The asymptotic strong Feller property and its application to stochastic differential equations with Lipschitz nonlinearities

Diplomarbeit

vorgelegt von
Sebastian Schwarzkopf

Fakultät für Mathematik
Universität Bielefeld

März 2009

Contents

Introduction

In the study of stochastic differential equations the questions of existence and uniqueness of an invariant measure for the associated Markov semigroup are crucial:

(1) If there exists an invariant measure μ for the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$, the semigroup can be extended uniquely to the space $L^p(\mathcal{H},\mu),\ p \geq 1$.

(2) If, in addition, the invariant measure μ is unique, the dynamical system is ergodic.

Hence it might be important to look for conditions under which existence respectively uniqueness of an invariant measure follows. Concerning the first there is the much celebrated theorem of Krylov and Bogoliubov, stating that existence of an invariant measure is a consequence of the Markov semigroup having the Feller property and fulfilling some tightness assumption. On the other hand uniqueness of the invariant measure is often derived from results due to Khasminskii and Doob, stating that the strong Feller property together with some irreducibility condition ensure uniqueness of the invariant measure.

The present work focuses on the first property. While in finite dimensions there is a sufficient condition for the strong Feller property to hold, e.g. Hörmander's Theorem (cf. Theorem 8.1 in [6]), in infinite-dimensional spaces no corresponding theorem is known. Moreover the strong Feller property often fails to hold in infinite-dimensional spaces. Only if the forcing noise is sufficiently rough, e.g. the covariance of the noise is nondegenerate, the Bismut-Elworthy formula allows to show the strong Feller property for a class of semilinear parabolic SPDE's with infinite-dimensional state space. But in cases where the noise is very weak, even this is not applicable.

Therefore it would be extremely convenient to have a weaker property that still allows to conclude uniqueness of the invariant Borel probability measure. This idea is pursued in [7] by introducing the 'asymptotic strong Feller property'. In fact, there is the following main result (cf. Corollary 3.20 below): if the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ is asymptotically strong Feller and there exists some point x which belongs to the support of every invariant Borel probability measure for $(\mathcal{P}_t)_{t\geq 0}$, then there is at most one invariant measure for this Markov semigroup.

How should one define this asymptotic strong Feller property? Since for a Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ on a Polish space $\mathcal{X}$ the strong Feller property is equivalent to the continuity of the function $x \mapsto \pi_t(x,\cdot)$ on $\mathcal{X}$ in the total variation norm (cf. Theorem 1.14 below), a suitable approach would be to find another (semi-)norm that generates a weaker topology on the space of all Borel probability measures on $\mathcal{X}$. Furthermore it is known that for any separable metric space the Wasserstein distance metrizes the weak topology (cf. [5], Theorem 11.3.3). Hence it should be at least plausible to work with the Wasserstein distance. And in fact, it turns out that this is the right choice, since the Wasserstein distances corresponding to an increasing sequence of (continuous) pseudo-metrics less or equal 1 (called a totally separating system of (continuous) pseudo-metrics) converge to the total variation distance. This leads to the definition of the asymptotic strong Feller property (cf. Definition 3.8 below).

Let us now give an overview of the structure of this Diploma Thesis:

In *Chapter 1* we first introduce well-known concepts for Markov semigroups, e.g. irreducibility, (strong) Feller property, regularity and ergodicity. Afterwards we will see that the combination of regularity properties (Feller property respectively strong Feller property) and topological concepts (compactness respectively irreducibility) guarantees existence (cf. Theorem 1.32 below) respectively uniqueness (cf. Theorem 1.35 below) of an invariant Borel probability measure for the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$. As a first step towards the above mentioned Corollary 3.20, in section 1.6 we present another way on how to derive uniqueness of the invariant measure.

Chapter 2 deals with finding a dual representation for the Wasserstein distance in terms of Lipschitz continuous functions. There we work - leaded by the book of Rachev - in a much more general framework than it is required by the application in chapter 3. More precisely, we first prove dual representations for the Monge-Kantorovich and the Kantorovich-Rubinstein problem respectively, meaning that the cost function is not neccessarily a metric. Then in section 2.3 we link both problems by showing that they coincide in case of a metric taking the role of the cost function. In particular, the Monge-Kantorovich problem equals the dual representation of the Kantorovich-Rubinstein problem when the cost function is a metric. Section 2.4 slightly generalizes this for a pseudo-metric. At this point it should be mentioned that chapter 2 is independent of the rest of this thesis and worth reading in its own. Unequivocal Rachev's book is a standard reference in the field of optimal transportation. In particular one is often refered to it for the proof of the (multi-dimensional) Kantorovich Theorem (cf. Theorem 5.2.1 in [12]). Nevertheless this proof is done quite rough and many details are left to the reader. So we decided to work out this part of his monograph in all details at least for the two-dimensional case (cf. Theorem 2.1 below).

The main content of *Chapter 3* is to introduce the asymptotic strong Feller property and to give a sufficient condition. As a basic preparation it is shown in Corollary 3.7 that the total variation distance of two Borel probability measures is the limit of a sequence of their Wasserstein distances corresponding to a totally separating system of (continuous) pseudo-metrics. Having this in mind, we can define what we mean by saying that a Markov semigroup is asymptotically strong Feller. While the rest of section 3.3 examines the relation to the strong Feller property, Theorem 3.13 provides a sufficient condition for the asymptotic strong Feller property to hold. Afterwards the power of the asymptotic strong Feller property is shown even in the finite-dimensional setting by considering two quite simple examples of stochastic differential equations whose Markov semigroups are asymptotically strong Feller but not strong Feller. As the final result of this chapter it is shown in Corollary 3.20 that the asymptotic strong Feller property combined with some kind of irreducibility condition implies the uniqueness of the invariant Borel probability measure for the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$.

Finally, in *Chapter 4* we show the asymptotic strong Feller property for (Markov semigroups associated to) stochastic differential equations of type

$$\begin{aligned} dX(t) &= (AX(t) + F(X(t)))\, dt + B\, dW(t) \\ X(0) &= x \end{aligned}$$

on a real separable Hilbert space, where F is assumed to be Lipschitz (cf. Theorem 4.4 below). Contrary to former works proving the strong Feller property for such equations we do *not* impose the operator B to be continuously invertible in order to apply the Bismut-Elworthy formula. Unfortunately we have to require that A is of (sufficiently large) negative type. Nonetheless this is a quite large class of semilinear SPDE's. For example A can be chosen to be the Laplacian Δ.

Note that usually the asymptotic strong Feller property is shown by transforming a regularity problem into a linear control problem via techniques from Malliavin calculus, see e.g. [7]. A very similar situation to the above is treated in [6]. Since even there the asymptotic strong Feller property is shown in that way, we have good reason to conjecture that at least the idea of our proof is entirely new. In particular, we do not use tools from Malliavin calculus in order to show the asymptotic strong Feller property. However the considerations in this chapter are meant to be of an exemplary character so that even more applications are expected.

Acknowledgements

I would like to thank Prof. Dr. Michael Röckner who led me to the study of stochastic analysis. While visiting his lectures I was infected by his enthusiasm for this interesting mathematical field. The opportunity to get an insight into the latest research motivated me many times. I am grateful for his encouragement and the patience in answering my questions during writing this thesis.

Furthermore, I would like to thank Dr. Oleksandr Kutovyi for useful suggestions and helpful discussions.

Moreover, I would like to thank my fellow students, especially Norbert Breimhorst, Dennis Hagedorn, Florian Knäble, Vitali Peil and Lukas Sowa.

Finally, special thanks are given to my parents, my brother Julian and my sister Katharina Maria.

1 Invariant measures for Markov semigroups

This chapter is a slightly modificated composition of chapters 2.1, 3, 4 in [11], chapters 5 and 7 in [10] as well as chapters 6 and 7 in [6].

1.1 Markov semigroups

Let $\mathcal{H}$ be a real separable Hilbert space and denote by $\mathcal{B}(\mathcal{H})$ its Borel field. Let $\mathcal{B}_b(\mathcal{H})$ be the space of all bounded, Borel measurable functions $\varphi\colon \mathcal{H} \to \mathbb{R}$ and denote by $\mathcal{C}_b(\mathcal{H})$ the subspace of all continuous and bounded functions on $\mathcal{H}$ with values in $\mathbb{R}$. Furthermore $L(\mathcal{B}_b(\mathcal{H}))$ denotes the space of all linear bounded operators from $\mathcal{B}_b(\mathcal{H})$ into itself.

Definition 1.1. *A function $\kappa\colon \mathcal{H} \times \mathcal{B}(\mathcal{H}) \to [0, \infty[$ is called a* transition kernel *on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ if*

(i) $x \mapsto \kappa(x, A)$ is measurable for every $A \in \mathcal{B}(\mathcal{H})$ and

(ii) $A \mapsto \kappa(x, A)$ is a measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ for every $x \in \mathcal{H}$.

The transition kernel κ is said to be Markovian *if $\kappa(x, \mathcal{H}) = 1$ for all $x \in \mathcal{H}$, e.g. $\kappa(x, \cdot)$ is a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ for every $x \in \mathcal{H}$.*

So a Markovian transition kernel κ on the measurable space $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ could be thought as a family $(\kappa(x, \cdot))_{x \in \mathcal{H}}$ of probability measures on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ which is measurable in the parameter $x \in \mathcal{H}$. If there is a (Markovian) transition kernel π_t for each time $t \geq 0$, we introduce the following

Definition 1.2. *A family $(\pi_t)_{t \geq 0}$ of (Markovian) transition kernels on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ is called a* (Markovian) semigroup of transition kernels *on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ if $\pi_{t+s} = \pi_t \pi_s$ for all $t, s \in [0, \infty[$, i.e.*

$$\pi_{t+s}(x, A) = \int_{\mathcal{H}} \pi_s(y, A)\, \pi_t(x, dy) \tag{1.1}$$

for all $t, s \geq 0$, $x \in \mathcal{H}$ and $A \in \mathcal{B}(\mathcal{H})$.

Equation (1.1) is called *Chapman-Kolmogorov equation.* A heuristic interpretation of this equality is given by: the probability for a particle starting at time 0 in $x \in \mathcal{H}$ to be in $A \subset \mathcal{H}$ at time $t + s$ (left hand side) is equal to the probability the particle starts at time 0 in $x \in \mathcal{H}$ and being in some infinitesimal small volume dy at time t and then starting new in $y \in \mathcal{H}$ at time t and being in the subset A at time $t + s$ integrated over all 'intermediate points' $y \in \mathcal{H}$ (right hand side).

Definition 1.3. *A family* $(\mathcal{P}_t)_{t\geq 0}$ *of linear bounded operators on* $B_b(\mathcal{H})$ *is called a* Markov semigroup *if*

(i) $\mathcal{P}_0 = 1$;

(ii) $\mathcal{P}_{t+s} = \mathcal{P}_t\mathcal{P}_s$ *for all* $t, s \geq 0$;

(iii) For any $t \geq 0$ *and* $x \in \mathcal{H}$ *there exists a probability measure* $\pi_t(x, \cdot)$ *on* $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ *such that*

$$\mathcal{P}_t\varphi(x) = \int_{\mathcal{H}} \varphi(y)\, \pi_t(x, dy) \tag{1.2}$$

for all $\varphi \in \mathcal{B}_b(\mathcal{H})$.

If, in addition, for any $\varphi \in \mathcal{C}_b(\mathcal{H})$ *and* $x \in \mathcal{H}$ *the mapping*

$$t \mapsto P_t\varphi(x) \tag{1.3}$$

is continuous on $[0, \infty[$, *the Markov semigroup is said to be* stochastically continuous.

Remark 1.4. *(1) For* $t = 0$ *(1.2) yields* $\pi_0(x, \cdot) = \delta_x(\cdot)$. *Furthermore by (1.2) it follows that for any* $A \in \mathcal{B}(\mathcal{H})$

$$\mathcal{P}_t\mathbf{1}_A(x) = \pi_t(x, A) \tag{1.4}$$

for all $t \geq 0$, $x \in \mathcal{H}$.

(2) Since for arbitrary $A \in \mathcal{B}(\mathcal{H})$

$$\pi_{t+s}(x, A) = \mathcal{P}_{t+s}\mathbf{1}_A(x) = \mathcal{P}_t\mathcal{P}_s\mathbf{1}_A(x) = \mathcal{P}_t\pi_s(\cdot, A)(x) = \int_{\mathcal{H}} \pi_s(y, A)\, \pi_t(x, dy)$$

for all $t, s \geq 0$, *the probability measures* $\pi_t(x, \cdot)$, $x \in \mathcal{H}$, $t \geq 0$, *realizing (1.2) will in fact form a Markovian semigroup of transition kernels. Hence (iii) in the above definition could be reformulated as:*

(iii') There exists a Markovian semigroup of transition kernels, $(\pi_t)_{t\geq 0}$, *on the measurable space* $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ *such that*

$$\mathcal{P}_t\varphi(x) = \int_{\mathcal{H}} \varphi(y)\, \pi_t(x, dy)$$

for all $\varphi \in \mathcal{B}_b(\mathcal{H})$ *and for any* $t \geq 0$, $x \in \mathcal{H}$.

(3) For any $t \geq 0$ *the operator* $\mathcal{P}_t$ *preserves positivity, i.e.* $\mathcal{P}_t\varphi \geq 0$ *for all* $\varphi \geq 0$. *In particular,* $\mathcal{P}_t\mathbf{1} = \mathbf{1}$ *for all* $t \geq 0$.

(4) Let $\varphi \in \mathcal{B}_b(\mathcal{H})$. *Since by definition of the supremum norm* $|\varphi(x)| \leq \|\varphi\|_\infty$ *for all* $x \in \mathcal{H}$, *we have* $|\mathcal{P}_t\varphi(x)| \leq \|\varphi\|_\infty$ *for all* $x \in \mathcal{H}$, $t \geq 0$. *Hence the operator norm* $\|\mathcal{P}_t\|_{L(\mathcal{B}_b(\mathcal{H}))} \leq 1$ *for all* $t \geq 0$, *that is* $(\mathcal{P}_t)_{t\geq 0}$ *is a semigroup of contractions on* $\mathcal{B}_b(\mathcal{H})$.

Let d be some metric generating the topology of $\mathcal{H}$ and denote by $B(x,\delta) := \{y \in \mathcal{H} \mid d(x,y) < \delta\}$ the open ball of radius $\delta > 0$ centered at $x \in \mathcal{H}$. Furthermore let $\mathcal{UC}_b(\mathcal{H})$ (respectively $\mathcal{L}_d(\mathcal{H})$) be the space of all bounded and uniformly continuous (respectively d-Lipschitz-continuous) functions on $\mathcal{H}$ with values in $\mathbb{R}$.

Lemma 1.5. *A Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ is stochastically continuous if and only if one of the following equivalent conditions holds:*

(i) $\lim_{t\to 0} \pi_t(x, B(x,\delta)) = 1$ *for all* $x \in \mathcal{H}$, $\delta > 0$;

(ii) $\lim_{t\to 0} \mathcal{P}_t\varphi(x) = \varphi(x)$ *for all* $\varphi \in \mathcal{UC}_b(\mathcal{H})$, $x \in \mathcal{H}$;

(iii) $\lim_{t\to 0} \mathcal{P}_t\varphi(x) = \varphi(x)$ *for all* $\varphi \in \mathcal{L}_d(\mathcal{H})$, $x \in \mathcal{H}$.

Proof. Obviously, (1.3) implies (ii) and (ii) implies (iii). So it is enough to show that (iii) implies (i) and (i) implies (1.3).

Let $\varphi \in \mathcal{C}_b(\mathcal{H})$, $x \in \mathcal{H}$. Then for each $\delta > 0$ we have

$$\begin{aligned}
&|\mathcal{P}_t\varphi(x) - \varphi(x)| \\
=& \left|\int_{\mathcal{H}} (\varphi(y) - \varphi(x))\ \pi_t(x,dy)\right| \\
=& \left|\int_{B(x,\delta)} (\varphi(y) - \varphi(x))\ \pi_t(x,dy) + \int_{(B(x,\delta))^c} (\varphi(y) - \varphi(x))\ \pi_t(x,dy)\right| \\
\leq& \sup_{y\in B(x,\delta)} |\varphi(y) - \varphi(x)| + 2 \cdot \|\varphi\|_\infty \cdot (1 - \pi_t(x, B(x,\delta)))\,.
\end{aligned}$$

Since by (i) $\lim_{t\to 0} \pi_t(x, B(x,\delta)) = 1$, the second summand tends to 0 for $t \to 0$. Furthermore, letting $\delta \to 0$ the first summand vanishes because $\varphi \in \mathcal{C}_b(\mathcal{H})$. Hence the continuity of $t \mapsto \mathcal{P}_t\varphi(x)$ in zero is proved and by a straightforward computation using the semigroup property we gain the continuity for an arbitrary time $t > 0$. Therefore (i) implies (1.3).

To show that (iii) implies (i), first note that if $\varphi, \psi \in \mathcal{L}_d(\mathcal{H})$ then also $c \cdot \varphi$, $\varphi \vee \psi \in \mathcal{L}_d(\mathcal{H})$ for all $c \in \mathbb{R}$. For arbitrary $x \in \mathcal{H}$ and $\delta > 0$ define

$$\varphi(y) := \begin{cases} 1 - \frac{d(y,x)}{\delta} & \text{if} \quad y \in B(x,\delta) \\ 0 & \text{if} \quad y \in (B(x,\delta))^c \end{cases} = \frac{1}{\delta} \cdot ((\delta - d(y,x)) \vee 0)\,.$$

Then $\varphi \in \mathcal{L}_d(\mathcal{H})$ and

$$\begin{aligned}
\varphi(x) - \mathcal{P}_t\varphi(x) &= 1 - \int_{\mathcal{H}} \varphi(y)\ \pi_t(x,dy) \\
&= 1 - \int_{B(x,\delta)} \varphi(y)\ \pi_t(x,dy) \geq 1 - \pi_t(x, B(x,\delta)).
\end{aligned}$$

Consequently (iii) implies (i), because the right hand side is greater or equal 0. □

For the rest of the chapter $(\mathcal{P}_t)_{t\geq 0}$ is assumed to be a stochastically continuous Markov semigroup.

Definition 1.6. *The Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ is called*

(i) regular at time $t > 0$ *if all probability measures $\pi_t(x,\cdot)$, $x \in \mathcal{H}$, are mutually equivalent. It is called* regular *if it is regular at all times $t > 0$.*

(ii) Feller at time $t \geq 0$ *if $\mathcal{P}_t\varphi \in \mathcal{C}_b(\mathcal{H})$ for any $\varphi \in \mathcal{C}_b(\mathcal{H})$. It is called* Feller *if it is* Feller *at all times $t \geq 0$.*

(iii) irreducible at time $t > 0$ *if $\mathcal{P}_t\mathbf{1}_{B(x_0,\delta)}(x) > 0$ for all $x, x_0 \in \mathcal{H}$, $\delta > 0$. It is called* irreducible *if it is irreducible at all times $t > 0$.*

Remark 1.7. *(1) Note that the definitions in (ii) and (iii) depend on the topology of $\mathcal{H}$.*

(2) If the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ is regular at time $s > 0$, i.e. the probability measures $\pi_s(x,\cdot)$, $x \in \mathcal{H}$, are mutually equivalent, then it is regular for all times $u \geq s$. Moreover, all probability measures $\pi_u(x,\cdot)$, $x \in \mathcal{H}$, $u \geq s$, are mutually equivalent.

Clearly, it would be enough to show the last assertion. So, let $x, y \in \mathcal{H}$ and first suppose $\pi_s(x, A) = 0$ for some $A \in \mathcal{B}(\mathcal{H})$. Then using (1.1)

$$\pi_u(y, A) = \pi_{(u-s)+s}(y, A) = \int_{\mathcal{H}} \pi_s(z, A)\, \pi_{u-s}(y, dz) = 0,$$

because $\pi_s(z,\cdot) \approx \pi_s(x,\cdot)$ for all $z \in \mathcal{H}$. Hence $\pi_s(x,\cdot) \gg \pi_u(y,\cdot)$. Now take $A \in \mathcal{B}(\mathcal{H})$ such that $\pi_u(y, A) = 0$. Then from the above equality we conclude that $\pi_s(z, A) = 0$ for $\pi_{u-s}(y,\cdot)$-a.e. $z \in \mathcal{H}$. But $\pi_s(z,\cdot)$, $z \in \mathcal{H}$, are mutually equivalent and thus $\pi_s(z, A) = 0$ for every $z \in \mathcal{H}$. In particular, $\pi_s(x, A) = 0$, that is $\pi_s(x,\cdot) \ll \pi_u(y,\cdot)$. Alltogether we have shown that $\pi_s(x,\cdot) \approx \pi_u(y,\cdot)$ for all $x, y \in \mathcal{H}$, $u \geq s$.

(3) If the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ is irreducible at time s, i.e. $\mathcal{P}_s\mathbf{1}_{B(x_0,\delta)}(x) > 0$ for all $x, x_0 \in \mathcal{H}$, $\delta > 0$, then it is irreducible for all times $u \geq s$: for $x, x_0 \in \mathcal{H}$, $\delta > 0$ and $u \geq s$ we obtain

$$\begin{aligned} \mathcal{P}_u\mathbf{1}_{B(x_0,\delta)}(x) &= \mathcal{P}_{(u-s)+s}\mathbf{1}_{B(x_0,\delta)}(x) \\ &= \mathcal{P}_{u-s}\mathcal{P}_s\mathbf{1}_{B(x_0,\delta)}(x) \\ &= \int_{\mathcal{H}} \underbrace{\mathcal{P}_s\mathbf{1}_{B(x_0,\delta)}(y)}_{>0}\ \pi_{u-s}(x, dy) \\ &> 0, \end{aligned}$$

because $(\mathcal{P}_t)_{t\geq 0}$ is irreducible at time s.

1.2 Strong Feller property

Since it is not true in general to have uniqueness of the invariant Borel probability measure from topological irreducibility combined with the Feller property - the counterexmaple is the Ising model (cf. [6], Example 7.4) - we have to replace the latter by a stronger regularity property. The right choice will be the strong Feller property (cf. Theorem 1.35 below).

Definition 1.8. *A Markov semigroup* $(\mathcal{P}_t)_{t\geq 0}$ *is called* strong Feller at time $t > 0$ *if* $\mathcal{P}_t\varphi \in \mathcal{C}_b(\mathcal{H})$ *for any* $\varphi \in \mathcal{B}_b(\mathcal{H})$. *It is called* strong Feller *if it is strong Feller at all times* $t > 0$.

Remark 1.9. *If the Markov semigroup* $(\mathcal{P}_t)_{t\geq 0}$ *is strong Feller at time* s, *i.e.* $\mathcal{P}_s\varphi \in \mathcal{C}_b(\mathcal{H})$ *for all* $\varphi \in \mathcal{B}_b(\mathcal{H})$, *then it is strong Feller for all times* $u \geq s$. *In fact, by the semigroup property for arbitrary* $\varphi \in \mathcal{B}_b(\mathcal{X})$ *we have*

$$\mathcal{P}_u\varphi = \mathcal{P}_{s+(u-s)}\varphi = \mathcal{P}_s\left(\mathcal{P}_{u-s}\varphi\right)$$

and this function is continuous, because $\mathcal{P}_{u-s}\varphi \in \mathcal{B}_b(\mathcal{H})$ *and* $(\mathcal{P}_t)_{t\geq 0}$ *is strong Feller at time* s.

Definition 1.10. *Let* μ *be a finite signed measure on* $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ *with Jordan decomposition* $\mu = \mu^+ - \mu^-$. *Then the* total variation norm *of* μ *is given by*

$$\|\mu\|_{TV} := \frac{1}{2} \cdot \left(\mu^+(\mathcal{H}) + \mu^-(\mathcal{H})\right).$$

In order to prove a sufficient condition for the strong Feller property, we need the following approximation result for continuous functions by twice continuously Frechet differentiable functions:

Lemma 1.11. *Let* $\varphi \in \mathcal{C}_b(\mathcal{H})$. *Then there exists a sequence* $(\varphi_m)_{m\in\mathbb{N}}, \varphi_m \in \mathcal{C}_b^2(\mathcal{H})$, *such that* $\|\varphi_m\|_\infty \leq \|\varphi\|_\infty$ *for all* $m \in \mathbb{N}$ *and* $\varphi_m \overset{m\to\infty}{\longrightarrow} \varphi$ *pointwisely.*

Proof. Let $\varphi \in \mathcal{C}_b(\mathcal{H})$ and $\{e_i\}_{i\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$. For $m \in \mathbb{N}$ define orthogonal projections

$$\begin{aligned} P_m\colon \mathcal{H} &\rightarrow P_m(\mathcal{H}) = \operatorname{span}\{e_1, \ldots, e_m\} \\ x &\mapsto P_m x := \sum_{i=1}^{m} \langle x, e_i\rangle \cdot e_i \end{aligned}$$

and corresponding functions

$$\begin{aligned} J_m\colon P_m(\mathcal{H}) &\rightarrow \mathbb{R}^m \\ \sum_{i=1}^{m} \langle x, e_i\rangle \cdot e_i &\mapsto (\langle x, e_1\rangle, \ldots, \langle x, e_m\rangle), \end{aligned}$$

which are bijections.

Note that $\varphi \circ J_m^{-1} \in \mathcal{C}_b(\mathbb{R}^m)$ for all $m \in \mathbb{N}$. Hence by Lemma A.2 in the appendix, for every $m \in \mathbb{N}$ there exists a sequence $(f_{m,k})_{k\in\mathbb{N}}$, $f_{m,k} \in \mathcal{C}_b^2(\mathbb{R}^m)$, such that $f_{m,k}(x) \overset{k\to\infty}{\longrightarrow} (\varphi \circ J_m^{-1})(x)$ for all $x \in \mathbb{R}^m$ and $\|f_{m,k}\|_\infty \leq \|\varphi \circ J_m^{-1}\|_\infty$ for all $k \in \mathbb{N}$. Replacing x by $J_m(x)$ yields $(f_{m,k} \circ J_m)(x) \overset{k\to\infty}{\longrightarrow} \varphi(x)$ for all $x \in P_m(\mathcal{H})$. Consequently $(f_{m,k} \circ J_m \circ P_m)(x) \overset{k\to\infty}{\longrightarrow} (\varphi \circ P_m)(x)$ for all $x \in \mathcal{H}$. Moreover $\|f_{m,k} \circ J_m \circ P_m\|_\infty \leq \|\varphi \circ P_m\|_\infty$ for all $k \in \mathbb{N}$. Since $\varphi \circ P_m \overset{m\to\infty}{\longrightarrow} \varphi$ pointwisely and $\|\varphi \circ P_m\|_\infty \leq \|\varphi\|_\infty$ for all $m \in \mathbb{N}$, a diagonal argument applies in order to obtain a subsequence $(\varphi_m)_{m\in\mathbb{N}}$, $\varphi_m := f_{m,m} \circ J_m \circ P_m \in \mathcal{C}_b^2(\mathcal{H})$, such that $\varphi_m(x) \overset{m\to\infty}{\longrightarrow} \varphi(x)$ for all $x \in \mathcal{H}$ and $\|\varphi_m\|_\infty \leq \|\varphi\|_\infty$ for all $m \in \mathbb{N}$. □

Theorem 1.12. *Let $(\mathcal{P}_t)_{t\geq 0}$ be a Markov semigroup on $\mathcal{B}_b(\mathcal{H})$ and let $c > 0$ and $t > 0$ be fixed. Then the following conditions are equivalent:*

(i) For all $\varphi \in \mathcal{C}_b^2(\mathcal{H})$ $|\mathcal{P}_t\varphi(x) - \mathcal{P}_t\varphi(y)| \leq c \cdot \|\varphi\|_\infty \cdot \|x - y\|$ for all $x, y \in \mathcal{H}$, that is $\mathcal{P}_t\varphi(\cdot)$ is Lipschitz continuous for all $\varphi \in \mathcal{C}_b^2(\mathcal{H})$.

(ii) For all $\varphi \in \mathcal{B}_b(\mathcal{H})$ $|\mathcal{P}_t\varphi(x) - \mathcal{P}_t\varphi(y)| \leq c \cdot \|\varphi\|_\infty \cdot \|x - y\|$ for all $x, y \in \mathcal{H}$, that is $\mathcal{P}_t\varphi(\cdot)$ is Lipschitz continuous for all $\varphi \in \mathcal{B}_b(\mathcal{H})$.

(iii) $\|\pi_t(x,\cdot) - \pi_t(y,\cdot)\|_{TV} \leq \frac{c}{2} \cdot \|x - y\|$ for all $x, y \in \mathcal{H}$, that is the Markovian transition kernel π_t is Lipschitz continuous in the first parameter with respect to the total variation norm $\|\cdot\|_{TV}$.

In particular, if one of the above conditions holds (for all $t > 0$), the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ is strong Feller.

Proof. *(iii)* ⇒ *(ii)*: Let $\varphi \in \mathcal{B}_b(\mathcal{H})$ and $x, y \in \mathcal{H}$. Then

$$\begin{aligned}
|\mathcal{P}_t\varphi(x) - \mathcal{P}_t\varphi(y)| &= \left| \int_{\mathcal{H}} \varphi(z)\ \pi_t(x, dz) - \int_{\mathcal{H}} \varphi(z)\ \pi_t(y, dz) \right| \\
&= \left| \int_{\mathcal{H}} \varphi(z)\ \ (\pi_t(x, dz) - \pi_t(y, dz)) \right| \\
&\leq \int_{\mathcal{H}} |\varphi(z)|\ \ |\pi_t(x, dz) - \pi_t(y, dz)| \\
&\leq 2 \cdot \|\varphi\|_\infty \cdot \|\pi_t(x,\cdot) - \pi_t(y,\cdot)\|_{TV} \\
&\leq c \cdot \|\varphi\|_\infty \cdot \|x - y\|.
\end{aligned}$$

(ii) ⇒ *(i)*: Obvious, because $\mathcal{C}_b^2(\mathcal{H}) \subset \mathcal{B}_b(\mathcal{H})$.

(i) ⇒ *(iii)*: Define

$$\mathcal{K}_1 := \{\varphi \in \mathcal{C}_b(\mathcal{H}) \mid \|\varphi\|_\infty \leq 1\}$$

and

$$\mathcal{K}_2 := \{\varphi \in \mathcal{C}_b^2(\mathcal{H}) \mid \|\varphi\|_\infty \leq 1\}.$$

Obviously, $\mathcal{K}_2 \subseteq \mathcal{K}_1$ and therefore $\sup_{\varphi\in\mathcal{K}_2}|\mathcal{P}_t\varphi(x)-\mathcal{P}_t\varphi(y)| \le \sup_{\varphi\in\mathcal{K}_1}|\mathcal{P}_t\varphi(x)-\mathcal{P}_t\varphi(y)|$. But by Lemma 1.11 for each $\varphi \in \mathcal{K}_1$ there exists a sequence $(\varphi_n)_{n\in\mathbb{N}}$, $\varphi_n \in \mathcal{K}_2$, such that $\varphi_n \overset{n\to\infty}{\longrightarrow} \varphi$ pointwisely. Hence by Lebesgues dominated convergence theorem

$$
\begin{aligned}
|\mathcal{P}_t\varphi(x) - \mathcal{P}_t\varphi(y)| &= \left|\int_{\mathcal{H}} \varphi(z)\ \left(\pi_t(x,dz) - \pi_t(y,dz)\right)\right| \\
&= \left|\int_{\mathcal{H}} \lim_{n\to\infty} \varphi_n(z)\ \left(\pi_t(x,dz) - \pi_t(y,dz)\right)\right| \\
&= \lim_{n\to\infty} \left|\int_{\mathcal{H}} \varphi_n(z)\ \left(\pi_t(x,dz) - \pi_t(y,dz)\right)\right| \\
&= \lim_{n\to\infty} |\mathcal{P}_t\varphi_n(x) - \mathcal{P}_t\varphi_n(y)| \\
&\le \sup_{n\in\mathbb{N}} |\mathcal{P}_t\varphi_n(x) - \mathcal{P}_t\varphi_n(y)| \\
&\le \sup_{\varphi\in\mathcal{K}_2} |\mathcal{P}_t\varphi(x) - \mathcal{P}_t\varphi(y)|
\end{aligned}
$$

for all $x, y \in \mathcal{H}$ and any $\varphi \in \mathcal{K}_1$. Therefore $\sup_{\varphi\in\mathcal{K}_1}|\mathcal{P}_t\varphi(x)-\mathcal{P}_t\varphi(y)| \le \sup_{\varphi\in\mathcal{K}_2}|\mathcal{P}_t\varphi(x)-\mathcal{P}_t\varphi(y)|$ for all $x, y \in \mathcal{H}$ and alltogether

$$
\sup_{\varphi\in\mathcal{K}_1} |\mathcal{P}_t\varphi(x) - \mathcal{P}_t\varphi(y)| = \sup_{\varphi\in\mathcal{K}_2} |\mathcal{P}_t\varphi(x) - \mathcal{P}_t\varphi(y)| \tag{1.5}
$$

for all $x, y \in \mathcal{H}$. Furthermore, as a simple consequence of the Hahn-Banach theorem we have

$$
\sup_{\varphi\in\mathcal{K}_1} |\mathcal{P}_t\varphi(x) - \mathcal{P}_t\varphi(y)| = 2 \cdot \|\pi_t(x,\cdot) - \pi_t(y,\cdot)\|_{TV}. \tag{1.6}
$$

Combining (1.5) and (1.6) and applying (i), the assertion in (iii) follows. □

In order to compare the later introduced asymptotic strong Feller property with the strong Feller property, we stress the following proposition, which is in fact part of the previous theorem:

Proposition 1.13. *Let $\mathcal{H}$ be a separable Hilbert space and $(\mathcal{P}_t)_{t\ge0}$ a Markov semigroup on $\mathcal{B}_b(\mathcal{H})$. If for all functions $\varphi \in \mathcal{B}_b(\mathcal{H})$*

$$
|\mathcal{P}_t\varphi(x) - \mathcal{P}_t\varphi(y)| \le C(\|x\| \vee \|y\|) \cdot \|\varphi\|_\infty \cdot \|x - y\| \tag{1.7}
$$

for all $x, y \in \mathcal{H}$, $t > 0$, where $C\colon \mathbb{R}_+ \to \mathbb{R}$ is a fixed nondecreasing function, then $(\mathcal{P}_t)_{t\ge0}$ is strong Feller.

Note that the equivalence of (ii) and (iii) in Theorem 1.12 remains valid for a Polish space $\mathcal{X}$. In particular we have the following characterization of the strong Feller property, which will be needed in chapter 3:

Theorem 1.14. *Let $\mathcal{X}$ be a Polish space and $(\mathcal{P}_t)_{t\ge0}$ a Markov semigroup on $\mathcal{B}_b(\mathcal{X})$. Then $(\mathcal{P}_t)_{t\ge0}$ is strong Feller if and only if for all $t > 0$ the transition probabilities $\pi_t(x,\cdot)$ are continuous in the parameter x with respect to the total variation norm $\|\cdot\|_{TV}$.*

Proof. Let $t > 0$ and $x \in \mathcal{X}$ fixed.

First assume that the transition probabilities $\pi_t(x, \cdot)$ are continuous in x with respect to the total variation norm $\|\cdot\|_{TV}$. Let $\varphi \in \mathcal{B}_b(\mathcal{X})$ and take a sequence $(x_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}$ such that $d(x_n, x) \longrightarrow 0$ as $n \to \infty$. Then

$$\begin{aligned} |\mathcal{P}_t\varphi(x_n) - \mathcal{P}_t\varphi(x)| &= \Big| \int_{\mathcal{X}} \varphi(y)\ \pi_t(x_n, dy) - \int_{\mathcal{X}} \varphi(y)\ \pi_t(x, dy) \Big| \\ &= \Big| \int_{\mathcal{X}} \varphi(y)\ \big(\pi_t(x_n, dy) - \pi_t(x, dy)\big) \Big| \\ &\leq \int_{\mathcal{X}} |\varphi(y)|\ |\pi_t(x_n, dy) - \pi_t(x, dy)| \\ &\leq \|\varphi\|_\infty \cdot \int_{\mathcal{X}} |\pi_t(x_n, dy) - \pi_t(x, dy)| \\ &= 2 \cdot \|\varphi\|_\infty \cdot \|\pi_t(x_n, \cdot) - \pi_t(x, \cdot)\|_{TV} \end{aligned}$$

and this tends to zero for $n \to \infty$, since the transition probabilities $\pi_t(x, \cdot)$ are continuous in x with respect to the total variation norm $\|\cdot\|_{TV}$.

The converse direction follows immediately by the following lemma. □

Lemma 1.15. *Let $\mathcal{X}$ be a Polish space and P, Q be two Markov operators on $\mathcal{B}_b(\mathcal{X})$ that are strong Feller. Then the product PQ is a Markov operator whose transition probabilities $\pi(x, \cdot)$ are continuous in x with respect to the total variation norm $\|\cdot\|_{TV}$.*

Proof. The proof is an immediate consequence of Theorem B.2 and Lemma B.5 in the appendix. □

1.3 Invariant measures

Definition 1.16. *A Borel probability measure μ on $\mathcal{H}$ is called* invariant *for the (stochastically continuous) Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ if*

$$\int_{\mathcal{H}} \mathcal{P}_t\varphi(x)\ \mu(dx) = \int_{\mathcal{H}} \varphi(x)\ \mu(dx) \tag{1.8}$$

for all $\varphi \in \mathcal{B}_b(\mathcal{H})$, $t \geq 0$. The set of all invariant measures for $(\mathcal{P}_t)_{t\geq 0}$ is denoted by $\mathcal{J}(\mathcal{P}_t)$.

Remark 1.17. *$\mathcal{J}(\mathcal{P}_t)$ is convex: Let $\mu, \nu \in \mathcal{J}(\mathcal{P}_t)$ and $\alpha \in [0, 1]$. Then*

$$\begin{aligned} \int_{\mathcal{H}} \mathcal{P}_t\varphi(x)\ \left(\alpha\mu + (1-\alpha)\nu\right)(dx) &= \alpha \int_{\mathcal{H}} \mathcal{P}_t\varphi(x)\ \mu(dx) + (1-\alpha) \int_{\mathcal{H}} \mathcal{P}_t\varphi(x)\ \nu(dx) \\ &= \alpha \int_{\mathcal{H}} \varphi(x)\ \mu(dx) + (1-\alpha) \int_{\mathcal{H}} \varphi(x)\ \nu(dx) \\ &= \int_{\mathcal{H}} \varphi(x)\ \left(\alpha\mu + (1-\alpha)\nu\right)(dx). \end{aligned}$$

for all $\varphi \in \mathcal{B}_b(\mathcal{H})$, $t \geq 0$. Therefore $\alpha\mu + (1-\alpha)\nu \in \mathcal{J}(\mathcal{P}_t)$.

Denote by $\mathcal{M}(\mathcal{H})$ the space of all Borel measures on $\mathcal{H}$ and define $\mathcal{M}_1(\mathcal{H}) := \{\mu \in \mathcal{M}(\mathcal{H}) \mid \mu(\mathcal{H}) = 1\}$ to be the subspace of all Borel probability measures on $\mathcal{H}$. Recall that there is a natural embedding of the space of all probability measures on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ into the space of all continuous linear functionals on $\mathcal{C}_b(\mathcal{H})$: $\mathcal{M}_1(\mathcal{H}) \hookrightarrow \mathcal{C}_b(\mathcal{H})^*$. Namely, for any $\mu \in \mathcal{M}_1(\mathcal{H})$ set

$$F_\mu(\varphi) := \int_{\mathcal{H}} \varphi(x)\, \mu(dx)$$

for all $\varphi \in \mathcal{C}_b(\mathcal{H})$. Identifying $\mu \in \mathcal{M}_1(\mathcal{H})$ with $F_\mu \in \mathcal{C}_b(\mathcal{H})^*$, we obtain an alternative characterization of invariant measures:

Lemma 1.18. *Let $(\mathcal{P}_t)_{t\geq 0}$ be a Feller Markov semigroup on $\mathcal{B}_b(\mathcal{H})$. Then $\mu \in \mathcal{J}(\mathcal{P}_t)$ if and only if*

$$\mathcal{P}_t^* \mu = \mu \tag{1.9}$$

for all $t \geq 0$, where $\mathcal{P}_t^*$ is the transpose operator of $\mathcal{P}_t$, defined as ${}_{\mathcal{C}_b(\mathcal{H})^*}\langle \mathcal{P}_t^* F, \varphi\rangle_{\mathcal{C}_b(\mathcal{H})} = {}_{\mathcal{C}_b(\mathcal{H})^*}\langle F, \mathcal{P}_t\varphi\rangle_{\mathcal{C}_b(\mathcal{H})}$ for all $\varphi \in \mathcal{C}_b(\mathcal{H})$, $F \in \mathcal{C}_b(\mathcal{H})^*$.

Proof. First, if μ is invariant for $(\mathcal{P}_t)_{t\geq 0}$ we have

$$\mathcal{P}_t^*\mu(\varphi) = \mathcal{P}_t^* F_\mu(\varphi) = F_\mu(\mathcal{P}_t\varphi) = \int_{\mathcal{H}} \mathcal{P}_t\varphi(x)\, \mu(dx) = \int_{\mathcal{H}} \varphi(x)\, \mu(dx) = F_\mu(\varphi) = \mu(\varphi)$$

for all $\varphi \in \mathcal{C}_b(\mathcal{H})$ and this is (1.9).

Conversly, suppose (1.9) holds. Changing the order in the above calculation shows (1.8) for all $\varphi \in \mathcal{C}_b(\mathcal{H})$. Now a monotone class argument applies in order to show it for all $\varphi \in \mathcal{B}_b(\mathcal{H})$. □

In particular, if $\mu \in \mathcal{M}_1(\mathcal{H})$ is invariant for $(\mathcal{P}_t)_{t\geq 0}$ we have

$$\mu(A) = \int_{\mathcal{H}} \pi_t(x, A)\, \mu(dx) \tag{1.10}$$

for all $A \in \mathcal{B}(\mathcal{H})$. Equality (1.10) should be interpreted by saying that the μ-mass of A is given by the spatial μ-average of all $\pi_t(x, \cdot)$-masses of A, $x \in \mathcal{H}$.

The next theorem is basic to the study of Markov semigroups.

Theorem 1.19. *Assume that μ is an invariant Borel probability measure for the (stochastically continuous) Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$. Then for all $t \geq 0$, $p \geq 1$ $\mathcal{P}_t$ is uniquely extendible to a linear bounded operator on $L^p(\mathcal{H}, \mu)$ that we still denote by $\mathcal{P}_t$. Moreover*

$$\|\mathcal{P}_t\|_{L(L^p(\mathcal{H},\mu))} \leq 1 \tag{1.11}$$

for all $t \geq 0$. Finally, $(\mathcal{P}_t)_{t\geq 0}$ is a strongly continuous semigroup in $L^p(\mathcal{H}, \mu)$.

