathcal{P}(X)$ and $\mathcal{P}(\Omega)$ we consider the p -Wasserstein metrics $W_{p,X}$, $W_{p,\Omega}$, respectively. Let*

$$f : \Omega \rightarrow \mathcal{P}(X), \quad \omega \mapsto f_\omega,$$

be such that $W_{p,X}(f_{\omega_1}, f_{\omega_2}) \leq M d_\Omega(\omega_1, \omega_2)$ with a fixed constant $M > 0$. Then f is a Borel map. Given $\nu_1, \nu_2 \in \mathcal{P}(\Omega)$ the following probability measures $\mu_i \in \mathcal{P}(X)$ are well defined:

$$\mu_i := \int_{\Omega} f_\omega d\nu_i(\omega) \quad (i = 1, 2).$$

Then $W_{p,X}(\mu_1, \mu_2) \leq M W_{p,\Omega}(\nu_1, \nu_2)$.

Proof. To see that f is a Borel map, note that if $\omega_n \rightarrow \omega$ in Ω , then f_{ω_n} converges to f_ω in $\mathcal{P}(X)$ with respect to $W_{p,X}$. This implies narrow convergence [1, Proposition 7.1.5]. Hence Lemma 5.9 shows that f is a Borel map.

Now, for $i = 1, 2$, let $\tilde{\mu}_i$ be the image measure of ν_i under f on $\mathcal{P}(X)$, i.e., for $U \in \mathcal{B}(\mathcal{P}(X))$ we define $\tilde{\mu}_i(U) = \nu_i(f^{-1}(U))$. Then, by Lemma 5.8,

$$\mu_i(A) = \int_{\Omega} f_\omega(A) d\nu_i(\omega) = \int_{\mathcal{P}(X)} \varrho(A) d\tilde{\mu}_i(\varrho)$$

for any $A \in \mathcal{B}(X)$. Theorem 3.2 now implies that

$$W_{p,X}(\mu_1, \mu_2) \leq W_{p,\mathcal{P}(X)}(\tilde{\mu}_1, \tilde{\mu}_2).$$

To estimate $W_{p,\mathcal{P}(X)}(\tilde{\mu}_1, \tilde{\mu}_2)$, let $\gamma \in \mathcal{P}(\Omega \times \Omega)$ be an optimal plan for (ν_1, ν_2) . Let $\tilde{\gamma} \in \mathcal{P}(\mathcal{P}(X) \times \mathcal{P}(X))$ be the image measure of γ under $f \times f$, i.e., for $V \in \mathcal{B}(\mathcal{P}(X) \times \mathcal{P}(X))$, we define $\tilde{\gamma}(V) = \gamma(\{(\omega_1, \omega_2) \in \Omega \times \Omega \mid (f(\omega_1), f(\omega_2)) \in V\})$. We claim that the marginals of $\tilde{\gamma}$ are $\pi_{\sharp}^i(\tilde{\gamma}) = \tilde{\mu}_i$. In fact, for $U \in \mathcal{B}(\mathcal{P}(X))$, we have

$$\begin{aligned} (\pi_{\sharp}^1 \tilde{\gamma})(U) &= \tilde{\gamma}(U \times \mathcal{P}(X)) = \gamma(f^{-1}(U) \times f^{-1}(\mathcal{P}(X))) \\ &= \pi_{\sharp}^1 \gamma(f^{-1}(U)) = \nu_1(f^{-1}(U)) = \tilde{\mu}_1(U). \end{aligned}$$

The second marginal is checked similarly. Therefore we can estimate

$$\begin{aligned} W_{p,\mathcal{P}(X)}^p(\tilde{\mu}_1, \tilde{\mu}_2) &\leq \int_{\mathcal{P}(X) \times \mathcal{P}(X)} W_{p,X}^p(\varrho_1, \varrho_2) d\tilde{\gamma}(\varrho_1, \varrho_2) \\ &= \int_{\Omega \times \Omega} W_{p,X}^p(f_{\omega_1}, f_{\omega_2}) d\gamma(\omega_1, \omega_2) \\ &\leq \int_{\Omega \times \Omega} M^p d_{\Omega}(\omega_1, \omega_2)^p d\gamma(\omega_1, \omega_2) = M^p W_{p,\Omega}^p(\nu_1, \nu_2). \quad \blacksquare \end{aligned}$$

4. An application. In this section let P_t denote the heat semigroup in $\mathbf{L}^2(\mathbb{R}^N)$. Its generator is the Laplacian $-A = \Delta$ defined on $\mathbf{W}^{2,2}(\mathbb{R}^N)$. We are interested in the so-called α -stable semigroup ($\alpha \in (0, 2)$), denoted by P_t^α , which is generated by $-A^{\alpha/2}$. In this section we will prove the following:

THEOREM 4.1. *Let N be a positive integer, $\alpha \in (1, 2)$. Then there exist probability measures $\sigma_t^\alpha \in \mathcal{P}(\mathbb{R}^N)$ such that*

$$[P_t^\alpha f](x) = \int_{\mathbb{R}^N} f(x - y) d\sigma_t^\alpha(y).$$

For each $p \in [1, \alpha)$ there exists $M > 0$ such that for $s, t \geq 0$,

$$(4.1) \quad W_p(\sigma_t^\alpha, \sigma_s^\alpha) \leq M |t^{1/\alpha} - s^{1/\alpha}|.$$

In particular, σ_t^α depends absolutely continuously on t with respect to the p -Wasserstein metric.