Proof. Let $\varphi \in \mathcal{C}_b(\mathcal{H})$. Using the Hölder inequality we obtain

$$|\mathcal{P}_t\varphi(x)|^p \leq \int_{\mathcal{H}} |\varphi(y)|^p \; \pi_t(x, dy) = \mathcal{P}_t(|\varphi|^p)(x).$$

Integrating both sides of this inequality with respect to μ over $\mathcal{H}$ yields

$$\|\mathcal{P}_t\varphi\|_p^p = \int_{\mathcal{H}} |\mathcal{P}_t\varphi(x)|^p \; \mu(dx) \leq \int_{\mathcal{H}} \mathcal{P}_t(|\varphi|^p)(x) \; \mu(dx) = \int_{\mathcal{H}} |\varphi(x)|^p \; \mu(dx) = \|\varphi\|_p^p$$

according to the invariance of μ. Since $\mathcal{C}_b(\mathcal{H})$ is dense in $L^p(\mathcal{H}, \mu)$, $\mathcal{P}_t$ is uniquely extendible to $L^p(\mathcal{H}, \mu)$ and (1.11) follows.

It remains to show that $(\mathcal{P}_t)_{t\geq 0}$ is strongly continuous in $L^p(\mathcal{H}, \mu)$. In order to do so, first let $\varphi \in \mathcal{C}_b(\mathcal{H})$. Since $(\mathcal{P}_t)_{t\geq 0}$ is stochastically continuous by assumption, we have that the function $t \mapsto \mathcal{P}_t\varphi(x)$ is continuous for any $x \in \mathcal{H}$. Consequently, by the dominated convergence theorem

$$\lim_{t\to 0} \int_{\mathcal{H}} (\mathcal{P}_t\varphi(x))^p \;\; \mu(dx) = \int_{\mathcal{H}} \lim_{t\to 0} (\mathcal{P}_t\varphi(x))^p \;\; \mu(dx) = \int_{\mathcal{H}} (\varphi(x))^p \;\; \mu(dx),$$

i.e. $\mathcal{P}_t\varphi \to \varphi$ in $L^p(\mathcal{H}, \mu)$ as $t \to 0$ for all $\varphi \in \mathcal{C}_b(\mathcal{H})$. Now let $\varphi \in L^p(\mathcal{H}, \mu)$. Since $\mathcal{C}_b(\mathcal{H}) \subset L^p(\mathcal{H}, \mu)$ densely with respect to $\|\cdot\|_p$, there exists a sequence $(\varphi_n)_{n\in\mathbb{N}}$, $\varphi_n \in \mathcal{C}_b(\mathcal{H})$, such that $\|\varphi_n - \varphi\|_p \overset{n\to\infty}{\longrightarrow} 0$. In particular, for $\varepsilon > 0$ we can find some index $N(\varepsilon) \in \mathbb{N}$ such that

$$\|\mathcal{P}_t\varphi - \mathcal{P}_t\varphi_n\|_p = \|\mathcal{P}_t\,(\varphi - \varphi_n)\|_p \leq \|\mathcal{P}_t\|_{L(L^p(\mathcal{H},\mu))} \cdot \|\varphi - \varphi_n\|_p \leq \|\varphi - \varphi_n\|_p \leq \frac{\varepsilon}{3}$$

for all $n \geq N(\varepsilon)$ by choice of the sequence $(\varphi_n)_{n\in\mathbb{N}}$. Moreover, according to the first part for all $n \in \mathbb{N}$ there is some $\delta(n, \varepsilon) > 0$ such that $\|\mathcal{P}_t\varphi_n - \varphi_n\|_p \leq \frac{\varepsilon}{3}$ for all $t \leq \delta(n, \varepsilon)$. Alltogether we then have for some (fixed) $n \geq N(\varepsilon)$

$$\begin{aligned} \|\mathcal{P}_t\varphi - \varphi\|_p &\leq \|\mathcal{P}_t\varphi - \mathcal{P}_t\varphi_n\|_p + \|\mathcal{P}_t\varphi_n - \varphi_n\|_p + \|\varphi_n - \varphi\|_p \\ &\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\ &= \varepsilon \end{aligned}$$

for all $t \leq \delta(n, \varepsilon)$. Therefore $(\mathcal{P}_t)_{t\geq 0}$ is strongly continuous. □

Now suppose that there exists an invariant Borel probability measure μ for the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$. Hence by the previous theorem the Markov semigroup is uniquely extendible to $L^2(\mathcal{H}, \mu)$ and the following definition is meaningful: Denote by Σ the set

$$\Sigma := \{\varphi \in L^2(\mathcal{H}, \mu) : \mathcal{P}_t\varphi = \varphi \;\; \mu\text{-a.s. for all } \; t \geq 0\}$$

of all *stationary* points of $(\mathcal{P}_t)_{t\geq 0}$. Note that in the definition of Σ the nullset depends up on t. Furthermore observe that by Remark 1.4(3) $\mathbf{1} \in \Sigma$. Moreover every function φ that is μ-a.s. constant is contained in Σ.

Remark 1.20. *$\Sigma \subset L^2(\mathcal{H},\mu)$ is closed; i.e. let $(\varphi_n)_{n\in\mathbb{N}} \subset \Sigma$ such that $\|\varphi_n - \varphi\|_2 \to 0$ as $n \to \infty$ for some $\varphi \in L^2(\mathcal{H},\mu)$. We have to prove that $\varphi \in \Sigma$. Let $t \geq 0$ be fixed. Since*

$$\begin{aligned} \|\mathcal{P}_t\varphi - \varphi\|_2 &\leq \|\mathcal{P}_t\varphi - \mathcal{P}_t\varphi_n\|_2 + \|\mathcal{P}_t\varphi_n - \varphi_n\|_2 + \|\varphi_n - \varphi\|_2 \\ &\leq 2\cdot\|\varphi_n - \varphi\|_2 + \underbrace{\|\mathcal{P}_t\varphi_n - \varphi_n\|_2}_{=0} \\ &= 2\cdot\|\varphi_n - \varphi\|_2 \end{aligned}$$

for all $n \in \mathbb{N}$, letting $n \to \infty$ yields $\|\mathcal{P}_t\varphi - \varphi\|_2 = 0$ according to the choice of $(\varphi_n)_{n\in\mathbb{N}}$. Therefore $\mathcal{P}_t\varphi = \varphi$ μ-a.s. for all $t \geq 0$, e.g. $\varphi \in \Sigma$.

Theorem 1.21 (Von Neumann)**.** *For $\varphi \in L^2(\mathcal{H},\mu)$, $T > 0$ define*

$$M(T)\varphi := \frac{1}{T}\int_0^T \mathcal{P}_t\varphi \, dt.$$

There exists the limit

$$M_\infty\varphi := \lim_{T\to\infty} M(T)\varphi \tag{1.12}$$

in $L^2(\mathcal{H},\mu)$. Moreover, M_∞ is a projection operator on Σ and

$$\int_{\mathcal{H}} M_\infty\varphi(x)\,\mu(dx) = \int_{\mathcal{H}} \varphi(x)\,\mu(dx). \tag{1.13}$$

Note that since by Theorem 1.19 the semigroup $(\mathcal{P}_t)_{t\geq 0}$ is strongly continuous in $L^2(\mathcal{H},\mu)$, the term $M(T)\varphi$ above is welldefined.

Proof. For any $T > 0$ we can find $n_T \in \mathbb{N}_0$, $r_T \in [0,1[$ such that $T = n_T + r_T$. According to the Fubini theorem we have for $\varphi \in L^2(\mathcal{H},\mu)$

$$\begin{aligned} M(T)\varphi &= \frac{1}{T}\int_0^T \mathcal{P}_s\varphi\,ds \\ &= \frac{1}{T}\sum_{k=0}^{n_T-1}\int_k^{k+1} \mathcal{P}_s\varphi\,ds + \frac{1}{T}\int_{n_T}^T \mathcal{P}_s\varphi\,ds \\ &= \frac{1}{T}\sum_{k=0}^{n_T-1}\int_0^1 \mathcal{P}_{s+k}\varphi\,ds + \frac{1}{T}\int_0^{r_T} \mathcal{P}_{s+n_T}\varphi\,ds \\ &= \frac{1}{T}\sum_{k=0}^{n_T-1}\int_0^1 \mathcal{P}_k(\mathcal{P}_s\varphi)\,ds + \frac{1}{T}\int_0^{r_T} \mathcal{P}_{n_T}(\mathcal{P}_s\varphi)\,ds \\ &= \frac{1}{T}\sum_{k=0}^{n_T-1}\int_0^1 \left(\int_{\mathcal{X}} \mathcal{P}_s\varphi(y)\,\pi_k(\cdot,dy)\right)ds + \frac{1}{T}\int_0^{r_T}\left(\int_{\mathcal{X}} \mathcal{P}_s\varphi(y)\,\pi_{n_T}(\cdot,dy)\right)ds \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{T}\sum_{k=0}^{n_T-1}\int_{\mathcal{X}}\left(\int_0^1 \mathcal{P}_s\varphi(y)\,ds\right)\ \pi_k(\cdot,dy) + \frac{1}{T}\int_{\mathcal{X}}\left(\int_0^{r_T}\mathcal{P}_s\varphi(y)\,ds\right)\ \pi_{n_T}(\cdot,dy)\\
&= \frac{1}{T}\sum_{k=0}^{n_T-1}\mathcal{P}_k\left(\int_0^1 \mathcal{P}_s\varphi\,ds\right)(\cdot) + \frac{1}{T}\mathcal{P}_{n_T}\left(\int_0^{r_T}\mathcal{P}_s\varphi\,ds\right)(\cdot)\\
&= \frac{1}{T}\sum_{k=0}^{n_T-1}(\mathcal{P}_1)^k\underbrace{\int_0^1\mathcal{P}_s\varphi\,ds}_{=M(1)\varphi} + \frac{r_T}{T}(\mathcal{P}_1)^{n_T}\underbrace{\frac{1}{r_T}\int_0^{r_T}\mathcal{P}_s\varphi\,ds}_{=M(r_T)\varphi}\\
&= \frac{n_T}{T}\cdot\underbrace{\frac{1}{n_T}\sum_{k=0}^{n_T-1}(\mathcal{P}_1)^k M(1)\varphi}_{(*)} + \frac{r_T}{T}(\mathcal{P}_1)^{n_T}M(r_T)\varphi.
\end{aligned}$$

Since $\lim_{T\to\infty}\frac{n_T}{T} = 1$ and $\lim_{T\to\infty}\frac{r_T}{T} = 0$, letting $T\to\infty$ the assertion in (1.12) follows from Theorem 5.11 in [10]. In fact, since $\sup_{n\in\mathbb{N}}\|(\mathcal{P}_1)^n\|_{L(L^2(\mathcal{H},\mu))} \le 1 < \infty$ and $M(1)\varphi\in L^2(\mathcal{H},\mu)$, the assumptions thereby are satisfied and we thus obtain existence of the L^2-limit of the term $(*)$ on the right hand side above as T tends to infinity. Furthermore the second term is bounded in the L^2-norm as it is easily checked.

In order to show that M_∞ is a projection operator on Σ, note that

$$\begin{aligned}
M_\infty(\mathcal{P}_t\varphi) &= \lim_{T\to\infty} M(T)(\mathcal{P}_t\varphi)\\
&= \lim_{T\to\infty}\frac{1}{T}\int_0^T \mathcal{P}_s(\mathcal{P}_t\varphi)\,ds\\
&= \lim_{T\to\infty}\frac{1}{T}\int_0^T \mathcal{P}_{t+s}\varphi\,ds\\
&= \lim_{T\to\infty}\frac{1}{T}\int_t^{t+T} \mathcal{P}_s\varphi\,ds\\
&= \lim_{T\to\infty}\frac{1}{T}\left(\int_0^T \mathcal{P}_s\varphi\,ds - \int_0^t \mathcal{P}_s\varphi\,ds + \int_T^{t+T}\mathcal{P}_s\varphi\,ds\right)\\
&= \lim_{T\to\infty} M(T)\varphi - \lim_{T\to\infty}\frac{1}{T}\int_0^t \mathcal{P}_s\varphi\,ds + \lim_{T\to\infty}\frac{1}{T}\int_T^{t+T}\mathcal{P}_s\varphi\,ds = M_\infty\varphi
\end{aligned}$$

for all $\varphi\in L^2(\mathcal{H},\mu)$ and all $t\ge 0$. Hence

$$M_\infty\mathcal{P}_t = \mathcal{P}_t M_\infty = M_\infty \tag{1.14}$$

for all $t\ge 0$ and $M_\infty\varphi\in\Sigma$ for all $\varphi\in L^2(\mathcal{H},\mu)$ follows. Note that equality in (1.14) can be checked by a similar calculation. Moreover, using (1.14) we have $M_\infty M(T) = M(T)M_\infty = M_\infty$, which yields, letting $T\to\infty$, $M_\infty^2 = M_\infty$. Therefore M_∞ is a projection operator on Σ. Again by the Fubini theorem and according to the invariance of μ the last assertion follows:

$$\begin{aligned}
\int_{\mathcal{H}} M_\infty\varphi(x)\ \mu(dx) &= \int_{\mathcal{H}} \lim_{T\to\infty} M(T)\varphi(x)\ \mu(dx) \\
&= \int_{\mathcal{H}} \lim_{T\to\infty} \left(\frac{1}{T}\int_0^T \mathcal{P}_t\varphi(x)\ dt\right)\ \mu(dx) \\
&= \lim_{T\to\infty} \frac{1}{T}\int_0^T \left(\int_{\mathcal{H}} \mathcal{P}_t\varphi(x)\ \mu(dx)\right)\ dt \\
&= \lim_{T\to\infty} \frac{1}{T}\int_0^T \left(\int_{\mathcal{H}} \varphi(x)\ \mu(dx)\right)\ dt = \int_{\mathcal{H}} \varphi(x)\ \mu(dx).
\end{aligned}$$

□

Definition 1.22. *An invariant Borel probability measure μ for $(\mathcal{P}_t)_{t\geq 0}$ is called* ergodic *if*

$$L^2\text{-}\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathcal{P}_t\varphi\ dt = \int_{\mathcal{H}} \varphi(x)\ \mu(dx) \tag{1.15}$$

for all $\varphi \in L^2(\mathcal{H},\mu)$. Denote by $\mathcal{E}(\mathcal{P}_t)$ the set of all ergodic measures for $(\mathcal{P}_t)_{t\geq 0}$.

The identity (1.15) is interpreted in physics by saying that the time average of the $\mathcal{P}_t\varphi$'s coincides with the spatial average of φ.

Proposition 1.23. *Let μ be an invariant Borel probability measure for $(\mathcal{P}_t)_{t\geq 0}$. Then μ is ergodic if and only if* $\dim(\Sigma) = 1$.

Proof. Suppose $\mu \in \mathcal{E}(\mathcal{P}_t)$ and let $\varphi \in \Sigma$. Then it follows from (1.15) that φ is μ-a.s. constant. Therefore $\dim(\Sigma) = 1$.

Conversely, assume that $\dim(\Sigma) = 1$ and define $F \in (L^2(\mathcal{H},\mu))^*$ by $F(\varphi) := \langle M_\infty\varphi, \mathbf{1}\rangle_\Sigma$ for all $\varphi \in L^2(\mathcal{H},\mu)$. Note that since $\mathbf{1} \in \Sigma$ and $M_\infty\varphi \in \Sigma$ for all $\varphi \in L^2(\mathcal{H},\mu)$ by Theorem 1.21, this is welldefined and we have

$$M_\infty\varphi = F(\varphi)\mathbf{1}. \tag{1.16}$$

By the Riesz representation theorem there exists a unique element $\varphi_0 \in L^2(\mathcal{H},\mu)$ such that $F(\varphi) = \langle\varphi,\varphi_0\rangle_{L^2(\mathcal{H},\mu)}$ for all $\varphi \in L^2(\mathcal{H},\mu)$. In order to show that μ is ergodic, we have to prove that $\varphi_0 = \mathbf{1}$. Integrating (1.16) with respect to μ over $\mathcal{H}$ yields

$$\begin{aligned}
\langle\varphi,\varphi_0\rangle_{L^2(\mathcal{H},\mu)} &= F(\varphi) \\
&= \int_{\mathcal{H}} F(\varphi)\mathbf{1}(x)\ \mu(dx) \\
&= \int_{\mathcal{H}} M_\infty\varphi(x)\ \mu(dx) \\
&= \int_{\mathcal{H}} \varphi(x)\ \mu(dx) \\
&= \int_{\mathcal{H}} \varphi(x)\cdot\mathbf{1}(x)\ \mu(dx) = \langle\varphi,\mathbf{1}\rangle_{L^2(\mathcal{H},\mu)}
\end{aligned}$$

for all $\varphi \in L^2(\mathcal{H},\mu)$ in view of (1.13). Therefore $\varphi_0 = \mathbf{1}$. □

Definition 1.24. *Let μ be an invariant Borel probability measure for the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$. A Borel set $\Gamma \in \mathcal{B}(\mathcal{H})$ is said to be* invariant *for $(\mathcal{P}_t)_{t\geq 0}$ if its characteristic function $\mathbf{1}_\Gamma \in \Sigma$. The set $\Gamma \in \mathcal{B}(\mathcal{H})$ is called* trivial *if $\mu(\Gamma) \in \{0, 1\}$.*

The next aim is to show that μ is ergodic if and only if all invariant sets are trivial.

Proposition 1.25. *Let $\varphi, \psi \in \Sigma$. Then the following statements hold:*

(i) $|\varphi| \in \Sigma$;

(ii) φ^+, $\varphi^- \in \Sigma$ [1];

(iii) $\varphi \vee \psi$, $\varphi \wedge \psi \in \Sigma$;

(iv) For any $a \in \mathbb{R}$ we have $\mathbf{1}_{\{x\in\mathcal{H} \mid \varphi(x)>a\}} \in \Sigma$.

Proof. See Proposition 5.14 in [10]. □

Theorem 1.26. *Let μ be an invariant Borel probability measure for $(\mathcal{P}_t)_{t\geq 0}$. Then μ is ergodic if and only if any invariant set is trivial.*

Proof. First, let $\mu \in \mathcal{E}(\mathcal{P}_t)$ and suppose Γ is invariant, e.g. $\mathbf{1}_\Gamma \in \Sigma$. Since by Proposition 1.23 $\dim(\Sigma) = 1$, the function $x \mapsto \mathbf{1}_\Gamma(x)$ must be (μ-a.s.) constant. Therefore $\mu(\Gamma) \in \{0, 1\}$, i.e. Γ is trivial.

Conversely, suppose $\mu \notin \mathcal{E}(\mathcal{P}_t)$. Hence according to Proposition 1.23 there exists a function $\varphi_0 \in \Sigma$ that is μ-a.s. not constant. So we can find some $a_0 \in \mathbb{R}$ such that $\mu(\{\varphi_0 > a_0\}) \notin \{0, 1\}$, i.e. the set $\{\varphi_0 > a_0\}$ is not trivial. On the other hand by Proposition 1.25 (iv) we have $\mathbf{1}_{\{\varphi_0>a_0\}} \in \Sigma$, e.g. $\{\varphi_0 > a_0\}$ is invariant. So there exists some invariant set that is not trivial and the assertion follows by contraposition. □

Now we are able to prove a sufficient condition for an invariant Borel probability measure to be ergodic:

Theorem 1.27. *Assume that there is a unique invariant Borel probability measure μ for $(\mathcal{P}_t)_{t\geq 0}$. Then μ is ergodic.*

Proof. Assume, by contradiction, that $\mu \notin \mathcal{E}(\mathcal{P}_t)$. Then by Theorem 1.26 there is a non-trivial invariant set Γ. Hence define for $A \in \mathcal{B}(\mathcal{H})$ the Borel probability measure μ_Γ by $\mu_\Gamma(A) := \frac{1}{\mu(\Gamma)} \cdot \mu(A \cap \Gamma)$. We will prove that $\mu_\Gamma \in \mathcal{J}(\mathcal{P}_t)$. Since $\mu_\Gamma \neq \mu$, this is a contradiction to the uniqueness of μ.

Since Γ is invariant for $(\mathcal{P}_t)_{t\geq 0}$, we have by definition $\mathbf{1}_\Gamma \in \Sigma$, i.e. $\mathcal{P}_t\mathbf{1}_\Gamma = \mathbf{1}_\Gamma$ μ-a.s. for all $t \geq 0$ and so $\pi_t(x, \Gamma) = \mathbf{1}_\Gamma(x)$ for μ-a.e. $x \in \mathcal{H}$ and all $t \geq 0$. Similarly, $\pi_t(x, \Gamma^c) = \mathbf{1}_{\Gamma^c}(x)$ for μ-a.e. $x \in \mathcal{H}$ and all $t \geq 0$. Consequently, for every $A \in \mathcal{B}(\mathcal{H})$ we

[1] As usual $\varphi^+ := \varphi \vee 0$ and $\varphi^- := (-\varphi) \vee 0$.

obtain $\pi_t(x, A \cap \Gamma) = 0$ for μ-a.e. $x \in \Gamma^c$ and $\pi_t(x, A \cap \Gamma^c) = 0$ for μ-a.e. $x \in \Gamma$ for all $t \geq 0$. Hence

$$\begin{aligned}\int_\Gamma \pi_t(x,A)\,\mu(dx) &= \int_\Gamma \pi_t(x, A\cap\Gamma)\,\mu(dx) + \int_\Gamma \pi_t(x, A\cap\Gamma^c)\,\mu(dx)\\ &= \int_\Gamma \pi_t(x, A\cap\Gamma)\,\mu(dx)\\ &= \int_\mathcal{H} \pi_t(x, A\cap\Gamma)\,\mu(dx)\\ &= \mu(A\cap\Gamma)\end{aligned}$$

for all $A \in \mathcal{B}(\mathcal{H})$ according to the invariance of μ. Dividing both sides by $\mu(\Gamma) \neq 0$ yields

$$\int_\mathcal{H} \mathcal{P}_t \mathbf{1}_A(x)\,\mu_\Gamma(dx) = \int_\mathcal{H} \pi_t(x,A)\,\mu_\Gamma(dx) = \mu_\Gamma(A) = \int_\mathcal{H} \mathbf{1}_A(x)\,\mu_\Gamma(dx)$$

for all $A \in \mathcal{B}(\mathcal{H})$, which means that μ_Γ is invariant for $(\mathcal{P}_t)_{t\geq 0}$. □

In order to show that the ergodic Borel probability measures for a Markov semigroup are exactly the extremal points of the set of all invariant measures, we need the following helping lemma:

Lemma 1.28. *Let $\mu \in \mathcal{E}(\mathcal{P}_t)$ and $\nu \in \mathcal{J}(\mathcal{P}_t)$ such that $\nu \ll \mu$. Then $\mu = \nu$.*

Proof. Let $\Gamma \in \mathcal{B}(\mathcal{H})$. By Definition 1.22 there exists a sequence $(T_n)_{n\in\mathbb{N}}$, $\lim_{n\to\infty} T_n = \infty$, such that

$$\lim_{n\to\infty} \frac{1}{T_n}\int_0^{T_n} \mathcal{P}_t \mathbf{1}_\Gamma\, dt = \mu(\Gamma) \quad \mu\text{-a.s.} \tag{1.17}$$

Since $\nu \ll \mu$, identity (1.17) holds also ν-a.s. Now integrating with respect to ν yields

$$\int_\mathcal{H} \left(\frac{1}{T_n}\int_0^{T_n} \mathcal{P}_t \mathbf{1}_\Gamma(x)\, dt\right)\,\nu(dx) = \frac{1}{T_n}\int_0^{T_n}\left(\int_\mathcal{H} \mathcal{P}_t \mathbf{1}_\Gamma(x)\,\nu(dx)\right)\,dt = \nu(\Gamma)$$

for all $n \in \mathbb{N}$ according to the invariance of ν. Hence letting $n \to \infty$ by Lebesgue and (1.17)

$$\nu(\Gamma) = \lim_{n\to\infty}\int_\mathcal{H} \left(\frac{1}{T_n}\int_0^{T_n} \mathcal{P}_t \mathbf{1}_\Gamma(x)\, dt\right)\,\nu(dx) = \int_\mathcal{H} \mu(\Gamma)\,\nu(dx) = \mu(\Gamma).$$

Now the assertion follows by the arbitrariness of $\Gamma \in \mathcal{B}(\mathcal{H})$. □

Theorem 1.29. *The set $\mathcal{E}(\mathcal{P}_t)$ of all ergodic Borel probability measures for $(\mathcal{P}_t)_{t\geq 0}$ coincides with the set $(\mathcal{J}(\mathcal{P}_t))_e$ of all extremal points of $\mathcal{J}(\mathcal{P}_t)$: $\mathcal{E}(\mathcal{P}_t) = (\mathcal{J}(\mathcal{P}_t))_e$.*

Proof. (i) $\mathcal{E}(\mathcal{P}_t) \subseteq (\mathcal{J}(\mathcal{P}_t))_e$: Let $\mu \in \mathcal{E}(\mathcal{P}_t)$ and assume, by contradiction, that $\mu \notin (\mathcal{J}(\mathcal{P}_t))_e$. Then there exist $\mu_1, \mu_2 \in \mathcal{J}(\mathcal{P}_t)$ with $\mu_1 \neq \mu_2$ and $\alpha \in]0,1[$ such that $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$. Hence $\mu_1 \ll \mu$ and $\mu_2 \ll \mu$ and this is a contradiction according to Lemma 1.28. Therefore $\mu \in (\mathcal{J}(\mathcal{P}_t))_e$.

(ii) $\mathcal{E}(\mathcal{P}_t) \supseteq (\mathcal{J}(\mathcal{P}_t))_e$: Conversely, let $\mu \in (\mathcal{J}(\mathcal{P}_t))_e$ and assume $\mu \notin \mathcal{E}(\mathcal{P}_t)$. Then by Theorem 1.26 there exists a non-trivial invariant set Γ. As in the proof of Theorem 1.27 one can show that $\mu_\Gamma, \mu_{\Gamma^c} \in \mathcal{J}(\mathcal{P}_t)$ (with $\mu_\Gamma, \mu_{\Gamma^c}$ defined as thereby). Since obviously $\mu_\Gamma \neq \mu_{\Gamma^c}$ and $\mu = \mu(\Gamma)\mu_\Gamma + (1-\mu(\Gamma))\mu_{\Gamma^c}$, μ is not extremal, in contradiction to above.

□

Theorem 1.30. *Let μ and ν, $\mu \neq \nu$, be two ergodic Borel probability measures for $(\mathcal{P}_t)_{t\geq 0}$. Then μ and ν are singular.*

Proof. Let $\Gamma \in \mathcal{B}(\mathcal{H})$ such that $\mu(\Gamma) \neq \nu(\Gamma)$. Since μ and ν are ergodic, by Definition 1.22 there exist a sequence $(T_n)_{n\in\mathbb{N}}$, $T_n \to \infty$, and sets $M, N \in \mathcal{B}(\mathcal{H})$ with $\mu(M) = \nu(N) = 1$ such that

$$\lim_{n\to\infty} \frac{1}{T_n} \int_0^{T_n} \mathcal{P}_t \mathbf{1}_\Gamma(x)\, dt = \int_{\mathcal{H}} \mathbf{1}_\Gamma(x)\, \mu(dx) = \mu(\Gamma)$$

for all $x \in M$ and

$$\lim_{n\to\infty} \frac{1}{T_n} \int_0^{T_n} \mathcal{P}_t \mathbf{1}_\Gamma(x)\, dt = \int_{\mathcal{H}} \mathbf{1}_\Gamma(x)\, \nu(dx) = \nu(\Gamma)$$

for all $x \in N$ respectively. Since $\mu(\Gamma) \neq \nu(\Gamma)$, this implies $M \cap N = \emptyset$ and so μ and ν are singular. □

Corollary 1.31. *If the set $\mathcal{J}(\mathcal{P}_t)$ of invariant Borel probability measures for $(\mathcal{P}_t)_{t\geq 0}$ contains more than one element, there exist at least two elements $\mu, \nu \in \mathcal{E}(\mathcal{P}_t) \subseteq \mathcal{J}(\mathcal{P}_t)$ such that μ and ν are mutually singular.*

Proof. Suppose $\mathcal{J}(\mathcal{P}_t)$ has at least two elements, $\lambda_1, \lambda_2 \in \mathcal{J}(\mathcal{P}_t)$ with $\lambda_1 \neq \lambda_2$. Then by Theorem 5.2.16 in [4] there exist probability measures ρ_{λ_1}, ρ_{λ_2} on $\mathcal{E}(\mathcal{P}_t)$ such that

$$\lambda_i(\cdot) = \int_{\mathcal{E}(\mathcal{P}_t)} \mu(\cdot)\, \rho_{\lambda_i}(d\mu)$$

for $i = 1, 2$. Assume $\mathcal{E}(\mathcal{P}_t) = \{\mu\}$. Then $\lambda_1 = \mu = \lambda_2$, in contradiction to the choice of λ_1, λ_2. Therefore $\mathcal{E}(\mathcal{P}_t)$ must contain at least two elements, $\mu, \nu \in \mathcal{E}(\mathcal{P}_t)$ with $\mu \neq \nu$. According to Theorem 1.30 μ and ν are mutually singular. □

Consequently, if $\mathcal{J}(\mathcal{P}_t)$ contains more than one element, the state space $\mathcal{H}$ can be partitioned into (at least) two disjoint parts, e.g. $\mathcal{H} = A \,\dot\cup\, B$, with the property that if the Markov process starts in A, then it will stay in A for all times $t \geq 0$ almost surely and the same is true for the complement B. (In particular, the zero-set does not depend on the point in time t.) The intuition that derives from this consideration, is that uniqueness of the invariant measure is a consequence of the process visiting a 'sufficiently large' portion of the state space, independently of its initial position.

1.4 Existence of an invariant measure

In Theorem 1.19 we have seen that, if there exists an invariant measure for the (stochastically continuous) Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ on $\mathcal{B}_b(\mathcal{H})$, then $(\mathcal{P}_t)_{t\geq 0}$ can be extended uniquely to $L^p(\mathcal{H},\mu)$, $p\geq 1$. So it might be important to ask, whether there exists an invariant measure. In the literature there is the following well-known result (cf. [10] or [11]):

Theorem 1.32 (Krylov-Bogoliubov Theorem)**.** *Let $(\mathcal{P}_t)_{t\geq 0}$ be a Feller Markov semigroup on $\mathcal{B}_b(\mathcal{H})$. Assume that there is some $\mu_0 \in \mathcal{M}_1(\mathcal{H})$ such that the sequence $(\mathcal{P}_t^*\mu_0)_{t\geq 0}$ is tight. Then there exists at least one invariant Borel probability measure μ_* for $(\mathcal{P}_t)_{t\geq 0}$.*

Proof. Let $(\mu_t)_{t\geq 0}$ be the set defined by

$$\mu_t(A) := \frac{1}{t}\int_0^t (\mathcal{P}_s^*\mu_0)(A)\, ds, \quad A\in\mathcal{B}(\mathcal{X}).$$

Since $(\mathcal{P}_t^*\mu_0)_{t\geq 0}$ is (uniformly) tight by assumption, for every $\varepsilon>0$ there exists some compact set K_ε such that $\sup_{t\geq 0}\mathcal{P}_t^*\mu_0(K_\varepsilon^c)<\varepsilon$. Taking the same compact set, it is straightforward to check that $(\mu_t)_{t\geq 0}$ is (uniformly) tight as well. Therefore by Prohorov (cf. [10], Theorem 6.7) there exists a subsequence $(\mu_{t_n})_{n\in\mathbb{N}}$ and a Borel probability measure μ_* such that $\mu_{t_n}\to\mu_*$ weakly as $n\to\infty$. Let $\varphi\in\mathcal{C}_b(\mathcal{H})$. Since according to the Feller property $\mathcal{P}_t\varphi\in\mathcal{C}_b(\mathcal{H})$, by weak convergence and Fubini we have

$$\begin{aligned}
& |(\mathcal{P}_t^*\mu_*)(\varphi)-\mu_*(\varphi)| \\
= \ & |(\mathcal{P}_t^*F_{\mu_*})(\varphi)-F_{\mu_*}(\varphi)| \\
= \ & |F_{\mu_*}(\mathcal{P}_t\varphi)-F_{\mu_*}(\varphi)| \\
= \ & \left|\int_{\mathcal{H}}\mathcal{P}_t\varphi(x)\,\mu_*(dx)-\int_{\mathcal{H}}\varphi(x)\,\mu_*(dx)\right| \\
= \ & \lim_{n\to\infty}\left|\int_{\mathcal{H}}\mathcal{P}_t\varphi(x)\,\mu_{t_n}(dx)-\int_{\mathcal{H}}\varphi(x)\,\mu_{t_n}(dx)\right| \\
= \ & \lim_{n\to\infty}|\mu_{t_n}(\mathcal{P}_t\varphi)-\mu_{t_n}(\varphi)| \\
= \ & \lim_{n\to\infty}\frac{1}{t_n}\left|\int_0^{t_n}(\mathcal{P}_s^*\mu_0)\,(\mathcal{P}_t\varphi)\;ds-\int_0^{t_n}(\mathcal{P}_s^*\mu_0)\,(\varphi)\;ds\right| \\
= \ & \lim_{n\to\infty}\frac{1}{t_n}\left|\int_0^{t_n}\left(\int_{\mathcal{H}}\mathcal{P}_s\mathcal{P}_t\varphi(x)\,\mu_0(dx)\right)\,ds-\int_0^{t_n}\left(\int_{\mathcal{H}}\mathcal{P}_s\varphi(x)\,\mu_0(dx)\right)\,ds\right| \\
= \ & \lim_{n\to\infty}\frac{1}{t_n}\left|\int_{\mathcal{H}}\left(\int_0^{t_n}\mathcal{P}_{s+t}\varphi(x)\,ds\right)\,\mu_0(dx)-\int_{\mathcal{H}}\left(\int_0^{t_n}\mathcal{P}_s\varphi(x)\,ds\right)\,\mu_0(dx)\right| \\
= \ & \lim_{n\to\infty}\frac{1}{t_n}\left|\int_{\mathcal{H}}\left(\int_t^{t_n+t}\mathcal{P}_s\varphi(x)\,ds\right)\,\mu_0(dx)-\int_{\mathcal{H}}\left(\int_0^{t_n}\mathcal{P}_s\varphi(x)\,ds\right)\,\mu_0(dx)\right| \\
= \ & \lim_{n\to\infty}\frac{1}{t_n}\left|\int_t^{t_n+t}\left(\int_{\mathcal{H}}\mathcal{P}_s\varphi(x)\,\mu_0(dx)\right)\,ds-\int_0^{t_n}\left(\int_{\mathcal{H}}\mathcal{P}_s\varphi(x)\,\mu_0(dx)\right)\,ds\right|
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \frac{1}{t_n} \Bigg| \int_{t_n}^{t_n+t} \Big(\int_{\mathcal{H}} \mathcal{P}_s\varphi(x)\ \mu_0(dx) \Big)\ ds - \int_{t_n}^{t} \Big(\int_{\mathcal{H}} \mathcal{P}_s\varphi(x)\ \mu_0(dx) \Big)\ ds \\
&\qquad\qquad - \int_0^{t_n} \Big(\int_{\mathcal{H}} \mathcal{P}_s\varphi(x)\ \mu_0(dx) \Big)\ ds \Bigg| \\
&= \lim_{n\to\infty} \frac{1}{t_n} \Bigg| \int_{t_n}^{t_n+t} \Big(\int_{\mathcal{H}} \mathcal{P}_s\varphi(x)\ \mu_0(dx) \Big)\ ds - \int_0^{t} \Big(\int_{\mathcal{H}} \mathcal{P}_s\varphi(x)\ \mu_0(dx) \Big)\ ds \Bigg| \\
&\le \lim_{n\to\infty} \frac{2t}{t_n} \cdot \|\varphi\|_\infty \\
&= 0
\end{aligned}
$$

for all $t \ge 0$. Now the conclusion follows according to Lemma 1.18. □

1.5 Uniqueness of the invariant measure

In view of Theorem 1.27 we have ergodicity of the system if there is a *unique* invariant Borel probability measure μ for the (stochastically continuous) Markov semigroup $(\mathcal{P}_t)_{t\ge0}$. Hence it might be useful to look for conditions under which uniqueness of the invariant measure is guaranteed. It turns out that the combination of the strong Feller property and irreducibility is the right requirement (cf. Theorem 1.35 below).

Proposition 1.33 (Doob)**.** *Let $(\mathcal{P}_t)_{t\ge0}$ be a (stochastically continuous) regular Markov semigroup and μ an invariant Borel probability measure for $(\mathcal{P}_t)_{t\ge0}$. Then μ is equivalent to $\pi_t(x,\cdot)$ for all $x \in \mathcal{H}$, $t > 0$. Moreover, μ is the unique invariant Borel probability measure for $(\mathcal{P}_t)_{t\ge0}$.*

Proof. Let $A \in \mathcal{B}(\mathcal{H})$ and $t > 0$ arbitrary. Since μ is an invariant measure for $(\mathcal{P}_t)_{t\ge0}$

$$\mu(A) = \int_{\mathcal{H}} \mathbf{1}_A(y)\ \mu(dy) = \int_{\mathcal{H}} \mathcal{P}_t\mathbf{1}_A(y)\ \mu(dy) = \int_{\mathcal{H}} \pi_t(y, A)\ \mu(dy). \tag{1.18}$$

Let $x \in \mathcal{H}$, $t > 0$.

First show that $\mu \ll \pi_t(x,\cdot)$. Let $A \in \mathcal{B}(\mathcal{H})$ such that $\pi_t(x, A) = 0$. Since $(\mathcal{P}_t)_{t\ge0}$ is regular by assumption, $\pi_t(y, A) = 0$ for all $y \in \mathcal{H}$. Therefore by (1.18) $\mu(A) = 0$. Hence $\mu \ll \pi_t(x,\cdot)$, e.g. μ is absolutely continuous with respect to $\pi_t(x,\cdot)$.

Conversely take $A \in \mathcal{B}(\mathcal{H})$ with $\mu(A) = 0$. Again using equation (1.18) gives $\pi_t(y, A) = 0$ for μ-a.e. $y \in \mathcal{H}$. By the regularity of $(\mathcal{P}_t)_{t\ge0}$ we obtain $\pi_t(y, A) = 0$ for all $y \in \mathcal{H}$. In particular, $\pi_t(x, A) = 0$. Hence $\pi_t(x,\cdot) \ll \mu$.

Alltogether we thus have shown that μ is equivalent to $\pi_t(x,\cdot)$ for arbitrary $x \in \mathcal{H}$ and $t > 0$.

It remains to prove the uniqueness of μ. For this assume that μ and ν, $\mu \ne \nu$, are two ergodic Borel probability measures for $(\mathcal{P}_t)_{t\ge0}$. Then μ and ν are singular by Theorem 1.30. Hence there exist $A, B \in \mathcal{B}(\mathcal{H})$, $A \cap B = \emptyset$, such that $\mu(A) = \nu(B) = 1$. Since it was shown in the first part of the proof that $\mu \approx \pi_t(x,\cdot)$ for all $t > 0$ and

$x \in \mathcal{H}$, we obtain $\pi_t(x, A) = 1$ for all $t > 0$ and $x \in \mathcal{H}$. Applying the same argument for the ergodic measure ν leads $\pi_t(x, B) = 1$ for all $t > 0$ and $x \in \mathcal{H}$. This implies $\pi_t(x, A \dot{\cup} B) = \pi_t(x, A) + \pi_t(x, B) = 1 + 1 = 2$ for all $t > 0$ and $x \in \mathcal{H}$, which obviously contradicts $\pi_t(x, \mathcal{H}) = 1$. Therefore μ is the unique invariant Borel probability measure for $(\mathcal{P}_t)_{t \geq 0}$. □

Proposition 1.34 (Khasminskii). *Let the Markov semigroup $(\mathcal{P}_t)_{t \geq 0}$ be strong Feller and irreducible. Then it is regular.*

Proof. Let $t > 0$ and $x_0 \in \mathcal{H}$ be arbitrary and fixed. We have to show that $\pi_t(x, \cdot) \approx \pi_t(x_0, \cdot)$ for all $x \in \mathcal{H}$, i.e. the null-sets of $\pi_t(x, \cdot)$ and $\pi_t(x_0, \cdot)$ coincide for all $x \in \mathcal{H}$. For this it would be enough to show: If $A \in \mathcal{B}(\mathcal{H})$ is taken in a way such that $\pi_t(x_0, A) > 0$, then $\pi_t(x, A) > 0$ for all $x \in \mathcal{H}$.

Suppose $A \in \mathcal{B}(\mathcal{H})$ such that $\pi_t(x_0, A) > 0$. Then by (1.1) for any $0 < s < t$ we have

$$\int_{\mathcal{H}} \pi_{t-s}(y, A)\, \pi_s(x_0, dy) = \pi_{s+(t-s)}(x_0, A) = \pi_t(x_0, A) > 0.$$

Therefore there exists at least one $y_0 \in \mathcal{H}$ such that $\pi_{t-s}(y_0, A) > 0$. Since by (1.4) $\pi_{t-s}(y_0, A) = \mathcal{P}_{t-s}\mathbf{1}_A(y_0)$ and $y \mapsto \mathcal{P}_{t-s}\mathbf{1}_A(y)$ is continuous (because $(\mathcal{P}_t)_{t \geq 0}$ is strong Feller by assumption), there exists $r > 0$ such that $\pi_{t-s}(y, A) = \mathcal{P}_{t-s}\mathbf{1}_A(y) > 0$ for all $y \in B(y_0, r)$. Consequently for arbitrary $x \in \mathcal{H}$ we obtain

$$\pi_t(x, A) = \int_{\mathcal{H}} \pi_{t-s}(y, A)\, \pi_s(x, dy) \geq \int_{B(y_0, r)} \pi_{t-s}(y, A)\, \pi_s(x, dy) > 0,$$

because $\pi_s(x, B(y_0, r)) > 0$ according to the irreducibility of $(\mathcal{P}_t)_{t \geq 0}$. So, $\pi_t(x, A) > 0$ and we have proved that $\pi_t(x, \cdot) \approx \pi_t(x_0, \cdot)$ as required. □

Taking Propositions 1.33 and 1.34 together we obtain

Theorem 1.35. *Let $(\mathcal{P}_t)_{t \geq 0}$ be a Markov semigroup that is strong Feller and irreducible. Then there is at most one invariant Borel probability measure for $(\mathcal{P}_t)_{t \geq 0}$.*

1.6 First step towards the asymptotic strong Feller property

We conclude the first chapter by proving an interesting property of two distinct ergodic measures concerning their supports if the Markov semigroup is strong Feller. For the proof we will need the following

Lemma 1.36. *Let μ be a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$. Then*

$$supp(\mu) = \{x \in \mathcal{H} \mid \mu(B(x, r)) > 0 \ \ \forall\, r > 0\}. \tag{1.19}$$

Moreover, if $\varphi \in \mathcal{C}_b(\mathcal{H})$, $\varphi \geq 0$, is such that

$$\int_{\mathcal{H}} \varphi(x)\, \mu(dx) = 0, \tag{1.20}$$

then $\varphi(x) = 0$ for all $x \in supp(\mu)$.

Proof. Let $x_0 \in \operatorname{supp}(\mu)$ and assume $x_0 \notin \{x \in \mathcal{H} \mid \mu(B(x,r)) > 0 \ \forall\, r > 0\}$. Then there is some $r > 0$ such that $\mu(B(x_0,r)) = 0$. Hence in view of the definition of the support of a probability measure as the intersection of all closed subsets having probability 1, $x_0 \notin \operatorname{supp}(\mu)$, in contradiction to above. Therefore $x_0 \in \{x \in \mathcal{H} \mid \mu(B(x,r)) > 0 \ \forall\, r > 0\}$. Conversely, suppose $x_0 \notin \operatorname{supp}(\mu)$. Then there exists some $r_0 > 0$ such that $\mu(B(x_0,r_0)) = 0$ and so $x_0 \notin \{x \in \mathcal{H} \mid \mu(B(x,r)) > 0 \ \forall\, r > 0\}$. Hence the assertion in (1.19) follows.

Now let $\varphi \in \mathcal{C}_b(\mathcal{H})$, $\varphi \geq 0$, such that $\int_{\mathcal{H}} \varphi(x)\ \mu(dx) = 0$. Then $\varphi(x) = 0$ for μ-a.e. $x \in \mathcal{H}$. Suppose $x_0 \in \mathcal{H}$ such that $\varphi(x_0) > 0$. Since $\varphi\colon \mathcal{H} \to \mathbb{R}$ is continuous, there exists some $r_0 > 0$ such that $\varphi(x) > 0$ for all $x \in B(x_0,r_0)$. According to (1.20) $\mu(B(x_0,r_0)) = 0$ then. Therefore by (1.19) $x_0 \notin \operatorname{supp}(\mu)$. Thus we have shown that $\varphi(x) = 0$ for all $x \in \operatorname{supp}(\mu)$. □

Proposition 1.37. *Let $(\mathcal{P}_t)_{t\geq 0}$ be a Markov semigroup with the strong Feller property. Let μ and ν, $\mu \neq \nu$, be two ergodic Borel probability measures for $(\mathcal{P}_t)_{t\geq 0}$. Then $supp(\mu) \cap supp(\nu) = \emptyset$.*

Proof. By Theorem 1.30 μ and ν are singular. Hence there are $A, B \in \mathcal{B}(\mathcal{H})$ such that $A \cap B = \emptyset$ and $\mu(A) = \nu(B) = 1$. Then for any $t > 0$ we have

$$0 = \mu(A^c) = \int_{\mathcal{H}} \mathbf{1}_{A^c}(x)\ \mu(dx) = \int_{\mathcal{H}} \mathcal{P}_t \mathbf{1}_{A^c}(x)\ \mu(dx) = \int_{\mathcal{H}} \pi_t(x, A^c)\ \mu(dx)$$

and

$$0 = \nu(B^c) = \int_{\mathcal{H}} \mathbf{1}_{B^c}(x)\ \nu(dx) = \int_{\mathcal{H}} \mathcal{P}_t \mathbf{1}_{B^c}(x)\ \nu(dx) = \int_{\mathcal{H}} \pi_t(x, B^c)\ \nu(dx)$$

respectively. Since $(\mathcal{P}_t)_{t\geq 0}$ is strong Feller, the functions $x \mapsto \mathcal{P}_t \mathbf{1}_{A^c}(x) = \pi_t(x, A^c)$ and $x \mapsto \mathcal{P}_t \mathbf{1}_{B^c}(x) = \pi_t(x, B^c)$ are continuous. Hence by Lemma 1.36 we have $\pi_t(x, A^c) = 0$ for all $x \in \operatorname{supp}(\mu)$ and $\pi_t(x, B^c) = 0$ for all $x \in \operatorname{supp}(\nu)$ respectively. Assume that there is some $x_0 \in \operatorname{supp}(\mu) \cap \operatorname{supp}(\nu)$. Then $\pi_t(x_0, A) = \pi_t(x_0, B) = 1$, which implies $\pi_t(x_0, A \dot{\cup} B) = \pi_t(x_0, A) + \pi_t(x_0, B) = 1 + 1 = 2$. This is a contradiction in comparison to $\pi_t(x, \mathcal{H}) = 1$. Therefore $\operatorname{supp}(\mu) \cap \operatorname{supp}(\nu) = \emptyset$. □

Actually, the proof of Proposition 1.37 suggests to introduce the notion of being strong Feller *at* some point $x \in \mathcal{H}$, e.g. the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ is called *strong Feller at* $x \in \mathcal{H}$, if the function $\mathcal{P}_t\varphi$ is continuous at x for all $\varphi \in \mathcal{B}_b(\mathcal{H})$, $t \geq 0$. With this notation, the same proof as above allows to conclude, that if the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ is strong Feller at the point $x \in \mathcal{H}$, then x can belong to the support $\operatorname{supp}(\mu)$ of at most one invariant Borel probability measure μ for $(P_t)_{t\geq 0}$. In particular, if $(\mathcal{P}_t)_{t\geq 0}$ is strong Feller (at every $x \in \mathcal{H}$) and there exists some point $x \in \mathcal{H}$ such that $x \in \operatorname{supp}(\mu)$ for every invariant Borel probability measure μ for $(\mathcal{P}_t)_{t\geq 0}$, then there exists at most one invariant measure μ.

This idea will be important when deriving uniqueness of the invariant Borel probability measure from the *asymptotic strong Feller property* later in this thesis.

2 Monge-Kantorovich duality

The field of mass transference problems and dual representations of it originates in the 1781 formulated Monge problem. Although there had been intensively study in the following centuries, there are open problems even today. For our purpose it would be enough to distinguish two main formulations of this famous problem (cf. [12]):

(1) The Monge-Kantorovich problem (One-stage problem). Suppose we are given two finite measures μ_1 and μ_2 on some space $\mathcal{X}$ with equal mass, e.g. $\mu_1(\mathcal{X}) = \mu_2(\mathcal{X})$, describing the masses of $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{X}$ respectively. While μ_1 is refered to the *initial* distribution, μ_2 is called the *final* distribution. A *transference plan* would be a finite measure μ on the product space $\mathcal{X} \times \mathcal{X}$ with marginals μ_1 and μ_2 respectively. The set of admissible transference plans is denoted by $\mathcal{C}(\mu_1, \mu_2)$. The amount of mass shipped from an infinitesimal small neighborhood dx of $x \in \mathcal{X}$ into another infinitesimal small neighborhood dy of $y \in \mathcal{X}$ is then proportional to $\mu(dx, dy)$. If the unit cost of shipment from x to y is denoted by $c(x, y)$, the total cost of shipment is given by

$$\int_{\mathcal{X}^2} c(x,y)\ \mu(dx,dy). \tag{2.1}$$

To minimize the transportation costs, we have to find some *optimal transference plan* $\mu^* \in \mathcal{C}(\mu_1, \mu_2)$ for which (2.1) is minimal, e.g.

$$\int_{\mathcal{X}^2} c(x,y)\ \mu^*(dx,dy) \le \int_{\mathcal{X}^2} c(x,y)\ \mu(dx,dy)$$

for all $\mu \in \mathcal{C}(\mu_1, \mu_2)$. Therefore we will consider the *Kantorovich functional*:

$$K_c(\mu_1,\mu_2) := \inf_{\mu \in \mathcal{C}(\mu_1,\mu_2)} \int_{\mathcal{X}^2} c(x,y)\ \mu(dx,dy). \tag{2.2}$$

(2) The Kantorovich-Rubinstein problem (Multi-stage problem). Contrary to the above we consider here the problem of transferring masses in cases where transits are permitted. Rather than shipping a mass from a certain subset $A \subseteq \mathcal{X}$ to another subset $B \subseteq \mathcal{X}$ in just one step, the shipment is made in n stages: Ship $A = A_1$ to the volume $A_2 \subseteq \mathcal{X}$, then transfer $A_2 \subseteq \mathcal{X}$ to $A_3 \subseteq \mathcal{X}$, ..., $A_{n-1} \subseteq \mathcal{X}$ to $A_n = B$. Let $\mathbb{B}(\mu_1 - \mu_2)$ be the space of finite measures on $(\mathcal{X}^2, \mathcal{B}(\mathcal{X}^2))$ having marginal difference $\mu_1 - \mu_2$. It can be shown (cf. [12], p. 91-92) that minimization of the involved transportation costs in this case is equivalent - at least if the infimum below is attained - to find some optimal transference plan $b^* \in \mathbb{B}(\mu_1 - \mu_2)$:

$$\int_{\mathcal{X}^2} c(x,y)\ b^*(dx,dy) = \inf_{b \in \mathbb{B}(\mu_1 - \mu_2)} \int_{\mathcal{X}^2} c(x,y)\ b(dx,dy) =: R_c(\mu_1,\mu_2). \tag{2.3}$$

The right hand side in (2.3) is called *Kantorovich-Rubinstein functional.* If $c = d$ for some metric d, $W(\mu_1, \mu_2) := R_d(\mu_1, \mu_2)$ is called the *Wasserstein norm* of the measures μ_1 and μ_2.

Instead of solving the above problems separately, e.g. finding measures μ^* and b^* that realize (2.2) and (2.3) respectively, we aim to formulate dual representations for both problems respectively. Afterwards it will be shown that the two primal problems coincide if and only if c is a metric (cf. Theorem 2.18 below). In particular, (2.32), which means equality of the Monge-Kantorovich problem and the dual problem of the Kantorovich-Rubinstein problem, holds in case of a metric d.

Before we do so, note that via suitable normalization, e.g. dividing by $\mu_1(\mathcal{X})$, it would be sufficient to consider probability measures P_1, P_2 on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ instead of finite Borel measures μ_1, μ_2. Hence the infima in the Monge-Kantorovich problem and the Kantorovich-Rubinstein problem are taken over the sets

$$\begin{aligned}\mathcal{C}(P_1, P_2) &:= \{P \in \mathcal{M}_1(\mathcal{X}^2) \ \mid \ P(A \times \mathcal{X}) = P_1(A),\ P(\mathcal{X} \times A) = P_2(A)\ \forall\ A \in \mathcal{B}(\mathcal{X})\}\\ &= \{P \in \mathcal{M}_1(\mathcal{X}^2) \ \mid \ T_1 P = P_1,\ \ T_2 P = P_2\},\end{aligned}$$

and

$$\begin{aligned}\mathbb{B}(P_1 - P_2) &:= \{P \in \mathcal{M}_1(\mathcal{X}^2) \ \mid \ P(A \times \mathcal{X}) - P(\mathcal{X} \times A) = (P_1 - P_2)(A)\ \forall\ A \in \mathcal{B}(\mathcal{X})\}\\ &= \{P \in \mathcal{M}_1(\mathcal{X}^2) \ \mid \ T_1 P - T_2 P = P_1 - P_2\},\end{aligned}$$

respectively, where $T_i P$ denotes the i-th marginal of P, $i = 1, 2$.

2.1 Dual representation for the Monge-Kantorovich problem

Now we turn to the duality theorem for the Monge-Kantorovich problem (2.2). Let $(\mathcal{X}, d)$ be a separable metric space and define

$$\mathbf{C} := \{c\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+ \mid \ \exists\ H \in \mathbf{H}\ :\ c(x,y) = H(d(x,y))\ \ \forall\ \ (x,y) \in \mathcal{X}^2\},$$

where $\mathbf{H}$ is given by

$$\mathbf{H} := \left\{ H \in \mathcal{L}(\mathbb{R}_+, \mathbb{R}_+) \mid H(0) = 0,\ H \text{ strictly increasing and convex},\ \sup_{t>0} \frac{H(2t)}{H(t)} < \infty \right\}.$$

Here $\mathcal{L}(\mathbb{R}_+, \mathbb{R}_+)$ denotes the set of all Lipschitz-continuous functions from $\mathbb{R}_+$ to $\mathbb{R}_+$. The last condition in the definition of $\mathbf{H}$ is known as *Orlicz' condition.* Furthermore define for $c \in \mathbf{C}$

$$\mathcal{G}_c(\mathcal{Y}) := \{(f,g) \mid f, g \in \mathcal{L}_d(\mathcal{Y}),\ \ f(x) + g(y) \le c(x,y)\ \ \forall\ \ x, y \in \mathcal{Y}\} \tag{2.4}$$

for any subset $\mathcal{Y} \subseteq \mathcal{X}$.

Theorem 2.1 (Kantorovich Theorem). *Let $(\mathcal{X}, d)$ be a separable metric space, $c \in \mathbf{C}$ and P_i such that $\int_{\mathcal{X}} c(x,a)\ P_i(dx) < \infty$ for some fixed $a \in \mathcal{X}$, $i = 1,2$. Then*

$$\begin{aligned} K_c(P_1,P_2) &= \inf_{P \in \mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} c(x,y)\ P(dx,dy) \\ &= \sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right). \end{aligned} \tag{2.5}$$

Moreover, if the Borel probability measures P_1, P_2 are tight, then the infimum on the left hand side is attained.

Proof. Obviously we have

$$\begin{aligned} \int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) &= \int_{\mathcal{X}^2} f(x)\ P(dx,dy) + \int_{\mathcal{X}^2} g(y)\ P(dx,dy) \\ &= \int_{\mathcal{X}^2} f(x) + g(y)\ P(dx,dy) \\ &\le \int_{\mathcal{X}^2} c(x,y)\ P(dx,dy) \end{aligned}$$

for all $(f,g) \in \mathcal{G}_c(\mathcal{X})$, $P \in \mathcal{C}(P_1,P_2)$. Hence

$$\sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \le \inf_{P\in\mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} c(x,y)\ P(dx,dy). \tag{2.6}$$

It remains to prove the converse inequality. This will be done stepwise with the help of the following preliminaries. □

First part: Separable metric space $(\mathcal{X}, d)$ with bounded metric d. Suppose that d is a bounded metric on $\mathcal{X}$ and define for $c \in \mathbf{C}$

$$\begin{aligned} \rho_1(x,x') &:= \sup_{y\in\mathcal{X}} |c(x,y) - c(x',y)|, \\ \rho_2(y,y') &:= \sup_{x\in\mathcal{X}} |c(x,y) - c(x,y')| \end{aligned}$$

for all $x, x' \in \mathcal{X}$ and for all $y, y' \in \mathcal{X}$ respectively.

Claim 1. *ρ_i is a bounded metric on $\mathcal{X}$ for $i = 1,2$.*

Proof. The assertion will be shown for $i = 1$ only.

(i) It is clear that $\rho_1(x,x') \ge 0$ for all $x, x' \in \mathcal{X}$.

(ii) For $x, x' \in \mathcal{X}$ we have

$$\begin{aligned} \rho_1(x,x') = 0 &\Leftrightarrow \sup_{y\in\mathcal{X}} |c(x,y) - c(x',y)| = 0 \\ &\Leftrightarrow |H(d(x,y)) - H(d(x',y))| = 0\ \forall\ y \in \mathcal{X} \\ &\Leftrightarrow d(x,y) = d(x',y)\ \forall\ y \in \mathcal{X} \\ &\Leftrightarrow x = x'. \end{aligned}$$

(iii)

$$\rho_1(x,x') = \sup_{y\in\mathcal{X}} |c(x,y) - c(x',y)| = \sup_{y\in\mathcal{X}} |c(x',y) - c(x,y)| = \rho_1(x',x)$$

for all $x, x' \in \mathcal{X}$.

(iv) For $x, x', x'' \in \mathcal{X}$ we have

$$\begin{aligned} \rho_1(x,x') &= \sup_{y\in\mathcal{X}} |c(x,y) - c(x',y)| \\ &\leq \sup_{y\in\mathcal{X}} |c(x,y) - c(x'',y)| + \sup_{y\in\mathcal{X}} |c(x'',y) - c(x',y)| \\ &= \rho_1(x,x'') + \rho_1(x'',x'). \end{aligned}$$

It remains to show that ρ_1 is bounded: Observe that since the metric d is bounded, we have $d(x,y) \leq K$ for all $x, y \in \mathcal{X}$ and some constant $K \in \mathbb{R}_+$. Hence we receive for all $(x,x') \in \mathcal{X}^2$

$$\rho_1(x,x') = \sup_{y\in\mathcal{X}} |c(x,y) - c(x',y)| = \sup_{y\in\mathcal{X}} |H(d(x,y)) - H(d(x',y))| \leq 2 \cdot H(K) < \infty.$$

□

For $c \in \mathbf{C}$ and any subset $\mathcal{Y} \subseteq \mathcal{X}$ define

$$\mathcal{G}'_c(\mathcal{Y}) := \{(f,g) \mid f, g \in \mathcal{B}_b(\mathcal{Y}),\ f(x) + g(y) \leq c(x,y)\ \ \forall\, x, y \in \mathcal{Y}\}$$

and

$$\begin{aligned} \mathcal{G}''_c(\mathcal{Y}) := \big\{(f,g) \in \mathcal{G}'_c(\mathcal{Y}) \mid\ & |f(x) - f(x')| \leq \rho_1(x,x') \quad , \\ & |g(y) - g(y')| \leq \rho_2(y,y') \quad \forall\, x, x', y, y' \in \mathcal{Y}\big\}. \end{aligned}$$

Claim 2. *For arbitrary $c \in \mathbf{C}$ and any subset $\mathcal{Y} \subseteq \mathcal{X}$ we have the chain of inclusions*

$$\mathcal{G}''_c(\mathcal{Y}) \subseteq \mathcal{G}_c(\mathcal{Y}) \subseteq \mathcal{G}'_c(\mathcal{Y}).$$

Proof. The second inclusion is obvious by definition. To prove the first inclusion, it would be enough to show $\rho_1(x,x') \leq K_1 \cdot d(x,x')$ for all $x, x' \in \mathcal{Y}$ and $\rho_2(y,y') \leq K_2 \cdot d(y,y')$ for all $y, y' \in \mathcal{Y}$ and suitable constants $K_1, K_2 \in \mathbb{R}_+$. But this follows immediately from the Lipschitz-continuity of H. In fact

$$\begin{aligned} \rho_1(x,x') &= \sup_{y\in\mathcal{X}} |c(x,y) - c(x',y)| \\ &= \|H(d(x,\cdot)) - H(d(x',\cdot))\|_\infty \\ &\leq \mathrm{Lip}(H) \cdot \|d(x,\cdot) - d(x',\cdot)\|_\infty \\ &\leq K_1 \cdot d(x,x') \end{aligned}$$

for all $x, x' \in \mathcal{Y}$. The assertion for the bounded metric ρ_2 follows in the same way. □

We will need the following

Lemma 2.2. *If $\mathcal{Y} \subseteq \mathcal{X}$ such that $P_i(\mathcal{Y}) = 1$ for $i = 1, 2$, then*

$$\sup_{(f,g)\in\mathcal{G}'_c(\mathcal{Y})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) = \sup_{(f,g)\in\mathcal{G}''_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right). \tag{2.7}$$

Proof. Let $(f, g) \in \mathcal{G}'_c(\mathcal{Y})$ and define for $x \in \mathcal{X}$

$$f^*(x) := \inf_{y\in\mathcal{Y}} (c(x, y) - g(y))$$

and for $y \in \mathcal{X}$

$$g^*(y) := \inf_{x\in\mathcal{X}} (c(x, y) - f^*(x)) .$$

Claim 3. $(f^*, g^*) \in \mathcal{G}''_c(\mathcal{X})$.

Proof. Since $c\colon \mathcal{X}^2 \to \mathbb{R}_+$ is continuous, $f^*, g^*\colon \mathcal{X} \to \mathbb{R}$ are upper semi-continuous. Indeed, suppose $x \in \mathcal{X}$ and $(x_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}$ such that $d(x_n, x) \overset{n\to\infty}{\longrightarrow} 0$. Since

$$f^*(x_n) = \inf_{y\in\mathcal{Y}}(c(x_n, y) - g(y)) \le c(x_n, y) - g(y)$$

for all $y \in \mathcal{Y}, n \in \mathbb{N}$, we have

$$\limsup_{n\to\infty} f^*(x_n) \le \lim_{n\to\infty} (c(x_n, y) - g(y)) = c(x, y) - g(y)$$

for all $y \in \mathcal{Y}$ and thus

$$\limsup_{n\to\infty} f^*(x_n) \le \inf_{y\in\mathcal{Y}} (c(x, y) - g(y)) = f^*(x),$$

that is f^* is upper semi-continuous in x. It is well-known that each upper semi-continuous function $h\colon \mathcal{X} \to \mathbb{R}$ is such that $h^{-1}((-\infty, \alpha)) \subseteq \mathcal{X}$ is open for all $\alpha \in \mathbb{R}$. As result f^* and g^* are Borel measurable. Furthermore

$$\begin{aligned} f^*(x) + g^*(y) &= f^*(x) + \inf_{x'\in\mathcal{X}} (c(x', y) - f^*(x')) \\ &\le f^*(x) + (c(x, y) - f^*(x)) \\ &= c(x, y) \end{aligned}$$

for all $x, y \in \mathcal{X}$. Since the boundedness of f^*, g^* is clear according to the choice of $(f, g) \in \mathcal{G}'_c(\mathcal{Y})$ and the boundedness of d, we already know that $(f^*, g^*) \in \mathcal{G}'_c(\mathcal{X})$.

In order to prove that $(f^*, g^*) \in \mathcal{G}''_c(\mathcal{X})$, observe that

$$\begin{aligned} f^*(x) - f^*(x') &= \inf_{y\in\mathcal{Y}} (c(x, y) - g(y)) - \inf_{y\in\mathcal{Y}} (c(x', y) - g(y)) \\ &= \inf_{y\in\mathcal{Y}} (c(x, y) - g(y)) + \sup_{y\in\mathcal{Y}} (g(y) - c(x', y)) \\ &\le \sup_{y\in\mathcal{Y}} (c(x, y) - c(x', y)) \\ &\le \sup_{y\in\mathcal{Y}} |c(x, y) - c(x', y)| \\ &= \rho_1(x, x') \end{aligned}$$

for all $x, x' \in \mathcal{X}$. Changing the roles of x, x' and using the symmetry of the metric ρ_1 we gain $|f^*(x) - f^*(x')| \leq \rho_1(x, x')$ for all $x, x' \in \mathcal{X}$. In the same way one can argue for g^*, e.g. let $y, y' \in \mathcal{X}$. Then

$$\begin{aligned} g^*(y) - g^*(y') &= \inf_{x\in\mathcal{X}} (c(x,y) - f^*(x)) - \inf_{x\in\mathcal{X}} (c(x,y') - f^*(x)) \\ &= \inf_{x\in\mathcal{X}} (c(x,y) - f^*(x)) + \sup_{x\in\mathcal{X}} (f^*(x) - c(x,y')) \\ &\leq \sup_{x\in\mathcal{X}} (c(x,y) - c(x,y')) \\ &\leq \sup_{x\in\mathcal{X}} |c(x,y) - c(x,y')| \\ &= \rho_2(y,y'). \end{aligned}$$

Hence we have $|g^*(y) - g^*(y')| \leq \rho_2(y, y')$ for all $y, y' \in \mathcal{X}$ according to the symmetry of ρ_2. Therefore $(f^*, g^*) \in \mathcal{G}_c''(\mathcal{X})$. □

Since $(f, g) \in \mathcal{G}_c'(\mathcal{Y})$, we have for fixed $x \in \mathcal{Y}$

$$f(x) \leq c(x,y) - g(y)$$

for all $y \in \mathcal{Y}$ and hence

$$f(x) \leq \inf_{y\in\mathcal{Y}} (c(x,y) - g(y)) = f^*(x).$$

Similarly, for fixed $y \in \mathcal{Y}$ we have:

$$\begin{aligned} g^*(y) &= \inf_{x\in\mathcal{X}} (c(x,y) - f^*(x)) \\ &= \inf_{x\in\mathcal{X}} \left(c(x,y) - \inf_{y'\in\mathcal{Y}} (c(x,y') - g(y')) \right) \\ &\geq \inf_{x\in\mathcal{X}} (c(x,y) - c(x,y) + g(y)) \\ &= g(y). \end{aligned}$$

Thus $f(x) \leq f^*(x)$ for all $x \in \mathcal{Y}$ and $g(y) \leq g^*(y)$ for all $y \in \mathcal{Y}$. Therefore

$$\int_{\mathcal{Y}} f(x)\ P_1(dx) \leq \int_{\mathcal{Y}} f^*(x)\ P_1(dx), \tag{2.8}$$

$$\int_{\mathcal{Y}} g(y)\ P_2(dy) \leq \int_{\mathcal{Y}} g^*(y)\ P_2(dy). \tag{2.9}$$

Summing up both sides of (2.8) and (2.9) respectively yields

$$\int_{\mathcal{Y}} f(x)\ P_1(dx) + \int_{\mathcal{Y}} g(y)\ P_2(dy) \leq \int_{\mathcal{Y}} f^*(x)\ P_1(dx) + \int_{\mathcal{Y}} g^*(y)\ P_2(dy)$$

for all $(f, g) \in \mathcal{G}_c'(\mathcal{Y})$. Since $P_1(\mathcal{X} \setminus \mathcal{Y}) = P_2(\mathcal{X} \setminus \mathcal{Y}) = 0$, we can replace $\mathcal{Y}$ by $\mathcal{X}$, e.g.

$$\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \leq \int_{\mathcal{X}} f^*(x)\ P_1(dx) + \int_{\mathcal{X}} g^*(y)\ P_2(dy)$$

for all $(f,g) \in \mathcal{G}_c'(\mathcal{Y})$. Hence, since as shown above $(f^*, g^*) \in \mathcal{G}_c''(\mathcal{X})$ for all $(f,g) \in \mathcal{G}_c'(\mathcal{Y})$,

$$\sup_{(f,g)\in\mathcal{G}_c'(\mathcal{Y})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \leq \sup_{(f,g)\in\mathcal{G}_c''(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right). \tag{2.10}$$

On the other hand, since $\mathcal{G}_c'(\mathcal{X}) \subseteq \mathcal{G}_c'(\mathcal{Y})$, we have

$$\sup_{(f,g)\in\mathcal{G}_c'(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \leq \sup_{(f,g)\in\mathcal{G}_c'(\mathcal{Y})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right)$$

and this yields

$$\sup_{(f,g)\in\mathcal{G}_c'(\mathcal{Y})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \geq \sup_{(f,g)\in\mathcal{G}_c''(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \tag{2.11}$$

because

$$\sup_{(f,g)\in\mathcal{G}_c''(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \leq \sup_{(f,g)\in\mathcal{G}_c'(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right).$$

Now the assertion follows from (2.10) and (2.11). □

We proceed in the proof of Theorem 2.1 by considering different cases for the space $\mathcal{X}$:

1) Let $\mathcal{X}$ be a finite space. So, there exists some $n \in \mathbb{N}$ such that $\mathcal{X} = \{x_1, ..., x_n\}$. Then

$$\begin{aligned} &\inf_{P\in\mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} c(x,y)\ P(dx,dy) \\ = &\inf_{\substack{P(x_{i_1},x_{i_2}) \geq 0\ \forall\, i_1,i_2, \\ \sum_{i_2=1}^n P(x_{i_1},x_{i_2}) = P_1(x_{i_1})\ \forall\, i_1, \\ \sum_{i_1=1}^n P(x_{i_1},x_{i_2}) = P_2(x_{i_2})\ \forall\, i_2}} \left(\sum_{i_1=1}^n \sum_{i_2=1}^n c(x_{i_1},x_{i_2}) \cdot P(x_{i_1},x_{i_2}) \right) \end{aligned}$$

and according to the duality principle in linear programming this is equal to

$$\begin{aligned} &\sup_{f(x_{i_1})+g(x_{i_2})\leq c(x_{i_1},x_{i_2})\ \forall\, i_1,i_2} \left(\sum_{i=1}^n \left(f(x_i)\cdot P_1(x_i) + g(x_i)\cdot P_2(x_i) \right) \right) \\ = &\sup_{(f,g)\in\mathcal{G}_c'(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right). \end{aligned}$$

Hence by Lemma 2.2 (with $\mathcal{Y} = \mathcal{X}$) we have

$$\begin{aligned} \inf_{P\in\mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} c(x,y)\ P(dx,dy) &= \sup_{(f,g)\in\mathcal{G}_c'(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \\ &= \sup_{(f,g)\in\mathcal{G}_c''(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \\ &\leq \sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right), \end{aligned}$$

because $\mathcal{G}''_c(\mathcal{X}) \subseteq \mathcal{G}_c(\mathcal{X})$. In particular, there exists some Borel probability measure $P^* \in \mathcal{C}(P_1, P_2)$ such that

$$\int_{\mathcal{X}^2} c(x,y)\ P^*(dx,dy) = \inf_{P \in \mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} c(x,y)\ P(dx,dy).$$

2) Let $\mathcal{X}$ be a compact set.

Claim 4. *For any $n \in \mathbb{N}$ there exist sets $A_1, \dots, A_{m_n} \in \mathcal{B}(\mathcal{X})$, $A_k \neq \emptyset$, $diam(A_k) < \frac{1}{n}$ for all $k = 1, \dots, m_n$ and $A_k \cap A_l = \emptyset$ for all $k \neq l$, such that*

$$\mathcal{X} = \dot{\bigcup}_{k=1}^{m_n} A_k.$$

Proof. Let $n \in \mathbb{N}$. Suppose $(\mathcal{O}_i)_{i\in\mathbb{N}}$ is an open cover of $\mathcal{X}$ such that $\operatorname{diam}(\mathcal{O}_i) < \frac{1}{n}$ for all $i \in \mathbb{N}$. Since $\mathcal{X}$ is compact, there exist $m_n \in \mathbb{N}$ and i_k, $k = 1, \dots, m_n$, such that $\mathcal{X} \subseteq \bigcup_{k=1}^{m_n} \mathcal{O}_{i_k}$. For every $k \in \{1, ..., m_n\}$ define $\mathcal{O}'_{i_k} := \mathcal{O}_{i_k} \setminus \dot{\bigcup}_{j=1}^{k-1} \mathcal{O}'_{i_j}$. Observe that $\mathcal{O}'_{i_k}$, $k = 1, \dots, m_n$, are disjoint and $\mathcal{X} \subseteq \dot{\bigcup}_{k=1}^{m_n} \mathcal{O}'_{i_k}$. Defining $A_k := \mathcal{O}'_{i_k} \cap \mathcal{X}$ for every $k = 1, \dots, m_n$, we finally obtain $\mathcal{X} = \dot{\bigcup}_{k=1}^{m_n} A_k$. □

Define a mapping $h_n \colon \mathcal{X} \to \mathcal{X}_n := \{x_1, ..., x_{m_n}\} \subseteq \mathcal{X}$ by $h_n(x) := x_k$ if $x \in A_k$ for some $x_k \in A_k$, $k = 1, \dots, m_n$. It is not difficult to show that $h_n(x) \to x$ as $n \to \infty$. Extend the Borel probability measures $P_i \circ h_n^{-1}$, $i = 1, 2$, to $\mathcal{X}$ in an obvious way, namely $(P_i \circ h_n^{-1})_e(A) := (P_i \circ h_n^{-1})(A \cap \mathcal{X}_n)$ for all $A \in \mathcal{B}(\mathcal{X})$. In order to simplify notation we will not distinguish between $(P_i \circ h_n^{-1})_e$ and $P_i \circ h_n^{-1}$. Note that the measures $P_1 \circ h_n^{-1}$ and $P_2 \circ h_n^{-1}$ are completely supported by $\mathcal{X}_n$. Hence using Lemma 2.2 with $\mathcal{Y} = \mathcal{X}_n$ and the Borel probability measures $P_i \circ h_n^{-1}$ instead of P_i for $i = 1, 2$ we gain

$$\begin{aligned}
& \sup_{(f,g)\in\mathcal{G}'_c(\mathcal{X}_n)} \left(\int_{\mathcal{X}} f(x)\ \left(P_1 \circ h_n^{-1}\right)(dx) + \int_{\mathcal{X}} g(y)\ \left(P_2 \circ h_n^{-1}\right)(dy) \right) \\
= & \sup_{(f,g)\in\mathcal{G}''_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ \left(P_1 \circ h_n^{-1}\right)(dx) + \int_{\mathcal{X}} g(y)\ \left(P_2 \circ h_n^{-1}\right)(dy) \right) \\
\leq & \sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ \left(P_1 \circ h_n^{-1}\right)(dx) + \int_{\mathcal{X}} g(y)\ \left(P_2 \circ h_n^{-1}\right)(dy) \right) \\
= & \sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})} \underbrace{\left(\int_{\mathcal{X}} f(h_n(x))\ P_1(dx) + \int_{\mathcal{X}} g(h_n(y))\ P_2(dy) \right)}_{=\left(\int_{\mathcal{X}} f(h_n(x)) - \frac{K}{n}\ P_1(dx) + \int_{\mathcal{X}} g(h_n(y))\ P_2(dy)\right) + \frac{K}{n}},
\end{aligned}$$

where we use the transformation theorem for the last equality.

Let $(f,g) \in \mathcal{G}_c(\mathcal{X})$. Denoting by $\|f\|_d$ and $\|g\|_d$ the Lipschitz-constant of f and g respectively, we then have

$$\begin{aligned} f(h_n(x)) + g(h_n(y)) &\leq |f(h_n(x)) - f(x)| + |g(h_n(y)) - g(y)| + f(x) + g(y) \\ &\leq \|f\|_d \cdot \underbrace{d(h_n(x), x)}_{<\frac{1}{n}} + \|g\|_d \cdot \underbrace{d(h_n(y), y)}_{<\frac{1}{n}} + c(x,y) \\ &< \frac{1}{n} \cdot (\|f\|_d + \|g\|_d) + c(x,y) \\ &= \frac{K}{n} + c(x,y) \end{aligned}$$

for all $x, y \in \mathcal{X}$ and some constant $K \in \mathbb{R}_+$ which is independent of n and x, y. Since, in addition, $f \circ h_n$ and $g \circ h_n$ are Borel measurable and bounded, we have $(f \circ h_n - \frac{K}{n}, g \circ h_n) \in \mathcal{G}'_c(\mathcal{X})$. Therefore, in order to lose the n-dependence of the supremum, we find

$$\begin{aligned} &\sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(h_n(x)) - \frac{K}{n}\ P_1(dx) + \int_{\mathcal{X}} g(h_n(y))\ P_2(dy) \right) + \frac{K}{n} \\ \leq\ &\sup_{(\tilde{f},\tilde{g})\in\mathcal{G}'_c(\mathcal{X})} \left(\int_{\mathcal{X}} \tilde{f}(x)\ P_1(dx) + \int_{\mathcal{X}} \tilde{g}(y)\ P_2(dy) \right) + \frac{K}{n}. \end{aligned} \tag{2.12}$$

Since $\mathcal{X}_n$ is finite, by 1) there exists a measure $P^{(n)} \in \mathcal{C}(P_1 \circ h_n^{-1}, P_2 \circ h_n^{-1})$ such that

$$\begin{aligned} &\int_{\mathcal{X}^2} c(x,y)\ P^{(n)}(dx,dy) \\ =\ &\int_{\mathcal{X}_n^2} c(x,y)\ P^{(n)}(dx,dy) \\ =\ &\inf_{Q\in\mathcal{C}(P_1\circ h_n^{-1}, P_2\circ h_n^{-1})} \int_{\mathcal{X}_n^2} c(x,y)\ Q(dx,dy) \\ =\ &\sup_{(f,g)\in\mathcal{G}'_c(\mathcal{X}_n)} \left(\int_{\mathcal{X}} f(x)\ \left(P_1 \circ h_n^{-1}\right)(dx) + \int_{\mathcal{X}} g(y)\ \left(P_2 \circ h_n^{-1}\right)(dy) \right). \end{aligned} \tag{2.13}$$

Claim 5. *$P_i \circ h_n^{-1} \longrightarrow P_i$ weakly on $\mathcal{X}$ as $n \to \infty$ for $i = 1, 2$.*

Proof. Let $f \in \mathcal{C}_b(\mathcal{X})$ and $i = 1, 2$. Applying the transformation theorem and Lebesgue yields

$$\begin{aligned} \lim_{n\to\infty} \int_{\mathcal{X}} f(x)\ \left(P_i \circ h_n^{-1}\right)(dx) &= \lim_{n\to\infty} \int_{\mathcal{X}} f(h_n(x))\ P_i(dx) \\ &= \int_{\mathcal{X}} \lim_{n\to\infty} f(h_n(x))\ P_i(dx) \\ &= \int_{\mathcal{X}} f(x)\ P_i(dx), \end{aligned}$$

because f is continuous and $h_n(x)$ tends to x as $n \to \infty$. Therefore $P_i \circ h_n^{-1}$ converges weakly to P_i as $n \to \infty$. □

Since $\mathcal{X}_n \subseteq \mathcal{X}$ for all $n \in \mathbb{N}$ and $\mathcal{X}$ is compact, $(P_i \circ h_n^{-1})_{n\in\mathbb{N}}$ is uniformly tight. But $\mathcal{X}^2$ is compact as well and $P^{(n)} \in \mathcal{C}(P_1 \circ h_n^{-1}, P_2 \circ h_n^{-1})$. Hence $(P^{(n)})_{n\in\mathbb{N}}$ is uniformly tight and thus relatively compact. Therefore there exists a Borel probability measure P^* on $\mathcal{X}^2$ and a subsequence $(P^{(n_k)})_{k\in\mathbb{N}}$ such that $P^{(n_k)} \longrightarrow P^*$ weakly as $k \to \infty$. Since $P_i \circ h_n^{-1} \longrightarrow P_i$ weakly as $n \to \infty$ for $i = 1, 2$, a monotone class argument applies in order to show that $P^* \in \mathcal{C}(P_1, P_2)$. Hence by (2.13) and (2.12)

$$\begin{aligned}
&\int_{\mathcal{X}^2} c(x,y)\ P^*(dx,dy) \\
=\ & \lim_{k\to\infty} \int_{\mathcal{X}^2} c(x,y)\ P^{(n_k)}(dx,dy) \\
=\ & \lim_{k\to\infty} \left(\sup_{(f,g)\in\mathcal{G}'_c(\mathcal{X}_{n_k})} \left(\int_{\mathcal{X}} f(x)\ \left(P_1 \circ h_{n_k}^{-1}\right)(dx) + \int_{\mathcal{X}} g(y)\ \left(P_2 \circ h_{n_k}^{-1}\right)(dy) \right) \right) \\
\leq\ & \lim_{k\to\infty} \left(\sup_{(f,g)\in\mathcal{G}'_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) + \frac{K}{n_k} \right) \\
=\ & \sup_{(f,g)\in\mathcal{G}'_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right).
\end{aligned}$$

Therefore, since $P^* \in \mathcal{C}(P_1, P_2)$, using Lemma 2.2 with $\mathcal{Y} = \mathcal{X}$ we obtain

$$\begin{aligned}
\inf_{P\in\mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} c(x,y)\ P(dx,dy) &\leq \int_{\mathcal{X}^2} c(x,y)\ P^*(dx,dy) \\
&\leq \sup_{(f,g)\in\mathcal{G}'_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \\
&= \sup_{(f,g)\in\mathcal{G}''_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \\
&\leq \sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right).
\end{aligned}$$

Moreover, according to (2.6) the coupling $P^* \in \mathcal{C}(P_1, P_2)$ is optimal for the primal problem.

3) Let $(\mathcal{X}, d)$ be a separable metric space with bounded metric d, i.e. there exists some constant $K \in \mathbb{R}_+$ such that $d(x,y) \leq K$ for all $x, y \in \mathcal{X}$. Since H is Lipschitz-continuous, it follows

$$\begin{aligned}
\int_{\mathcal{X}} \rho_1(x,a)\ P_1(dx) &= \int_{\mathcal{X}} \sup_{y\in\mathcal{X}} |c(x,y) - c(a,y)|\ P_1(dx) \\
&= \int_{\mathcal{X}} \sup_{y\in\mathcal{X}} |H(d(x,y)) - H(d(a,y))|\ P_1(dx) \\
&= \int_{\mathcal{X}} \underbrace{\|H(d(x,\cdot)) - H(d(a,\cdot))\|_\infty}_{\leq \mathrm{Lip}(H)\cdot\|d(x,\cdot)-d(a,\cdot)\|_\infty}\ P_1(dx)
\end{aligned}$$

$$\begin{aligned} &\leq \operatorname{Lip}(H) \cdot \int_{\mathcal{X}} \underbrace{d(a,x)}_{\leq K} \; P_1(dx) \\ &\leq \operatorname{Lip}(H) \cdot K \\ &< \infty \end{aligned}$$

and in the same way $\int_{\mathcal{X}} \rho_2(y,a) \; P_2(dy) < \infty$ for some fixed $a \in \mathcal{X}$.
3.1) First of all, let P_1, P_2 be tight probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Then for every $n \in \mathbb{N}$ there exists a compact set $K_n \subseteq \mathcal{X}$ such that

$$\sup_{i=1,2} \int_{K_n^c} \rho_i(x,a) \; P_i(dx) \leq \frac{1}{n}. \tag{2.14}$$

For $A \in \mathcal{B}(\mathcal{X})$ define

$$P_{i,n}(A) := P_i(A \cap K_n) + P_i(K_n^c) \cdot \delta_a(A).$$

Since

$$P_{i,n}(K_n \cup \{a\}) = P_i((K_n \cup \{a\}) \cap K_n) + P_i(K_n^c) \cdot \delta_a(K_n \cup \{a\}) = P_i(K_n) + P_i(K_n^c) = 1$$

for $i = 1,2$, according to Lemma 2.2 (applied to $\mathcal{Y} = K_n \cup \{a\}$ and the Borel probability measures $P_{1,n}, P_{2,n}$) we have

$$\begin{aligned} &\sup_{(f,g) \in \mathcal{G}_c'(K_n \cup \{a\})} \left(\int_{\mathcal{X}} f(x) \; P_{1,n}(dx) + \int_{\mathcal{X}} g(y) \; P_{2,n}(dy) \right) \\ = &\sup_{(f,g) \in \mathcal{G}_c''(\mathcal{X})} \left(\int_{\mathcal{X}} f(x) \; P_{1,n}(dx) + \int_{\mathcal{X}} g(y) \; P_{2,n}(dy) \right). \end{aligned}$$

Observe that for the first integral on the right hand side we have

$$\begin{aligned} \int_{\mathcal{X}} f(x) \; P_{1,n}(dx) &= \int_{K_n} f(x) \; P_1(dx) + P_1(K_n^c) \cdot \int_{\mathcal{X}} f(x) \; \delta_a(dx) \\ &= \int_{\mathcal{X}} f(x) \; P_1(dx) - \int_{K_n^c} f(x) \; P_1(dx) + P_1(K_n^c) \cdot f(a) \\ &= \int_{\mathcal{X}} f(x) \; P_1(dx) + \int_{K_n^c} \underbrace{f(a) - f(x)}_{\leq |f(a)-f(x)| \leq \rho_1(a,x)} \; P_1(dx) \\ &\leq \int_{\mathcal{X}} f(x) \; P_1(dx) + \int_{K_n^c} \rho_1(a,x) \; P_1(dx) \end{aligned}$$

and analogously for the second integral above

$$\int_{\mathcal{X}} g(y) \; P_{2,n}(dy) \leq \int_{\mathcal{X}} g(y) \; P_2(dy) + \int_{K_n^c} \rho_2(a,y) \; P_2(dy).$$

Hence by (2.14)

$$\begin{aligned}
&\sup_{(f,g)\in\mathcal{G}_c''(\mathcal{X})}\left(\int_{\mathcal{X}} f(x)\ P_{1,n}(dx) + \int_{\mathcal{X}} g(y)\ P_{2,n}(dy)\right)\\
\leq\ &\sup_{(f,g)\in\mathcal{G}_c''(\mathcal{X})}\left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy)\right.\\
&\qquad\left. + \int_{K_n^c} \rho_1(a,x)\ P_1(dx) + \int_{K_n^c} \rho_2(a,y)\ P_2(dy)\right)\\
=\ &\sup_{(f,g)\in\mathcal{G}_c''(\mathcal{X})}\left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy)\right)\\
&\qquad + \underbrace{\int_{K_n^c} \rho_1(a,x)\ P_1(dx)}_{\leq\frac{1}{n}} + \underbrace{\int_{K_n^c} \rho_2(a,y)\ P_2(dy)}_{\leq\frac{1}{n}}\\
\leq\ &\sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})}\left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy)\right) + \frac{2}{n},
\end{aligned}\tag{2.15}$$

because $\mathcal{G}_c''(\mathcal{X}) \subseteq \mathcal{G}_c(\mathcal{X})$. Since $K_n \cup \{a\}$ is compact, according to the case 2) for all $n \in \mathbb{N}$ there exists a measure $P^{(n)} \in \mathcal{C}(P_{1,n}, P_{2,n})$ such that

$$\int_{\mathcal{X}^2} c(x,y)\ P^{(n)}(dx,dy) \ \leq \sup_{(f,g)\in\mathcal{G}_c'(K_n\cup\{a\})}\left(\int_{\mathcal{X}} f(x)\ P_{1,n}(dx) + \int_{\mathcal{X}} g(y)\ P_{2,n}(dy)\right). \tag{2.16}$$

Since each $P_{i,n}, n \in \mathbb{N}$, is tight (as mentioned above $P_{i,n}(K_n \cup \{a\}) = 1$ for all $n \in \mathbb{N}$) and $P_{i,n} \overset{n\to\infty}{\longrightarrow} P_i$ (weakly), where P_i is tight as well by assumption, $i = 1, 2$, by Theorem 8 in Appendix III of [2], $\{P_{i,n} \mid n \in \mathbb{N}\} \cup \{P_i\}$ is uniformly tight.