The proof will be divided into several lemmas.

DEFINITION 4.2. Let $A : D(A) \subset E \rightarrow E$ be a linear operator in a (real or complex) Banach space $(E, |\cdot|_E)$. The operator A is called *accretive* (and $-A$ *dissipative*) if for all $\lambda > 0$ and all $x \in D(A)$ we have $|x|_E \leq |x + \lambda Ax|_E$. An accretive (resp. dissipative) operator A is called *m-accretive* (resp. *m-dissipative*) if the range of $\text{id} + \lambda A$ (resp. $\text{id} - \lambda A$) is E for all (equivalently for some) $\lambda > 0$.

It follows from the Hille version of the Hille–Yosida theorem that for a densely defined m-accretive operator A in a Banach space E , for all $t > 0$ and $x \in E$ the following limit exists:

$$e^{-tA}x := \lim_{n \rightarrow \infty} \left(\text{id} + \frac{t}{n} A \right)^{-n} x.$$

Moreover, with $e^{0A}x := x$, the family of operators $\{e^{-tA} \mid t \geq 0\}$ is a C_0 -semigroup on E , called the *semigroup generated by $-A$* . We recall the following result for linear semigroups, which we will utilize with $\beta = \alpha/2$:

LEMMA 4.3. *Let A be a densely defined, linear m-accretive operator on a Banach space E , and let e^{-tA} be the semigroup generated by $-A$. Let $\beta \in (0, 1)$. Then the fractional power A^β generates a semigroup e^{-tA^β} with the representation*

$$(4.2) \quad e^{-tA^\beta} x = \int_0^\infty f_{t,\beta}(s) e^{-sA} x ds.$$

The integral is to be understood as a strong Bochner integral. Here, the function $f_{t,\beta}$ is given by

$$(4.3) \quad f_{t,\beta}(s) = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} e^{zs-z^\beta t} dz$$

(with $\sigma > 0$, $t > 0$, $s \geq 0$ and the branch of z^β such that $\Re(z^\beta) > 0$ if $\Re(z) > 0$). It has the following properties:

$$(4.4) \quad \int_0^\infty f_{t,\beta}(s) ds = 1,$$

$$(4.5) \quad f_{t,\beta}(s) \geq 0 \quad \text{for all } s \geq 0.$$

The Laplace transform with respect to s of $f_{t,\beta}(s)$ is

$$(4.6) \quad \widehat{f}_{t,\beta}(\sigma) = e^{-t\sigma^\beta}.$$

For $t > 0$, the function $f_{t,\beta}$ solves the integro-differential equation

$$(4.7) \quad \frac{\partial}{\partial t} f_{t,\beta}(s) + \frac{\partial}{\partial s} \int_0^s g_{1-\beta}(s-\sigma) f_{t,\beta}(\sigma) d\sigma = 0$$

with $g_{1-\beta}(s) = s^{-\beta}/\Gamma(1-\beta)$. For $\gamma > 0$ we have the rescaling identity

$$(4.8) \quad f_{t,\beta}(s) = \gamma^{1/\beta} f_{\gamma t,\beta}(\gamma^{1/\beta} s).$$

Proof. These results are proved in detail in [12, Chapter IX, 11]. In particular, (4.3) is equation (1), (4.4) is equation (14), (4.6) is Proposition 1, and (4.5) is Proposition 2 there. The differential equation (4.7) is easily checked using Laplace transforms with respect to s , in particular (4.6). Finally, to check (4.8) we use (4.3) and make a change of variables $y = z\gamma^{1/\beta}$:

$$\begin{aligned} \gamma^{1/\beta} f_{\gamma t,\beta}(\gamma^{1/\beta} s) &= \gamma^{1/\beta} \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} e^{z\gamma^{1/\beta} s - z^\beta \gamma t} dz \\ &= \frac{1}{2\pi i} \int_{\sigma\gamma^{1/\beta}-i\infty}^{\sigma\gamma^{1/\beta}+i\infty} e^{ys - y^\beta t} dy = f_{t,\beta}(s). \quad \blacksquare \end{aligned}$$

To treat our special case, we have to introduce the heat kernels:

DEFINITION 4.4. Let $N \in \mathbb{N}$ and Q be a positive definite symmetric $N \times N$ matrix. Then \mathcal{N}_Q denotes the centered Gaussian measure in \mathbb{R}^N with covariance matrix Q .

LEMMA 4.5. For $f \in \mathbf{L}^2(\mathbb{R}^N)$,

$$(4.9) \quad P_t^\alpha f = \sigma_t^\alpha * f$$

with

$$(4.10) \quad \sigma_t^\alpha = \int_0^\infty \mathcal{N}_{s^{2I}} g_{t,\alpha}(s) ds,$$

$$(4.11) \quad g_{t,\alpha}(s) = t^{-2/\alpha} s f_{1,\alpha/2}(t^{-2/\alpha} s^2/2).$$

Proof. It is well known that the heat semigroup is given by

$$[P_t f](x) = \int_{\mathbb{R}^N} f(x-y) d\mathcal{N}_{2tI}(y).$$

Lemma 4.3 implies immediately that

$$[P_t^\alpha f](x) = \int_0^\infty \int_{\mathbb{R}^N} f(x-y) d\sigma_t^\alpha(y) \quad \text{with} \quad \sigma_t^\alpha = \int_0^\infty \mathcal{N}_{2rI} f_{t,\alpha/2}(r) dr.$$

We first rescale t using (4.8) and then make a change of variable $s^2 = 2r$ to obtain

$$\begin{aligned} \int_0^\infty \mathcal{N}_{2rI} f_{t,\alpha/2}(r) dr &= \int_0^\infty \mathcal{N}_{2rI} t^{-2/\alpha} f_{1,\alpha/2}(t^{-2/\alpha} r) dr \\ &= \int_0^\infty \mathcal{N}_{s^2I} t^{-2/\alpha} f_{1,\alpha/2}(t^{-2/\alpha} s^2/2) s ds. \blacksquare \end{aligned}$$

LEMMA 4.6. *Let X be a random variable with values in \mathbb{R}^N and $p \in [1, \infty)$ be such that $\mathbb{E}(|X|^p) = M < \infty$. Let $c_1, c_2 \geq 0$ and let μ_i be the law of $c_i X$. Then $W_p(\mu_1, \mu_2) \leq M^{1/p} |c_1 - c_2|$. In particular, for centered Gaussian measures we have*

$$W_p(\mathcal{N}_{s_1I}, \mathcal{N}_{s_2I}) \leq M |\sqrt{s_1} - \sqrt{s_2}|,$$

where M is a constant depending only on p and N .

Proof. We utilize the characterization (2.2) with the random variables $X_i = c_i X$. Then

$$W_p^p(\mu_1, \mu_2) \leq \mathbb{E}(|c_1 X - c_2 X|^p) = |c_1 - c_2|^p \mathbb{E}(|X|^p). \blacksquare$$

We recall that a probability measure $\mu \in \mathcal{P}(\mathbb{R})$ is called β -stable if the following holds: If X_i ($i = 1, \dots, n$) are independent random variables with the same law μ , then there exists a constant b_n such that μ is also the law of $n^{1/\beta} \sum_{i=1}^n X_i - b_n$.

LEMMA 4.7. *Let $\beta \in (0, 1)$ and $f_{t,\beta}$ be as in Lemma 4.3, continued by 0 to the interval $(-\infty, 0)$. Then the probability measure $\kappa_{t,\beta}$ with the density $f_{t,\beta}$ is β -stable. In particular, for $\delta \in (0, \beta)$,*

$$(4.12) \quad \int_0^\infty s^\delta f_{t,\beta}(s) ds < \infty.$$

For $\delta > \beta$, the integral in (4.12) is infinite.

Proof. By (4.6), the characteristic function of the measure $\kappa_{t,\beta}$ is

$$\int_{-\infty}^\infty e^{izs} f_{t,\beta}(s) ds = e^{-t|z|^\beta (\cos(\beta\pi/2) - i \operatorname{sgn}(z) \sin(\beta\pi/2))},$$

which is a special case of the Lévy–Khinchin representation of a β -stable measure (see [9, p. 347]). It is well known that β -stable measures for $\beta < 2$ have finite moments of any order $\delta < \beta$ (see [9, p. 341, Problem 6] or the estimates in [8, p. 187]), but the moments of order $\delta > \beta$ are infinite (see [11]). \blacksquare

LEMMA 4.8. *The function $g_{t,\alpha}$ given by (4.11) is the density function of a probability measure ν_t^α . For $0 < s < t$ and $1 \leq p < \alpha$, the following*

estimate holds with some constant $M > 0$:

$$(4.13) \quad W_p(\nu_s^\alpha, \nu_t^\alpha) \leq M|t^{2/\alpha} - s^{2/\alpha}|.$$