Now analogously to the case 2) we conclude that $(P^{(n)})_{n\in\mathbb{N}}$ is uniformly tight and thus relatively compact. Therefore there exists a measure P^* and a subsequence $(P^{(n_k)})_{k\in\mathbb{N}}$ such that $P^{(n_k)} \overset{k\to\infty}{\longrightarrow} P^*$ weakly. Since $P_{i,n}$ converges weakly to P_i, in order to verify $P^* \in \mathcal{C}(P_1, P_2)$, we proceed in the same way as in the case 2) above, e.g. by applying a monotone class argument. Therefore using (2.16) and (2.15) yields

$$\begin{aligned}
&\int_{\mathcal{X}^2} c(x,y)\ P^*(dx,dy)\\
=\ &\lim_{k\to\infty}\int_{\mathcal{X}^2} c(x,y)\ P^{(n_k)}(dx,dy)\\
\leq\ &\lim_{k\to\infty}\left(\sup_{(f,g)\in\mathcal{G}_c'(K_{n_k}\cup\{a\})}\left(\int_{\mathcal{X}} f(x)\ P_{1,n_k}(dx) + \int_{\mathcal{X}} g(y)\ P_{2,n_k}(dy)\right)\right)\\
\leq\ &\lim_{k\to\infty}\left(\sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})}\left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy)\right) + \frac{2}{n_k}\right)\\
=\ &\sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})}\left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy)\right).
\end{aligned}$$

In particular, since $P^* \in \mathcal{C}(P_1, P_2)$, we have

$$\begin{aligned}\inf_{P\in\mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} c(x,y)\ P(dx,dy) &\leq \int_{\mathcal{X}^2} c(x,y)\ P^*(dx,dy) \\ &\leq \sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy)\right).\end{aligned}$$

According to (2.6) the coupling $P^* \in \mathcal{C}(P_1, P_2)$ is optimal for the primal problem.

3.2) Now let P_1, P_2 be Borel probability measures on $\mathcal{X}$ that are not necessarily tight and denote by $\bar{\mathcal{X}}$ the completion of $\mathcal{X}$ with respect to d. For $\varepsilon > 0$ choose the largest set $A := A(\varepsilon) \subseteq \bar{\mathcal{X}}$ such that $d(x,y) \geq \frac{\varepsilon}{2}$ for all $x, y \in A,\ x \neq y$, that is, if $z \in A^c$, then there exists some $x \in A$ such that $d(x,z) < \frac{\varepsilon}{2}$. The existence of such a set follows from the Lemma of Zorn (cf. Theorem I.2 in [13]). In fact, let $\mathbb{M}$ be the collection of all subsets of $\bar{\mathcal{X}}$ which fulfill the distance condition, namely their elements have distance greater or equal $\frac{\varepsilon}{2}$. Then $\mathbb{M} \neq \emptyset$ and it is clear that $(\mathbb{M}, \subseteq)$ is partially ordered. Furthermore every chain $(M_n)_{n\in\mathbb{N}}$, $M_n \subseteq M_{n+1}$ for all $n \in \mathbb{N}$, has an upper bound. (Just take the union $M := \bigcup_{n=1}^{\infty} M_n$.) Hence the Lemma of Zorn is applicable and yields at least one maximal element $A \in \mathbb{M}$. Moreover, the set A is countable. In fact, since $\bar{\mathcal{X}}$ is separable, we can find some countable set $\{y_n \mid n \in \mathbb{N}\}$ which is dense in $\bar{\mathcal{X}}$ with respect to d. Hence $\left(\bar{B}_{\frac{\varepsilon}{4}}(y_n)\right)_{n\in\mathbb{N}}$, where $\bar{B}_\delta(y_n) := \{x \in \bar{\mathcal{X}} \mid d(x, y_n) < \delta\}$, is an open cover of $\bar{\mathcal{X}}$. Assume, by contradiction, that $A \subseteq \bar{\mathcal{X}}$ is not countable. Then there is at least some $n \in \mathbb{N}$ such that the open ball $\bar{B}_{\frac{\varepsilon}{4}}(y_n)$ contains uncountable many elements of A, in particular more than two. This contradicts the definition of A. Hence the set A must be countable. So, let $A = \{x_n \mid n \in \mathbb{N}\}$.

For every $n \in \mathbb{N}$ define $\bar{A}_n \subseteq \bar{\mathcal{X}}$ by $\bar{A}_n := \bar{B}_{\frac{\varepsilon}{2}}(x_n) \setminus \left(\bigcup_{j=1}^{n-1} \bar{B}_{\frac{\varepsilon}{2}}(x_j)\right)$ and put $A_n := \bar{A}_n \cap \mathcal{X}$. Note that $\bar{A}_n \in \mathcal{B}(\bar{\mathcal{X}})$ for all $n \in \mathbb{N}$ and $\bar{A}_m \cap \bar{A}_n = \emptyset$ for all $m \neq n$. Furthermore $\bigcup_{n=1}^{\infty} \bar{A}_n = \bar{\mathcal{X}}$. Indeed, suppose there is some $x \in \bar{\mathcal{X}} \setminus \left(\bigcup_{n=1}^{\infty} \bar{A}_n\right)$. Hence $x \notin \bar{A}_n$ for all $n \in \mathbb{N}$ and consequently $x \notin \bar{B}_{\frac{\varepsilon}{2}}(x_n)$ for all $n \in \mathbb{N}$, i.e. $d(x, x_n) \geq \frac{\varepsilon}{2}$ for all $n \in \mathbb{N}$, in contradiction to the choice of A. Therefore $\bar{\mathcal{X}} \subseteq \bigcup_{n=1}^{\infty} \bar{A}_n$. The converse inclusion is obvious.

Let $\bar{P}_i$ be the probability measure on $(\bar{\mathcal{X}}, \mathcal{B}(\bar{\mathcal{X}}))$ generated by P_i for $i = 1, 2$, i.e. $\bar{P}_i(\bar{A}) = P_i(\bar{A} \cap \mathcal{X})$ for all $\bar{A} \in \mathcal{B}(\bar{\mathcal{X}})$. According to Lemma C.1 in the appendix $\bar{P}_1, \bar{P}_2$ are tight and hence by 3.1) there exists a probability measure $\bar{P} \in \mathcal{C}(\bar{P}_1, \bar{P}_2)$ (on $(\bar{\mathcal{X}}^2, \mathcal{B}(\bar{\mathcal{X}}^2))$ such that

$$\int_{\bar{\mathcal{X}}^2} c(x,y)\ \bar{P}(dx,dy) = \sup_{(f,g)\in\mathcal{G}_c(\bar{\mathcal{X}})} \left(\int_{\bar{\mathcal{X}}} f(x)\ \bar{P}_1(dx) + \int_{\bar{\mathcal{X}}} g(y)\ \bar{P}_2(dy)\right). \tag{2.17}$$

Let $P_{i,m}$, $i = 1, 2$, $m \in \mathbb{N}$, be the restriction of the Borel probability measure P_i to the set A_m, i.e. $P_{i,m}(B) := P_i(B \cap A_m)$ for all $B \in \mathcal{B}(\mathcal{X})$. To any dual index (m_1, m_2), $m_1, m_2 \in \mathbb{N}$, define the product measure $\mu_{(m_1,m_2)}$ on $(\mathcal{X}^2, \mathcal{B}(\mathcal{X}^2))$ by

$\mu_{(m_1,m_2)} := c_{(m_1,m_2)} \cdot (P_{1,m_1} \otimes P_{2,m_2})$, where the constant $c_{(m_1,m_2)}$ is chosen such that $\mu_{(m_1,m_2)}(A_{m_1} \times A_{m_2}) = \bar{P}(\bar{A}_{m_1} \times \bar{A}_{m_2})$. Let $P_\varepsilon := \sum_{(m_1,m_2)} \mu_{(m_1,m_2)}$. Then for any $B \in \mathcal{B}(\mathcal{X})$ we have

$$
\begin{aligned}
& P_\varepsilon(B \times \mathcal{X}) \\
= & \sum_{(m_1,m_2)} \mu_{(m_1,m_2)}(B \times \mathcal{X}) \\
= & \sum_{(m_1,m_2)} c_{(m_1,m_2)} \cdot \left(P_{1,m_1} \otimes P_{2,m_2}\right)(B \times \mathcal{X}) \\
= & \sum_{(m_1,m_2)} c_{(m_1,m_2)} \cdot P_{1,m_1}(B) \cdot P_{2,m_2}(\mathcal{X}) \\
= & \sum_{(m_1,m_2)} c_{(m_1,m_2)} \cdot P_1(B \cap A_{m_1}) \cdot P_2(\mathcal{X} \cap A_{m_2}) \\
= & \sum_{(m_1,m_2)} c_{(m_1,m_2)} \cdot P_1(B \cap A_{m_1}) \cdot P_2(A_{m_2}) \\
= & \sum_{(m_1,m_2):\ P_{i,m_i}(A_{m_i})>0\ \forall\ i=1,2} \frac{\bar{P}(\bar{A}_{m_1} \times \bar{A}_{m_2})}{P_{1,m_1}(A_{m_1}) \cdot P_{2,m_2}(A_{m_2})} \cdot P_1(B \cap A_{m_1}) \cdot P_2(A_{m_2}) \\
= & \sum_{m_1 \in \mathbb{N}:\ P_{1,m_1}(A_{m_1})>0} \frac{\bar{P}(\bar{A}_{m_1} \times (\bigcup_{m_2 \in \mathbb{N}} \bar{A}_{m_2}))}{P_{1,m_1}(A_{m_1})} \cdot P_1(B \cap A_{m_1}) \\
= & \sum_{m_1 \in \mathbb{N}:\ P_{1,m_1}(A_{m_1})>0} \frac{\bar{P}(\bar{A}_{m_1} \times \bar{\mathcal{X}})}{P_{1,m_1}(A_{m_1})} \cdot P_1(B \cap A_{m_1}) \\
= & \sum_{m_1 \in \mathbb{N}:\ P_{1,m_1}(A_{m_1})>0} \frac{\bar{P}_1(\bar{A}_{m_1})}{P_{1,m_1}(A_{m_1})} \cdot P_1(B \cap A_{m_1}) \\
= & P_1\Big(B \cap \underbrace{\bigcup_{m_1 \in \mathbb{N}} A_{m_1}}_{=\mathcal{X}}\Big) \\
= & P_1(B).
\end{aligned}
$$

Similarly, one can show

$$P_\varepsilon(\mathcal{X} \times B) = P_2(B)$$

for all $B \in \mathcal{B}(\mathcal{X})$. Therefore $P_\varepsilon \in \mathcal{C}(P_1, P_2)$.

We will prove

$$\limsup_{\varepsilon \to 0} \int_{\mathcal{X}^2} c(x,y)\ P_\varepsilon(dx,dy) \le \sup_{(f,g) \in \mathcal{G}_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right).$$

In order to achieve this, let $\alpha > 0$ and observe, that if $(y_1, y_2) \in A_{m_1} \times A_{m_2}$ such that

$d(y_1, y_2) > \alpha + 2\varepsilon$, we have using the triangle inequality $d(x_{m_1}, x_{m_2}) > \alpha + \varepsilon$. Hence

$$\begin{aligned} P_\varepsilon(d(y_1,y_2) > \alpha + 2\varepsilon) &= \sum_{(m_1,m_2)} \mu_{(m_1,m_2)}(\{(y_1,y_2) \in A_{m_1} \times A_{m_2} \mid d(y_1,y_2) > \alpha + 2\varepsilon\}) \\ &\leq \sum_{(m_1,m_2):\ d(x_{m_1},x_{m_2})>\alpha+\varepsilon} \underbrace{\mu_{(m_1,m_2)}(A_{m_1} \times A_{m_2})}_{=\bar{P}(\bar{A}_{m_1}\times\bar{A}_{m_2})} \\ &= \sum_{(m_1,m_2):\ d(x_{m_1},x_{m_2})>\alpha+\varepsilon} \bar{P}\left(\bar{A}_{m_1} \times \bar{A}_{m_2}\right). \end{aligned}$$

Applying the triangle inequality a second time, e.g. $\alpha+\varepsilon < d(x_{m_1}, x_{m_2}) \leq d(x_{m_1}, y_1) + d(y_1, y_2) + d(y_2, x_{m_2}) < d(y_1, y_2) + \varepsilon$, the right hand side above is less or equal to

$$\begin{aligned} &\sum_{(m_1,m_2)} \bar{P}\left(\{(y_1,y_2) \in \bar{A}_{m_1} \times \bar{A}_{m_2} \mid d(y_1,y_2) > \alpha\}\right) \\ &= \bar{P}\left(\{(y_1,y_2) \in \bar{\mathcal{X}}^2 \mid d(y_1,y_2) > \alpha\}\right) \\ &= \bar{P}(d(y_1,y_2) > \alpha). \end{aligned}$$

Therefore $P_\varepsilon(d > \alpha + 2\varepsilon) \leq \bar{P}(d > \alpha)$ for all $\alpha > 0$. Since $H\colon \mathbb{R}_+ \to \mathbb{R}_+$ is strictly increasing, by the fundamental theorem of calculus and Fubini this implies

$$\begin{aligned} &\int_{\mathcal{X}^2} c(x,y)\ P_\varepsilon(dx,dy) \\ =\ &\int_{\mathcal{X}^2} H(d(x,y))\ P_\varepsilon(dx,dy) \\ =\ &\int_{\mathcal{X}^2} \underbrace{\left(\int_0^{d(x,y)} 1\ dH(t)\right)}_{=\int_0^\infty \mathbf{1}_{[0,d(x,y)[}(t)\ dH(t)} P_\varepsilon(dx,dy) \\ =\ &\int_0^\infty \int_{\mathcal{X}^2} \underbrace{\mathbf{1}_{[0,d(x,y)[}(t)}_{=\mathbf{1}_{\{(x,y)\in\mathcal{X}^2:\ d(x,y)>t\}}(x,y)} P_\varepsilon(dx,dy)\ dH(t) \\ =\ &\int_0^\infty \underbrace{P_\varepsilon(d(x,y) > t)}_{=P_\varepsilon(d(x,y)+2\varepsilon>t+2\varepsilon)}\ dH(t) \\ \leq\ &\int_0^\infty \underbrace{\bar{P}(d(x,y) + 2\varepsilon > t)}_{=\bar{P}(d(x,y)>t-2\varepsilon)}\ dH(t) \\ =\ &\int_0^\infty \bar{P}(d(x,y) > t)\ dH(t+2\varepsilon) + H(2\varepsilon) \\ =\ &\int_0^\infty \left(\int_{\bar{\mathcal{X}}^2} \underbrace{\mathbf{1}_{\{(x,y)\in\bar{\mathcal{X}}^2:\ d(x,y)>t\}}(x,y)}_{=\mathbf{1}_{[0,d(x,y)[}(t)}\ \bar{P}(dx,dy)\right) dH(t+2\varepsilon) + H(2\varepsilon) \end{aligned}$$

$$
\begin{aligned}
&= \int_{\bar{\mathcal{X}}^2} \int_0^\infty \mathbf{1}_{[0,d(x,y)[}(t)\ dH(t+2\varepsilon)\ \bar{P}(dx,dy) + H(2\varepsilon) \\
&= \int_{\bar{\mathcal{X}}^2} \int_0^{d(x,y)} 1\ dH(t+2\varepsilon)\ \bar{P}(dx,dy) + H(2\varepsilon) \\
&= \int_{\bar{\mathcal{X}}^2} H(d(x,y)+2\varepsilon) - \underbrace{H(2\varepsilon)}_{\geq 0}\ \bar{P}(dx,dy) + H(2\varepsilon) \\
&\leq \int_{\bar{\mathcal{X}}^2} \underbrace{H(d(x,y))}_{=c(x,y)}\ \bar{P}(dx,dy) + \int_{\bar{\mathcal{X}}^2} H(d(x,y)+2\varepsilon) - H(d(x,y))\ \bar{P}(dx,dy) \\
&\qquad +H(2\varepsilon). \qquad (2.18)
\end{aligned}
$$

Consider the second integral on the right hand side:

$$
\begin{aligned}
&\int_{\bar{\mathcal{X}}^2} H(d(x,y)+2\varepsilon) - H(d(x,y))\ \bar{P}(dx,dy) \\
= &\int_{\{(x,y)\in\bar{\mathcal{X}}^2:\ d(x,y)\leq 2p\}} H(d(x,y)+2\varepsilon) - H(d(x,y))\ \bar{P}(dx,dy) \\
&+ \int_{\{(x,y)\in\bar{\mathcal{X}}^2:\ d(x,y)> 2p\}} H(d(x,y)+2\varepsilon) - H(d(x,y))\ \bar{P}(dx,dy) \\
\leq &\sup_{t\in[0,2p]} (H(t+2\varepsilon) - H(t)) + \int_{\{(x,y)\in\bar{\mathcal{X}}^2:\ d(x,y)> 2p\}} H(d(x,y)+2\varepsilon)\ \bar{P}(dx,dy).
\end{aligned}
$$

Since $d(x,y) \leq d(x,a) + d(y,a)$ for some fixed $a \in \mathcal{X}$, applying two times the Orlicz' condition and the convexity to the integrand yields

$$
\begin{aligned}
H(d(x,y)+2\varepsilon) &\leq K_1 \cdot H\left(\frac{1}{2}\cdot d(x,y) + \frac{1}{2}\cdot 2\varepsilon\right) \\
&\leq \frac{K_1}{2}\cdot (H(d(x,y)) + H(2\varepsilon)) \\
&\leq \frac{K_1}{2}\cdot \left(K_2 \cdot H\left(\frac{1}{2}\cdot d(x,a) + \frac{1}{2}\cdot d(y,a)\right) + H(2\varepsilon)\right) \\
&\leq \frac{K_1}{2}\cdot \left(\frac{K_2}{2}\cdot (H(d(x,a)) + H(d(y,a))) + H(2\varepsilon)\right) \\
&= \frac{K_1K_2}{4}\cdot (H(d(x,a)) + H(d(y,a))) + \frac{K_1}{2}\cdot H(2\varepsilon)
\end{aligned}
$$

for suitable constants $K_1, K_2 \in \mathbb{R}_+$. Hence

$$\begin{aligned}
&\int_{\{(x,y)\in\bar{\mathcal{X}}^2:\, d(x,y)>2p\}} H(d(x,y)+2\varepsilon)\, \bar{P}(dx,dy) \\
\leq\ & \frac{K_1K_2}{4}\int_{\{(x,y)\in\bar{\mathcal{X}}^2:\, d(x,a)+d(y,a)>2p\}} H(d(x,a))+H(d(y,a))\, \bar{P}(dx,dy)+\frac{K_1}{2}\cdot H(2\varepsilon) \\
\leq\ & K\left(\int_{\{x\in\bar{\mathcal{X}}:\, d(x,a)>p\}} H(d(x,a))\, \bar{P}_1(dx)+\int_{\{y\in\bar{\mathcal{X}}:\, d(y,a)>p\}} H(d(y,a))\, \bar{P}_2(dy)\right) \\
&\qquad +\frac{K_1}{2}\cdot H(2\varepsilon),
\end{aligned}$$

because $\{(x,y)\in\bar{\mathcal{X}}^2 : d(x,a)+d(y,a)>2p\} \subseteq \{(x,y)\in\bar{\mathcal{X}}^2 : d(x,a)>p\}\cup\{(x,y)\in\bar{\mathcal{X}}^2 : d(y,a)>p\}$, where we put $K := \frac{K_1K_2}{4}$. Therefore

$$\begin{aligned}
&\int_{\bar{\mathcal{X}}^2} H(d(x,y)+2\varepsilon)-H(d(x,y))\, \bar{P}(dx,dy) \\
\leq\ & \sup_{t\in[0,2p]}(H(t+2\varepsilon)-H(t))+K\cdot\Bigg(\int_{\bar{\mathcal{X}}} H(d(x,a))\mathbf{1}_{\{x\in\bar{\mathcal{X}}:\, d(x,a)>p\}}(x)\, \bar{P}_1(dx) \\
&\qquad +\int_{\bar{\mathcal{X}}} H(d(y,a))\mathbf{1}_{\{y\in\bar{\mathcal{X}}:\, d(y,a)>p\}}(y)\, \bar{P}_2(dy)\Bigg)+\frac{K_1}{2}\cdot H(2\varepsilon)
\end{aligned}$$

for some constants $K, K_1 > 0$ which are independent of ε and p. Therefore letting first $\varepsilon \to 0$ and then $p \to \infty$ in (2.18) yields

$$\begin{aligned}
\limsup_{\varepsilon\to 0}\int_{\mathcal{X}^2} c(x,y)\, P_\varepsilon(dx,dy) &\leq \int_{\bar{\mathcal{X}}^2} c(x,y)\, \bar{P}(dx,dy) \\
&= \sup_{(f,g)\in\mathcal{G}_c(\bar{\mathcal{X}})}\left(\int_{\bar{\mathcal{X}}} f(x)\, \bar{P}_1(dx)+\int_{\bar{\mathcal{X}}} g(y)\, \bar{P}_2(dy)\right) \\
&\leq \sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})}\left(\int_{\mathcal{X}} f(x)\, P_1(dx)+\int_{\mathcal{X}} g(y)\, P_2(dy)\right)
\end{aligned}$$

according to (2.17) and the fact that $\mathcal{G}_c(\bar{\mathcal{X}}) \subseteq \mathcal{G}_c(\mathcal{X})$. Since $P_\varepsilon \in \mathcal{C}(P_1,P_2)$ for all $\varepsilon > 0$, the conclusion follows.

Second part: Separable metric space $(\mathcal{X}, d)$ *with unbounded metric* d. Now let $(\mathcal{X}, d)$ be any separable metric space. So in general the metric d is not bounded any more.

As before, first suppose P_1, P_2 to be tight Borel probability measures. For each $n \in \mathbb{N}$ define the bounded metric $d_n\colon \mathcal{X}\times\mathcal{X}\to\mathbb{R}_+$ by

$$d_n(x,y) := d(x,y)\wedge n$$

for all $(x,y)\in\mathcal{X}^2$. Since $(\mathcal{X}, d_n)$ is a separable metric space with bounded metric d_n for every $n\in\mathbb{N}$, by the case 3.1) of the first part there exists a measure $P^{(n)}\in\mathcal{C}(P_1,P_2)$

such that

$$
\begin{aligned}
& \int_{\mathcal{X}^2} H(d_n(x,y))\ P^{(n)}(dx,dy) \\
= & \inf_{P \in \mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} H(d_n(x,y))\ P(dx,dy) \\
= & \sup_{(f,g) \in \mathcal{G}_c((\mathcal{X},d_n))} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right),
\end{aligned}
\tag{2.19}
$$

where $\mathcal{G}_c((\mathcal{X},d_n))$ is defined in an obvious way, namely as $\mathcal{G}_c(\mathcal{Y})$ in (2.4) with d_n replacing the metric d and $\mathcal{X}$ replacing $\mathcal{Y}$ respectively.

Since $(P^{(n)})_{n\in\mathbb{N}}$ is uniformly tight ($P^{(n)} \in \mathcal{C}(P_1,P_2)$ for all $n \in \mathbb{N}$ and the measures P_1, P_2 are tight by assumption; the same assertion appears later in this thesis - cf. the proof of Lemma 3.4 - and the exact 2ε-argument is given there), it is relatively compact and so there exists a subsequence $(P^{(n_k)})_{k\in\mathbb{N}}$ and a measure $P^{(0)} \in \mathcal{C}(P_1,P_2)$ such that $P^{(n_k)}$ converges weakly to $P^{(0)}$ as $k \to \infty$. By the Skorokhod Theorem (cf. [5], Theorem 11.7.2) there is a probability space (Ω,μ) and a sequence $(X_k)_{k\in\mathbb{N}}$ of random variables $X_k\colon \Omega \to \mathcal{X}^2$ such that X_k has distribution $P^{(n_k)}$ for all $k \in \mathbb{N}$ and $X_k \longrightarrow X_0$ μ-a.s. as $k \to \infty$.

Hence using (2.19), the transformation theorem and Fatou yields

$$
\begin{aligned}
& \liminf_{k\to\infty} \sup_{(f,g)\in\mathcal{G}_c((\mathcal{X},d_{n_k}))} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \\
= & \liminf_{k\to\infty} \int_{\mathcal{X}^2} H(d_{n_k}(x,y))\ P^{(n_k)}(dx,dy) \\
= & \liminf_{k\to\infty} \int_{\mathcal{X}^2} H(d_{n_k}(x,y))\ \left(\mu \circ X_k^{-1}\right)(dx,dy) \\
= & \liminf_{k\to\infty} \int_{\Omega} H(d_{n_k}(X_k(\omega)))\ \mu(d\omega) \\
= & \liminf_{k\to\infty} \mathbb{E}_\mu[H(d_{n_k}(X_k))] \\
\geq & \mathbb{E}_\mu[\liminf_{k\to\infty} H(d_{n_k}(X_k))] \\
= & \mathbb{E}_\mu[H(d(X_0)) + \liminf_{k\to\infty} (H(d_{n_k}(X_k)) - H(d(X_0)))] \\
= & \mathbb{E}_\mu[H(d(X_0)) - \limsup_{k\to\infty} (H(d(X_0)) - H(d_{n_k}(X_k)))] \\
\geq & \mathbb{E}_\mu[H(d(X_0)) - \limsup_{k\to\infty} |H(d(X_0)) - H(d_{n_k}(X_k))|] \\
= & \mathbb{E}_\mu[H(d(X_0)) - \limsup_{k\to\infty} |H(d_{n_k}(X_k)) - H(d(X_0))|] \\
= & \mathbb{E}_\mu[H(d(X_0))] - \mathbb{E}_\mu[\limsup_{k\to\infty} |H(d_{n_k}(X_k)) - H(d(X_0))|] \\
= & \mathbb{E}_\mu[H(d(X_0))],
\end{aligned}
\tag{2.20}
$$

because

$$\begin{aligned}
&|H(d_{n_k}(X_k)) - H(d(X_0))| \\
\leq\ & \underbrace{|H(d_{n_k}(X_k)) - H(d_{n_k}(X_0))|}_{\stackrel{k\to\infty}{\longrightarrow} 0 \ \ \mu\text{-a.s.}} + \underbrace{|H(d_{n_k}(X_0)) - H(d(X_0))|}_{\stackrel{k\to\infty}{\longrightarrow} 0, \text{ because } d_{n_k} \nearrow d} \stackrel{k\to\infty}{\longrightarrow} 0 \quad \mu\text{-a.s.}
\end{aligned}$$

Furthermore, since $X_k \stackrel{k\to\infty}{\longrightarrow} X_0$ μ-a.s., by the transformation theorem, the Orlicz' condition and the convexity of H

$$\begin{aligned}
& \mathbb{E}_\mu \left[\limsup_{k\to\infty} \left(H(d_{n_k}(X_k)) + H(d(X_0)) \right) \right] \\
=\ & \int_\Omega \limsup_{k\to\infty} H(d_{n_k}(X_k(\omega)))\, \mu(d\omega) + \int_\Omega H(d(X_0(\omega)))\, \mu(d\omega) \\
\leq\ & \int_\Omega \limsup_{k\to\infty} H(d(X_k(\omega)))\, \mu(d\omega) + \int_\Omega H(d(X_0(\omega)))\, \mu(d\omega) \\
=\ & 2 \cdot \int_\Omega H(d(X_0(\omega)))\, \mu(d\omega) \\
=\ & 2 \cdot \int_{\mathcal{X}^2} \underbrace{H(d(x,y))}_{\leq \frac{K}{2}\cdot(H(d(x,a))+H(d(y,a)))} P^{(0)}(dx,dy) \\
\leq\ & K \cdot \int_{\mathcal{X}^2} H(d(x,a)) + H(d(y,a))\ P^{(0)}(dx,dy) \\
=\ & K \cdot \left(\int_{\mathcal{X}} H(d(x,a))\ P_1(dx) + \int_{\mathcal{X}} H(d(y,a))\ P_2(dy) \right) \\
=\ & K \cdot \left(\int_{\mathcal{X}} c(x,a)\ P_1(dx) + \int_{\mathcal{X}} c(y,a)\ P_2(dy) \right) \\
<\ & \infty.
\end{aligned}$$

Since $\mathcal{G}_c((\mathcal{X}, d_{n_k})) \subseteq \mathcal{G}_c(\mathcal{X})$ for all $k \in \mathbb{N}$, we have

$$\begin{aligned}
& \sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \\
\geq\ & \sup_{(f,g)\in\mathcal{G}_c((\mathcal{X},d_{n_k}))} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right)
\end{aligned}$$

for all $k \in \mathbb{N}$ and therefore by (2.20)

$$\begin{aligned}
& \sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \\
\geq\ & \liminf_{k\to\infty} \sup_{(f,g)\in\mathcal{G}_c((\mathcal{X},d_{n_k}))} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \\
\geq\ & \mathbb{E}_\mu \left[H(d(X_0)) \right] \\
=\ & \int_\Omega H(d(X_0(\omega)))\, \mu(d\omega)
\end{aligned}$$

$$\begin{aligned} &= \int_{\mathcal{X}^2} H(d(x,y))\; P^{(0)}(dx,dy) \\ &= \int_{\mathcal{X}^2} c(x,y)\; P^{(0)}(dx,dy) \\ &\geq \inf_{P\in\mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} c(x,y)\; P(dx,dy). \end{aligned}$$

In particular, according to (2.6) the measure $P^{(0)} \in \mathcal{C}(P_1, P_2)$ is optimal for the primal problem, i.e.

$$\int_{\mathcal{X}^2} c(x,y)\; P^{(0)}(dx,dy) = \inf_{P\in\mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} c(x,y)\; P(dx,dy). \tag{2.21}$$

It remains to show equality (2.5) for not necessarily tight measures P_1, P_2. Since the boundedness of the metric d is not used in 3.2) of the first part, this could be done in the same way as thereby with the only exception that the existence of some Borel probability measure $\bar{P} \in \mathcal{C}(\bar{P}_1, \bar{P}_2)$ such that

$$\int_{\bar{\mathcal{X}}^2} c(x,y)\; \bar{P}(dx,dy) = \sup_{(f,g)\in\mathcal{G}(\bar{\mathcal{X}})} \left(\int_{\bar{\mathcal{X}}} f(x)\; \bar{P}_1(dx) + \int_{\bar{\mathcal{X}}} g(y)\; \bar{P}_2(dy) \right)$$

follows from (2.21) here.

2.2 Dual representation for the Kantorovich-Rubinstein problem

Similarly as for the Monge-Kantorovich problem, we now go on by proving a dual representation for the Kantorovich-Rubinstein problem (2.3).

Let $(\mathcal{X}, d)$ be a separable metric space and suppose $c\colon \mathcal{X}\times\mathcal{X} \to \mathbb{R}_+$ and $\lambda\colon \mathcal{X} \to \mathbb{R}_+$ to be measurable functions such that

(C1) $c(x,y) = 0$ if and only if $x = y$;

(C2) $c(x,y) = c(y,x)$ for all $(x,y) \in \mathcal{X}^2$;

(C3) $c(x,y) \leq \lambda(x) + \lambda(y)$ for all $x, y \in \mathcal{X}$;

(C4) λ maps bounded sets to bounded sets, that is $\operatorname{diam}(\lambda(B)) := \sup_{r,s\in\lambda(B)}|r-s| \leq K_{\lambda(B)}$ for some constant $K_{\lambda(B)} \in \mathbb{R}_+$ for all sets $B \subseteq \mathcal{X}$ with $\operatorname{diam}(B) := \sup_{x,y\in B} d(x,y) \leq K_B$ for some constant $K_B \in \mathbb{R}_+$;

(C5)

$$\sup_{x,y\in B(a,r)\colon d(x,y)\leq\delta} c(x,y) \stackrel{\delta\to 0}{\longrightarrow} 0$$

for each $a \in \mathcal{X}$ and $r > 0$, where $B(a,r) := \{x \in \mathcal{X} \mid d(x,a) < r\}$ denotes the open ball with radius $r > 0$ centered at $a \in \mathcal{X}$.

Note that any function $c \in \mathbf{C}$ satisfies (C1) to (C5). In particular, the metric d satisfies all the above conditions (C1) to (C5) with λ given by $\lambda(x) := d(x,a)$ for some fixed $a \in \mathcal{X}$.

Denote by $\mathcal{L}_c(\mathcal{X},\mathbb{R})$ the linear space of all c-Lipschitz continuous functions φ from $\mathcal{X}$ to $\mathbb{R}$, that is there exists some constant $K \in \mathbb{R}_+$ such that $|\varphi(x) - \varphi(y)| \leq K \cdot c(x,y)$ for all $x, y \in \mathcal{X}$, and define $\|\cdot\|_c \colon \mathcal{L}_c(\mathcal{X},\mathbb{R}) \to \mathbb{R}_+$ by

$$\|\varphi\|_c := \sup_{x,y\in\mathcal{X}\colon x\neq y} \frac{|\varphi(x) - \varphi(y)|}{c(x,y)}. \tag{2.22}$$

Note that for $\varphi \in \mathcal{L}_c(\mathcal{X},\mathbb{R})$ we have

$$|\varphi(x)-\varphi(y)| = \frac{|\varphi(x) - \varphi(y)|}{c(x,y)} \cdot c(x,y) \leq \left(\sup_{x,y\in\mathcal{X}\colon x\neq y} \frac{|\varphi(x) - \varphi(y)|}{c(x,y)} \right) \cdot c(x,y) = \|\varphi\|_c \cdot c(x,y)$$

for all $x, y \in \mathcal{X}$, $x \neq y$. Moreover, $\|\varphi\|_c$ is the smallest such constant $K \in \mathbb{R}_+$, called the *Lipschitz constant* of φ. In particular, $\|\varphi\|_c < \infty$. Hence by condition (C5) we conclude that each $\varphi \in \mathcal{L}_c(\mathcal{X},\mathbb{R})$ is continuous (with respect to d) and thus measurable.

Lemma 2.3. *$\|\cdot\|_c$ is a seminorm on $\mathcal{L}_c(\mathcal{X},\mathbb{R})$.*

Proof. For $\varphi = 0 \in \mathcal{L}_c(\mathcal{X},\mathbb{R})$ we obtain $\|\varphi\|_c = 0$.

Let $\alpha \in \mathbb{R}$, $\varphi \in \mathcal{L}_c(\mathcal{X},\mathbb{R})$. Then

$$\begin{aligned}
\|\alpha\varphi\|_c &= \sup_{x,y\in\mathcal{X}\colon x\neq y} \frac{|(\alpha\varphi)(x) - (\alpha\varphi)(y)|}{c(x,y)} \\
&= \sup_{x,y\in\mathcal{X}\colon x\neq y} \frac{|\alpha(\varphi(x) - \varphi(y))|}{c(x,y)} \\
&= |\alpha| \cdot \sup_{x,y\in\mathcal{X}\colon x\neq y} \frac{|\varphi(x) - \varphi(y)|}{c(x,y)} \\
&= |\alpha| \cdot \|\varphi\|_c.
\end{aligned}$$

Let $\varphi, \psi \in \mathcal{L}_c(\mathcal{X},\mathbb{R})$. Then

$$\begin{aligned}
\|\varphi + \psi\|_c &= \sup_{x,y\in\mathcal{X}\colon x\neq y} \frac{|(\varphi+\psi)(x) - (\varphi+\psi)(y)|}{c(x,y)} \\
&= \sup_{x,y\in\mathcal{X}\colon x\neq y} \frac{|\varphi(x) + \psi(x) - (\varphi(y) + \psi(y))|}{c(x,y)} \\
&= \sup_{x,y\in\mathcal{X}\colon x\neq y} \frac{|\varphi(x) - \varphi(y) + \psi(x) - \psi(y)|}{c(x,y)} \\
&\leq \sup_{x,y\in\mathcal{X}\colon x\neq y} \frac{|\varphi(x) - \varphi(y)|}{c(x,y)} + \sup_{x,y\in\mathcal{X}\colon x\neq y} \frac{|\psi(x) - \psi(y)|}{c(x,y)} \\
&= \|\varphi\|_c + \|\psi\|_c.
\end{aligned}$$

□

Since $\|\varphi\|_c = 0$ if and only if φ is constant, $\|\cdot\|_c$ is not a norm on $\mathcal{L}_c(\mathcal{X},\mathbb{R})$ at this stage. To obtain a norm we have to consider equivalence classes with respect to $\|\cdot\|_c$. For this define

$$\mathcal{N} := \{\varphi \in \mathcal{L}_c(\mathcal{X},\mathbb{R}) \mid \|\varphi\|_c = 0\} = \{\varphi \in \mathcal{L}_c(\mathcal{X},\mathbb{R}) \mid \exists\ u \in \mathbb{R} : \varphi(x) = u\ \forall\ x \in \mathcal{X}\}$$

and set

$$\mathbb{L}_c(\mathcal{X},\mathbb{R}) := \mathcal{L}_c(\mathcal{X},\mathbb{R})/\mathcal{N}.$$

Now define $\|\cdot\|_c : \mathbb{L}_c(\mathcal{X},\mathbb{R}) \to \mathbb{R}_+$ by $\|[\varphi]\|_c := \|\varphi\|_c$ where $\varphi \in [\varphi]$ is an arbitrary representative of $[\varphi]$.

Remark 2.4. *$\|\cdot\|_c : \mathbb{L}_c(\mathcal{X},\mathbb{R}) \to \mathbb{R}_+$ is welldefined, since for $\varphi, \psi \in [\varphi]$ two representatives, there exists some $u \in \mathbb{R}$ such that $\psi(x) = \varphi(x) + u$ for all $x \in \mathcal{X}$ and thus we have*

$$\begin{aligned}\|\varphi\|_c &= \sup_{x,y\in\mathcal{X}:\ x\neq y} \frac{|\varphi(x)-\varphi(y)|}{c(x,y)} \\ &= \sup_{x,y\in\mathcal{X}:\ x\neq y} \frac{|\varphi(x)+u-(\varphi(y)+u)|}{c(x,y)} \\ &= \sup_{x,y\in\mathcal{X}:\ x\neq y} \frac{|\psi(x)-\psi(y)|}{c(x,y)} \\ &= \|\psi\|_c.\end{aligned}$$

Lemma 2.5. *$\|\cdot\|_c : \mathbb{L}_c(\mathcal{X},\mathbb{R}) \to \mathbb{R}_+$ is a norm.*

Proof. This is clear, because $\|[\varphi]\|_c = 0$ if and only if $[\varphi] = \mathcal{N} = [0]$. □

Although elements of $\mathbb{L}_c(\mathcal{X},\mathbb{R})$ are equivalence classes $[\varphi]$, according to Remark 2.4 and to shorten notation we shall write φ instead of $[\varphi]$.

Define the linear space

$$\mathcal{M}_\lambda(\mathcal{X}) := \left\{ m \text{ is a finite signed measure on } (\mathcal{X},\mathcal{B}(\mathcal{X})) \ \middle|\ m(\mathcal{X}) = 0,\ \int_{\mathcal{X}} \lambda(x)\,|m|(dx) < \infty \right\},$$

where $|m|(A) := m^+(A) + m^-(A)$ $(= 2\cdot\|m\|_{TV}(A))$, $A \in \mathcal{B}(\mathcal{X})$, is the total variation of the finite signed Borel measure m with Jordan decomposition $m = m^+ - m^-$. Furthermore for $m \in \mathcal{M}_\lambda(\mathcal{X})$ set

$$\begin{aligned}\mathbb{B}(m) &:= \{b \in \mathcal{M}(\mathcal{X}^2) \mid\ b(\mathcal{X}^2) < \infty,\ b(A\times\mathcal{X}) - b(\mathcal{X}\times A) = m(A)\ \forall\ A \in \mathcal{B}(\mathcal{X})\} \\ &= \{b \in \mathcal{M}(\mathcal{X}^2) \mid\ b(\mathcal{X}^2) < \infty,\ T_1 b - T_2 b = m\},\end{aligned}$$

where $T_i b$ denotes the i-th marginal of b, $i = 1, 2$.

Remark 2.6. *For each $m \in \mathcal{M}_\lambda(\mathcal{X})$ the corresponding $\mathbb{B}(m) \neq \emptyset$, because $\frac{m^+ \otimes m^-}{m^+(\mathcal{X})} \in \mathbb{B}(m)$. In fact, for $A \in \mathcal{B}(\mathcal{X})$ we have*

$$\begin{aligned} & \frac{m^+ \otimes m^-}{m^+(\mathcal{X})}(A \times \mathcal{X}) - \frac{m^+ \otimes m^-}{m^+(\mathcal{X})}(\mathcal{X} \times A) \\ = \; & \frac{m^+(A) \cdot m^-(\mathcal{X})}{m^+(\mathcal{X})} - \frac{m^+(\mathcal{X}) \cdot m^-(A)}{m^+(\mathcal{X})} \\ = \; & m^+(A) - m^-(A) \\ = \; & m(A), \end{aligned}$$

where we use $m^+(\mathcal{X}) = m^-(\mathcal{X})$ for the second equality.

Hence the following definition is meaningful: let $\|\cdot\|_w \colon \mathcal{M}_\lambda(\mathcal{X}) \to \mathbb{R}_+$ be given by

$$\|m\|_w := \inf_{b \in \mathbb{B}(m)} \int_{\mathcal{X}^2} c(x,y)\, b(dx,dy). \tag{2.23}$$

Then by the above remark and (C3) we have

$$\begin{aligned} \|m\|_w &= \inf_{b \in \mathbb{B}(m)} \int_{\mathcal{X}^2} c(x,y)\, b(dx,dy) \\ &\leq \int_{\mathcal{X}^2} c(x,y) \left(\frac{m^+ \otimes m^-}{m^+(\mathcal{X})} \right)(dx,dy) \\ &\leq \int_{\mathcal{X}^2} (\lambda(x) + \lambda(y)) \left(\frac{m^+ \otimes m^-}{m^+(\mathcal{X})} \right)(dx,dy) \\ &= \int_{\mathcal{X}} \lambda(x)\, m^+(dx) + \int_{\mathcal{X}} \lambda(y)\, m^-(dy) \\ &= \int_{\mathcal{X}} \lambda(x)\, |m|(dx) \\ &< \infty \end{aligned}$$

according to the definition of $\mathcal{M}_\lambda(\mathcal{X})$.

Note that for any two Borel probability measures P_1, P_2 on $\mathcal{X}$ with $\int_{\mathcal{X}} \lambda(x)\, P_i(dx) < \infty$, $i = 1, 2$, we have $P_1 - P_2 \in \mathcal{M}_\lambda(\mathcal{X})$. Hence $\|P_1 - P_2\|_w$ is welldefined and the final result of this section (cf. Theorem 2.16 below) claims that

$$R_c(P_1, P_2) = \|P_1 - P_2\|_w = \sup_{\|\varphi\|_c = 1} \int_{\mathcal{X}} \varphi(x)\, (P_1 - P_2)(dx).$$

In order to achieve this dual representation for the Kantorovich-Rubinstein problem, we start by proving the following

Lemma 2.7. *$\|\cdot\|_w$ is a seminorm on $\mathcal{M}_\lambda(\mathcal{X})$.*

Proof. Let $m \in \mathcal{M}_\lambda(\mathcal{X})$. Since by choice $c(x,y) \geq 0$ for all $(x,y) \in \mathcal{X}^2$ and b is a finite Borel measure on $\mathcal{X}^2$, we obviously have $\|m\|_w \geq 0$, where equality holds if $m = 0$.

In order to show $\|\alpha \cdot m\|_w = |\alpha| \cdot \|m\|_w$ for all $\alpha \in \mathbb{R}$, we distinguish the two cases $\alpha > 0$ and $\alpha < 0$:

If $\alpha > 0$ we have

$$\begin{aligned}
\|\alpha \cdot m\|_w &= \inf_{b \in \mathbb{B}(\alpha \cdot m)} \int_{\mathcal{X}^2} c(x,y)\; b(dx,dy) \\
&= \inf_{T_1 b - T_2 b = \alpha \cdot m} \int_{\mathcal{X}^2} c(x,y)\; b(dx,dy) \\
&= \inf_{T_1\left(\frac{1}{\alpha}b\right) - T_2\left(\frac{1}{\alpha}b\right) = m} \int_{\mathcal{X}^2} c(x,y)\; b(dx,dy) \\
&= \alpha \cdot \inf_{T_1\left(\frac{1}{\alpha}b\right) - T_2\left(\frac{1}{\alpha}b\right) = m} \int_{\mathcal{X}^2} c(x,y)\; \left(\frac{1}{\alpha} b\right)(dx,dy) \\
&= \alpha \cdot \inf_{\left(\frac{1}{\alpha}b\right) \in \mathbb{B}(m)} \int_{\mathcal{X}^2} c(x,y)\; \left(\frac{1}{\alpha} b\right)(dx,dy) \\
&= \alpha \cdot \|m\|_w.
\end{aligned}$$

Similarly, for $\alpha < 0$ we have, defining the finite Borel measure $\tilde{b}$ on $\mathcal{X}^2$ by $\tilde{b}(A_1 \times A_2) := b(A_2 \times A_1)$ for all $A_1, A_2 \in \mathcal{B}(\mathcal{X})$ and using the symmetry of c

$$\begin{aligned}
\|\alpha \cdot m\|_w &= \inf_{b \in \mathbb{B}(\alpha \cdot m)} \int_{\mathcal{X}^2} c(x,y)\; b(dx,dy) \\
&= \inf_{T_1 b - T_2 b = \alpha \cdot m} \int_{\mathcal{X}^2} c(x,y)\; b(dx,dy) \\
&= \inf_{T_1\left(\frac{1}{\alpha}b\right) - T_2\left(\frac{1}{\alpha}b\right) = m} \int_{\mathcal{X}^2} c(x,y)\; b(dx,dy) \\
&= \inf_{T_2\left(-\frac{1}{\alpha}b\right) - T_1\left(-\frac{1}{\alpha}b\right) = m} \int_{\mathcal{X}^2} c(x,y)\; b(dx,dy) \\
&= \inf_{T_1\left(-\frac{1}{\alpha}\tilde{b}\right) - T_2\left(-\frac{1}{\alpha}\tilde{b}\right) = m} \int_{\mathcal{X}^2} c(y,x)\; \tilde{b}(dy,dx) \\
&= |\alpha| \cdot \inf_{T_1\left(-\frac{1}{\alpha}\tilde{b}\right) - T_2\left(-\frac{1}{\alpha}\tilde{b}\right) = m} \int_{\mathcal{X}^2} c(y,x)\; \left(\frac{1}{|\alpha|} \tilde{b}\right)(dy,dx) \\
&= |\alpha| \cdot \inf_{T_1\left(-\frac{1}{\alpha}\tilde{b}\right) - T_2\left(-\frac{1}{\alpha}\tilde{b}\right) = m} \int_{\mathcal{X}^2} c(y,x)\; \left(-\frac{1}{\alpha} \tilde{b}\right)(dy,dx) \\
&= |\alpha| \cdot \inf_{\left(-\frac{1}{\alpha}\tilde{b}\right) \in \mathbb{B}(m)} \int_{\mathcal{X}^2} c(y,x)\; \left(-\frac{1}{\alpha} \tilde{b}\right)(dy,dx) \\
&= |\alpha| \cdot \|m\|_w.
\end{aligned}$$

It remains to show the triangle inequality. So, let $m_i \in \mathcal{M}_\lambda(\mathcal{X})$ and $b_i \in \mathbb{B}(m_i)$ for

$i = 1, 2$. Then for $b := b_1 + b_2$ we have

$$\int_{\mathcal{X}^2} c(x,y)\, b(dx,dy) = \int_{\mathcal{X}^2} c(x,y)\, b_1(dx,dy) + \int_{\mathcal{X}^2} c(x,y)\, b_2(dx,dy)$$

and

$$\begin{aligned} T_1 b - T_2 b &= T_1 (b_1 + b_2) - T_2 (b_1 + b_2) \\ &= T_1 b_1 + T_1 b_2 - T_2 b_1 - T_2 b_2 \\ &= (T_1 b_1 - T_2 b_1) + (T_1 b_2 - T_2 b_2)\,. \end{aligned}$$

Therefore

$$\begin{aligned} &\|m_1 + m_2\|_w \\ = &\inf_{b \in \mathbb{B}(m_1+m_2)} \int_{\mathcal{X}^2} c(x,y)\, b(dx,dy) \\ = &\inf_{T_1 b - T_2 b = m_1 + m_2} \int_{\mathcal{X}^2} c(x,y)\, b(dx,dy) \\ = &\inf_{(T_1 b_1 - T_2 b_1)+(T_1 b_2 - T_2 b_2) = m_1+m_2} \left(\int_{\mathcal{X}^2} c(x,y)\, b_1(dx,dy) + \int_{\mathcal{X}^2} c(x,y)\, b_2(dx,dy) \right) \\ \leq &\inf_{T_1 b_1 - T_2 b_1 = m_1} \int_{\mathcal{X}^2} c(x,y)\, b_1(dx,dy) + \inf_{T_1 b_2 - T_2 b_2 = m_2} \int_{\mathcal{X}^2} c(x,y)\, b_2(dx,dy) \\ = &\inf_{b_1 \in \mathbb{B}(m_1)} \int_{\mathcal{X}^2} c(x,y)\, b_1(dx,dy) + \inf_{b_2 \in \mathbb{B}(m_2)} \int_{\mathcal{X}^2} c(x,y)\, b_2(dx,dy) \\ = &\|m_1\|_w + \|m_2\|_w. \end{aligned}$$

□

Let $m \in \mathcal{M}_\lambda(\mathcal{X})$, $\varphi \in \mathcal{L}_c(\mathcal{X}, \mathbb{R})$ and $a \in \mathcal{X}$ fixed. Then

$$|\varphi(x)| \leq |\varphi(x) - \varphi(a)| + |\varphi(a)| \leq \|\varphi\|_c \cdot c(x,a) + |\varphi(a)| = K_1 \cdot c(x,a) + K_2$$

for all $x \in \mathcal{X}$ and some constants $K_1, K_2 \in \mathbb{R}_+$. Hence by definition of $\mathcal{M}_\lambda(\mathcal{X})$

$$\begin{aligned} \int_{\mathcal{X}} |\varphi(x)|\, |m|(dx) &\leq K_1 \int_{\mathcal{X}} c(x,a)\, |m|(dx) + K_2 \cdot |m|(\mathcal{X}) \\ &\leq K_1 \int_{\mathcal{X}} \lambda(x)\, |m|(dx) + (K_1 \lambda(a) + K_2) \cdot |m|(\mathcal{X}) < \infty, \end{aligned}$$

i.e. $\varphi \in \mathcal{L}_c(\mathcal{X}, \mathbb{R})$ is $|m|$-integrable, and induces a linear form

$$\begin{aligned} F_\varphi \colon \mathcal{M}_\lambda(\mathcal{X}) &\to \mathbb{R} \\ m &\mapsto F_\varphi(m) := \int_{\mathcal{X}} \varphi(x)\, m(dx). \end{aligned}$$

Remark 2.8. *Suppose that φ and ψ are in the same equivalence class $[\varphi]$ of $\mathbb{L}_c(\mathcal{X},\mathbb{R})$. Then $F_\varphi = F_\psi$. In fact, for $m \in \mathcal{M}_\lambda(\mathcal{X})$ we have*

$$F_\varphi(m) = \int_{\mathcal{X}} \varphi(x)\; m(dx) = \int_{\mathcal{X}} (\psi(x) + u)\;\; m(dx) = F_\psi(m) + u \cdot m(\mathcal{X}) = F_\psi(m).$$

Hence there is no reason to distinguish $\mathcal{L}_c(\mathcal{X},\mathbb{R})$ and $\mathbb{L}_c(\mathcal{X},\mathbb{R})$ when considering the functional F_φ on $\mathcal{M}_\lambda(\mathcal{X})$. Since by definition of $\mathbb{B}(m)$ the difference of the marginals of b coincides with m, we obtain

$$\begin{aligned}
|F_\varphi(m)| &= \left|\int_{\mathcal{X}} \varphi(x)\; m(dx)\right| \\
&= \left|\int_{\mathcal{X}} \varphi(x)\;\; (T_1 b - T_2 b)\,(dx)\right| \\
&= \left|\int_{\mathcal{X}} \varphi(x)\; (T_1 b)(dx) - \int_{\mathcal{X}} \varphi(y)\; (T_2 b)(dy)\right| \\
&= \left|\int_{\mathcal{X}^2} \varphi(x)\; b(dx,dy) - \int_{\mathcal{X}^2} \varphi(y)\; b(dx,dy)\right| \\
&= \left|\int_{\mathcal{X}^2} (\varphi(x) - \varphi(y))\;\; b(dx,dy)\right| \\
&\le \int_{\mathcal{X}^2} |\varphi(x) - \varphi(y)|\; b(dx,dy) \\
&\le \|\varphi\|_c \cdot \int_{\mathcal{X}^2} c(x,y)\; b(dx,dy)
\end{aligned}$$

for all $b \in \mathbb{B}(m)$. Hence

$$|F_\varphi(m)| \le \|\varphi\|_c \cdot \inf_{b\in\mathbb{B}(m)} \int_{\mathcal{X}^2} c(x,y)\; b(dx,dy) = \|\varphi\|_c \cdot \|m\|_w.$$

Therefore $F_\varphi \colon \mathcal{M}_\lambda(\mathcal{X}) \to \mathbb{R}$ is a continuous linear functional for each $\varphi \in \mathbb{L}_c(\mathcal{X},\mathbb{R})$, i.e. $F_\varphi \in \mathcal{M}_\lambda(\mathcal{X})^*$, with dual norm

$$\|F_\varphi\|_w^* \le \|\varphi\|_c. \tag{2.24}$$

Define the continuous linear transformation

$$\begin{aligned}
D\colon (\mathbb{L}_c(\mathcal{X},\mathbb{R}), \|\cdot\|_c) &\;\to\; (\mathcal{M}_\lambda(\mathcal{X})^*, \|\cdot\|_w^*) \\
\varphi &\;\mapsto\; D\varphi = F_\varphi.
\end{aligned}$$

Lemma 2.9. *The transformation $D\colon (\mathbb{L}_c(\mathcal{X},\mathbb{R}), \|\cdot\|_c) \to (\mathcal{M}_\lambda(\mathcal{X})^*, \|\cdot\|_w^*)$ is an isometry, i.e. $\|\varphi\|_c = \|F_\varphi\|_w^*$ for all $\varphi \in \mathbb{L}_c(\mathcal{X},\mathbb{R})$.*

Proof. For $x, y \in \mathcal{X}$ set $m_{xy} := \delta_x - \delta_y$ $(\in \mathcal{M}_\lambda(\mathcal{X}))$. Then

$$\|m_{xy}\|_w = \inf_{b\in\mathbb{B}(m_{xy})} \int_{\mathcal{X}^2} c(u,v)\; b(du,dv) \le \int_{\mathcal{X}^2} c(u,v)\; (\delta_x \otimes \delta_y)(du,dv) = c(x,y). \tag{2.25}$$

Hence for each $\varphi \in \mathbb{L}_c(\mathcal{X},\mathbb{R})$

$$\begin{aligned}
\|\varphi\|_c &= \sup_{x,y\in\mathcal{X}:\, x\neq y} \frac{|\varphi(x)-\varphi(y)|}{c(x,y)} \\
&= \sup_{x,y\in\mathcal{X}:\, x\neq y} \frac{\left|\int_{\mathcal{X}} \varphi(z)\,(\delta_x-\delta_y)(dz)\right|}{c(x,y)} \\
&= \sup_{x,y\in\mathcal{X}:\, x\neq y} \frac{|F_\varphi(m_{xy})|}{c(x,y)} \\
&\leq \|F_\varphi\|_w^* \cdot \sup_{x,y\in\mathcal{X}:\, x\neq y} \frac{\|m_{xy}\|_w}{c(x,y)} \\
&\leq \|F_\varphi\|_w^*.
\end{aligned}$$

Therefore we have $\|\varphi\|_c \leq \|F_\varphi\|_w^*$ and so according to (2.24) $\|\varphi\|_c = \|F_\varphi\|_w^*$, e.g. D is an isometry. □

We now pursue to show that D is an isometric isomorphism between the normed linear space $(\mathbb{L}_c(\mathcal{X},\mathbb{R}), \|\cdot\|_c)$ and the Banach space $(\mathcal{M}_\lambda(\mathcal{X})^*, \|\cdot\|_w^*)$. For this the following preliminaries will be useful: Define

$$\begin{aligned}
\mathcal{M}_0(\mathcal{X}) := \Big\{ & m \text{ is finite signed measure on } (\mathcal{X},\mathcal{B}(\mathcal{X})) \;\Big|\; \exists\ m_1, m_2 \in \mathcal{M}(\mathcal{X}) \text{ finite} \\
& \text{and with bounded supports such that } m_1(\mathcal{X}) = m_2(\mathcal{X}) \text{ and } m = m_1 - m_2 \Big\}.
\end{aligned}$$

Note that condition (C4) on λ implies that $\mathcal{M}_0(\mathcal{X}) \subseteq \mathcal{M}_\lambda(\mathcal{X})$.

Lemma 2.10. $\mathcal{M}_0(\mathcal{X}) \subseteq \mathcal{M}_\lambda(\mathcal{X})$ *is a dense subspace with respect to* $\|\cdot\|_w$.

Proof. Let $m \in \mathcal{M}_\lambda(\mathcal{X})$, $m \neq 0$, and $a \in \mathcal{X}$ fixed. Define $B_n := B(a,n) := \{x \in \mathcal{X} \mid d(x,a) < n\}$ for all $n \in \mathbb{N}$. Then there exists $n_0 \in \mathbb{N}$ such that $m^+(B_n)\cdot m^-(B_n) > 0$ for all $n \geq n_0$, where m^+, m^- denote as usual the measures of the Jordan decomposition $m = m^+ - m^-$ of m. For these $n \geq n_0$ define

$$\begin{aligned}
m_n \colon \mathcal{B}(\mathcal{X}) &\to \mathbb{R} \\
A &\mapsto m_n(A) := m^+(\mathcal{X}) \cdot \left(\frac{m^+(A\cap B_n)}{m^+(B_n)} - \frac{m^-(A\cap B_n)}{m^-(B_n)} \right)
\end{aligned}$$

and set $\delta_n := \frac{m^-(\mathcal{X})}{m^-(B_n)} - 1 \geq 0$, $\varepsilon_n := \frac{m^+(\mathcal{X})}{m^+(B_n)} - 1 \geq 0$. Since the open balls B_n exhaust $\mathcal{X}$ as $n \to \infty$, we have $\lim_{n\to\infty} \delta_n = \lim_{n\to\infty} \varepsilon_n = 0$. Note that $m_n \in \mathcal{M}_0(\mathcal{X})$ for all $n \in \mathbb{N}$. We will show that $(m_n)_{n\in\mathbb{N}}$ converges to m with respect to $\|\cdot\|_w$. First observe that for every $A \in \mathcal{B}(\mathcal{X})$ we obtain the following expression for $(m - m_n)(A)$:

$$\begin{aligned}
& (m - m_n)(A) \\
= \; & m(A) - m_n(A) \\
= \; & m(A) - m^+(\mathcal{X}) \cdot \left(\frac{m^+(A \cap B_n)}{m^+(B_n)} - \frac{m^-(A \cap B_n)}{m^-(B_n)} \right) \\
= \; & m(A) - \left(\frac{m^+(\mathcal{X})}{m^+(B_n)} - 1 \right) \cdot m^+(A \cap B_n) - m^+(A \cap B_n) \\
& \quad + \left(\frac{m^-(\mathcal{X})}{m^-(B_n)} - 1 \right) \cdot m^-(A \cap B_n) + m^-(A \cap B_n) \\
= \; & m(A) - \varepsilon_n \cdot m^+(A \cap B_n) + \delta_n \cdot m^-(A \cap B_n) - \big(m^+(A \cap B_n) - m^-(A \cap B_n) \big) \\
= \; & m(A) - \varepsilon_n \cdot m^+(A \cap B_n) + \delta_n \cdot m^-(A \cap B_n) - m(A \cap B_n) \\
= \; & m(A \setminus B_n) - \varepsilon_n \cdot m^+(A \cap B_n) + \delta_n \cdot m^-(A \cap B_n).
\end{aligned}$$

Define $\mu_n, \nu_n \colon \mathcal{B}(\mathcal{X}) \to \mathbb{R}_+$ by

$$\mu_n(A) := m^+(A \setminus B_n) + \delta_n \cdot m^-(A \cap B_n) \tag{2.26}$$

and

$$\nu_n(A) := m^-(A \setminus B_n) + \varepsilon_n \cdot m^+(A \cap B_n) \tag{2.27}$$

respectively. Note that μ_n, ν_n are finite with $m - m_n = \mu_n - \nu_n$. Moreover $\mu_n \ll |m|$ and $\nu_n \ll |m|$. In fact, for $A \in \mathcal{B}(\mathcal{X})$ with $0 = |m|(A) = m^+(A) + m^-(A)$ we have by definition of μ_n and ν_n, $\mu_n(A) = 0$ and $\nu_n(A) = 0$ respectively. Hence the Radon-Nikodym densities $\frac{d\mu_n}{d|m|}$ and $\frac{d\nu_n}{d|m|}$ exist. Since $m = m^+ - m^-$ is the Jordan decomposition of m, we can find sets $P, N \in \mathcal{B}(\mathcal{X})$ such that $P \cup N = \mathcal{X}$, $P \cap N = \emptyset$ and $m^+(N) = 0$, $m^-(P) = 0$. Therefore by (2.26) respectively (2.27) the Radon-Nikodym densities look like

$$\begin{aligned}
\frac{d\mu_n}{d|m|}(x) &= \begin{cases} 1 & \text{if } x \in P \setminus B_n \\ \delta_n & \text{if } x \in N \cap B_n \\ 0 & \text{if } x \in (N \setminus B_n) \cup (B_n \setminus N) = N \Delta B_n \end{cases} \\
\frac{d\nu_n}{d|m|}(y) &= \begin{cases} 1 & \text{if } y \in N \setminus B_n \\ \varepsilon_n & \text{if } y \in P \cap B_n \\ 0 & \text{if } y \in (P \setminus B_n) \cup (B_n \setminus P) = P \Delta B_n \end{cases} .
\end{aligned}$$

Next we claim that $b_n := \frac{\mu_n \otimes \nu_n}{\mu_n(\mathcal{X})} \in \mathbb{B}(m - m_n)$ for each $n \in \mathbb{N}$. In fact, for $A \in \mathcal{B}(\mathcal{X})$

we have

$$\begin{aligned} b_n(A \times \mathcal{X}) - b_n(\mathcal{X} \times A) &= \frac{\mu_n \otimes \nu_n}{\mu_n(\mathcal{X})}(A \times \mathcal{X}) - \frac{\mu_n \otimes \nu_n}{\mu_n(\mathcal{X})}(\mathcal{X} \times A) \\ &= \frac{\mu_n(A) \cdot \nu_n(\mathcal{X})}{\mu_n(\mathcal{X})} - \frac{\mu_n(\mathcal{X}) \cdot \nu_n(A)}{\mu_n(\mathcal{X})} \\ &= \mu_n(A) - \nu_n(A) \\ &= (\mu_n - \nu_n)(A) \\ &= (m - m_n)(A), \end{aligned}$$

where we use $\mu_n(\mathcal{X}) = \nu_n(\mathcal{X})$ for the third equality. (This can be shown by an elementary calculation.)

Noting that

$$\nu_n(\mathcal{X}) = \mu_n(\mathcal{X}) = m^+(\mathcal{X} \setminus B_n) + \big(m^-(\mathcal{X}) - m^-(B_n)\big) = |m|(\mathcal{X} \setminus B_n) = |m|(B_n^c),$$

we define the Radon-Nikodym derivative

$$f_n(x,y) := \frac{db_n}{d(|m| \otimes |m|)}(x,y) = \frac{1}{|m|(B_n^c)} \cdot \frac{d\mu_n}{d|m|}(x) \cdot \frac{d\nu_n}{d|m|}(y).$$

Claim 6. *The function $g(x,y) := (\sup_{n\in\mathbb{N}} f_n(x,y)) \cdot c(x,y)$ is $|m| \otimes |m|$-integrable.*

Proof. It will be shown that the function g is $|m| \otimes |m|$-integrable over various disjoint subsets of $\mathcal{X} \times \mathcal{X}$.

(i) g is $|m| \otimes |m|$-integrable over $P \times N$: Suppose $(x,y) \in P \times N$. Then

$$\begin{aligned} g(x,y) &= \left(\sup_{n\in\mathbb{N}} f_n(x,y)\right) \cdot c(x,y) \\ &= \left(\sup_{n\in\mathbb{N}} \frac{1}{|m|(B_n^c)} \cdot \frac{d\mu_n}{d|m|}(x) \cdot \frac{d\nu_n}{d|m|}(y)\right) \cdot c(x,y) \\ &= \left(\sup_{n\in\mathbb{N}} \frac{c(x,y)}{|m|(B_n^c)} \cdot \mathbf{1}_{B_n^c \times B_n^c}(x,y)\right) \\ &= \left(\sup_{n\in\mathbb{N}} \frac{c(x,y)}{|m|(B_n^c)} \cdot \mathbf{1}_{\dot{\bigcup}_{i=n}^{\infty}(B_i^c \times B_i^c)\setminus(B_{i+1}^c \times B_{i+1}^c)}(x,y)\right) \\ &= \left(\sup_{n\in\mathbb{N}} \underbrace{\frac{c(x,y)}{|m|(B_n^c)}}_{\leq \frac{c(x,y)}{|m|(B_i^c)}\ \forall\, i \geq n} \cdot \sum_{i=n}^{\infty} \mathbf{1}_{(B_i^c \times B_i^c)\setminus(B_{i+1}^c \times B_{i+1}^c)}(x,y)\right) \\ &\leq \left(\sup_{n\in\mathbb{N}} \sum_{i=n}^{\infty} \frac{c(x,y)}{|m|(B_i^c)} \cdot \mathbf{1}_{(B_i^c \times B_i^c)\setminus(B_{i+1}^c \times B_{i+1}^c)}(x,y)\right) \\ &= \sum_{i=1}^{\infty} \frac{c(x,y)}{|m|\,(B_i^c)} \cdot \mathbf{1}_{C_i}(x,y), \end{aligned}$$

where $C_i := (B_i^c \times B_i^c) \setminus \left(B_{i+1}^c \times B_{i+1}^c\right)$. According to Levi's monotone convergence theorem we obtain

$$\begin{aligned}
&\int_{P\times N} g(x,y)\ \ (|m|\otimes|m|)\,(dx,dy) \\
\leq\ & \int_{P\times N} \sum_{i=1}^{\infty} \frac{c(x,y)}{|m|\,(B_i^c)} \cdot \mathbf{1}_{C_i}(x,y)\ \ (|m|\otimes|m|)\,(dx,dy) \\
=\ & \sum_{i=1}^{\infty} \frac{1}{|m|\,(B_i^c)} \int_{P\times N} c(x,y)\cdot \mathbf{1}_{C_i}(x,y)\ \ (|m|\otimes|m|)\,(dx,dy) \\
\leq\ & \sum_{i=1}^{\infty} \frac{1}{|m|\,(B_i^c)} \int_{C_i} (\lambda(x)+\lambda(y))\ \ (|m|\otimes|m|)\,(dx,dy) \\
\leq\ & \sum_{i=1}^{\infty} \frac{2}{|m|\,(B_i^c)} \int_{\left(B_i^c\setminus B_{i+1}^c\right)\times B_i^c} \lambda(x)\ \ (|m|\otimes|m|)\,(dx,dy) \\
=\ & 2\cdot\sum_{i=1}^{\infty} \int_{B_i^c\setminus B_{i+1}^c} \lambda(x)\ |m|(dx) \\
=\ & 2\cdot\int_{B_1^c} \lambda(x)\ |m|(dx) \\
\leq\ & 2\cdot\int_{\mathcal{X}} \lambda(x)\ |m|(dx) \\
<\ & \infty,
\end{aligned}$$

because $m \in \mathcal{M}_\lambda(\mathcal{X})$.

(ii) g is $|m|\otimes|m|$-integrable over $P\times P$: show that $g(x,y) \leq \frac{1}{m^+(B_1)}\cdot c(x,y)$ on $P\times P$. Suppose $(x,y)\in P\times P$. Then

$$\begin{aligned}
g(x,y) &= \left(\sup_{n\in\mathbb{N}} f_n(x,y)\right)\cdot c(x,y) \\
&= \left(\sup_{n\in\mathbb{N}} \frac{1}{|m|\,(B_n^c)}\cdot\frac{d\mu_n}{d|m|}(x)\cdot\frac{d\nu_n}{d|m|}(y)\right)\cdot c(x,y) \\
&= \left(\sup_{n\in\mathbb{N}} \frac{1}{|m|\,(B_n^c)}\cdot 1\cdot\varepsilon_n\cdot\mathbf{1}_{B_n^c\times B_n}(x,y)\right)\cdot c(x,y) \\
&\leq \sup_{n\in\mathbb{N}} \frac{\varepsilon_n\cdot c(x,y)}{|m|(B_n^c)} \\
&= \sup_{n\in\mathbb{N}} \left(\frac{c(x,y)}{m^+(B_n)}\cdot(m^+(\mathcal{X})-m^+(B_n))\cdot\frac{1}{|m|(B_n^c)}\right) \\
&= \sup_{n\in\mathbb{N}} \left(\underbrace{\frac{m^+(B_n^c)}{|m|(B_n^c)}}_{\leq 1}\cdot\frac{c(x,y)}{m^+(B_n)}\right) \\
&\leq \frac{c(x,y)}{m^+(B_1)}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\int_{P\times P} g(x,y)\ \ (|m|\otimes|m|)\,(dx,dy)\\
\leq\ & \frac{1}{m^+(B_1)}\cdot\int_{P\times P} c(x,y)\ \ (|m|\otimes|m|)\,(dx,dy)\\
\leq\ & \frac{1}{m^+(B_1)}\cdot\int_{\mathcal{X}^2} \lambda(x)+\lambda(y)\ \ (|m|\otimes|m|)\,(dx,dy)\\
=\ & \frac{|m|(\mathcal{X})}{m^+(B_1)}\cdot\left(\int_{\mathcal{X}}\lambda(x)\ |m|(dx)+\int_{\mathcal{X}}\lambda(y)\ |m|(dy)\right)\\
<\ & \infty,
\end{aligned}$$

because $m\in\mathcal{M}_\lambda(\mathcal{X})$.

(iii) g is $|m|\otimes|m|$-integrable over $N\times N$: show that $g(x,y)\leq\frac{1}{m^-(B_1)}\cdot c(x,y)$ on $N\times N$. Let $(x,y)\in N\times N$. Then

$$\begin{aligned}
g(x,y) &= \left(\sup_{n\in\mathbb{N}} f_n(x,y)\right)\cdot c(x,y)\\
&= \left(\sup_{n\in\mathbb{N}}\frac{1}{|m|\,(B_n^c)}\cdot\frac{d\mu_n}{d|m|}(x)\cdot\frac{d\nu_n}{d|m|}(y)\right)\cdot c(x,y)\\
&= \left(\sup_{n\in\mathbb{N}}\frac{1}{|m|\,(B_n^c)}\cdot\delta_n\cdot 1\cdot\mathbf{1}_{B_n\times B_n^c}(x,y)\right)\cdot c(x,y)\\
&\leq \sup_{n\in\mathbb{N}}\frac{\delta_n\cdot c(x,y)}{|m|(B_n^c)}\\
&= \left(\sup_{n\in\mathbb{N}}\frac{c(x,y)}{m^-(B_n)}\cdot\left(m^-(\mathcal{X})-m^-(B_n)\right)\cdot\frac{1}{|m|(B_n^c)}\right)\\
&= \left(\sup_{n\in\mathbb{N}}\underbrace{\frac{m^-(B_n^c)}{|m|(B_n^c)}}_{\leq 1}\cdot\frac{c(x,y)}{m^-(B_n)}\right)\\
&\leq \frac{c(x,y)}{m^-(B_1)}.
\end{aligned}$$

Now the assertion follows in the same way as in (ii).

(iv) g is $|m|\otimes|m|$-integrable over $N\times P$: show that $g(x,y)\leq\frac{m^+(B_1^c)}{m^-(B_1)\cdot m^+(B_1)}\cdot c(x,y)$ on

$N \times P$. Let $(x, y) \in N \times P$. Then

$$\begin{aligned}
& g(x,y) \\
= \ & \left(\sup_{n\in\mathbb{N}} f_n(x,y)\right) \cdot c(x,y) \\
= \ & \left(\sup_{n\in\mathbb{N}} \frac{1}{|m|\,(B_n^c)} \cdot \frac{d\mu_n}{d|m|}(x) \cdot \frac{d\nu_n}{d|m|}(y)\right) \cdot c(x,y) \\
= \ & \left(\sup_{n\in\mathbb{N}} \frac{1}{|m|(B_n^c)} \cdot \delta_n \cdot \varepsilon_n \cdot \mathbf{1}_{B_n\times B_n}(x,y)\right) \cdot c(x,y) \\
= \ & \left(\sup_{n\in\mathbb{N}} \frac{1}{|m|(B_n^c)} \cdot \left(\frac{m^-(\mathcal{X})}{m^-(B_n)} - 1\right) \cdot \left(\frac{m^+(\mathcal{X})}{m^+(B_n)} - 1\right) \cdot \mathbf{1}_{B_n\times B_n}(x,y)\right) \cdot c(x,y) \\
= \ & \left(\sup_{n\in\mathbb{N}} \frac{1}{|m|(B_n^c)} \cdot \frac{1}{m^-(B_n)} \cdot \big(m^-(\mathcal{X}) - m^-(B_n)\big)\right. \\
& \left. \cdot \frac{1}{m^+(B_n)} \cdot \big(m^+(\mathcal{X}) - m^+(B_n)\big) \cdot \mathbf{1}_{B_n\times B_n}(x,y)\right) \cdot c(x,y) \\
= \ & \left(\sup_{n\in\mathbb{N}} \frac{1}{|m|(B_n^c)} \cdot \frac{m^-(B_n^c)}{m^-(B_n)} \cdot \frac{m^+(B_n^c)}{m^+(B_n)} \cdot \mathbf{1}_{B_n\times B_n}(x,y)\right) \cdot c(x,y) \\
\leq \ & \sup_{n\in\mathbb{N}} \frac{m^+(B_n^c)}{m^-(B_n) \cdot m^+(B_n)} \cdot c(x,y) \\
= \ & \frac{m^+(B_1^c)}{m^-(B_1) \cdot m^+(B_1)} \cdot c(x,y)
\end{aligned}$$

and the assertion follows as above.

Combining (i) to (iv) establishes the claim. □

Since $b_n \in \mathbb{B}(m - m_n)$,

$$\begin{aligned}
\|m - m_n\|_w &= \inf_{b\in\mathbb{B}(m-m_n)} \int_{\mathcal{X}^2} c(x,y)\, b(dx,dy) \\
&\leq \int_{\mathcal{X}^2} c(x,y)\, b_n(dx,dy) \\
&= \int_{\mathcal{X}^2} c(x,y) f_n(x,y)\, (|m| \otimes |m|)(dx,dy).
\end{aligned}$$

Since $f_n(x,y) \stackrel{n\to\infty}{\longrightarrow} 0$ for all $(x,y) \in \mathcal{X}^2$ (to see this, recall that $\lim_{n\to\infty} \delta_n = \lim_{n\to\infty} \varepsilon_n = 0$), applying Lebesgue's dominated convergence theorem yields

$$\begin{aligned}
\lim_{n\to\infty} \|m - m_n\|_w &\leq \lim_{n\to\infty} \int_{\mathcal{X}^2} c(x,y) f_n(x,y)\, (|m| \otimes |m|)(dx,dy) \\
&= \int_{\mathcal{X}^2} c(x,y) \cdot \lim_{n\to\infty} f_n(x,y)\, (|m| \otimes |m|)(dx,dy) \\
&= 0.
\end{aligned}$$

So we have shown that $\mathcal{M}_0(\mathcal{X}) \subseteq \mathcal{M}_\lambda(\mathcal{X})$ densely with respect to $\|\cdot\|_w$. □

Definition 2.11. *A signed measure m on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is called* simple *if there exist $N \in \mathbb{N}$, $\alpha_1, \ldots, \alpha_N \in \mathbb{R}$ and $x_1, \ldots, x_N$, $y_1, \ldots, y_N \in \mathcal{X}$ such that $m = \sum_{i=1}^N \alpha_i (\delta_{x_i} - \delta_{y_i})$. Denote by $\mathcal{S}(\mathcal{X})$ the set of all simple measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.*

Lemma 2.12. $\mathcal{S}(\mathcal{X}) \subseteq \mathcal{M}_\lambda(\mathcal{X})$.

Proof. Let $m \in \mathcal{S}(\mathcal{X})$. Then by definition of a simple measure there exist $N \in \mathbb{N}$, $\alpha_1, \ldots, \alpha_N \in \mathbb{R}$ and $x_1, \ldots, x_N$, $y_1, \ldots, y_N \in \mathcal{X}$ such that $m = \sum_{i=1}^N \alpha_i (\delta_{x_i} - \delta_{y_i})$. Hence

$$m(\mathcal{X}) = \left(\sum_{i=1}^N \alpha_i (\delta_{x_i} - \delta_{y_i}) \right)(\mathcal{X}) = \sum_{i=1}^N \alpha_i (\delta_{x_i} - \delta_{y_i})(\mathcal{X}) = \sum_{i=1}^N \alpha_i (\delta_{x_i}(\mathcal{X}) - \delta_{y_i}(\mathcal{X})) = 0$$

and since $m_{x_i y_i} = \delta_{x_i} - \delta_{y_i} \in \mathcal{M}_\lambda(\mathcal{X})$ for all $i = 1, \ldots, N$,

$$\begin{aligned} \int_{\mathcal{X}} \lambda(x)\, |m|(dx) &= \int_{\mathcal{X}} \lambda(x) \left| \sum_{i=1}^N \alpha_i (\delta_{x_i} - \delta_{y_i}) \right| (dx) \\ &\leq \int_{\mathcal{X}} \lambda(x) \left(\sum_{i=1}^N |\alpha_i| \cdot \underbrace{|\delta_{x_i} - \delta_{y_i}|}_{=m_{x_i y_i}} \right) (dx) \\ &= \sum_{i=1}^N |\alpha_i| \int_{\mathcal{X}} \lambda(x)\, |m_{x_i y_i}|(dx) \\ &< \infty, \end{aligned}$$

that is, $m \in \mathcal{M}_\lambda(\mathcal{X})$. Since $m \in \mathcal{S}(\mathcal{X})$ arbitrary, $\mathcal{S}(\mathcal{X}) \subseteq \mathcal{M}_\lambda(\mathcal{X})$. □

In order to prove that the simple measures are dense in $\mathcal{M}_\lambda(\mathcal{X})$, we will need the following

Theorem 2.13. *Let $(\mathcal{X}, d)$ be a separable metric space and suppose P, P_n, $n \in \mathbb{N}$, to be Borel probability measures on $\mathcal{X}$ such that $P_n \overset{n\to\infty}{\longrightarrow} P$ weakly. Then for all $\varepsilon > 0$, $\delta > 0$ there exists $N(\varepsilon) \in \mathbb{N}$ such that for all $n \geq N(\varepsilon)$ exists a probability measure b_n on $(\mathcal{X}^2, \mathcal{B}(\mathcal{X}^2))$ with marginals P_n and P, e.g. $b_n \in \mathcal{C}(P_n, P)$, such that*

$$b_n \left(\{(x,y) \in \mathcal{X}^2 \mid d(x,y) > \delta\} \right) < \varepsilon.$$

Proof. Since π metrizes the weak topology in $\mathcal{M}_1(\mathcal{X})$ (cf. [5], Theorem 11.3.3), for all $\varepsilon > 0$ we can find some $N(\varepsilon) \in \mathbb{N}$ such that $\pi(P_n, P) < \varepsilon$ for all $n \geq N(\varepsilon)$. Let $n \geq N(\varepsilon)$ fixed. Observe that it would be enough to show

$$b_n \left(\{(x,y) \in \mathcal{X}^2 \mid d(x,y) > \varepsilon\} \right) < \varepsilon. \tag{2.28}$$

In [12], Corollary 7.4.2, it is shown that the Prohorov metric π is minimal with respect to the Ky Fan metric K, that is $\pi(Q_1, Q_2) = \inf_{Q \in \mathcal{C}(Q_1, Q_2)} \{K(Q)\}$, where $K(Q) := \inf\{\varepsilon > 0 \mid Q(\{(x,y) \in \mathcal{X}^2 : d(x,y) > \varepsilon\}) < \varepsilon\}$. Hence

$$\inf_{\tilde{b}_n \in \mathcal{C}(P_n, P)} K(\tilde{b}_n) < \varepsilon.$$

Therefore we can find some $b_n \in \mathcal{C}(P_n, P)$ such that $K(b_n) < \varepsilon$. Let $\eta \in]0, \frac{\varepsilon - K(b_n)}{2}[$. Since $K(b_n) + \eta < \varepsilon$, there exists $\varepsilon(\eta) < \varepsilon - \eta$ such that

$$b_n\left(\{(x,y) \in \mathcal{X}^2 \mid d(x,y) > \varepsilon(\eta)\}\right) < \varepsilon(\eta).$$

According to $\varepsilon(\eta) < \varepsilon$ the inequality in (2.28) follows. □

Lemma 2.14. *$\mathcal{S}(\mathcal{X}) \subseteq \mathcal{M}_\lambda(\mathcal{X})$ densely with respect to $\|\cdot\|_w$, i.e. the simple measures are dense in $(\mathcal{M}_\lambda(\mathcal{X}), \|\cdot\|_w)$.*

Proof. Let $m \in \mathcal{M}_\lambda(\mathcal{X})$ be arbitrary and fixed. According to Lemma 2.10 without loss of generality take $m \in \mathcal{M}_0(\mathcal{X})$, e.g. $m = m_1 - m_2$ for m_1, m_2 two finite measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ with $m_1(\mathcal{X}) = m_2(\mathcal{X})$, whose supports $\mathrm{supp}(m_1), \mathrm{supp}(m_2) \subseteq \mathcal{X}_0$ for some bounded subset $\mathcal{X}_0 \subseteq \mathcal{X}$. In fact, (via suitable normalization) take m_1, m_2 to be Borel probability measures. Then by the Varadarajan theorem (cf. [5], Theorem 11.4.1) there exist probability measures $m_1^{(n)}, m_2^{(n)}$, $n \in \mathbb{N}$, on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that $m_1^{(n)}(\mathcal{X}_0) = m_2^{(n)}(\mathcal{X}_0) = 1$, $m_1^{(n)} - m_2^{(n)} \in \mathcal{S}(\mathcal{X})$ for all $n \in \mathbb{N}$ and $m_1^{(n)} \stackrel{n\to\infty}{\longrightarrow} m_1$ respectively $m_2^{(n)} \stackrel{n\to\infty}{\longrightarrow} m_2$ weakly. (Let $(\Omega, \mathcal{A}, P)$ be a probability space and X_j, $j \in \mathbb{N}$, independent random variables with values in $\mathcal{X}_0$ and distribution m_1. Define $m_1^{(n)}(\cdot)(\omega) := \frac{1}{n} \cdot \sum_{j=1}^n \delta_{X_j(\omega)}(\cdot)$ for every $\omega \in \Omega$. In this situation the Varadarajan theorem states that, for P-a.e. $\omega \in \Omega$, $m_1^{(n)}(\cdot)(\omega)$ converges weakly to m_1 as $n \to \infty$. Hence take one of this "good" ω's and put $m_1^{(n)} := m_1^{(n)}(\cdot)(\omega)$. Analogously one can find Borel probability measures $m_2^{(n)}$, $n \in \mathbb{N}$, converging weakly to m_2. Obviously, $m_1^{(n)} - m_2^{(n)} \in \mathcal{S}(\mathcal{X})$ for all $n \in \mathbb{N}$.) To prove the lemma it would be enough to show that $\|m_1^{(n)} - m_1\|_w \stackrel{n\to\infty}{\longrightarrow} 0$, cause then by the triangle inequality we have $\|\left(m_1^{(n)} - m_2^{(n)}\right) - \underbrace{(m_1 - m_2)}_{=m}\|_w \le \|m_1^{(n)} - m_1\|_w + \|m_2^{(n)} - m_2\|_w \stackrel{n\to\infty}{\longrightarrow} 0$.