Proof. Let X be a random variable such that $f_{1,\alpha/2}$ is a density function for X . Then by a change of variables, g_t is the density function of

$$Y_t := t^{1/\alpha} \sqrt{2X} = t^{1/\alpha} Y_1.$$

By Lemma 4.7, X has moments of any order less than $\alpha/2$, and consequently Y_t has moments of all orders $p < \alpha$. Let ν_t^α be the law of Y_t . Lemma 4.6 now implies that for $p \in [1, \alpha)$ we have $W_p(\nu_t^\alpha, \nu_s^\alpha) \leq M|t^{1/\alpha} - s^{1/\alpha}|$. ■

Proof of Theorem 4.1. Let $g_{t,\alpha}$ be the density function given by (4.11) and let ν_t^α be the corresponding probability measure on $[0, \infty)$. By Lemma 4.5 we have

$$\varrho_\alpha(t) = \int_0^\infty \mathcal{N}_{s^2 I} d\nu_t^\alpha(s).$$

By Lemma 4.6, $W_p(\mathcal{N}_{s_1^2 I}, \mathcal{N}_{s_2^2 I}) \leq M_1|s_1 - s_2|$ with some fixed $M_1 > 0$. By Lemma 4.8, $W_p(\nu_t^\alpha, \nu_s^\alpha) \leq M_2|t^{1/\alpha} - s^{1/\alpha}|$ with some $M_2 > 0$. Thus Theorem 3.3 yields (4.1) with $M = M_1 M_2$. ■

Finally, we may also define the α -stable semigroup P_t^α on the space $\mathcal{P}(\mathbb{R}^N)$ by

$$P_t^\alpha \varrho = \sigma_t^\alpha * \varrho.$$

We show that the trajectories of P_t^α in $\mathcal{P}(\mathbb{R}^N)$ are also absolutely continuous with respect to the W_p -Wasserstein metric, if $1 < p < \alpha$:

LEMMA 4.9. *Let $\sigma_1, \sigma_2, \varrho \in \mathcal{P}_p(\mathbb{R}^N)$ where $p \in [1, \infty)$. Then $\varrho * \sigma_i \in \mathcal{P}_p(\mathbb{R}^N)$ for $i = 1, 2$, and*

$$W_p(\varrho * \sigma_1, \varrho * \sigma_2) \leq W_p(\sigma_1, \sigma_2).$$

Proof. Let $\varepsilon > 0$ and X_1, X_2 be \mathbb{R}^N -valued random variables whose laws are σ_1, σ_2 , respectively, and $\mathbb{E}(|X_1 - X_2|^p) \leq W_p^p(\sigma_1, \sigma_2) + \varepsilon$. Let Y be a random variable independent of X_1 and X_2 such that ϱ is the law of Y . Then $\sigma_i * \varrho$ is the law of $X_i + Y$. Consequently, $\sigma_i * \varrho \in \mathcal{P}_p(\mathbb{R}^d)$ and

$$\begin{aligned} W_p^p(\sigma_1 * \varrho, \sigma_2 * \varrho) &\leq \mathbb{E}(|(X_1 + Y) - (X_2 + Y)|^p) = \mathbb{E}(|X_1 - X_2|^p) \\ &\leq W_p^p(\sigma_1, \sigma_2) + \varepsilon. \quad \blacksquare \end{aligned}$$

COROLLARY 4.10. *Let $\varrho \in \mathcal{P}_p(\mathbb{R}^d)$ and let $P_t^\alpha \varrho = \sigma_t * \varrho$ with σ_t as in Lemma 4.5. Let $p \in (1, \alpha)$. Then the map*

$$[0, \infty) \rightarrow \mathcal{P}_p(\mathbb{R}^N), \quad t \mapsto P_t^\alpha \varrho,$$

is absolutely continuous.

5. Appendix. In Section 2 we have stated that some of our definitions are slightly more general than the standard ones found in the literature, for example when we allow the Wasserstein distance $W_{p,X}$ to be infinite, or when we define a Wasserstein distance on $\mathcal{P}(\mathcal{P}(X))$ again by integrating $W_{p,X}^p$, where $\mathcal{P}(X)$ is topologized by another metric, namely β . Yet the basic results still hold, without the need of changing the proofs significantly. In this appendix we illustrate this fact by several examples. In fact, the results stated here are not new, and their proofs are given only to emphasize that the more general setting does not cause serious difficulties.

The existence of optimal plans follows from a more general result from [1, p. 133]. The lower semicontinuity of the Wasserstein distance is found in [1, Proposition 7.1.3]. We recall the reasoning given in [1] to see that there is no need to confine the result to $\mathcal{P}_p(X)$, and that the same works on $\mathcal{P}(\mathcal{P}(X))$. For shorthand we will adopt the following notation:

DEFINITION 5.1. Let (X, d_X) be a separable metric space, let $\mu_1, \mu_2 \in \mathcal{P}(X)$. We define

$$\Gamma(\mu_1, \mu_2) = \{\gamma \in \mathcal{P}(X \times X) \mid \pi_{\#}^i \gamma = \mu_i \ (i = 1, 2)\}.$$

We recall that $\Gamma(\mu_1, \mu_2)$ is closed in $\mathcal{P}(X \times X)$ with respect to the narrow topology, and compact by [1, Lemma 5.2.2], if (X, d) has the Radon property, i.e., every probability measure on $\mathcal{B}(X)$ is tight ([1, Definition 5.1.4]).

LEMMA 5.2. Let (X, d_X) be a separable Radon metric space. Let $c : X \times X \rightarrow [0, \infty]$ be lower semicontinuous and for $\mu_1, \mu_2 \in \mathcal{P}(X)$ let

$$(5.1) \quad C(\mu_1, \mu_2) = \inf_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{X \times X} c(x, y) d\gamma(x, y).$$

Then the following assertions hold:

- (1) $C(\mu_1, \mu_2)$ is in fact a minimum, i.e., there exists some optimizing $\gamma \in \Gamma(\mu_1, \mu_2)$ for (5.1).
- (2) $C(\mu_1, \mu_2)$ is lower semicontinuous on $\mathcal{P}(X) \times \mathcal{P}(X)$ with respect to the topology of narrow convergence.

Proof. Let $\mu_1, \mu_2 \in \mathcal{P}(X)$. Then $\Gamma(\mu_1, \mu_2)$ is compact with respect to the topology of narrow convergence. By [1, (5.1.15)] the map

$$J : \mathcal{P}(X \times X) \rightarrow [0, \infty], \quad \gamma \mapsto \int_{X \times X} c(x, y) d\gamma(x, y),$$

is lower semicontinuous with respect to narrow convergence on $\mathcal{P}(X \times X)$. Hence J assumes a minimum (possibly infinite) on $\Gamma(\mu_1, \mu_2)$, proving (1).

To prove (2) let $\mu_i^n \rightarrow \mu_i^0$ narrowly for $i = 1, 2$. The sets $\{\mu_i^n \mid n = 0, 1, \dots\}$ are relatively compact, hence tight. Thus $\bigcup_{n=0}^{\infty} \Gamma(\mu_1^n, \mu_2^n)$ is also tight and hence relatively compact. Let γ_n be optimal for (μ_1^n, μ_2^n) . There

exists a convergent subsequence. Without loss of generality we may assume that γ_n converges narrowly to some γ . Then evidently $\gamma \in \Gamma(\mu_1^0, \mu_2^0)$. Moreover, by [1, (5.1.15)],

$$\begin{aligned} C(\mu_1^0, \mu_2^0) &\leq \int_{X \times X} c(x, y) d\gamma(x, y) \\ &\leq \liminf_{n \rightarrow \infty} \int_{X \times X} c(x, y) d\gamma_n(x, y) = \liminf C(\mu_1^n, \mu_2^n). \blacksquare \end{aligned}$$

As an immediate consequence we have

COROLLARY 5.3. *Let (X, d_X) be a Radon separable metric space. Let $\mu_1, \mu_2 \in \mathcal{P}(X)$.*

- (1) *There exists an optimal plan for the Wasserstein metric $W_p(\mu_1, \mu_2)$.*
- (2) *The Wasserstein distance is lower semicontinuous on $\mathcal{P}(X) \times \mathcal{P}(X)$ with respect to the topology of narrow convergence.*

To reiterate the procedure and investigate the Wasserstein distance for measures in $\mathcal{P}(\mathcal{P}(X))$, we require that (X, d_X) is complete. This implies that $(\mathcal{P}(X), \beta)$ is complete, hence Radon, and we can again apply Lemma 5.2:

COROLLARY 5.4. *Let (X, d_X) be a complete separable metric space. On $\mathcal{P}(X)$ consider the topology of narrow convergence (metrized by β) and the corresponding Borel sigma-algebra. Let $W_{p,X}$ be the Wasserstein distance on $\mathcal{P}(X)$. We define the Wasserstein distance for $\tilde{\mu}_i \in \mathcal{P}(\mathcal{P}(X))$ by*

$$(5.2) \quad W_{p,\mathcal{P}(X)}(\tilde{\mu}_1, \tilde{\mu}_2) = \inf_{\tilde{\gamma} \in \Gamma(\tilde{\mu}_1, \tilde{\mu}_2)} \int_{\mathcal{P}(X) \times \mathcal{P}(X)} W_{p,X}^p(\varrho_1, \varrho_2) d\tilde{\gamma}(\varrho_1, \varrho_2).$$

Then for any $\tilde{\mu}_1, \tilde{\mu}_2 \in \mathcal{P}(\mathcal{P}(X))$ there exists an optimal plan $\tilde{\gamma} \in \Gamma(\tilde{\mu}_1, \tilde{\mu}_2)$ minimizing (5.2).