Let $\varepsilon > 0$. Since $\mathcal{X}_0$ is bounded, according to condition (C5) we can find $\delta > 0$ such that $c(x,y) < \frac{\varepsilon}{2}$ for all $x, y \in \mathcal{X}_0$ with $d(x,y) \le \delta$. Define $C := \sup\{\lambda(x) \mid x \in \mathcal{X}_0\}$. By Theorem 2.13 there is $N(\varepsilon) \in \mathbb{N}$ such that for all $n \ge N(\varepsilon)$ exists a probability measure b_n on $(\mathcal{X}^2, \mathcal{B}(\mathcal{X}^2))$ with marginals $m_1^{(n)}$ and m_1 such that $b_n(\{(x,y) \in \mathcal{X}^2 \mid d(x,y) > \delta\}) < \frac{\varepsilon}{4C}$. Set $A := \{(x,y) \in \mathcal{X}^2 \mid d(x,y) > \delta\}$. Then using (C3) we get

$$\begin{aligned}
\|m_1^{(n)} - m_1\|_w &= \inf_{b \in B\left(m_1^{(n)} - m_1\right)} \int_{\mathcal{X}^2} c(x,y)\, b(dx,dy) \\
&\le \int_{\mathcal{X}^2} c(x,y)\, b_n(dx,dy) \\
&= \int_A c(x,y)\, b_n(dx,dy) + \int_{A^c} c(x,y)\, b_n(dx,dy) \\
&\le \int_A (\lambda(x) + \lambda(y))\, b_n(dx,dy) + \frac{\varepsilon}{2} \\
&\le 2C \cdot b_n(A) + \frac{\varepsilon}{2} \\
&< \varepsilon
\end{aligned}$$

for all $n \ge N(\varepsilon)$. □

Lemma 2.15. *The linear transformation* $D\colon (\mathbb{L}_c(\mathcal{X},\mathbb{R}), \|\cdot\|_c) \to (\mathcal{M}_\lambda(\mathcal{X})^*, \|\cdot\|_w^*)$ *is an isometric isomorphism.*

Proof. According to Lemma 2.9 it is enough to show that D is an isomorphism. To do so, first observe that D is injective: Let $\varphi, \psi \in \mathbb{L}_c(\mathcal{X},\mathbb{R})$ such that $D\varphi = D\psi$, e.g. $F_\varphi(m) = F_\psi(m)$ for all $m \in \mathcal{M}_\lambda(\mathcal{X})$. Without loss of generality assume that there is some $a \in \mathcal{X}$ such that $\varphi(a) = 0$. Then $\varphi(x) = \varphi(x) - \varphi(a) = \int_\mathcal{X} \varphi(u)\ (\delta_x - \delta_a)(du) = F_\varphi(\delta_x - \delta_a) = F_\psi(\delta_x - \delta_a) = \int_\mathcal{X} \psi(u)\ (\delta_x - \delta_a)(du) = \psi(x) - \psi(a)$ for all $x \in \mathcal{X}$. Since $\psi(a)$ is constant, φ and ψ belong to the same equivalence class of $\mathbb{L}_c(\mathcal{X},\mathbb{R})$, i.e. $\varphi = \psi$, and this means that D is injective. Hence it remains to prove that D is surjective. So, let $F \in \mathcal{M}_\lambda(\mathcal{X})^*$, i.e. $F\colon \mathcal{M}_\lambda(\mathcal{X}) \to \mathbb{R}$ is continuous and linear. Fix $a \in \mathcal{X}$ and define $\varphi\colon \mathcal{X} \to \mathbb{R}$ by $\varphi(x) := F(\delta_x - \delta_a) = F(m_{xa})$. Then for any $x, y \in \mathcal{X}$ using (2.25) we gain

$$\begin{aligned} |\varphi(x) - \varphi(y)| &= |F(\delta_x - \delta_a) - F(\delta_y - \delta_a)| \\ &= |F(\delta_x - \delta_y)| \\ &\le \|F\|_w^* \cdot \|\delta_x - \delta_y\|_w \\ &= \|F\|_w^* \cdot \|m_{xy}\|_w \\ &\le \|F\|_w^* \cdot c(x,y). \end{aligned}$$

Since $\|\varphi\|_c$ is the smallest such constant, we conclude that $\|\varphi\|_c \le \|F\|_w^* < \infty$, e.g. $\varphi \in \mathbb{L}_c(\mathcal{X},\mathbb{R})$. A straightforward calculation shows $F(m) = F_\varphi(m)$ for $m = \delta_x - \delta_y$ and therefore $F(m) = F_\varphi(m)$ for all $m \in \mathcal{S}(\mathcal{X})$ and by the previous Lemma 2.14 for all $m \in \mathcal{M}_\lambda(\mathcal{X})$. Hence by definition of D, $F = D\varphi$ and so D is surjective. □

Since D is surjective, for $F \in \mathcal{M}_\lambda(\mathcal{X})^*$ we can find some $\varphi \in \mathbb{L}_c(\mathcal{X},\mathbb{R})$ such that $F = D\varphi = F_\varphi$. Hence

$$\|m\|_w = \sup_{\|F\|_w^*=1} |F(m)| = \sup_{\|F_\varphi\|_w^*=1} |F_\varphi(m)| = \sup_{\|\varphi\|_c=1} \left| \int_\mathcal{X} \varphi(x)\ m(dx) \right|.$$

Theorem 2.16. *Let* $m \in \mathcal{M}_\lambda(\mathcal{X})$. *Then*

$$\|m\|_w = \sup_{\|\varphi\|_c=1} \int_\mathcal{X} \varphi(x)\ m(dx).$$

In particular, for P_1, P_2 *two Borel probability measures on* $\mathcal{X}$ *such that* $\int_\mathcal{X} \lambda(x)\ P_i(dx) < \infty$, $i = 1, 2$, *we have*

$$R_c(P_1, P_2) = \|P_1 - P_2\|_w = \sup_{\|\varphi\|_c=1} \int_\mathcal{X} \varphi(x)\ (P_1 - P_2)(dx).$$

We now show that the supremum in Theorem 2.16 is attained for some optimal φ.

Theorem 2.17. *Let* $m \in \mathcal{M}_\lambda(\mathcal{X})$. *Then there is some* $\varphi \in \mathbb{L}_c(\mathcal{X},\mathbb{R})$ *with* $\|\varphi\|_c = 1$ *such that* $\|m\|_w = \int_\mathcal{X} \varphi(x)\ m(dx)$.

Proof. Using the Hahn-Banach theorem, choose a linear functional $F \in \mathcal{M}_\lambda(\mathcal{X})^*$ with $\|F\|_w^* = 1$ such that $F(m) = \|m\|_w$. By Lemma 2.15 we have $F = F_\varphi$ for some $\varphi \in \mathbb{L}_c(\mathcal{X},\mathbb{R})$ with $\|\varphi\|_c = \|F_\varphi\|_w^* = \|F\|_w^* = 1$. □

2.3 Link between the functionals K_c and R_c

Since $\mathcal{C}(P_1, P_2) \subseteq \mathbb{B}(P_1 - P_2)$ only, we already know that $K_c(P_1, P_2) \geq R_c(P_1, P_2)$. In order to obtain the final result (2.32) - equality of the Monge-Kantorovich problem (primal problem) and the dual problem of the Kantorovich-Rubinstein problem - in case of a separable metric space $(\mathcal{X}, d)$, we have to verify that

$$\begin{aligned} K_d(P_1, P_2) &= \inf_{P \in \mathcal{C}(P_1, P_2)} \int_{\mathcal{X}^2} d(x, y)\, P(dx, dy) \\ &= \inf_{b \in \mathbb{B}(P_1 - P_2)} \int_{\mathcal{X}^2} d(x, y)\, b(dx, dy) = R_d(P_1, P_2). \end{aligned}$$

This will be guaranteed by the following

Theorem 2.18 (Neveu and Dudley)**.** *Let $(\mathcal{X}, d)$ be a separable metric space and $c \in \mathbf{C}$. Then*

$$K_c(P_1, P_2) = R_c(P_1, P_2) \tag{2.29}$$

for all Borel probability measures P_1, P_2 on $\mathcal{X}$ with $\int_{\mathcal{X}} c(x, a)\, P_i(dx) < \infty$ for some fixed $a \in \mathcal{X}$, $i = 1, 2$, if and only if c is a metric.

In particular, equality (2.29) holds if H is the identity, i.e. $c = d$.

Proof. Suppose (2.29) holds and put $P_1 := \delta_x$, $P_2 := \delta_y$ for some $x, y \in \mathcal{X}$. Then $\mathcal{C}(P_1, P_2) = \{\delta_{(x,y)}\}$, i.e. the set of all laws P on $(\mathcal{X}^2, \mathcal{B}(\mathcal{X}^2))$ with marginals P_1 and P_2 only contains the dirac-measure in (x, y). Hence by Theorem 2.16

$$\begin{aligned} c(x, y) &= \int_{\mathcal{X}^2} c(u, v)\, \delta_{(x,y)}(du, dv) \\ &= \inf_{P \in \mathcal{C}(P_1, P_2)} \int_{\mathcal{X}^2} c(u, v)\, P(du, dv) \\ &= K_c(P_1, P_2) \\ &= R_c(P_1, P_2) \\ &= \sup_{\|\varphi\|_c = 1} \int_{\mathcal{X}} \varphi(u)\, \left(P_1 - P_2\right)(du) \\ &= \sup_{\|\varphi\|_c = 1} \left(\varphi(x) - \varphi(y)\right) \\ &\leq \sup_{\|\varphi\|_c = 1} |\varphi(x) - \varphi(y)| \\ &\leq \sup_{\|\varphi\|_c = 1} |\varphi(x) - \varphi(z)| + \sup_{\|\varphi\|_c = 1} |\varphi(z) - \varphi(y)| \\ &\leq c(x, z) + c(z, y) \end{aligned}$$

for all $x, y, z \in \mathcal{X}$. Furthermore, since $c \in \mathbf{C}$, there is some $H \in \mathbf{H}$ such that $c(x, y) = H(d(x, y))$ for all $(x, y) \in \mathcal{X}^2$, and so we have

$$c(x, y) = 0 \Longleftrightarrow H(d(x, y)) = 0 \Longleftrightarrow d(x, y) = 0 \Longleftrightarrow x = y$$

and

$$c(x,y) = H(d(x,y)) = H(d(y,x)) = c(y,x)$$

for all $x, y \in \mathcal{X}$. Therefore c is a metric.

To show the converse, first recall that

$$\mathcal{G}_c(\mathcal{X}) = \{(f,g) \mid f,g \in \mathcal{L}_d(\mathcal{X}),\ f(x)+g(y) \le c(x,y)\ \ \forall\ \ x,y \in \mathcal{X}\}.$$

Suppose c is a metric and let $(f,g) \in \mathcal{G}_c(\mathcal{X})$. Define a function $h : \mathcal{X} \to \mathbb{R}$ by

$$h(x) := \inf_{y\in\mathcal{X}} \left(c(x,y) - g(y)\right).$$

Then

$$\begin{aligned} h(x) - h(y) &= \inf_{v\in\mathcal{X}} \left(c(x,v) - g(v)\right) - \inf_{v\in\mathcal{X}} \left(c(y,v) - g(v)\right) \\ &= \inf_{v\in\mathcal{X}} \left(c(x,v) - g(v)\right) + \sup_{v\in\mathcal{X}} \left(g(v) - c(y,v)\right) \\ &\le \sup_{v\in\mathcal{X}} \left(c(x,v) - g(v) + g(v) - c(y,v)\right) \\ &= \sup_{v\in\mathcal{X}} \left(c(x,v) - c(y,v)\right) \\ &\le c(x,y) \end{aligned}$$

for all $x, y \in \mathcal{X}$ and according to the symmetry of c we thus have $\|h\|_c \le 1$.

Since for fixed $y \in \mathcal{X}$ the function $c(x,y) - g(y)$ is continuous in $x \in \mathcal{X}$, h is upper semi-continuous. (This can be seen in the same way as in the proof of Claim 3.) Moreover $f(x) \le h(x) \le -g(x)$ for all $x \in \mathcal{X}$. In fact, since $(f,g) \in \mathcal{G}_c(\mathcal{X})$, we know $f(x) \le c(x,y) - g(y)$ for all $x, y \in \mathcal{X}$ which implies $f(x) \le \inf_{y\in\mathcal{X}} (c(x,y) - g(y)) = h(x)$ for all $x \in \mathcal{X}$. By definition of h we have $h(x) = \inf_{y\in\mathcal{X}} (c(x,y) - g(y)) \le c(x,x) - g(x) = -g(x)$ for all $x \in \mathcal{X}$. Hence for any two Borel probability measures P_1, P_2 on $\mathcal{X}$ satisfying $\int_{\mathcal{X}} c(x,a)\ P_1(dx) < \infty$ respectively $\int_{\mathcal{X}} c(y,a)\ P_2(dy) < \infty$ for some fixed $a \in \mathcal{X}$ we have

$$\begin{aligned} \int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) &= \int_{\mathcal{X}} \underbrace{f(x)}_{\le h(x)}\ P_1(dx) - \int_{\mathcal{X}} \underbrace{-g(y)}_{\ge h(y)}\ P_2(dy) \\ &\le \int_{\mathcal{X}} h(x)\ P_1(dx) - \int_{\mathcal{X}} h(y)\ P_2(dy) \\ &= \int_{\mathcal{X}} h(x)\ \left(P_1 - P_2\right)(dx) \end{aligned}$$

so that according to Theorem 2.1 and Theorem 2.16 it follows

$$\begin{aligned} K_c(P_1,P_2) &= \sup_{(f,g)\in\mathcal{G}_c(\mathcal{X})} \left(\int_{\mathcal{X}} f(x)\ P_1(dx) + \int_{\mathcal{X}} g(y)\ P_2(dy) \right) \\ &\le \sup_{\|h\|_c\le 1} \int_{\mathcal{X}} h(x)\ \left(P_1 - P_2\right)(dx) = R_c(P_1,P_2). \end{aligned}$$

Therefore $K_c(P_1,P_2) = R_c(P_1,P_2)$. □

Corollary 2.19. *Let $(\mathcal{X}, d)$ be a separable metric space and $a \in \mathcal{X}$. Let P_1, P_2 be two Borel probability measures on $\mathcal{X}$ such that $\int_{\mathcal{X}} d(x,a)\ P_i(dx) < \infty$ for $i = 1,2$. Then*

$$\inf_{P \in \mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} d(x,y)\ P(dx,dy) = \sup_{\|\varphi\|_d = 1} \int_{\mathcal{X}} \varphi(x)\ \ (P_1 - P_2)\,(dx). \tag{2.30}$$

Moreover, the supremum is attained for some optimal φ^* with $\|\varphi^*\|_d = 1$.

If P_1, P_2 are tight, there are some $P^* \in \mathcal{C}(P_1,P_2)$ and $\varphi^* \colon \mathcal{X} \to \mathbb{R}$ with $\|\varphi^*\|_d = 1$ such that

$$K_d(P_1,P_2) = \int_{\mathcal{X}^2} d(x,y)\ P^*(dx,dy) = \int_{\mathcal{X}} \varphi^*(x)\ \ (P_1 - P_2)\,(dx),$$

where $\varphi^*(x) - \varphi^*(y) = d(x,y)$ for P^*-a.e. $(x,y) \in \mathcal{X}^2$.

Proof. Choose $c = d$ and recall that for this choice the conditions (C1) to (C5) are satisfied with $\lambda(x) = d(x,a)$ for some fixed $a \in \mathcal{X}$. Hence Theorem 2.16 is applicable and we get

$$W(P_1,P_2) = R_d(P_1,P_2) = \sup_{\|\varphi\|_d = 1} \int_{\mathcal{X}} \varphi(x)\ \ (P_1 - P_2)\,(dx).$$

Already by definition we have

$$K_d(P_1,P_2) = \inf_{P \in \mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} d(x,y)\ P(dx,dy).$$

Since d is a metric, Theorem 2.18 yields the assertion, namely

$$\inf_{P \in \mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} d(x,y)\ P(dx,dy) = \sup_{\|\varphi\|_d = 1} \int_{\mathcal{X}} \varphi(x)\ \ (P_1 - P_2)\,(dx).$$

The existence of some φ^* follows from Theorem 2.17.

If P_1 and P_2 are tight, then by Theorem 2.1 there exists $P^* \in \mathcal{C}(P_1,P_2)$ such that

$$K_d(P_1,P_2) = \int_{\mathcal{X}^2} d(x,y)\ P^*(dx,dy),$$

i.e. P^* is optimal for the Monge-Kantorovich problem. Integrating both sides of the inequality $\varphi^*(x) - \varphi^*(y) \le d(x,y)$ with respect to P^* yields

$$\int_{\mathcal{X}} \varphi^*(x)\ \ (P_1 - P_2)\,(dx) \le \int_{\mathcal{X}^2} d(x,y)\ P^*(dx,dy).$$

However, since P^* and φ^* are optimal for the primal and the dual problem respectively, we know from the first part of the proof that we have equality of these integrals. Therefore $\varphi^*(x) - \varphi^*(y) = d(x,y)$ for P^*-a.e. $(x,y) \in \mathcal{X}^2$. □

2.4 The Monge-Kantorovich duality in case of a pseudo-metric d

Here we slightly generalize Corollary 2.19 by replacing the metric thereby by a pseudo-metric.

Definition 2.20. *A function* $d\colon \mathcal{X}^2 \to \mathbb{R}_+$ *is called a* pseudo-metric *if*

(i) $x = y \Longrightarrow d(x,y) = 0$,

(ii) $d(x,y) = d(y,x)$ *for all* $(x,y) \in \mathcal{X}^2$,

(iii) $d(x,y) \leq d(x,z) + d(z,y)$ *for all* $x,y,z \in \mathcal{X}$.

Remark 2.21. *Note that the only difference to a metric is the missing converse direction in (i).*

So, let d be a pseudo-metric on $\mathcal{X}$. Similarly to section 2.2 denote by $\mathcal{L}_d(\mathcal{X},\mathbb{R})$ the linear space of all d-Lipschitz continuous functions φ from $\mathcal{X}$ to $\mathbb{R}$ and define the seminorm $\|\cdot\|_d\colon \mathcal{L}_d(\mathcal{X},\mathbb{R}) \to \mathbb{R}_+$ by

$$\|\varphi\|_d := \sup_{x,y\in\mathcal{X}\colon x\neq y} \frac{|\varphi(x)-\varphi(y)|}{d(x,y)}. \tag{2.31}$$

Again for $\varphi \in \mathcal{L}_d(\mathcal{X},\mathbb{R})$ we have $|\varphi(x)-\varphi(y)| \leq \|\varphi\|_d \cdot d(x,y)$ for all $x,y \in \mathcal{X}$. In order to obtain a norm, set

$$\mathbb{L}_d(\mathcal{X},\mathbb{R}) := \mathcal{L}_d(\mathcal{X},\mathbb{R})/\mathcal{N},$$

where

$$\mathcal{N} := \{\varphi \in \mathcal{L}_d(\mathcal{X},\mathbb{R}) \mid \|\varphi\|_d = 0\} = \{\varphi \in \mathcal{L}_d(\mathcal{X},\mathbb{R}) \mid \exists\ u \in \mathbb{R} : \varphi(x) = u \ \ \forall\ x \in \mathcal{X}\}.$$

Define $\|\cdot\|_d : \mathbb{L}_d(\mathcal{X},\mathbb{R}) \to \mathbb{R}_+$ by $\|[\varphi]\|_d := \|\varphi\|_d$, where $\varphi \in [\varphi]$ is an arbitrary representative of $[\varphi]$. Analogously to section 2.2 one can show that $\|\cdot\|_d$ is welldefined and a norm on $\mathbb{L}_d(\mathcal{X},\mathbb{R})$. Again use φ and $[\varphi]$ interchangeable, emphasizing that the latter is an equivalence class of Lipschitz functions rather than a single function.

As the final result of this chapter we formulate:

Theorem 2.22. *Let* $\mathcal{X}$ *be a Polish space and* $d\colon \mathcal{X}^2 \to \mathbb{R}_+$ *a (continuous) pseudo-metric. Let* P_1, P_2 *be two probability measures on* $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ *such that* $\int_{\mathcal{X}} d(x,a)\ P_i(dx) < \infty$ *for* $i = 1,2$ *and some fixed* $a \in \mathcal{X}$. *Then*

$$K_d(P_1,P_2) = \inf_{P\in\mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} d(x,y)\ P(dx,dy) = \sup_{\|\varphi\|_d=1} \int_{\mathcal{X}} \varphi(x)\ \ (P_1-P_2)\,(dx). \tag{2.32}$$

Proof. If $(\mathcal{X}, d)$ is a separable metric space, the assertion is just part of Corollary 2.19.

Now consider the case where d is an arbitrary (continuous) pseudo-metric. Define an equivalence relation on $\mathcal{X}$ by

$$x \sim y :\Leftrightarrow d(x, y) = 0.$$

and set $\mathcal{X}_d := \mathcal{X}/\sim$. Notice that elements of $\mathcal{X}_d$ are equivalence classes $[x]$. Then d is defined on $\mathcal{X}_d$ in an obvious way, e.g. $d([x], [y]) := d(x, y)$ where $x \in [x]$, $y \in [y]$ are arbitrary representatives. It is clear that d is well-defined. Observe that $(\mathcal{X}_d, d)$ is a separable metric space (although it may no longer be complete). Hence the assertion is right for $(\mathcal{X}_d, d)$. Defining the map $\pi\colon \mathcal{X} \to \mathcal{X}_d$ by $\pi(x) := [x]$, the result follows from the Monge-Kantorovich duality in $(\mathcal{X}_d, d)$ and the fact that both sides of (2.32) do not change if the Borel probability measures P_i are replaced by $P_i \circ \pi^{-1}$:

$$\begin{aligned}
K_d(P_1, P_2) &= \inf_{P \in \mathcal{C}(P_1, P_2)} \int_{\mathcal{X}^2} d(x, y)\; P(dx, dy) \\
&= \inf_{P \in \mathcal{C}(P_1, P_2)} \int_{\mathcal{X}^2} d(\pi(x), \pi(y))\; P(dx, dy) \\
&= \inf_{Q \in \mathcal{C}(P_1 \circ \pi^{-1}, P_2 \circ \pi^{-1})} \int_{\mathcal{X}_d^2} d(x, y)\; Q(dx, dy) \\
&= \sup_{\|\varphi\|_d = 1} \int_{\mathcal{X}_d} \varphi(x)\; \left(P_1 \circ \pi^{-1} - P_2 \circ \pi^{-1}\right)(dx) \\
&= \sup_{\|\varphi\|_d = 1} \int_{\mathcal{X}_d} \varphi(x)\; \left((P_1 - P_2) \circ \pi^{-1}\right)(dx) \\
&= \sup_{\|\varphi\|_d = 1} \int_{\mathcal{X}} \varphi(\pi(x))\; (P_1 - P_2)(dx) \\
&= \sup_{\|\varphi\|_d = 1} \int_{\mathcal{X}} \varphi(x)\; (P_1 - P_2)(dx).
\end{aligned}$$

□

3 Asymptotic strong Feller property

This chapter is an extended version of chapter 3 in [7].

3.1 Totally separating systems

Let $\mathcal{X}$ be a Polish space, i.e. $\mathcal{X}$ is metrizable, complete and separable.

Definition 3.1. *A (pseudo-)metric d_2 is said to be* larger *than d_1 if $d_2(x,y) \geq d_1(x,y)$ for all $(x,y) \in \mathcal{X}^2$. A sequence $(d_n)_{n\in\mathbb{N}}$ of (pseudo-)metrics is called* increasing *if d_{n+1} is larger than d_n for all $n \in \mathbb{N}$.*

Consider the (trivial) metric d_{TV} given by

$$d_{TV}(x,y) := \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases} . \tag{3.1}$$

This is a metric that totally separates the points of $\mathcal{X}$ and therefore loses completely all information about the topology of $\mathcal{X}$.

Definition 3.2. *An increasing sequence $(d_n)_{n\in\mathbb{N}}$ of (pseudo-)metrics on a Polish space $\mathcal{X}$ is called a* totally separating system of (pseudo-)metrics *for $\mathcal{X}$ if $\lim_{n\to\infty} d_n(x,y) = 1$ for all $(x,y) \in \mathcal{X}^2$, $x \neq y$, i.e. $(d_n)_{n\in\mathbb{N}}$ converges pointwisely to the (trivial) metric d_{TV}.*

In the next section we will see how to use such a totally separating system of pseudo-metrics in order to approximate the total variation distance of two Borel probability measures by a sequence of their Wasserstein distances. Before we do so, let us give few examples of totally separating systems of (pseudo-)metrics:

Example 3.3. *Let $(a_n)_{n\in\mathbb{N}}$, $a_n \in \mathbb{R}_+$, be an increasing sequence such that $\lim_{n\to\infty} a_n = \infty$.*

(1) Let $\mathcal{X}$ be a Polish space and $d\colon \mathcal{X}^2 \to \mathbb{R}_+$ an arbitrary metric on $\mathcal{X}$. For every $n \in \mathbb{N}$ define $d_n(x,y) := 1 \wedge a_n \cdot d(x,y)$. Then $(d_n)_{n\in\mathbb{N}}$ is a totally separating system of metrics for $\mathcal{X}$.

(2) Let $\mathcal{X} := \mathcal{C}_0(\mathbb{R})$ be the space of continuous functions x on $\mathbb{R}$ such that $\lim_{|s|\to\infty} x(s) = 0$. Since $(\mathcal{C}_b(\mathbb{R}), \|\cdot\|_\infty)$ is a Banach space and $\mathcal{C}_0(\mathbb{R}) \subseteq \mathcal{C}_b(\mathbb{R})$ is closed with respect to $\|\cdot\|_\infty$ (cf. [14], p. 6), $\mathcal{X}$ is complete. Furthermore one can show that $\mathcal{X}$ is separable. For any $x,y \in \mathcal{X}$ set $d_n(x,y) := 1 \wedge \big(a_n \cdot \sup_{s\in[-n,n]} |x(s) - y(s)|\big)$. Then $(d_n)_{n\in\mathbb{N}}$ is a totally separating system of metrics for $\mathcal{X}$.

(3) *Let $\mathcal{X} := l^2(\mathbb{R})$, i.e. $\sum_{k=1}^{\infty}|x_k|^2 < \infty$ for all sequences $x = (x_k)_{k\in\mathbb{N}} \in \mathcal{X}$. First of all, $\mathcal{X}$ is separable by example 2.17.2 in [1] and complete with respect to the norm $\|x\|_{l^2} := (\sum_{k=1}^{\infty}|x_k|^2)^{\frac{1}{2}}$ (cf. [1], 0.18.4). For $n \in \mathbb{N}$ define $d_n(x,y) := 1 \wedge a_n \cdot \sum_{k=1}^{n}|x_k - y_k|^2$ for any two $x, y \in \mathcal{X}$. Then $(d_n)_{n\in\mathbb{N}}$ is a totally separating system of metrics for $\mathcal{X}$.*

3.2 The key lemma

Let d be a bounded and continuous pseudo-metric. For clarity we do not overtake the (topic specific) notation from chapter 2, but simply set

$$\|P_1 - P_2\|_d := K_d(P_1, P_2) = \inf_{P\in\mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} d(x,y)\; P(dx,dy) \tag{3.2}$$

for the Wasserstein distance of the Borel probability measures P_1, P_2 on $\mathcal{X}$ and

$$|||P_1 - P_2|||_d := \sup_{\|\varphi\|_d = 1} \int_{\mathcal{X}} \varphi(x)\; (P_1 - P_2)(dx) \tag{3.3}$$

for the seminorm of their difference on the space $\mathcal{M}_\lambda(\mathcal{X})$ with λ given by $\lambda(\cdot) = d(\cdot, a)$ for some fixed $a \in \mathcal{X}$.

The following lemma is crucial to the approach of the asymptotic strong Feller property, introduced in the next section.

Lemma 3.4. *Let $(d_n)_{n\in\mathbb{N}}$ be a bounded and increasing sequence of continuous pseudo-metrics on a Polish space $\mathcal{X}$. Define $d(x,y) := \lim_{n\to\infty} d_n(x,y)$ for all $(x,y) \in \mathcal{X}^2$. Let P_1, P_2 be two probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Then*

$$\lim_{n\to\infty} \|P_1 - P_2\|_{d_n} = \|P_1 - P_2\|_d.$$

Proof. By assumption the sequence $(d_n)_{n\in\mathbb{N}}$ of pseudo-metrics on $\mathcal{X}$ is bounded and increasing, i.e. there exists $K \in \mathbb{R}_+$ such that $d_n(x,y) \le K$ and $d_n(x,y) \le d_{n+1}(x,y)$ for all $(x,y) \in \mathcal{X}^2$, $n \in \mathbb{N}$. Therefore the limit $d(x,y) := \lim_{n\to\infty} d_n(x,y)$ exists for all $(x,y) \in \mathcal{X}^2$. Since

$$\begin{aligned}
\|P_1 - P_2\|_{d_n} &= \inf_{P\in\mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} d_n(x,y)\; P(dx,dy) \\
&\le \inf_{P\in\mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} d_{n+1}(x,y)\; P(dx,dy) = \|P_1 - P_2\|_{d_{n+1}}
\end{aligned}$$

for all $n \in \mathbb{N}$, the sequence $(\|P_1 - P_2\|_{d_n})_{n\in\mathbb{N}}$ is increasing. Furthermore for arbitrary

$n \in \mathbb{N}$

$$\begin{aligned}
\|P_1 - P_2\|_{d_n} &= \inf_{P \in \mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} d_n(x,y)\ P(dx,dy) \\
&\leq K \cdot \inf_{P \in \mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} P(dx,dy) \\
&= K \cdot \inf_{P \in \mathcal{C}(P_1,P_2)} P(\mathcal{X}^2) \\
&= K \cdot P_1(\mathcal{X}) \\
&= K \\
&< \infty.
\end{aligned}$$

Thus the sequence $(\|P_1 - P_2\|_{d_n})_{n\in\mathbb{N}}$ is bounded. Hence the limit $L := \lim_{n\to\infty} \|P_1 - P_2\|_{d_n}$ exists. It remains to show that $L = \|P_1 - P_2\|_d$.

(1) $L \leq \|P_1 - P_2\|_d$: For each $n \in \mathbb{N}$

$$\begin{aligned}
\|P_1 - P_2\|_{d_n} &= \inf_{P \in \mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} d_n(x,y)\ P(dx,dy) \\
&\leq \inf_{P \in \mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} d(x,y)\ P(dx,dy) = \|P_1 - P_2\|_d.
\end{aligned}$$

Hence $L = \lim_{n\to\infty} \|P_1 - P_2\|_{d_n} \leq \|P_1 - P_2\|_d$.

(2) $L \geq \|P_1 - P_2\|_d$: For arbitrary $n \in \mathbb{N}$ let $P^{(n)} \in \mathcal{C}(P_1, P_2)$ such that

$$\|P_1 - P_2\|_{d_n} = \int_{\mathcal{X}^2} d_n(x,y)\ P^{(n)}(dx,dy),$$

i.e. $P^{(n)}$ realizes (3.2) for the continuous pseudo-metric d_n. Such a measure is shown to exist in Corollary 2.19. (Recall that since $\mathcal{X}$ is Polish, by Lemma C.1 in the appendix the Borel probability measures P_1 and P_2 are tight.)

The sequence $(P^{(n)})_{n\in\mathbb{N}}$ is tight, since its marginals P_1 and P_2 are constant. In fact, let $\varepsilon > 0$. Since P_1 and P_2 are tight, there exist compact sets $K_1(\varepsilon), K_2(\varepsilon) \subseteq \mathcal{X}$ such that $P_1(K_1(\varepsilon)^c) < \frac{\varepsilon}{2}$ and $P_2(K_2(\varepsilon)^c) < \frac{\varepsilon}{2}$. Hence for all $n \in \mathbb{N}$

$$\begin{aligned}
P^{(n)}((K_1(\varepsilon) \times K_2(\varepsilon))^c) &\leq P^{(n)}(K_1(\varepsilon)^c \times \mathcal{X}) + P^{(n)}(\mathcal{X} \times K_2(\varepsilon)^c) \\
&= P_1(K_1(\varepsilon)^c) + P_2(K_2(\varepsilon)^c) \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon.
\end{aligned}$$

Therefore by Prohorov (cf. [10], Theorem 6.7) there exists a subsequence $\{P^{(n_k)} \mid k \in \mathbb{N}\} \subseteq \{P^{(n)} \mid n \in \mathbb{N}\}$ such that $P^{(n_k)} \to P^\infty$ weakly as $k \to \infty$ for some

$P^\infty \in \mathcal{C}(P_1, P_2)$. Furthermore for arbitrary $l \geq k$ it follows $d_{n_k}(x,y) \leq d_{n_l}(x,y)$ for all $(x,y) \in \mathcal{X}^2$ and thus

$$\int_{\mathcal{X}^2} d_{n_k}(x,y)\; P^{(n_l)}(dx,dy) \leq \int_{\mathcal{X}^2} d_{n_l}(x,y)\; P^{(n_l)}(dx,dy) = \|P_1 - P_2\|_{d_{n_l}} \leq L$$

for all $l \geq k$ by choice of $P^{(n)}$ and because $(\|P_1 - P_2\|_{d_n})_{n\in\mathbb{N}}$ is increasing. Hence, since $d_{n_k}\colon \mathcal{X}^2 \to \mathbb{R}_+$ is bounded and continuous, by weak convergence

$$\int_{\mathcal{X}^2} d_{n_k}(x,y)\; P^\infty(dx,dy) = \lim_{l\to\infty} \int_{\mathcal{X}^2} d_{n_k}(x,y)\; P^{(n_l)}(dx,dy) \leq L$$

for all $k \in \mathbb{N}$. Finally letting $k \to \infty$ and using Lebesgue yields

$$\int_{\mathcal{X}^2} d(x,y)\; P^\infty(dx,dy) = \lim_{k\to\infty} \int_{\mathcal{X}^2} d_{n_k}(x,y)\; P^\infty(dx,dy) \leq L.$$

Since $P^\infty \in \mathcal{C}(P_1, P_2)$, this implies $\|P_1 - P_2\|_d \leq L$.

□

In order to approximate the total variation distance of two Borel probability measures on $\mathcal{X}$ by a sequence of their Wasserstein distances corresponding to a totally separating system of pseudo-metrics, recall the following

Definition 3.5. *Let μ be a finite signed measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ with Jordan decomposition $\mu = \mu^+ - \mu^-$. Then the* total variation norm *of μ is given by*

$$\|\mu\|_{TV} := \frac{1}{2} \cdot \left(\mu^+(\mathcal{X}) + \mu^-(\mathcal{X})\right).$$

Remark 3.6. *The factor $\frac{1}{2}$ is chosen in order to guarantee that the total variation distance of two mutually singular Borel probability measures is normalized to* 1.

With this notion at hand we receive:

Corollary 3.7. *Let $(d_n)_{n\in\mathbb{N}}$ be a totally separating system of continuous pseudo-metrics for the Polish space $\mathcal{X}$ and P_1, P_2 two probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Then*

$$\|P_1 - P_2\|_{TV} = \lim_{n\to\infty} \|P_1 - P_2\|_{d_n}.$$

Proof. By definition the totally separating system of continuous pseudo-metrics $(d_n)_{n\in\mathbb{N}}$ is bounded (by 1) and increasing. According to the previous Lemma 3.4 it therefore suffices to show that $\|P_1 - P_2\|_{TV} = \|P_1 - P_2\|_{d_{TV}}$ with d_{TV} given by (3.1). In fact, by Corollary 2.19 we have

$$\begin{aligned} \|P_1 - P_2\|_{d_{TV}} &= \inf_{P\in\mathcal{C}(P_1,P_2)} \int_{\mathcal{X}^2} d_{TV}(x,y)\; P(dx,dy) \\ &= \sup_{\|\varphi\|_{d_{TV}}=1} \int_{\mathcal{X}} \varphi(x)\; (P_1 - P_2)(dx) \\ &= \sup_{\|\varphi\|_\infty=1} \int_{\mathcal{X}} \varphi(x)\; (P_1 - P_2)(dx) = \|P_1 - P_2\|_{TV}. \end{aligned}$$

□

To stress the main part of the proof, notice that the total variation distance of the two Borel probability measures P_1 and P_2 is given by the Wasserstein distance of these measures corresponding to the trivial metric d_{TV}.

3.3 Definition and classification

Now we introduce the asymptotic strong Feller property which instead of prescribing a smoothing property at some fixed time $t > 0$, prescribes some kind of smoothing property 'at time ∞'.

Definition 3.8. *Let $\mathcal{X}$ be a Polish space and denote by $\mathcal{U}_x$ the collection of all open sets $U \subseteq \mathcal{X}$ containing x. A Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ on $\mathcal{B}_b(\mathcal{X})$ is called* asymptotically strong Feller at $x \in \mathcal{X}$ *if there exist a totally separating system of continuous pseudo-metrics $(d_n)_{n\in\mathbb{N}}$ for $\mathcal{X}$ and a nondecreasing sequence $(t_n)_{n\in\mathbb{N}}$, $t_n > 0$, such that*

$$\inf_{U\in\mathcal{U}_x} \limsup_{n\to\infty} \sup_{y\in U} \|\pi_{t_n}(x,\cdot) - \pi_{t_n}(y,\cdot)\|_{d_n} = 0. \tag{3.4}$$

It is called asymptotically strong Feller *if it is asymptotically strong Feller at every $x \in \mathcal{X}$.*

Remark 3.9. *Let d be some metric defining the topology of $\mathcal{X}$. (Such a metric exists, because $\mathcal{X}$ is a Polish space, so in particular metrizable.) Define $B(x,\gamma) := \{y \in \mathcal{X} \mid d(x,y) < \gamma\}$, e.g. $B(x,\gamma)$ denotes the open ball of radius $\gamma > 0$ centered at $x \in \mathcal{X}$. Then it is immediate that (3.4) is equivalent to*

$$\lim_{\gamma\to 0} \limsup_{n\to\infty} \sup_{y\in B(x,\gamma)} \|\pi_{t_n}(x,\cdot) - \pi_{t_n}(y,\cdot)\|_{d_n} = 0. \tag{3.5}$$

Proposition 3.10. *Suppose the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ is asymptotically strong Feller at $x \in \mathcal{X}$ and $t_n = t$ in (3.4) for all $n \in \mathbb{N}$ and some fixed $t > 0$. Then the Markovian transition kernel $(\pi_t(x,\cdot))_{x\in\mathcal{X}}$ is continuous in x in the total variation norm $\|\cdot\|_{TV}$.*

Proof. Let $t > 0$ and $(x_k)_{k\in\mathbb{N}} \subseteq \mathcal{X}$ with $\lim_{k\to\infty} d(x_k, x) = 0$. Then

$$\|\pi_t(x,\cdot) - \pi_t(x_l,\cdot)\|_{d_n} \leq \sup_{k\geq l} \|\pi_t(x,\cdot) - \pi_t(x_k,\cdot)\|_{d_n}$$

for all $l, n \in \mathbb{N}$. Hence letting $n \to \infty$ by Corollary 3.7

$$\|\pi_t(x,\cdot) - \pi_t(x_l,\cdot)\|_{TV} = \lim_{n\to\infty} \|\pi_t(x,\cdot) - \pi_t(x_l,\cdot)\|_{d_n} \leq \limsup_{n\to\infty} \sup_{k\geq l} \|\pi_t(x,\cdot) - \pi_t(x_k,\cdot)\|_{d_n}.$$

for all $l \in \mathbb{N}$. Therefore letting $l \to \infty$

$$\lim_{l\to\infty} \|\pi_t(x,\cdot) - \pi_t(x_l,\cdot)\|_{TV} \leq \lim_{l\to\infty} \limsup_{n\to\infty} \sup_{k\geq l} \|\pi_t(x,\cdot) - \pi_t(x_k,\cdot)\|_{d_n}.$$

Since by assumption the right hand side is equal to 0, the assertion follows, e.g. the Markovian transition kernel $(\pi_t(x,\cdot))_{x\in\mathcal{X}}$ is continuous in x in the total variation norm $\|\cdot\|_{TV}$. □

Remark 3.11. *If $(\mathcal{P}_t)_{t\geq 0}$ is asymptotically strong Feller and $t_n = t$ in (3.4) for all $n \in \mathbb{N}$ and some fixed $t > 0$, according to the previous proposition and Theorem 1.14 $(\mathcal{P}_t)_{t\geq 0}$ is strong Feller at time t and therefore by Remark 1.9 strong Feller for all times $s \geq t$.*

The following proposition states the converse in some sense:

Proposition 3.12. *Let $\mathcal{X}$ be a Polish space and $(\mathcal{P}_t)_{t\geq 0}$ a Markov semigroup on $\mathcal{B}_b(\mathcal{X})$ that is strong Feller. Then $(\mathcal{P}_t)_{t\geq 0}$ is asymptotically strong Feller.*

Proof. Let $x \in \mathcal{X}$ arbitrary and fixed. Let $(d_n)_{n\in\mathbb{N}}$ be a totally separating system of continuous pseudo-metrics for $\mathcal{X}$, in particular d_n converges pointwisely to the trivial metric d_{TV} as $n \to \infty$. Suppose the sequence $(t_n)_{n\in\mathbb{N}}$ is given by $t_n = t$ for some fixed $t > 0$ and all $n \in \mathbb{N}$. Then applying Corollary 3.7 yields

$$\|\pi_t(x,\cdot) - \pi_t(y,\cdot)\|_{d_n} \leq \|\pi_t(x,\cdot) - \pi_t(y,\cdot)\|_{d_{TV}} = \|\pi_t(x,\cdot) - \pi_t(y,\cdot)\|_{TV}$$

for all $n \in \mathbb{N}$, $y \in \mathcal{X}$. Therefore

$$\sup_{y\in B(x,\gamma)} \|\pi_t(x,\cdot) - \pi_t(y,\cdot)\|_{d_n} \leq \sup_{y\in B(x,\gamma)} \|\pi_t(x,\cdot) - \pi_t(y,\cdot)\|_{TV}$$

for all $\gamma > 0$, $n \in \mathbb{N}$. Hence letting $n \to \infty$

$$\limsup_{n\to\infty} \sup_{y\in B(x,\gamma)} \|\pi_t(x,\cdot) - \pi_t(y,\cdot)\|_{d_n} \quad\leq\quad \sup_{y\in B(x,\gamma)} \|\pi_t(x,\cdot) - \pi_t(y,\cdot)\|_{TV}$$

for all $\gamma > 0$. Finally, letting $\gamma \to 0$ gives

$$\lim_{\gamma\to 0} \limsup_{n\to\infty} \sup_{y\in B(x,\gamma)} \|\pi_t(x,\cdot) - \pi_t(y,\cdot)\|_{d_n} \leq \lim_{\gamma\to 0} \sup_{y\in B(x,\gamma)} \|\pi_t(x,\cdot) - \pi_t(y,\cdot)\|_{TV}. \tag{3.6}$$

Since $\mathcal{P}_{\frac{t}{2}}$ is strong Feller, by Lemma 1.15 the transition probabilities $\pi_t(x,\cdot)$, $x \in \mathcal{X}$, are continuous in x in the total variation norm $\|\cdot\|_{TV}$. Hence the right hand side of (3.6) equals 0. So $(\mathcal{P}_t)_{t\geq 0}$ is asymptotically strong Feller at $x \in \mathcal{X}$. Since $x \in \mathcal{X}$ arbitrary, the assertion follows. □

3.4 Sufficient condition

Another way of seeing the connection to the strong Feller property, is to recall that a standard criterion for $(\mathcal{P}_t)_{t\geq 0}$ to be strong Feller is given in Proposition 1.13. A sufficient condition of similar type for a Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ to be asymptotically strong Feller is given by

Theorem 3.13. *Let $\mathcal{H}$ be a separable Hilbert space. Let $(t_n)_{n\in\mathbb{N}}$ and $(\delta_n)_{n\in\mathbb{N}}$ be two positive sequences such that $t_n \leq t_{n+1}$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} \delta_n = 0$. If for all functions $\varphi \in \mathcal{C}_b^1(\mathcal{H})$*

$$|\mathcal{P}_{t_n}\varphi(x) - \mathcal{P}_{t_n}\varphi(y)| \leq C(\|x\| \vee \|y\|) \cdot (\|\varphi\|_\infty + \delta_n \cdot \|\nabla\varphi\|_\infty) \cdot \|x - y\| \tag{3.7}$$

for all $x, y \in \mathcal{H}$, $n \in \mathbb{N}$, where $C\colon \mathbb{R}_+ \to \mathbb{R}$ is a fixed nondecreasing function, then the semigroup $(\mathcal{P}_t)_{t\geq 0}$ is asymptotically strong Feller.

Remark 3.14. *In applications one typically has* $\lim_{n\to\infty} t_n = \infty$.

For the proof we will need the following two Lemmas.

Lemma 3.15. *Let* $\varphi \in \mathcal{C}_b^1(\mathcal{H})$. *Then*

$$\|\nabla\varphi\|_\infty = \|\varphi\|_d$$

with d *given by* $d(x,y) := \|x-y\|$ *for all* $x, y \in \mathcal{H}$.

Proof. Define a function $g\colon [0,1] \to \mathcal{H}$ by $g(s) := y + s\cdot(x-y)$. Using the fundamental theorem of calculus, the chain rule and the Cauchy-Schwartz inequality we have for all $x, y \in \mathcal{H}$

$$\begin{aligned}
|\varphi(x) - \varphi(y)| &= |\varphi(g(1)) - \varphi(g(0))| \\
&= \left| \int_0^1 \frac{d}{ds}\varphi\left(g(s)\right)\, ds \right| \\
&= \left| \int_0^1 D\varphi(g(s))g'(s)\, ds \right| \\
&= \left| \int_0^1 \langle \nabla\varphi\left(y + s\cdot(x-y)\right), x-y\rangle\, ds \right| \\
&\leq \int_0^1 |\langle \nabla\varphi\left(y + s\cdot(x-y)\right), x-y\rangle|\, ds \\
&\leq \int_0^1 \|\nabla\varphi\left(y + s\cdot(x-y)\right)\| \cdot \|x-y\|\, ds \\
&\leq \|\nabla\varphi\|_\infty \cdot \|x-y\| \\
&= \|\nabla\varphi\|_\infty \cdot d(x,y),
\end{aligned}$$

where the term $D\varphi(g(s))g'(s)$ should be interpreted as the Frechet derivative of φ at the point $g(s) \in \mathcal{H}$ applied to the element $g'(s) = x - y \in \mathcal{H}$. Hence

$$\frac{|\varphi(x) - \varphi(y)|}{d(x,y)} \leq \|\nabla\varphi\|_\infty$$

for all $x, y \in \mathcal{H}$, $x \neq y$, and therefore

$$\|\varphi\|_d = \sup_{x,y\in\mathcal{H}\colon x\neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x,y)} \leq \|\nabla\varphi\|_\infty.$$

To show the converse inequality, let $x, v \in \mathcal{H}$ with $\|v\| = 1$ and $(h_n)_{n\in\mathbb{N}}$ a sequence converging to 0. Then

$$\begin{aligned}
|\langle \nabla\varphi(x), v\rangle| &= \left| \lim_{n\to\infty} \frac{\varphi(x+h_n v)-\varphi(x)}{h_n} \right| \\
&= \lim_{n\to\infty} \frac{|\varphi(x+h_n v)-\varphi(x)|}{|h_n|} \\
&\leq \sup_{n\in\mathbb{N}} \frac{|\varphi(x+h_n v)-\varphi(x)|}{|h_n|} \\
&= \sup_{n\in\mathbb{N}} \frac{|\varphi(x+h_n v)-\varphi(x)|}{|h_n|\cdot\|v\|}
\end{aligned}$$

$$\begin{aligned}
&= \sup_{n\in\mathbb{N}} \frac{|\varphi(x+h_n v)-\varphi(x)|}{\|(x+h_n v)-x\|} \\
&\leq \sup_{x,y\in\mathcal{H}\colon x\neq y} \frac{|\varphi(x)-\varphi(y)|}{\|x-y\|} \\
&= \sup_{x,y\in\mathcal{H}\colon x\neq y} \frac{|\varphi(x)-\varphi(y)|}{d(x,y)} \\
&= \|\varphi\|_d.
\end{aligned}$$

In particular, for $v = \frac{\nabla\varphi(x)}{\|\nabla\varphi(x)\|}$

$$\|\nabla\varphi(x)\| = \frac{1}{\|\nabla\varphi(x)\|}\cdot|\langle\nabla\varphi(x),\nabla\varphi(x)\rangle| = |\langle\nabla\varphi(x),\frac{\nabla\varphi(x)}{\|\nabla\varphi(x)\|}\rangle| \leq \|\varphi\|_d$$

for all $x \in \mathcal{H}$. Therefore

$$\|\nabla\varphi\|_\infty = \sup_{x\in\mathcal{H}}\|\nabla\varphi(x)\| \leq \|\varphi\|_d.$$

□

Lemma 3.16. *Let $\varphi \in \mathcal{L}_d(\mathcal{H})$. Then there exists a sequence $(\varphi_m)_{m\in\mathbb{N}}$, $\varphi_m \in \mathcal{C}_b^\infty(\mathcal{H})$, such that*

(i) $\varphi_m \overset{m\to\infty}{\longrightarrow} \varphi$ pointwisely,

(ii) $\|\varphi_m\|_\infty \leq \|\varphi\|_\infty$ for all $m \in \mathbb{N}$,

(iii) $\|\varphi_m\|_d \leq \|\varphi\|_d$ for all $m \in \mathbb{N}$.

Proof. Let $\varphi \in \mathcal{L}_d(\mathcal{H})$ and $\{e_i\}_{i\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$. For $m \in \mathbb{N}$ define orthogonal projections

$$\begin{aligned}
P_m\colon \mathcal{H} &\to P_m(\mathcal{H}) = \operatorname{span}\{e_1,\dots e_m\} \\
x &\mapsto P_m x := \sum_{i=1}^m \langle x, e_i\rangle\cdot e_i
\end{aligned}$$

and corresponding bijections

$$\begin{aligned} J_m\colon P_m(\mathcal{H}) &\to \mathbb{R}^m \\ \sum_{i=1}^m \langle x, e_i\rangle \cdot e_i &\mapsto (\langle x, e_1\rangle, \ldots, \langle x, e_m\rangle)\,. \end{aligned}$$

Note that $\varphi \circ J_m^{-1} \in \mathcal{L}_d(\mathbb{R}^m)$ with Lipschitz constant $\|\varphi \circ J_m^{-1}\|_d \leq \|\varphi\|_d$ for all $m \in \mathbb{N}$. Hence by Lemma A.1 in the appendix, for every $m \in \mathbb{N}$ there exists a sequence $(f_{m,k})_{k\in\mathbb{N}}$, $f_{m,k} \in \mathcal{C}_b^\infty(\mathbb{R}^m)$, such that $f_{m,k}(x) \stackrel{k\to\infty}{\longrightarrow} (\varphi\circ J_m^{-1})(x)$ for all $x \in \mathbb{R}^m$, $\|f_{m,k}\|_\infty \leq \|\varphi\circ J_m^{-1}\|_\infty$ and $\|f_{m,k}\|_d \leq \|\varphi \circ J_m^{-1}\|_d$ for all $k \in \mathbb{N}$. Replacing x by $J_m(x)$ yields $(f_{m,k} \circ J_m)(x) \stackrel{k\to\infty}{\longrightarrow} \varphi(x)$ for all $x \in P_m(\mathcal{H})$. Cosequently $(f_{m,k} \circ J_m \circ P_m)(x) \stackrel{k\to\infty}{\longrightarrow} (\varphi \circ P_m)(x)$ for all $x \in \mathcal{H}$. Moreover $\|f_{m,k}\circ J_m\circ P_m\|_\infty \leq \|\varphi\circ P_m\|_\infty$ and $\|f_{m,k}\circ J_m\circ P_m\|_d \leq \|\varphi\circ P_m\|_d$ for all $k \in \mathbb{N}$. Since $\varphi \circ P_m \stackrel{m\to\infty}{\longrightarrow} \varphi$ pointwisely, $\|\varphi \circ P_m\|_\infty \leq \|\varphi\|_\infty$ and $\|\varphi \circ P_m\|_d \leq \|\varphi\|_d$, a diagonal argument applies in order to obtain a subsequence $(\varphi_m)_{m\in\mathbb{N}}$, $\varphi_m := f_{m,m} \circ J_m \circ P_m \in \mathcal{C}_b^\infty(\mathcal{H})$, such that $\varphi_m(x) \stackrel{m\to\infty}{\longrightarrow} \varphi(x)$ for all $x \in \mathcal{H}$, $\|\varphi_m\|_\infty \leq \|\varphi\|_\infty$ and $\|\varphi_m\|_d \leq \|\varphi\|_d$ for all $m \in \mathbb{N}$. □

Now we turn to the proof of Theorem 3.13.

Proof. For $\varepsilon > 0$ define on $\mathcal{H}$ the metric

$$\begin{aligned} d_\varepsilon\colon \mathcal{H}\times\mathcal{H} &\to \mathbb{R}_+ \\ (w_1, w_2) &\mapsto d_\varepsilon(w_1, w_2) := 1 \wedge \frac{1}{\varepsilon} \cdot \|w_1 - w_2\|. \end{aligned}$$

It is clear that this is a metric on $\mathcal{H}$. In fact, the triangle inequality follows from $(1 \wedge a) + (1 \wedge b) \geq 1 \wedge (a + b)$ for all $a, b \geq 0$. Denote by $\|\cdot\|_\varepsilon := \|\cdot\|_{d_\varepsilon}$ the corresponding seminorms on functions and on measures given by (2.31) and (3.2) respectively. Since $(\delta_n)_{n\in\mathbb{N}}$ converges to 0 from above, extracting a subsequence if necessary, we have $\delta_n \geq \delta_{n+1}$ for all $n \in \mathbb{N}$. Then $(d_{\delta_n})_{n\in\mathbb{N}}$ is a totally separating system of continuous metrics for $\mathcal{H}$: first, $d_{\delta_n}(w_1, w_2) = 1 \wedge \frac{1}{\delta_n} \cdot \|w_1 - w_2\| \leq 1 \wedge \frac{1}{\delta_{n+1}} \cdot \|w_1 - w_2\| = d_{\delta_{n+1}}(w_1, w_2)$ for all $(w_1, w_2) \in \mathcal{H}^2$, $n \in \mathbb{N}$. Furthermore, $\lim_{n\to\infty} d_{\delta_n}(w_1, w_2) = \lim_{n\to\infty} 1 \wedge \frac{1}{\delta_n} \cdot \|w_1 - w_2\| = 1$ for all $(w_1, w_2) \in \mathcal{H}^2$, $w_1 \neq w_2$. Hence $(d_{\delta_n})_{n\in\mathbb{N}}$ is a totally separating system of continuous metrics for $\mathcal{H}$.

Since

$$\|\varphi\|_d = \frac{1}{\varepsilon} \cdot \sup_{x,y\in\mathcal{H}\colon x\neq y} \frac{|\varphi(x) - \varphi(y)|}{\varepsilon^{-1} \cdot d(x, y)} \leq \frac{1}{\varepsilon} \cdot \sup_{x,y\in\mathcal{H}\colon x\neq y} \frac{|\varphi(x) - \varphi(y)|}{\underbrace{1 \wedge \varepsilon^{-1} \cdot d(x, y)}_{=d_\varepsilon(x,y)}} = \frac{1}{\varepsilon} \cdot \|\varphi\|_\varepsilon,$$

it follows immediately from (3.7) and Lemma 3.15 that for every Frechet differentiable function $\varphi\colon \mathcal{H} \to \mathbb{R}$ we have

$$\begin{aligned}
&\left|\int_{\mathcal{H}} \varphi(w)\ \left(\pi_{t_n}(w_1, dw) - \pi_{t_n}(w_2, dw)\right)\right| \\
=\ & |\mathcal{P}_{t_n}\varphi(w_1) - \mathcal{P}_{t_n}\varphi(w_2)| \\
\leq\ & C(\|w_1\| \vee \|w_2\|) \cdot (\|\varphi\|_\infty + \delta_n \cdot \|\nabla\varphi\|_\infty) \cdot \|w_1 - w_2\| \\
=\ & C(\|w_1\| \vee \|w_2\|) \cdot (\|\varphi\|_\infty + \delta_n \cdot \|\varphi\|_d) \cdot \|w_1 - w_2\| \\
\leq\ & C(\|w_1\| \vee \|w_2\|) \cdot \left(\|\varphi\|_\infty + \frac{\delta_n}{\varepsilon} \cdot \|\varphi\|_\varepsilon\right) \cdot \|w_1 - w_2\|. \qquad (3.8)
\end{aligned}$$

Now take a d_ε-Lipschitz continuous function $\varphi\colon \mathcal{H} \to \mathbb{R}$ with $\|\varphi\|_\varepsilon \leq 1$. Without loss of generality assume $\|\varphi\|_\infty \leq 1$. Then by Lemma 3.16 there exists a sequence $(\varphi_m)_{m\in\mathbb{N}}$ of Frechet differentiable functions $\varphi_m\colon \mathcal{H} \to \mathbb{R}$ such that $\varphi_m \longrightarrow \varphi$ pointwisely as $m \to \infty$, $\|\varphi_m\|_\infty \leq \|\varphi\|_\infty \leq 1$ and $\|\varphi_m\|_\varepsilon \leq \|\varphi\|_\varepsilon \leq 1$ for all $m \in \mathbb{N}$. Therefore by the dominated convergence theorem and (3.8)

$$\begin{aligned}
&\left|\int_{\mathcal{H}} \varphi(w)\ \left(\pi_{t_n}(w_1, dw) - \pi_{t_n}(w_2, dw)\right)\right| \\
=\ & \lim_{m\to\infty} \left|\int_{\mathcal{H}} \varphi_m(w)\ \left(\pi_{t_n}(w_1, dw) - \pi_{t_n}(w_2, dw)\right)\right| \\
\leq\ & \lim_{m\to\infty} C(\|w_1\| \vee \|w_2\|) \cdot \left(\|\varphi_m\|_\infty + \frac{\delta_n}{\varepsilon} \cdot \|\varphi_m\|_\varepsilon\right) \cdot \|w_1 - w_2\| \\
\leq\ & C(\|w_1\| \vee \|w_2\|) \cdot \left(\|\varphi\|_\infty + \frac{\delta_n}{\varepsilon} \cdot \|\varphi\|_\varepsilon\right) \cdot \|w_1 - w_2\| \\
\leq\ & C(\|w_1\| \vee \|w_2\|) \cdot \left(1 + \frac{\delta_n}{\varepsilon}\right) \cdot \|w_1 - w_2\|.
\end{aligned}$$

Hence

$$\begin{aligned}
|||\pi_{t_n}(w_1, \cdot) - \pi_{t_n}(w_2, \cdot)|||_\varepsilon &= \sup_{\|\varphi\|_\varepsilon = 1} \left|\int_{\mathcal{H}} \varphi(w)\ \left(\pi_{t_n}(w_1, dw) - \pi_{t_n}(w_2, dw)\right)\right| \\
&\leq C(\|w_1\| \vee \|w_2\|) \cdot \left(1 + \frac{\delta_n}{\varepsilon}\right) \cdot \|w_1 - w_2\|.
\end{aligned}$$

Applying Corollary 2.19 yields

$$\|\pi_{t_n}(w_1, \cdot) - \pi_{t_n}(w_2, \cdot)\|_\varepsilon \leq C(\|w_1\| \vee \|w_2\|) \cdot \left(1 + \frac{\delta_n}{\varepsilon}\right) \cdot \|w_1 - w_2\|.$$

Choosing $\varepsilon = a_n = \sqrt{\delta_n}$, we obtain

$$\|\pi_{t_n}(w_1, \cdot) - \pi_{t_n}(w_2, \cdot)\|_{a_n} \leq C(\|w_1\| \vee \|w_2\|) \cdot (1 + a_n) \cdot \|w_1 - w_2\|,$$

for all $n \in \mathbb{N}$, which in turn implies that $(\mathcal{P}_t)_{t\geq 0}$ is asymptotically strong Feller, since $a_n \to 0$ for $n \to \infty$. □

3.5 Examples

The following two examples will demonstrate the powerness of the asymptotic strong Feller property even in the finite-dimensional setting.

Example 3.17. *Consider the two-dimensional SDE*

$$\begin{pmatrix} dx(t) \\ dy(t) \end{pmatrix} = -\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} dt + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} dW(t) \tag{3.9}$$

with initial condition

$$\begin{pmatrix} x(0) \\ y(0) \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

where

$$W(t) = \begin{pmatrix} W^{(1)}(t) \\ W^{(2)}(t) \end{pmatrix}$$

is a two-dimensional real-valued Wiener process on some probability space $(\Omega, \mathcal{F}, P)$.

The solution is given by

$$\begin{pmatrix} x(t, x_0) \\ y(t, y_0) \end{pmatrix} = \begin{pmatrix} e^{-t}x_0 + \int_0^t e^{-(t-s)}\, dW^{(1)}(s) \\ e^{-t}y_0 \end{pmatrix}.$$

Note that the first component is an Ornstein-Uhlenbeck process.

We claim that the corresponding Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$*, defined by*

$$\mathcal{P}_t\varphi\left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\right) := \mathbb{E}\left[\varphi\left(\begin{pmatrix} x(t, x_0) \\ y(t, y_0) \end{pmatrix}\right)\right]$$

for all $\varphi \in \mathcal{B}_b(\mathbb{R}^2)$, $(x_0, y_0) \in \mathbb{R}^2$, $t \geq 0$, *is not strong Feller but asymptotically strong Feller.*

To see that $\mathcal{P}_t$ *is not strong Feller, let* $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$ *be given by*

$$\varphi\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = sgn(y) := \begin{cases} 1 & if\ y \geq 0 \\ -1 & if\ y < 0 \end{cases}$$

and observe that for this choice we obtain

$$\begin{aligned}
\mathcal{P}_t\varphi\left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\right) &= \mathbb{E}\left[\varphi\left(\begin{pmatrix} x(t, x_0) \\ y(t, y_0) \end{pmatrix}\right)\right] \\
&= \mathbb{E}\left[\varphi\left(\begin{pmatrix} e^{-t}x_0 + \int_0^t e^{-(t-s)}\, dW^{(1)}(s) \\ e^{-t}y_0 \end{pmatrix}\right)\right] \\
&= \mathbb{E}\left[sgn(e^{-t}y_0)\right] \\
&= sgn\left(e^{-t}y_0\right) \\
&= sgn(y_0) \\
&= \varphi\left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\right)
\end{aligned}$$

for all $(x_0, y_0) \in \mathbb{R}^2$, $t \geq 0$. *Since* $\varphi \in \mathcal{B}_b(\mathbb{R}^2)$ *but* $\varphi \notin \mathcal{C}_b(\mathbb{R}^2)$, *the system is not strong Feller.*

Denote by $\binom{u(t)}{v(t)}$ *the linearization of equation (3.9), that is* $du(t) = -u(t)dt$ *and* $dv(t) = -v(t)dt$ *with initial conditions* $u(0) = \xi_1$ *and* $v(0) = \xi_2$ *respectively. As before, define a function* $g\colon [0,1] \to \mathbb{R}^2$ *by*

$$g(s) := \begin{pmatrix} x(t,x_0) \\ y(t,y_0) \end{pmatrix} + s \cdot \begin{pmatrix} u(t,\xi_1) \\ v(t,\xi_2) \end{pmatrix}.$$

In order to show that the system is asymptotically strong Feller, observe that for any differentiable function $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$ *and any direction* $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$ *with* $\|\xi\| = 1$ *we have, using the fundamental theorem of calculus, the chain rule and the Cauchy-Schwartz inequality*

$$\begin{aligned}
&\left| \mathcal{P}_t\varphi\left(\begin{pmatrix} x_0+\xi_1 \\ y_0+\xi_2 \end{pmatrix}\right) - \mathcal{P}_t\varphi\left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\right)\right| \\
= &\left| \mathbb{E}\left[\varphi\left(\begin{pmatrix} x(t,x_0+\xi_1) \\ y(t,y_0+\xi_2) \end{pmatrix}\right)\right] - \mathbb{E}\left[\varphi\left(\begin{pmatrix} x(t,x_0) \\ y(t,y_0) \end{pmatrix}\right)\right]\right| \\
= &\left| \mathbb{E}\left[\varphi\left(\begin{pmatrix} x(t,x_0+\xi_1) \\ y(t,y_0+\xi_2) \end{pmatrix}\right) - \varphi\left(\begin{pmatrix} x(t,x_0) \\ y(t,y_0) \end{pmatrix}\right)\right]\right| \\
= &\left| \mathbb{E}\left[\varphi\left(\begin{pmatrix} x(t,x_0)+u(t,\xi_1) \\ y(t,y_0)+v(t,\xi_2) \end{pmatrix}\right) - \varphi\left(\begin{pmatrix} x(t,x_0) \\ y(t,y_0) \end{pmatrix}\right)\right]\right| \\
= &\left| \mathbb{E}\left[\varphi\left(\begin{pmatrix} x(t,x_0) \\ y(t,y_0) \end{pmatrix} + \begin{pmatrix} u(t,\xi_1) \\ v(t,\xi_2) \end{pmatrix}\right) - \varphi\left(\begin{pmatrix} x(t,x_0) \\ y(t,y_0) \end{pmatrix}\right)\right]\right| \\
= &\left| \mathbb{E}\left[\varphi\left(g(1)\right) - \varphi\left(g(0)\right)\right]\right| \\
= &\left| \mathbb{E}\left[\int_0^1 \frac{d}{ds}\varphi\left(g(s)\right)\, ds\right]\right| \\
= &\left| \mathbb{E}\left[\int_0^1 \left(\frac{\partial}{\partial x}\varphi(g(s)), \frac{\partial}{\partial y}\varphi(g(s))\right) \cdot g'(s)\, ds\right]\right| \\
= &\left| \mathbb{E}\left[\int_0^1 \langle \nabla\varphi\left(\begin{pmatrix} x(t,x_0) \\ y(t,y_0) \end{pmatrix} + s\cdot\begin{pmatrix} u(t,\xi_1) \\ v(t,\xi_2) \end{pmatrix}\right), \begin{pmatrix} u(t,\xi_1) \\ v(t,\xi_2) \end{pmatrix}\rangle\, ds\right]\right| \\
\leq\ &\mathbb{E}\left[\int_0^1 \left|\langle \nabla\varphi\left(\begin{pmatrix} x(t,x_0) \\ y(t,y_0) \end{pmatrix} + s\cdot\begin{pmatrix} u(t,\xi_1) \\ v(t,\xi_2) \end{pmatrix}\right), \begin{pmatrix} u(t,\xi_1) \\ v(t,\xi_2) \end{pmatrix}\rangle\right|\, ds\right] \\
\leq\ &\mathbb{E}\left[\int_0^1 \left\|\nabla\varphi\left(\begin{pmatrix} x(t,x_0) \\ y(t,y_0) \end{pmatrix} + s\cdot\begin{pmatrix} u(t,\xi_1) \\ v(t,\xi_2) \end{pmatrix}\right)\right\| \cdot \left\|\begin{pmatrix} u(t,\xi_1) \\ v(t,\xi_2) \end{pmatrix}\right\|\, ds\right] \\
\leq\ &\|\nabla\varphi\|_\infty \cdot \mathbb{E}\left[\left\|\begin{pmatrix} u(t,\xi_1) \\ v(t,\xi_2) \end{pmatrix}\right\|\right] \\
= &\ \|\nabla\varphi\|_\infty \cdot \left\|\begin{pmatrix} e^{-t}\cdot u(0) \\ e^{-t}\cdot v(0) \end{pmatrix}\right\| \\
= &\ \|\nabla\varphi\|_\infty \cdot \left|e^{-t}\right| \cdot \left\|\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}\right\| \\
= &\ \|\nabla\varphi\|_\infty \cdot e^{-t}.
\end{aligned}$$

Let $(t_n)_{n\in\mathbb{N}}$ *be a positive nondecreasing sequence with* $\lim_{n\to\infty} t_n = \infty$ *and define* $\delta_n := e^{-t_n}$ *for all* $n \in \mathbb{N}$*. Then the conclusion follows by Theorem 3.13 taking the function* C *to be constant and equal* 1*.*

Example 3.18. *Now consider the two-dimensional SDE*

$$\begin{pmatrix} dx(t) \\ dy(t) \end{pmatrix} = \left[\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} - \begin{pmatrix} x(t)^3 \\ 0 \end{pmatrix} \right] dt + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} dW(t) \tag{3.10}$$

with initial condition

$$\begin{pmatrix} x(0) \\ y(0) \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

where again

$$W(t) = \begin{pmatrix} W^{(1)}(t) \\ W^{(2)}(t) \end{pmatrix}$$

is a two-dimensional real-valued Wiener process on some probability space $(\Omega, \mathcal{F}, P)$*. Again denote by* $\binom{u(t)}{v(t)}$ *the linearization of equation (3.10).*

As before the function $\varphi(\binom{x}{y}) = sgn(y)$ *is invariant under* $\mathcal{P}_t$*, implying that the system is not strong Feller. However, in the contrast to the previous example, it is not globally contractive. Therefore the situation is in some kind a little delicate and we will need the following fact (cf. Lemma 4.10 in [9]):*

$$\left| \frac{\partial}{\partial x} \mathcal{P}_t \varphi \left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \right) \right| \leq C(|x_0|) \cdot \|\varphi\|_\infty$$

for some nondecreasing function $C \colon \mathbb{R}_+ \to \mathbb{R}$ *and all* $t \geq 1$*.*

Hence applying the mean value theorem to both variables respectively we gain

$$\begin{aligned}
&\left| \mathcal{P}_t \varphi \left(\begin{pmatrix} x_0 + \xi_1 \\ y_0 + \xi_2 \end{pmatrix} \right) - \mathcal{P}_t \varphi \left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \right) \right| \\
= \; &\left| \mathbb{E} \left[\varphi \left(\begin{pmatrix} x(t, x_0 + \xi_1) \\ y(t, y_0 + \xi_2) \end{pmatrix} \right) - \varphi \left(\begin{pmatrix} x(t, x_0) \\ y(t, y_0) \end{pmatrix} \right) \right] \right| \\
= \; &\left| \mathbb{E} \left[\varphi \left(\begin{pmatrix} x(t, x_0 + \xi_1) \\ y(t, y_0) + v(t, \xi_2) \end{pmatrix} \right) - \varphi \left(\begin{pmatrix} x(t, x_0) \\ y(t, y_0) \end{pmatrix} \right) \right] \right| \\
= \; &\Bigg| \mathbb{E} \left[\varphi \left(\begin{pmatrix} x(t, x_0 + \xi_1) \\ y(t, y_0) + v(t, \xi_2) \end{pmatrix} \right) - \varphi \left(\begin{pmatrix} x(t, x_0 + \xi_1) \\ y(t, y_0) \end{pmatrix} \right) \right] \\
&\qquad + \mathbb{E} \left[\varphi \left(\begin{pmatrix} x(t, x_0 + \xi_1) \\ y(t, y_0)) \end{pmatrix} \right) - \varphi \left(\begin{pmatrix} x(t, x_0) \\ y(t, y_0) \end{pmatrix} \right) \right] \Bigg| \\
= \; &\Bigg| \mathbb{E} \left[\int_0^1 \frac{\partial}{\partial y} \varphi \left(\begin{pmatrix} x(t, x_0 + \xi_1) \\ y(t, y_0) + s \cdot v(t, \xi_2) \end{pmatrix} \right) \, ds \cdot v(t, \xi_2) \right] \\
&\qquad + \mathcal{P}_t \varphi \left(\begin{pmatrix} x_0 + \xi_1 \\ y_0 \end{pmatrix} \right) - \mathcal{P}_t \varphi \left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \right) \Bigg|
\end{aligned}$$

$$= \left| \mathbb{E}\left[\int_0^1 \frac{\partial}{\partial y}\varphi\left(\begin{pmatrix} x(t,x_0+\xi_1) \\ y(t,y_0)+s\cdot v(t,\xi_2)\end{pmatrix}\right)\, ds \cdot v(t,\xi_2)\right] + \int_0^1 \frac{\partial}{\partial x}\mathcal{P}_t\varphi\left(\begin{pmatrix} x_0+r\cdot\xi_1 \\ y_0\end{pmatrix}\right)\, dr\cdot \xi_1 \right|$$

$$\begin{aligned}
&\leq \mathbb{E}\left[\int_0^1 \underbrace{\left|\frac{\partial}{\partial y}\varphi\left(\begin{pmatrix} x(t,x_0+\xi_1) \\ y(t,y_0)+s\cdot v(t,\xi_2)\end{pmatrix}\right)\right|}_{\leq\|\nabla\varphi\|_\infty}\, ds\cdot |v(t,\xi_2)|\right] \\
&\qquad + \int_0^1 \underbrace{\left|\frac{\partial}{\partial x}\mathcal{P}_t\varphi\left(\begin{pmatrix} x_0+r\cdot\xi_1 \\ y_0\end{pmatrix}\right)\right|}_{\leq C(|x_0+r\cdot\xi_1|)\cdot\|\varphi\|_\infty\leq C(|x_0|+1)\cdot\|\varphi\|_\infty}\, dr\cdot|\xi_1| \\
&\leq \|\nabla\varphi\|_\infty\cdot\mathbb{E}\left[|v(t,\xi_2)|\right] + C(|x_0|+1)\cdot\|\varphi\|_\infty \\
&= \|\nabla\varphi\|_\infty\cdot e^{-t}\cdot|\xi_2| + C(|x_0|+1)\cdot\|\varphi\|_\infty \\
&\leq (C(|x_0|+1)+1)\cdot\left(\|\varphi\|_\infty + e^{-t}\cdot\|\nabla\varphi\|_\infty\right).
\end{aligned}$$

Taking a positive nondecreasing sequence $(t_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} t_n = \infty$ and defining $\delta_n := e^{-t_n}$ for all $n\in\mathbb{N}$, the assertion follows by Theorem 3.13.

3.6 Uniqueness of the invariant measure

We conclude this chapter by proving in some sense the analogue of Proposition 1.37.

Theorem 3.19. *Let $\mathcal{X}$ be a Polish space, $(\mathcal{P}_t)_{t\geq 0}$ a Markov semigroup on $\mathcal{B}_b(\mathcal{X})$ and μ,ν, $\mu\neq\nu$, two ergodic Borel probability measures for $(\mathcal{P}_t)_{t\geq 0}$. If $(\mathcal{P}_t)_{t\geq 0}$ is asymptotically strong Feller at $x\in\mathcal{X}$, then $x\notin supp(\mu)\cap supp(\nu)$.*

Proof. First of all μ and ν are singular by Theorem 1.30. Hence we obtain for their difference the total variation

$$\|\mu-\nu\|_{TV} = \frac{1}{2}\cdot\left((\mu-\nu)^+(\mathcal{X}) + (\mu-\nu)^-(\mathcal{X})\right) = \frac{1}{2}\cdot(\mu(\mathcal{X})+\nu(\mathcal{X})) = 1.$$

For every $A\in\mathcal{B}(\mathcal{X})$, $t>0$ and every pseudo-metric d on $\mathcal{X}$ with $d\leq 1$ the triangle inequality for $\|\cdot\|_d$ implies

$$\|\mu-\nu\|_d \leq 1-\min\{\mu(A),\nu(A)\}\cdot\left(1-\max_{y,z\in A}\|\pi_t(z,\cdot)-\pi_t(y,\cdot)\|_d\right). \tag{3.11}$$

To see this, set $\alpha := \min\{\mu(A),\nu(A)\}$ and distinguish the following two cases:

Case 1 ($\alpha = 0$): For $\alpha = 0$ we obtain by definition of the Wasserstein distance

$$\begin{aligned}
\|\mu - \nu\|_d &= \inf_{\eta \in \mathcal{C}(\mu,\nu)} \int_{\mathcal{X}^2} d(x,y)\, \eta(dx,dy) \\
&\leq \inf_{\eta \in \mathcal{C}(\mu,\nu)} \int_{\mathcal{X}^2} 1\, \eta(dx,dy) \\
&\leq \int_{\mathcal{X}^2} 1\, (\mu \otimes \nu)(dx,dy) \\
&= (\mu \otimes \nu)(\mathcal{X}^2) \\
&= \mu(\mathcal{X}) \cdot \nu(\mathcal{X}) \\
&= 1 \\
&= 1 - \alpha \cdot \left(1 - \max_{y,z \in A} \|\pi_t(z,\cdot) - \pi_t(y,\cdot)\|_d\right).
\end{aligned}$$

Case 2 ($\alpha > 0$): Clearly, there exist probability measures $\mu_A, \bar{\mu}, \nu_A, \bar{\nu}$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ with $\nu_A(A) = \mu_A(A) = 1$ such that $\mu = (1-\alpha)\cdot\bar{\mu} + \alpha\cdot\mu_A$ and $\nu = (1-\alpha)\cdot\bar{\nu} + \alpha\cdot\nu_A$. In fact, without loss of generality assume $\alpha = \mu(A)$ and take, e.g. $\mu_A(B) := \frac{\mu(B\cap A)}{\alpha}$, $\bar{\mu}(B) := \frac{\mu(B\cap A^c)}{1-\alpha}$, $\nu_A(B) := \frac{\nu(B\cap A)}{\nu(A)}$ and $\bar{\nu}(B) := \frac{\nu(B) - \alpha\cdot\nu_A(B)}{1-\alpha}$ for all $B \in \mathcal{B}(\mathcal{X})$. From the invariance of the measures μ and ν and the triangle inequality this implies

$$\begin{aligned}
& \|\mu - \nu\|_d \\
=\ & \|\mathcal{P}_t^*\mu - \mathcal{P}_t^*\nu\|_d \\
=\ & \|\mathcal{P}_t^*\left((1-\alpha)\cdot\bar{\mu} + \alpha\cdot\mu_A\right) - \mathcal{P}_t^*\left((1-\alpha)\cdot\bar{\nu} + \alpha\cdot\nu_A\right)\|_d \\
=\ & \|(1-\alpha)\cdot\mathcal{P}_t^*\bar{\mu} + \alpha\cdot\mathcal{P}_t^*\mu_A - (1-\alpha)\cdot\mathcal{P}_t^*\bar{\nu} - \alpha\cdot\mathcal{P}_t^*\nu_A\|_d \\
=\ & \|(1-\alpha)\cdot(\mathcal{P}_t^*\bar{\mu} - \mathcal{P}_t^*\bar{\nu}) + \alpha\cdot(\mathcal{P}_t^*\mu_A - \mathcal{P}_t^*\nu_A)\|_d \\
\leq\ & (1-\alpha)\cdot\|\mathcal{P}_t^*\bar{\mu} - \mathcal{P}_t^*\bar{\nu}\|_d + \alpha\cdot\|\mathcal{P}_t^*\mu_A - \mathcal{P}_t^*\nu_A\|_d \\
\leq\ & (1-\alpha) + \alpha\cdot\left\|\int_{\mathcal{X}} \pi_t(z,\cdot)\,\mu_A(dz) - \int_{\mathcal{X}} \pi_t(y,\cdot)\,\nu_A(dy)\right\|_d \\
=\ & (1-\alpha) + \alpha\cdot\left\|\int_{A} \pi_t(z,\cdot)\,\mu_A(dz) - \int_{A} \pi_t(y,\cdot)\,\nu_A(dy)\right\|_d \\
=\ & (1-\alpha) + \alpha\cdot\left\|\iint_{A\times A} \pi_t(z,\cdot)\,\mu_A(dz)\,\nu_A(dy) - \iint_{A\times A} \pi_t(y,\cdot)\,\mu_A(dz)\,\nu_A(dy)\right\|_d \\
=\ & (1-\alpha) + \alpha\cdot\left\|\iint_{A\times A} \pi_t(z,\cdot) - \pi_t(y,\cdot)\,\mu_A(dz)\,\nu_A(dy)\right\|_d \\
\leq\ & (1-\alpha) + \alpha\cdot\iint_{A\times A} \|\pi_t(z,\cdot) - \pi_t(y,\cdot)\|_d\,\mu_A(dz)\,\nu_A(dy) \\
\leq\ & 1 - \alpha\cdot\left(1 - \max_{y,z\in A}\|\pi_t(z,\cdot) - \pi_t(y,\cdot)\|_d\right).
\end{aligned}$$

Continuing with the proof of the theorem, we see that, by definition of the asymptotic strong Feller property, there exist a constant $N > 0$, a totally separating system $(d_n)_{n\in\mathbb{N}}$

of continuous pseudo-metrics and an open set U containing x such that $\|\pi_{t_n}(z,\cdot) - \pi_{t_n}(y,\cdot)\|_{d_n} \leq \frac{1}{2}$ for every $n > N$ and every $y, z \in U$. (Note that by definition of a totally separating system, the pseudo-metrics d_n are less or equal 1.)

Assume, by contradiction, that $x \in \text{supp}(\mu) \cap \text{supp}(\nu)$. Hence $\alpha = \min\{\mu(U), \nu(U)\} > 0$ according to (1.19). Taking $A = U$, $d = d_n$ and $t = t_n$ in (3.11), we then get $\|\mu - \nu\|_{d_n} \leq 1 - \frac{\alpha}{2}$ for every $n > N$. Therefore $\|\mu - \nu\|_{TV} \leq 1 - \frac{\alpha}{2}$ by Corollary 3.7, in contradiction to $\|\mu - \nu\|_{TV} = 1$. □

As an immediate consequence we have

Corollary 3.20. *Let $(\mathcal{P}_t)_{t\geq 0}$ be an asymptotically strong Feller Markov semigroup on $\mathcal{B}_b(\mathcal{X})$ and assume that there exists a point $x \in \mathcal{X}$ such that $x \in supp(\mu)$ for every invariant Borel probability measure μ for $(\mathcal{P}_t)_{t\geq 0}$. Then there exists at most one invariant Borel probability measure μ for $(\mathcal{P}_t)_{t\geq 0}$.*

Proof. Suppose there is more than one invariant Borel probability measure for the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$. Then by Corollary 1.31 there exist at least two distinct ergodic Borel probability measures μ and ν for $(\mathcal{P}_t)_{t\geq 0}$. Since $(\mathcal{P}_t)_{t\geq 0}$ is asymptotically strong Feller (at every $x \in \mathcal{X}$), by the previous Theorem 3.19 $x \notin \text{supp}(\mu) \cap \text{supp}(\nu)$ for all $x \in \mathcal{X}$, i.e. $\text{supp}(\mu) \cap \text{supp}(\nu) = \emptyset$. Hence, if there exists a point x as required in the formulation of the corollary, e.g. $x \in \text{supp}(\mu)$ for every invariant probability measure μ, then there is at most one invariant Borel probability measure μ for $(\mathcal{P}_t)_{t\geq 0}$. □

Remark 3.21. *According to Theorem 1.27 μ is ergodic then.*

4 Application to stochastic differential equations with Lipschitz nonlinearities

The following situation is treated in [9].

We are given two real separable Hilbert spaces H and U. Suppose $\{e_k\}_{k\in\mathbb{N}}$ is an orthonormal basis in U and $\{\beta_k\}_{k\in\mathbb{N}}$ is a sequence of mutually independent real-valued standard Brownian motions on a fixed probability space $(\Omega, \mathcal{F}, P)$. Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration such that $\mathcal{F}_t$ is generated by $\beta_k(s)$, $s \leq t$, $k \in \mathbb{N}$.