One of the problems involved in the definition of the Wasserstein distance on $\mathcal{P}(\mathcal{P}(X))$ is that $(\mathcal{P}(X), W_{p,X})$ is not quite a metric space, since $W_{p,X}(\mu_1, \mu_2)$ may be infinite. However, we can cope with this problem by a simple approximation procedure. We illustrate this by the example of proving the triangle inequality. The triangle inequality for the Wasserstein distance is in fact a nontrivial result, which has been proved only for measures in $\mathcal{P}_p(X)$, where, in particular, the Wasserstein distance is finite.

LEMMA 5.5. *Let $X \neq \emptyset$ be a set and $d : X \times X \rightarrow [0, \infty]$ be a function satisfying, for all $x, y, z \in X$,*

$$(5.3) \quad d(x, y) = d(y, x),$$

$$(5.4) \quad d(x, z) \leq d(x, y) + d(y, z),$$

$$(5.5) \quad d(x, y) = 0 \Rightarrow x = y.$$

For $n \in \mathbb{N}$ let d_n be metrics on X such that $d_n \leq d_{n+1}$ and $d_n(x, y)$ converges to $d(x, y)$ for all $x, y \in X$.

Suppose that for all $n \in \mathbb{N}$ and all sequences $x_k, y_k \in X$ such that $d_n(x_k, y_k) \rightarrow 0$ we have $d(x_k, y_k) \rightarrow 0$, so that each metric d_n induces the same topology, hence the same Borel sigma-algebra on X . Suppose moreover that this topological space is separable and has the Radon property.

Let W_p be the Wasserstein distance on $\mathcal{P}(X)$, which can be defined by (2.1) although d is not a metric. Let $W_{p,n}$ be the Wasserstein distance on $\mathcal{P}(X)$ defined by (2.1) with d replaced by d_n .

Then, for all $\mu_1, \mu_2 \in \mathcal{P}(X)$, $W_{p,n}(\mu_1, \mu_2)$ converges monotonically to $W_p(\mu_1, \mu_2)$.

Proof. Let $\mu_1, \mu_2 \in \mathcal{P}(X)$. Evidently, $W_{p,n}(\mu_1, \mu_2) \leq W_{p,n+1}(\mu_1, \mu_2) \leq W_p(\mu_1, \mu_2)$. Thus $\sup_{n \in \mathbb{N}} W_{p,n}(\mu_1, \mu_2)$ is well defined and bounded by $W_p(\mu_1, \mu_2)$. Again, let

$$\Gamma(\mu_1, \mu_2) = \{\gamma \in \mathcal{P}(X \times X) \mid \pi_{\#}^i \gamma = \mu_i \ (i = 1, 2)\}.$$

Let γ_n be an optimal plan in the sense that

$$W_{p,n}^p(\mu_1, \mu_2) = \int_{X \times X} d_n(x, y)^p d\gamma_n$$

(see Lemma 5.2). By compactness there exists a convergent subsequence. Without loss of generality assume that γ_n converges narrowly to some γ , which is evidently in $\Gamma(\mu_1, \mu_2)$.

Let $M < \int_{X \times X} d(x, y)^p d\gamma$. By monotone convergence there exists some $m \in \mathbb{N}$ such that $M < \int_{X \times X} d_m(x, y)^p d\gamma$. We now use the lower semicontinuity of the integral with respect to narrow convergence of measures ([1, (5.1.15)]) and the fact that $d_m \leq d_n$ for large n :

$$\begin{aligned} M &< \int_{X \times X} d_m(x, y)^p d\gamma \leq \liminf_{n \rightarrow \infty} \int_{X \times X} d_m(x, y)^p d\gamma_n \\ &\leq \liminf_{n \rightarrow \infty} \int_{X \times X} d_n(x, y)^p d\gamma_n = \liminf_{n \rightarrow \infty} W_{p,n}^p(\mu_1, \mu_2). \end{aligned}$$

Hence we have

$$W_p(\mu_1, \mu_2) \leq \int_{X \times X} d(x, y)^p d\gamma \leq \liminf_{n \rightarrow \infty} W_{p,n}^p(\mu_1, \mu_2). \quad \blacksquare$$

COROLLARY 5.6. *Let $X \neq \emptyset$ be a set and $d : X \times X \rightarrow [0, \infty]$ be a function satisfying (5.3)–(5.5). Let $d_n(x, y) = \min(d(x, y), n)$ and suppose that (X, d_1) is a separable Radon metric space. Then (X, d_n) satisfies all assumptions of Lemma 5.5. In particular, the Wasserstein distances $W_{p,n}(\mu_1, \mu_2)$ with respect to d_n converge monotonically to the Wasserstein distance $W_p(\mu_1, \mu_2)$ with respect to d .*

We are now in a position to generalize the triangle inequality to general measures in $\mathcal{P}(X)$:

COROLLARY 5.7. *Let $X \neq \emptyset$ be a set and $d : X \times X \rightarrow [0, \infty]$ be a function satisfying (5.3)–(5.5). Let $d_n(x, y) = \min(d(x, y), n)$ and suppose that (X, d_1) is a separable Radon metric space. We define the Wasserstein distance on (X, d) by (2.1). Then the triangle inequality holds for all $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(X)$:*

$$W_p(\mu_1, \mu_3) \leq W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3).$$

Proof. Approximate d as in Corollary 5.6 by $d_n(x, y) = \min(n, d(x, y))$. Let

$$\mathcal{P}_{p,n}(X) = \{\mu \in \mathcal{P}(X) \mid (\exists x \in X) W_{p,n}(\mu, \delta_x) < \infty\}.$$

Here, δ_x denotes the point measure at x . It is well known that $W_{p,n}$ is a metric on $\mathcal{P}_{p,n}(X)$. Since d_n is bounded on $X \times X$, we easily infer that $\mathcal{P}_{p,n}(X) = \mathcal{P}(X)$. Therefore the triangle inequality holds for $W_{p,n}$ on the whole space $\mathcal{P}(X)$. By monotone approximation, using Lemma 5.5, we obtain the triangle inequality also for W_p . ■

The rest of this appendix is devoted to the machinery of integrating measurable functions.

LEMMA 5.8. *Let $(\Omega, \mathcal{F}, \theta)$ be a probability space, and (X, d_X) a complete separable metric space. Let*

$$f : \Omega \rightarrow \mathcal{P}(X), \quad \omega \mapsto f_\omega,$$

be a function. Then the following assertions hold:

- (1) *The function f is measurable in the sense of Definition 2.2 if for all closed (or all open) $A \subset X$, the map*

$$f_A : \Omega \rightarrow \mathbb{R}, \quad \omega \mapsto f_\omega(A),$$

is measurable from (Ω, \mathcal{F}) to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

- (2) *If f is measurable, then the map $\nu = \int_\Omega f_\omega d\theta(\omega)$ defined by*

$$\nu : \mathcal{B}(X) \rightarrow \mathbb{R}, \quad A \mapsto \int_\Omega f_\omega(A) d\theta(\omega),$$

is a probability measure.

- (3) *If f is measurable and $g : X \rightarrow [0, \infty)$ is a nonnegative Borel measurable function, and $\nu = \int_\Omega f_\omega d\theta(\omega)$, then*

$$\int_X g(x) d\nu(x) = \int_\Omega \int_X g(x) df_\omega(x) d\theta(\omega).$$

Proof. (1) This is a standard technique from measure theory. Let

$$\mathcal{D} = \{A \subset X \mid f_A : (\Omega, \mathcal{F}) \rightarrow (X, \mathcal{B}(X)) \text{ is measurable}\}.$$

We show that \mathcal{D} is a Dynkin system. In fact:

- (a) $X \in \mathcal{D}$, since f_X is the constant function, $f_X(\omega) = 1$.
- (b) If $A_i \in \mathcal{D}$ are mutually disjoint and $A = \bigcup_{i=1}^{\infty} A_i$, then $f_A(\omega) = \sum_{i=1}^{\infty} f_{A_i}(\omega)$, hence f_A is a sum of measurable functions and $A \in \mathcal{D}$.
- (c) If $A, B \in \mathcal{D}$ with $B \subset A$, then $f_{A \setminus B} = f_B - f_A$ is measurable, hence $A \setminus B \in \mathcal{D}$.

By assumption, \mathcal{D} contains all closed sets. Since the intersection of closed sets is closed again, [4, Theorem 1.6.1] shows that \mathcal{D} contains the sigma-algebra generated by the closed sets, i.e. all $A \in \mathcal{B}(X)$ have f_A measurable.

(2) Evidently, $\nu_A \geq 0$. For $A = X$ we have

$$\nu(X) = \int_{\Omega} f_{\omega}(X) d\theta(\omega) = \int_{\Omega} 1 d\theta(\omega) = 1.$$

Finally, if A_i are mutually disjoint and $A = \bigcup_{i=1}^{\infty} A_i$, then

$$\begin{aligned} \nu(A) &= \int_{\Omega} f_{\omega}(A) d\theta(\omega) = \int_{\Omega} \sum_{i=1}^{\infty} f_{\omega}(A_i) d\theta(\omega) \\ &= \sum_{i=1}^{\infty} \int_{\Omega} f_{\omega}(A_i) d\theta(\omega) = \sum_{i=1}^{\infty} \nu(A_i). \end{aligned}$$