We consider the SDE

$$\begin{aligned} dX(t) &= (AX(t) + F(X(t)))\, dt + B\, dW(t) \\ X(0) &= x \end{aligned} \tag{4.1}$$

where $A\colon D(A) \subset H \to H$ and $B\colon U \to H$ are linear operators, $F\colon H \to H$ is a nonlinear function and W is a cylindrical Wiener process in U, formally defined by

$$W(t) = \sum_{k=1}^{\infty} \beta_k(t) e_k, \ t \geq 0.$$

From now on we assume the following two Hypothesis (cf. Hypothesis 2.1 and 3.1 in [9]):

Hypothesis 4.1. *(i) $A\colon D(A) \subset H \to H$ is the infinitesimal generator of a strongly continuous semigroup $\left(e^{tA}\right)_{t\geq 0}$.*

(ii) $B \in L(U, H)$.

(iii) For any $t > 0$ the linear operator Q_t, defined as

$$Q_t x = \int_0^t e^{sA} C e^{sA^*} x\, ds, \ x \in H,$$

where $C = BB^*$, is of trace class.

By the Hille-Yosida theorem it follows that there exist $M \geq 0$ and $\omega \in \mathbb{R}$ such that

$$\|e^{tA}\|_{L(H)} \leq M e^{\omega t}$$

for all $t \geq 0$.

Hypothesis 4.2. *F is Lipschitz continuous: There exists some constant $K > 0$ such that $\|F(x) - F(y)\| \leq K \cdot \|x - y\|$ for all $x, y \in H$.*

Denote by $C_W([0,T]; H)$ the space of all continuous, $(\mathcal{F}_t)$-adapted mappings $Z\colon [0,T] \to L^2(\Omega, \mathcal{F}, P; H)$. Observe that $C_W([0,T]; H)$ endowed with the norm $\|\cdot\|_{C_W([0,T];H)}$ given by

$$\|Z\|_{C_W([0,T];H)} := \left(\sup_{t \in [0,T]} \mathbb{E}[|Z(t)|^2] \right)^{\frac{1}{2}}$$

is a Banach space. It is called the space of all *mean square continuous adapted processes* on $[0,T]$ taking values in H.

Definition 4.1. *By a* mild solution *of problem (4.1) on $[0,T]$ we mean a stochastic process $X \in C_W([0,T]; H)$ such that*

$$X(t) = e^{tA}x + \int_0^t e^{(t-s)A} F(X(s))\, ds + \int_0^t e^{(t-s)A} B\, dW(s)$$

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

It is well known that there exists a unique mild solution to (4.1), provided that Hypotheses 4.1 and 4.2 hold (cf. Theorem 3.2 in [9]). Moreover, if in addition $F \in \mathcal{C}_b^2(H, H)$, the mild solution $X(t,x)$ of (4.1) is differentiable with respect to the initial condition x P-a.s. and for any $h \in H$ we have $DX(t,x)h = \eta^h(t,x)$ P-a.s. where $\eta^h(t,x)$ is the mild solution of the equation

$$\begin{aligned} \frac{d}{dt}\eta^h(t,x) &= A\eta^h(t,x) + DF(X(t,x))\eta^h(t,x) \\ \eta^h(0,x) &= h, \end{aligned}$$

that is, $\eta^h(t,x)$ is the solution of the integral equation

$$\eta^h(t,x) = e^{tA}h + \int_0^t e^{(t-s)A} DF(X(s,x))\eta^h(s,x)\, ds, \ t \geq 0 \tag{4.2}$$

(cf. Theorem 3.6 in [9]).

Denote by $(\mathcal{P}_t)_{t\geq 0}$ the Markov semigroup corresponding to the SDE in (4.1), e.g. $\mathcal{P}_t\varphi(x) = \mathbb{E}[\varphi(X(t,x))]$ for all $\varphi \in \mathcal{B}_b(H)$, $x \in H$, $t \geq 0$.

Lemma 4.2. *Assume that Hypothesis 4.1 holds for some $\omega < -MK \leq 0$ and let $F \in \mathcal{C}_b^2(H, H)$. Then the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ is asymptotically strong Feller.*

Proof. According to (4.2) we have

$$\|\eta^h(t,x)\| \leq Me^{\omega t}\|h\| + MK \int_0^t e^{\omega(t-s)} \|\eta^h(s,x)\|\, ds,$$

which is equivalent to

$$e^{-\omega t}\|\eta^h(t,x)\| \leq M\|h\| + MK\int_0^t e^{-\omega s}\|\eta^h(s,x)\|\,ds.$$

Hence by the Gronwall lemma

$$e^{-\omega t}\|\eta^h(t,x)\| \leq M\|h\| \cdot e^{MKt},$$

i.e.

$$\|\eta^h(t,x)\| \leq M\|h\| \cdot e^{MKt}e^{\omega t} = Me^{(\omega+MK)t}\|h\|. \tag{4.3}$$

In order to show that the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ is asymptotically strong Feller, by Theorem 3.13 we have to find two positive sequences $(t_n)_{n\in\mathbb{N}}$ and $(\delta_n)_{n\in\mathbb{N}}$, $t_n \leq t_{n+1}$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty}\delta_n = 0$, such that

$$|\mathcal{P}_{t_n}\varphi(x) - \mathcal{P}_{t_n}\varphi(y)| \leq C(\|x\| \vee \|y\|) \cdot (\|\varphi\|_\infty + \delta_n \cdot \|\nabla\varphi\|_\infty) \cdot \|x-y\| \tag{4.4}$$

for all $\varphi \in \mathcal{C}_b^1(H)$, $x, y \in H$, $n \in \mathbb{N}$ and some fixed nondecreasing function $C\colon \mathbb{R}_+ \to \mathbb{R}$.

First of all note that by (4.3) for any $h, z \in H$ we have $\|DX(t,z)h\| = \|\eta^h(t,z)\| \leq Me^{(\omega+MK)t}\|h\|$ P-a.s. for all $t \geq 0$. Hence $\|DX(t,z)\|_{L(H)} \leq Me^{(\omega+MK)t}$ P-a.s. for all $t \geq 0$ and all $z \in H$. Therefore

$$\begin{aligned}
& |\mathcal{P}_t\varphi(x) - \mathcal{P}_t\varphi(y)| \\
\leq\ & \mathbb{E}\left[|\varphi(X(t,x)) - \varphi(X(t,y))|\right] \\
\leq\ & \|\nabla\varphi\|_\infty \cdot \mathbb{E}\left[\|X(t,x) - X(t,y)\|\right] \\
\leq\ & \|\nabla\varphi\|_\infty \cdot \mathbb{E}\left[\sup_{\alpha\in[0,1]}\|DX(t, y+\alpha(x-y))\|_{L(H)}\right] \cdot \|x-y\| \\
\leq\ & \|\nabla\varphi\|_\infty \cdot Me^{(\omega+MK)t} \cdot \|x-y\|
\end{aligned} \tag{4.5}$$

for all $\varphi \in \mathcal{C}_b^1(H)$, $x, y \in H$, $t \geq 0$. Now let $(t_n)_{n\in\mathbb{N}}$ be a positive nondecreasing sequence such that $\lim_{n\to\infty} t_n = \infty$ and define $\delta_n := e^{(\omega+MK)t_n}$ for all $n \in \mathbb{N}$. Since $\omega < -MK \leq 0$, we conclude $\delta_n \overset{n\to\infty}{\longrightarrow} 0$ and inequality (4.4) is valid taking the function C to be constant and equal M. □

Our next aim is to show that the Lemma remains valid when weaken the assumption $F \in \mathcal{C}_b^2(H,H)$, e.g. taking F just Lipschitz continuous. To this purpose set $F^N(x) := F(x) \wedge N$ for all $x \in H$, $N \in \mathbb{N}$ and introduce a regularization F_β^N of F^N by setting, for any $h \in H$,

$$\left\langle F_\beta^N(x), h\right\rangle = \int_H \left\langle F^N\left(e^{\beta S}x + y\right), e^{\beta S}h\right\rangle\ N_{\frac{1}{2}S^{-1}\left(e^{2\beta S}-1\right)}(dy),\ \beta > 0,$$

where $S\colon D(S) \subset H \to H$ is a given self-adjoint, negative definite operator such that S^{-1} is of trace class. Note that the definition of F_β^N corresponds to an Ornstein-Uhlenbeck semigroup $(U_t)_{t\geq 0}$, given by

$$U_t\varphi(x) = \int_H \varphi(e^{tS}x + y)\ N_{\frac{1}{2}S^{-1}(e^{2tS}-1)}(dy)$$

for all $\varphi \in \mathcal{B}_b(H)$, $x \in H$, $t > 0$. Since in particular $\mathrm{Tr}[\frac{1}{2}S^{-1}(e^{2tS}-1)] < \infty$ (cf. Theorem VI.19 in [13]), Hypothesis 4.1 is fulfilled and so by Proposition 2.17(iv) in [9] we know that the mapping $[0,T] \times H \to \mathbb{R}$, $(t,x) \mapsto U_t\varphi(x)$ is continuous for all $\varphi \in \mathcal{C}_b(H)$. Furthermore, according to Remark 2.25 in [9] we have $e^{tS}(H) \subset (\frac{1}{2}S^{-1}(e^{2tS}-1))^{\frac{1}{2}}(H)$ for all $t > 0$. Hence in view of Proposition 2.28 in [9], $U_t\varphi \in \mathcal{C}_b^\infty(H)$ for all $\varphi \in \mathcal{B}_b(H)$.

Concerning the properties of F_β^N this means

$$\lim_{\beta \to 0} F_\beta^N(x) = F^N(x)$$

for all $x \in H$ and

$$F_\beta^N \in \mathcal{C}_b^\infty(H,H) \tag{4.6}$$

for all $N \in \mathbb{N}$, $\beta > 0$.

Furthermore we have

Lemma 4.3. *F_β^N is Lipschitz continuous with Lipschitz constant $K(N,\beta) \le K$.*

Proof. Let $x, y \in H$ and $\beta > 0$. Then we have

$$\begin{aligned}
& \left\|F_\beta^N(x) - F_\beta^N(y)\right\|^2 \\
= \ & \left|\langle F_\beta^N(x) - F_\beta^N(y), F_\beta^N(x) - F_\beta^N(y)\rangle\right| \\
\le \ & \int_H \underbrace{\left|\langle F^N\left(e^{\beta S}x+z\right) - F^N\left(e^{\beta S}y+z\right), e^{\beta S}\left(F_\beta^N(x) - F_\beta^N(y)\right)\rangle\right|}_{\le \left\|F^N\left(e^{\beta S}x+z\right)-F^N\left(e^{\beta S}y+z\right)\right\| \cdot \left\|e^{\beta S}\right\|_{L(H)} \left\|F_\beta^N(x)-F_\beta^N(y)\right\|} N_{\frac{1}{2}S^{-1}\left(e^{2\beta S}-1\right)}(dz) \\
\le \ & K \left\|e^{\beta S}\right\|_{L(H)}^2 \cdot \|x-y\| \cdot \left\|F_\beta^N(x) - F_\beta^N(y)\right\|
\end{aligned}$$

according to the Lipschitz continuity of F^N. Dividing both sides by $\left\|F_\beta^N(x) - F_\beta^N(y)\right\|$ yields the first assertion. Moreover, since S is a self-adjoint, negative definite operator, from [8] we know that $\|e^{\beta S}\|_{L(H)} \le 1$. Hence $K(N,\beta) \le K$. □

Now let $N \in \mathbb{N}$, $\beta > 0$. Similarly as above one can show that the problem

$$\begin{aligned}
dX_\beta^N(t) &= \left(AX_\beta^N(t) + F_\beta^N(X_\beta^N(t))\right) dt + B\,dW(t) \\
X_\beta^N(0) &= x
\end{aligned}$$

has a unique mild solution $X_\beta^N(t,x)$. Moreover, it is not difficult to check that

$$\lim_{N\to\infty} \lim_{\beta \to 0} X_\beta^N(t,x) = X(t,x) \tag{4.7}$$

for all $x \in H$, $t > 0$.

Defining

$$\mathcal{P}_t^{N,\beta}\varphi(x) := \mathbb{E}[\varphi(X_\beta^N(t,x))]$$

for all $\varphi \in \mathcal{B}_b(H)$, according to (4.7) and Lebesgue we receive

$$\lim_{N\to\infty} \lim_{\beta\to 0} \mathcal{P}_t^{N,\beta}\varphi(x) = \lim_{N\to\infty} \lim_{\beta\to 0} \mathbb{E}[\varphi(X_\beta^N(t,x))] = \mathbb{E}[\varphi(X(t,x))] = \mathcal{P}_t\varphi(x) \tag{4.8}$$

for all $\varphi \in \mathcal{C}_b(H)$, $x \in H$, $t > 0$.

By the help of these preparations we are able to prove the final result of this chapter:

Theorem 4.4. *Assume that Hypothesis 4.1 holds for some $\omega < -MK \leq 0$ and Hypothesis 4.2 is fulfilled. Then the Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ is asymptotically strong Feller.*

Proof. Let $N \in \mathbb{N}$, $\beta > 0$. Since according to (4.6) $F_\beta^N \in \mathcal{C}_b^2(H, H)$, by Lemma 4.2 the assertion follows for $(\mathcal{P}_t^{N,\beta})_{t\geq 0}$. In particular, observe that the sequence $(t_n)_{n\in\mathbb{N}}$ in (4.5) (and consequently the sequence $(\delta_n)_{n\in\mathbb{N}}$) can be chosen independently of $N \in \mathbb{N}$ and $\beta > 0$. Since $K(N, \beta) \leq K$, we achieve

$$|\mathcal{P}_{t_n}^{N,\beta}\varphi(x) - \mathcal{P}_{t_n}^{N,\beta}\varphi(y)| \leq \|\nabla\varphi\|_\infty \cdot M \cdot \delta_n \cdot \|x - y\| \tag{4.9}$$

for all $\varphi \in \mathcal{C}_b^1(H)$, $x, y \in H$, $n \in \mathbb{N}$ with δ_n given as in the proof of Lemma 4.2. Since the right hand side in (4.9) is independent of $N \in \mathbb{N}$ and $\beta > 0$, letting first $\beta \to 0$ and then $N \to \infty$ the assertion for $(\mathcal{P}_t)_{t\geq 0}$ follows immediately from (4.8). □

Remark 4.5. *Observe that we can dispense with the assumption that B is continuously invertible, which is needed in [9] in order to show that $(\mathcal{P}_t)_{t\geq 0}$ is strong Feller (cf. Theorem 3.11). In particular, we do not apply the Bismut-Elworthy formula. Unfortunately we have to take $\omega < -MK \leq 0$.*

A Approximation via convolution

Let $\eta \in \mathcal{C}_0^\infty(\mathbb{R}^m)$ such that $\eta \geq 0$ and $\int_{\mathbb{R}^m} \eta(x)\, dx = 1$. Define a sequence $(\eta_\varepsilon)_{\varepsilon>0}$ by

$$\eta_\varepsilon(x) := \frac{1}{\varepsilon^m} \cdot \eta\left(\frac{x}{\varepsilon^m}\right).$$

Note that via the substitution $y := \frac{x}{\varepsilon^m}$ we have

$$\int_{\mathbb{R}^m} \eta_\varepsilon(x)\, dx = \int_{\mathbb{R}^m} \frac{1}{\varepsilon^m} \cdot \eta\left(\frac{x}{\varepsilon^m}\right)\, dx = \int_{\mathbb{R}^m} \eta(y)\, dy = 1$$

and moreover for every $\rho > 0$

$$\int_{\mathbb{R}^m \setminus B_\rho(0)} \eta_\varepsilon(x)\, dx \overset{\varepsilon \to 0}{\longrightarrow} 0.$$

In particular, for $\varepsilon = \frac{1}{n}$ we obtain with $\delta_n := \eta_{\frac{1}{n}}$

$$\int_{\mathbb{R}^m} \delta_n(x)\, dx = 1$$

and

$$\int_{\mathbb{R}^m \setminus B_\rho(0)} \delta_n(x)\, dx \overset{n \to \infty}{\longrightarrow} 0 \tag{A.1}$$

for all $\rho > 0$.

Let $f \in L^p(\mathbb{R}^m)$. The *convolution* of f and δ_n is defined by

$$\left(f \star \delta_n\right)(x) := \int_{\mathbb{R}^m} f(z)\delta_n(x-z)\, dz = \int_{\mathbb{R}^m} f(x-z)\delta_n(z)\, dz.$$

Now we are able to prove the following

Lemma A.1. *Let $f \in \mathcal{L}_d(\mathbb{R}^m)$. Then there exists a sequence $\left(f_n\right)_{n\in\mathbb{N}}$, $f_n \in \mathcal{C}_b^\infty(\mathbb{R}^m)$, such that*

(i) $f_n \overset{n\to\infty}{\longrightarrow} f$ *pointwisely,*

(ii) $\|f_n\|_\infty \leq \|f\|_\infty$ *for each* $n \in \mathbb{N}$,

(iii) $\|f_n\|_d \leq \|f\|_d$ *for each* $n \in \mathbb{N}$.

Proof. Let $f \in \mathcal{L}_d(\mathbb{R}^m)$ and define functions $f_n \colon \mathbb{R}^m \to \mathbb{R}$ by $f_n(x) := (f \star \delta_n)(x)$.

Let $x \in \mathbb{R}^m$. Observe that using the substitution $z := x - v$ we have

$$\begin{aligned}
(f \star \delta_n)(x) - f(x) &= \int_{\mathbb{R}^m} f(z)\delta_n(x-z)\,dz - f(x) \\
&= \int_{\mathbb{R}^m} f(z)\delta_n(x-z)\,dz - f(x) \cdot \int_{\mathbb{R}^m} \delta_n(x-z)\,dz \\
&= \int_{\mathbb{R}^m} (f(z) - f(x))\,\delta_n(x-z)\,dz \\
&= \int_{\mathbb{R}^m} -(f(x-v) - f(x))\,\delta_n(v)\,dv \\
&= \int_{\mathbb{R}^m} (f(x) - f(x-v))\,\delta_n(v)\,dv.
\end{aligned}$$

Hence

$$\begin{aligned}
& |f_n(x) - f(x)| \\
= {} & |(f \star \delta_n)(x) - f(x)| \\
\leq {} & \left| \int_{B_\rho(0)} (f(x) - f(x-v)) \cdot \delta_n(v)\,dv \right| + \left| \int_{\mathbb{R}^m \setminus B_\rho(0)} (f(x) - f(x-v)) \cdot \delta_n(v)\,dv \right| \\
\leq {} & \underbrace{\left(\int_{B_\rho(0)} \delta_n(v)\,dv \right)}_{\leq 1} \cdot \sup_{\|v\| \leq \rho} |f(x) - f(x-v)| \\
& + \left(\int_{\mathbb{R}^m \setminus B_\rho(0)} \delta_n(v)\,dv \right) \cdot \underbrace{\sup_{v \in \mathbb{R}^m} |f(x) - f(x-v)|}_{\leq 2 \cdot \|f\|_\infty < \infty}
\end{aligned}$$

for all $\rho > 0$. Therefore, letting first $n \to \infty$ and then $\rho \to 0$ yields

$$\lim_{n \to \infty} f_n(x) = f(x)$$

for all $x \in \mathbb{R}^m$ according to (A.1) and the continuity of f, i.e. f_n converges pointwisely to f as n tends to infinity.

Furthermore, since

$$\begin{aligned}
|f_n(x)| &= \left| \int_{\mathbb{R}^m} f(z)\delta_n(x-z)\,dz \right| \\
&\leq \int_{\mathbb{R}^m} |f(z)|\delta_n(x-z)\,dz \\
&\leq \|f\|_\infty \cdot \int_{\mathbb{R}^m} \delta_n(x-z)\,dz \\
&= \|f\|_\infty
\end{aligned}$$

for all $x \in \mathbb{R}^m$, $n \in \mathbb{N}$, we have $\|f_n\|_\infty \le \|f\|_\infty$ for all $n \in \mathbb{N}$ and this is (ii).

Similarly,

$$\begin{aligned} |f_n(x) - f_n(y)| &\le \int_{\mathbb{R}^m} |f(x-z) - f(y-z)| \cdot \delta_n(z)\, dz \\ &\le \|f\|_d \cdot \int_{\mathbb{R}^m} \delta_n(z)\, dz \cdot \|x-y\| \\ &= \|f\|_d \cdot \|x-y\|. \end{aligned}$$

Hence $f_n \in \mathcal{L}_d(\mathbb{R}^m)$ and $\|f_n\|_d \le \|f\|_d$ for all $n \in \mathbb{N}$.

It remains to show $f_n \in \mathcal{C}_b^\infty(\mathbb{R}^m)$ for all $n \in \mathbb{N}$. Observe that $\delta_n \in \mathcal{C}_0^\infty(\mathbb{R}^m)$ and for $v \in \mathbb{R}^m$ we have

$$\begin{aligned} & \frac{\partial}{\partial v} (f \star \delta_n)(x_0) \\ = & \lim_{h \to 0} \frac{1}{h} \cdot ((f \star \delta_n)(x_0 + hv) - (f \star \delta_n)(x_0)) \\ = & \lim_{h \to 0} \frac{1}{h} \cdot \int_{\mathbb{R}^m} (\delta_n(x_0 + hv - z) - \delta_n(x_0 - z)) \cdot f(z)\, dz \\ = & \int_{\mathbb{R}^m} \lim_{h \to 0} \frac{1}{h} \cdot (\delta_n((x_0 - z) + hv) - \delta_n(x_0 - z)) \cdot f(z)\, dz \\ = & \int_{\mathbb{R}^m} \frac{\partial}{\partial v} \delta_n(x_0 - z) \cdot f(z)\, dz \\ = & \left(f \star \left(\frac{\partial}{\partial v} \delta_n \right) \right)(x_0). \end{aligned}$$

By iterating this argument the assertion follows. □

In the same way one can prove:

Lemma A.2. *Let $f \in \mathcal{C}_b(\mathbb{R}^m)$. Then there exists a sequence $(f_n)_{n \in \mathbb{N}}$, $f_n \in \mathcal{C}_b^2(\mathbb{R}^m)$, such that $\|f_n\|_\infty \le \|f\|_\infty$ for all $n \in \mathbb{N}$ and $f_n \longrightarrow f$ pointwisely as $n \to \infty$.*

B Regularization of Markovian kernels by composition

The following results are taken from chapter IX.1 in [3].

Let $(E, \mathcal{B}(E))$, $(F, \mathcal{B}(F))$ and $(G, \mathcal{B}(G))$ be measurable spaces, $M\colon \mathcal{B}_b(F) \to \mathcal{B}_b(E)$, $N\colon \mathcal{B}_b(G) \to \mathcal{B}_b(F)$ Markov operators and π^M, π^N the corresponding Markovian kernels from $(E, \mathcal{B}(E))$ to $(F, \mathcal{B}(F))$ respectively from $(F, \mathcal{B}(F))$ to $(G, \mathcal{B}(G))$. It is clear, that P defined by $P := MN$ is a Markov operator from $\mathcal{B}_b(G)$ to $\mathcal{B}_b(E)$ and $\pi^P := \pi^N \pi^M$ the corresponding Markovian kernel from $(E, \mathcal{B}(E))$ to $(G, \mathcal{B}_b(G))$.

Definition B.1. *A kernel κ from $(E, \mathcal{B}(E))$ to $(F, \mathcal{B}(F))$ is called* basic *if there exists some Borel probability measure μ on F such that $\kappa(x, \cdot) \ll \mu(\cdot)$ for all $x \in E$. In this case the Borel probability measure μ is called the* base.

Theorem B.2. *Let E be metrizable, $M\colon \mathcal{B}_b(F) \to \mathcal{B}_b(E)$ strong Feller and π^N basic. Then the kernel π^P from $(E, \mathcal{B}(E))$ to $(G, \mathcal{B}_b(G))$ is continuous in x in the total variation norm, that is $\|\pi^P(x_n, \cdot) - \pi^P(x, \cdot)\|_{TV} \overset{n\to\infty}{\longrightarrow} 0$ for every sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \overset{n\to\infty}{\longrightarrow} x$. Moreover, if E is compact, P is compact as well.*

Proof. It is sufficient to check that if (E, d_E) is a compact metric space, then $P = MN$ is compact and its transition probabilities $\pi^P(x, \cdot)$ are continuous in x in the total variation norm, cause then the assertion follows for every compact subset $K \subseteq E$. Applying this to the compact subset $K = \{x_n \mid n \in \mathbb{N}\} \cup \{x\}$ formed by the convergent sequence $(x_n)_{n\in\mathbb{N}}$ with limit x, the assertion follows for E.

Let $U := N(B_{\mathcal{B}_b(G)}(0,1)) \subseteq \mathcal{B}_b(F)$, where $B_{\mathcal{B}_b(G)}(0,1) := \{g \in \mathcal{B}_b(G) \mid \|g\|_\infty \leq 1\}$ denotes the closed unit ball in $\mathcal{B}_b(G)$, and $V := M(U) \subseteq \mathcal{B}_b(E)$. Hence $V = MN(B_{\mathcal{B}_b(G)}(0,1)) = P(B_{\mathcal{B}_b(G)}(0,1))$. In order to prove that $P\colon \mathcal{B}_b(G) \to \mathcal{B}_b(E)$ is compact, we have to show that $V \subseteq \mathcal{B}_b(E)$ is relatively compact for the topology of uniform convergence on E.

So, let $(e_n)_{n\in\mathbb{N}} \subseteq V$. By definition of V there exist $f_n \in U$ such that $e_n = Mf_n$ for all $n \in \mathbb{N}$. According to Remark B.4 below there exist a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ and $f \in U$ such that $f_{n_k}(y) \overset{k\to\infty}{\longrightarrow} f(y)$ for all $y \in F$. Now it suffices to show that $(e_{n_k})_{k\in\mathbb{N}}$ converges uniformly to $e = Mf$, that is $\|e_{n_k} - e\|_\infty \overset{k\to\infty}{\longrightarrow} 0$. Without loss of generality take $f = 0$, e.g. $f(y) = 0$ for all $y \in F$. Hence $Mf = 0$.

Define $h_k(y) := \sup_{l\geq k} |f_{n_l}(y)|$ for all $y \in F$. Obviously $(h_k(y))_{k\in\mathbb{N}}$ is monotonically decreasing and converges to $f(y) = 0$ for all $y \in F$, because $f_{n_l}(y) \to f(y) = 0$ for all $y \in F$ as l tends to infinity. Since $h_k \in \mathcal{B}_b(F)$ for all $k \in \mathbb{N}$ and $M\colon \mathcal{B}_b(F) \to \mathcal{B}_b(E)$ is strong Feller by assumption, we have $Mh_k \in \mathcal{C}_b(E)$ for all $k \in \mathbb{N}$. Moreover the sequence

$(Mh_k(x))_{k\in\mathbb{N}}$ is monotonically decreasing for all $x \in E$, because

$$Mh_{k+1}(x) = \int_F h_{k+1}(y)\, \pi^M(x,dy) \le \int_F h_k(y)\, \pi^M(x,dy) = Mh_k(x)$$

for all $k \in \mathbb{N}$ and all $x \in E$. Since $Mh_k(x) \ge 0$ for all $k \in \mathbb{N}$, $x \in E$ and $(Mh_k(x))_{k\in\mathbb{N}}$ is monotonically decreasing for all $x \in E$, we have: $\inf_{k\in\mathbb{N}} Mh_k(x)$ exists for all $x \in E$ and $\inf_{k\in\mathbb{N}} Mh_k(x) = \lim_{k\to\infty} Mh_k(x) = Mf(x) = 0$ for all $x \in E$. Furthermore the function $Mf = 0$ is continuous.

Therefore all assumptions of Dini's theorem are satisfied and so we get that $(Mh_k)_{k\in\mathbb{N}}$ converges uniformly to $Mf = 0$, i.e. $\|Mh_k - 0\|_\infty \overset{k\to\infty}{\longrightarrow} 0$. But

$$\begin{aligned} |Mf_{n_k}(x)| &= \left| \int_F f_{n_k}(y)\, \pi^M(x,dy) \right| \\ &\le \int_F |f_{n_k}(y)|\, \pi^M(x,dy) \\ &\le \int_F h_k(y)\, \pi^M(x,dy) \\ &= Mh_k(x) \end{aligned}$$

for all $k \in \mathbb{N}$ and all $x \in E$, thus $\|Mf_{n_k}\|_\infty \le \|Mh_k\|_\infty$ for all $k \in \mathbb{N}$. So, $\|e_{n_k} - 0\|_\infty = \|Mf_{n_k} - 0\|_\infty \overset{k\to\infty}{\longrightarrow} 0$. Hence V is relatively compact and therefore $P\colon \mathcal{B}_b(G) \to \mathcal{B}_b(E)$ is a compact operator.

Since $V \subseteq \mathcal{B}_b(E)$ is relatively compact, we further conclude by Arzela-Ascoli's theorem that V is equicontinuous, i.e.

$$\begin{gathered} \forall\, \varepsilon > 0 \;\; \forall\, x \in E \;\; \exists\, \delta = \delta(x,\varepsilon) > 0 : \;\; \forall\, x' \in E : \\ d_E(x,x') < \delta \implies \sup_{e\in V} |e(x) - e(x')| \le \varepsilon. \end{gathered} \tag{B.1}$$

But by the definition of V we can choose some function $g \in B_{\mathcal{B}_b(G)}(0,1)$ such that $e = Pg$. Therefore (B.1) could be written as

$$\begin{gathered} \forall\, \varepsilon > 0 \;\; \forall\, x \in E \;\; \exists\, \delta = \delta(x,\varepsilon) > 0 : \;\; \forall\, x' \in E : \\ d_E(x,x') < \delta \implies \sup_{g\in B_{\mathcal{B}_b(G)}(0,1)} |Pg(x) - Pg(x')| \le \varepsilon. \end{gathered}$$

In particular, since $g = \mathbf{1}_A \in B_{\mathcal{B}_b(G)}(0,1)$ for arbitrary $A \in \mathcal{B}(G)$, we obtain

$$|P\mathbf{1}_A(x) - P\mathbf{1}_A(x')| \le \varepsilon.$$

Therefore

$$\|\pi^P(x,\cdot) - \pi^P(x',\cdot)\|_{TV} = \sup_{A\in\mathcal{B}(G)} |\pi^P(x,A) - \pi^P(x',A)| \le \varepsilon. \tag{B.2}$$

□

To complete the proof of the above theorem, it remains to prove the existence of some subsequence $(f_{n_k})_{k\in\mathbb{N}}$.

Lemma B.3. *Suppose π^N is basic. Then $U := N(B_{\mathcal{B}_b(G)}(0,1))$ is compact for the topology of pointwise convergence. Moreover, if in addition the σ-algebra $\mathcal{B}(G)$ is separable, U is compact metrizable.*

Proof. Since π^N is basic, there exists some probability measure μ on $(G,\mathcal{B}(G))$ (the base) such that $\pi^N(y,\cdot) \ll \mu(\cdot)$ for all $y \in F$.

First of all, N defines a bounded operator from $L^\infty(G,\mu)$ to $\mathcal{B}_b(F)$. In fact, let $g, \tilde{g} \in \mathcal{B}_b(G)$ with $g(z) = \tilde{g}(z)$ for μ-a.e. $z \in G$. Since $\pi^N(y,\cdot) \ll \mu(\cdot)$ for all $y \in F$, we have $g(z) = \tilde{g}(z)$ for $\pi^N(y,\cdot)$-a.e. $z \in G$ for all $y \in F$. Hence

$$Ng(y) = \int_G g(z)\,\pi^N(y,dz) = \int_G \tilde{g}(z)\,\pi^N(y,dz) = N\tilde{g}(y)$$

for all $y \in F$. Therefore Ng only depends on the equivalence class $[g] \in L^\infty(G,\mu)$ of g. Furthermore, since $g \in \mathcal{B}_b(G)$ we gain

$$|Ng(y)| \le \int_G |g(z)|\,\pi^N(y,dz) \le \|g\|_\infty \cdot \pi^N(y,G) = \|g\|_\infty$$

for all $y \in F$ and thus $\|Ng\|_\infty \le \|g\|_\infty$. So, $\|N\|_{L(L^\infty(G,\mu),\mathcal{B}_b(F))} \le 1$, e.g. N is a bounded operator from $L^\infty(G,\mu)$ to $\mathcal{B}_b(F)$ (even a contraction).

Denote by B the unit ball in $L^\infty(G,\mu)$. Since $\pi^N(y,\cdot) \ll \mu(\cdot)$ for all $y \in F$, by the Radon-Nikodym theorem there exists some density $\rho_y \in L^1(G,\mu)$, $\rho_y \ge 0$, such that $\int_A \rho_y(z)\ \mu(dz) = \pi^N(y,A)$ for all $A \in \mathcal{B}(G)$, $y \in F$. Hence $N\colon L^\infty(G,\mu) \to \mathcal{B}_b(F)$ is continuous with respect to the weak topology $\sigma(L^\infty, L^1)$ on $L^\infty(G,\mu)$ and the topology of pointwise convergence on $\mathcal{B}_b(F)$. In fact, for $g_n \overset{n\to\infty}{\longrightarrow} g$ with respect to $\sigma(L^\infty, L^1)$, that is $\int_G g_n(z)h(z)\ \mu(dz) \overset{n\to\infty}{\longrightarrow} \int_G g(z)h(z)\ \mu(dz)$ for all $h \in L^1(G,\mu)$, we receive

$$\begin{aligned}
\lim_{n\to\infty} Ng_n(y) &= \lim_{n\to\infty} \int_G g_n(z)\,\pi^N(y,dz) \\
&= \lim_{n\to\infty} \int_G g_n(z)\rho_y(z)\ \mu(dz) \\
&= \int_G g(z)\rho_y(z)\ \mu(dz) \\
&= \int_G g(z)\,\pi^N(y,dz) \\
&= Ng(y)
\end{aligned}$$

for all $y \in F$.

By the Banach-Alaoglu theorem (cf. [1], p. 215) the ball B is compact for the topology $\sigma(L^\infty, L^1)$. Since N is continuous with respect to the topologies mentioned above, $U = N(B)$ is compact for the topology of pointwise convergence.

Moreover, if $\mathcal{B}(G)$ is separable, B is metrizable and so is U. In fact, if $\mathcal{B}(G)$ is separable, $L^1(G,\mu)$ is separable and so there exists a countable dense subset $\{h_i \mid i \in \mathbb{N}\}$. Without loss of generality assume $h_i \neq 0$ for all $i \in \mathbb{N}$ and set $\tilde{h}_i := \frac{h_i}{\int_G |h_i| \, d\mu}$. Consider the map

$$\begin{aligned} T\colon B \subseteq L^\infty(G,\mu) &\to [-1,+1]^{\mathbb{N}} \\ g &\mapsto \left(\int_G g(z)\tilde{h}_i(z)\, \mu(dz) \right)_{i\in\mathbb{N}} . \end{aligned}$$

Claim 7. *T is injective.*

Proof. Let $g, \tilde{g} \in B$ such that $T(g) = T(\tilde{g})$. Then

$$\int_G (g(z) - \tilde{g}(z))\, \tilde{h}_i(z)\, \mu(dz) = 0$$

for all $i \in \mathbb{N}$. Multiplying both sides of the above equality with $\int_G |h_i| \, d\mu$ yields

$$\int_G (g(z) - \tilde{g}(z))\, h_i(z)\, \mu(dz) = 0$$

for all $i \in \mathbb{N}$. Since $\{h_i \mid i \in \mathbb{N}\} \subseteq L^1(G,\mu)$ densely,

$$\int_G (g(z) - \tilde{g}(z))\, h(z)\, \mu(dz) = 0$$

for all $h \in L^1(G,\mu)$. Therefore $g(z) = \tilde{g}(z)$ for μ-a.e. $z \in G$, e.g. g and $\tilde{g}$ form the same equivalence class in B with respect to μ. □

Claim 8. *T is continuous with respect to the topology $\sigma(L^\infty, L^1)$ on L^∞ and the product topology on $[-1,+1]^{\mathbb{N}}$.*

Proof. First observe that it would be enough to show the assertion for each coordinate. But this is clear according to the definition of convergence with respect to the topology $\sigma(L^\infty, L^1)$. □

Combining both claims we receive that $T\colon B \to T(B) \subseteq [-1,+1]^{\mathbb{N}}$ is homeomorphic. Therefore, since $[-1,+1]^{\mathbb{N}}$ is a metric space, $T(B)$ is a metric space and thus also B. Furthermore the same is true for its image U under the continuous map N. (Note that $N\colon B \to U$ is bijective.) □

Remark B.4. *Let $(f_n)_{n\in\mathbb{N}} \subseteq U$. Then there exists a sequence $(g_n)_{n\in\mathbb{N}}$, $g_n \in B_{\mathcal{B}_b(G)}(0,1)$, such that $f_n = Ng_n$ for all $n \in \mathbb{N}$. Consider the probability measures $\pi^N(y,\cdot)$, $y \in F$, only on $\sigma(g_n \mid n \in \mathbb{N}) \subseteq \mathcal{B}(G)$. Since π^N is basic and $\sigma(g_n \mid n \in \mathbb{N})$ is separable, according to Lemma B.3 $\{f_n \mid n \in \mathbb{N}\} = \{Ng_n \mid n \in \mathbb{N}\} \subseteq U$ is compact for the topology of pointwise convergence and metrizable. Therefore we can find a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ of $(f_n)_{n\in\mathbb{N}}$ and some $f \in U$ such that $f_{n_k}(y) \overset{k\to\infty}{\longrightarrow} f(y)$ for all $y \in F$.*

For the application in chapter 1 we further need the following relation between strong Feller and basic Markov operators:

Lemma B.5. *Let $\mathcal{X}$ be separable and let N be a strong Feller Markov operator on $\mathcal{B}_b(\mathcal{X})$. Then π^N is basic.*

Proof. Since $\mathcal{X}$ is separable, there exists a countable dense subset $\{x_k \mid k \in \mathbb{N}\} \subseteq \mathcal{X}$. It would be enough to show

$$\pi^N(x, \cdot) \ll \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot \pi^N(x_k, \cdot) =: \mu(\cdot) \tag{B.3}$$

for all $x \in \mathcal{X}$.

So, fix $x \in \mathcal{X}$ and let $A \in \mathcal{B}(\mathcal{X})$ such that $\mu(A) = \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot \pi^N(x_k, A) = 0$. Hence $N\mathbf{1}_A(x_k) = \pi^N(x_k, A) = 0$ for all $k \in \mathbb{N}$. Since $\mathbf{1}_A \in \mathcal{B}_b(\mathcal{X})$ and N is strong Feller, we get $N\mathbf{1}_A \in \mathcal{C}_b(\mathcal{X})$. Since $x \in \mathcal{X}$ and $\{x_k \mid k \in \mathbb{N}\} \subseteq \mathcal{X}$ is dense, there exists a subsequence $(x_{k_l})_{l \in \mathbb{N}}$ such that $x_{k_l} \overset{l\to\infty}{\longrightarrow} x$. Therefore we conclude

$$\pi^N(x, A) = N\mathbf{1}_A(x) = \lim_{l\to\infty} N\mathbf{1}_A(x_{k_l}) = 0,$$

i.e. $\pi^N(x, \cdot) \ll \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot \pi^N(x_k, \cdot) = \mu(\cdot)$. □

C Probability measures on Polish spaces

The following result is well-known (cf. Theorem 1.4 in [2]):

Lemma C.1. *Let $(\mathcal{X}, d)$ be a complete and separable metric space. Then any probability measure P on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is tight.*

Proof. Since $\mathcal{X}$ is separable there exists a countable dense subset $\{x_i \mid i \in \mathbb{N}\}$. For any $x \in \mathcal{X}$ and $\delta > 0$ denote by $\overline{B(x,\delta)} := \{y \in \mathcal{X} \mid d(x,y) \leq \delta\}$ the closed ball with center x and radius $\delta > 0$.

Let $\varepsilon > 0$. For each $n \in \mathbb{N}$ choose $i_n \in \mathbb{N}$ such that $P(\mathcal{X} \setminus (\bigcup_{i \leq i_n} \overline{B(x_i, \frac{1}{n})})) < \frac{\varepsilon}{2^n}$. Define $K := \bigcap_{n\in\mathbb{N}} \bigcup_{i \leq i_n} \overline{B(x_i, \frac{1}{n})}$. It is clear that K is closed and totally bounded. In fact, for given $\varepsilon > 0$ there exists $n(\varepsilon) \in \mathbb{N}$ such that $\frac{1}{n(\varepsilon)} < \varepsilon$. Hence $K \subseteq \bigcup_{i \leq i_{n(\varepsilon)}} \overline{B(x_i, \varepsilon)}$. Since $\mathcal{X}$ is complete and $K \subseteq \mathcal{X}$ is closed, K is complete. Therefore by Theorem 2.3.1 in [5] K is compact. But

$$P(K^c) = P\left(\bigcup_{n\in\mathbb{N}}\left(\bigcup_{i\leq i_n}\overline{B\left(x_i, \frac{1}{n}\right)}\right)^c\right) \leq \sum_{n=1}^{\infty} P\left(\left(\bigcup_{i\leq i_n}\overline{B\left(x_i, \frac{1}{n}\right)}\right)^c\right) < \sum_{n=1}^{\infty}\frac{\varepsilon}{2^n} = \varepsilon$$

and so the assertion follows. □

Remark C.2. *While the assumption of completeness could be weakened to topological completeness, e.g. there exists an equivalent metric $\tilde{d}$ on $\mathcal{X}$ such that $\mathcal{X}$ is complete with respect to $\tilde{d}$, the separability could be replaced by the reqirement that P has separable support (cf. [2], p. 234).*

Bibliography

[1] H. W. Alt. *Lineare Funktionalanalysis.* Springer, 1999.

[2] P. Billingsley. *Convergence of Probability Measures.* Wiley, 1968.

[3] C. Dellacherie and P.-A. Meyer. *Probabilities and Potential C.* North-Holland, 1988.

[4] J.-D. Deuschel and D. Stroock. *Large Deviations.* Academic Press, 1989.

[5] R. M. Dudley. *Real Analysis and probability.* Chapman and Hall, 1989.

[6] M. Hairer. Ergodic theory for Stochastic PDEs. Unpublished lecture notes. 2008. http://www.hairer.org/notes/Imperial.pdf.

[7] M. Hairer and J. C. Mattingly. Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing. *Annals of Mathematics*, 2007.

[8] Z.-M. Ma and M. Röckner. *Introduction to the Theory of (Non-Symmetric) Dirichlet Forms.* Springer, 1992.

[9] G. Da Prato. *Kolmogorov Equations for Stochastic PDEs.* Birkhäuser, 2004.

[10] G. Da Prato. *An Introduction to Infinite-Dimensional Analysis.* Springer, 2006.

[11] G. Da Prato and J. Zabczyk. *Ergodicity for Infinite Dimensional Systems.* Cambridge University Press, 1996.

[12] S. T. Rachev. *Probability Metrics and the Stability of Stochastic Models.* Wiley, 1991.

[13] M. Reed and B. Simon. *Methods of Modern Mathematical Physics. Volume I Functional Analysis.* Academic Press, 1980.

[14] D. Werner. *Funktionalanalysis.* Springer, 2000.