(3) Let $g_n : X \rightarrow \mathbb{R}$ be a monotone sequence of simple functions converging to g pointwise. By monotone convergence it is sufficient to prove

$$\int_X g_n(x) d\nu(x) = \int_{\Omega} \int_X g_n(x) df_{\omega}(x) d\theta(\omega).$$

Let $g_n(x) = \sum_{j=1}^m \alpha_j \chi_{A_j}(x)$, where $\alpha_j \geq 0$, $A_j \in \mathcal{B}(X)$, and χ_{A_j} is the indicator function of A_j . Then

$$\begin{aligned} \int_X g_n(x) d\nu(x) &= \sum_{j=1}^m \alpha_j \nu(A_j) = \sum_{j=1}^m \alpha_j \int_{\Omega} f_{\omega}(A_j) d\theta(\omega) \\ &= \int_{\Omega} \sum_{j=1}^m \alpha_j f_{\omega}(A_j) d\theta(\omega) = \int_{\Omega} \int_X g_n(x) df_{\omega}(x) d\theta(\omega). \blacksquare \end{aligned}$$

LEMMA 5.9. *Let (Ω, d_{Ω}) and (X, d_X) be metric spaces. Let $f : \Omega \rightarrow \mathcal{P}(X)$ be continuous in the sense that $\omega_n \rightarrow \omega$ in Ω implies that f_{ω_n} converges narrowly to f_{ω} . Then f is a Borel map.*

Proof. For $A \in \mathcal{B}(X)$ we define $f_A(\omega) = f_{\omega}(A)$. By Lemma 5.8(1) it is sufficient to show that $f_A : \Omega \rightarrow \mathbb{R}$ is Borel measurable if A is an open set. In this case, however, f_A is lower semicontinuous by [6, 11.1.1]. \blacksquare

The point of the last technical lemma is that we allow for functions which may take ∞ as a value, hence they cannot be subtracted from each other:

LEMMA 5.10. *Let (X, d) be a metric space, and let $h_1, h_2 : X \rightarrow [0, \infty]$ be lower semicontinuous functions. Then $\{x \in X \mid h_1(x) = h_2(x)\}$ is a Borel set.*

Proof. For $n \in \mathbb{N}$, $i = 1, 2$, we define lower semicontinuous functions $h_i^n(x) = \min(h_i(x), n)$. Then the function $h_1^n - h_2^n$ is Borel measurable, hence $U_n := \{x \in X \mid h_1^n(x) - h_2^n(x) = 0\}$ is a Borel set. Consequently, so is $\{x \in X \mid h_1(x) = h_2(x)\} = \bigcap_{n=1}^{\infty} U_n$. ■

References

- [1] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures Math. ETH Zürich, Birkhäuser, Basel, 2005.
- [2] J. D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem*, Numer. Math. 84 (2000), 375–393.
- [3] P. Clément and W. Desch, *An elementary proof of the triangle inequality for the Wasserstein metric*, Proc. Amer. Math. Soc. 136 (2008), 333–339.
- [4] D. L. Cohn, *Measure Theory*, Birkhäuser, Boston, 1980.
- [5] C. Dellacherie and P.-A. Meyer, *Probabilities and Potential*, North-Holland Math. Stud. 29, North-Holland, Amsterdam, 1978.
- [6] R. M. Dudley, *Real Analysis and Probability*, Stud. Adv. Math. 74, Cambridge Univ. Press, Cambridge, 2002.
- [7] O. Johnson and R. Samworth, *Central limit theorems and convergence to stable laws in Mallows distance*, Bernoulli 11 (2005), 829–845; Acknowledgement of priority, *ibid.* 12 (2006), 191.
- [8] Yu. V. Prokhorov and Yu. A. Rozanov, *Probability Theory*, Grundlehren Math. Wiss. 157, Springer, Berlin, 1969.
- [9] A. N. Shiryaev, *Probability*, Grad. Texts in Math. 95, Springer, New York, 1984.
- [10] S. M. Srivastava, *A Course on Borel Sets*, Grad. Texts in Math. 180, Springer, New York, 1998.
- [11] H. G. Tucker, *On moments of distribution functions attracted to stable laws*, Houston J. Math. 1 (1975), 149–152.
- [12] K. Yosida, *Functional Analysis*, 5th ed., Grundlehren Math. Wiss. 123, Springer, Berlin, 1978.
- [13] J. Zabczyk, *Topics in Stochastic Processes*, Quad. Scuola Norm. Sup., Pisa, 2004.

Mathematical Institute
Leiden University
P.O. Box 9512
NL-2300 RA Leiden, The Netherlands
E-mail: philippeclém@gmail.com

Institut für Mathematik
und Wissenschaftliches Rechnen
Karl-Franzens-Universität Graz
Heinrichstrasse 36
8010 Graz, Austria
E-mail: georg.desch@uni-graz.at

Received September 20, 2007
Revised version May 30, 2008

(6218)