

WASSERSTEIN PERTURBATIONS OF MARKOVIAN TRANSITION SEMIGROUPS

SVEN FUHRMANN, MICHAEL KUPPER, AND MAX NENDEL

ABSTRACT. In this paper, we deal with a class of time-homogeneous continuous-time Markov processes with transition probabilities bearing a nonparametric uncertainty. The uncertainty is modelled by considering perturbations of the transition probabilities within a proximity in Wasserstein distance. As a limit over progressively finer time periods, on which the level of uncertainty scales proportionally, we obtain a convex semigroup satisfying a nonlinear PDE in a viscosity sense. A remarkable observation is that, in standard situations, the nonlinear transition operators arising from nonparametric uncertainty coincide with the ones related to parametric drift uncertainty. On the level of the generator, the uncertainty is reflected as an additive perturbation in terms of a convex functional of first order derivatives. We additionally provide sensitivity bounds for the convex semigroup relative to the reference model. The results are illustrated with Wasserstein perturbations of Lévy processes, infinite-dimensional Ornstein-Uhlenbeck processes, geometric Brownian motions, and Koopman semigroups.

Key words: Markov process, Wasserstein distance, nonparametric uncertainty, convex semigroup, nonlinear PDE, viscosity solution

AMS 2020 Subject Classification: Primary 60J35; 47H20; Secondary 60G65; 90C31; 62G35

1. INTRODUCTION

When considering stochastic processes for the modeling of real world phenomena, a major issue is so-called *model uncertainty* or *epistemic uncertainty*. The latter refer to the impossibility of perfectly capturing information about the future in a single stochastic framework. The relevance of model uncertainty, in a stochastic and nonstochastic setting, is widely recognised within many fields such as statistics, operations research, finance, and economic theory, and one typically differentiates between parametric and nonparametric uncertainty.

Parametric uncertainty relates to the lack of information regarding certain parameters of a model while taking other model-specific assumptions as given. In a dynamic setting, the construction of consistent families of nonlinear transition semigroups related to parameter uncertainty has recently received a lot of attention, cf. Coquet et al. [12], Denk et al. [15], Fadina et al. [19], Hu and Peng [23], Kühn [26], Neufeld and Nutz [29], and Peng [32]. Nonlinear semigroups are intimately related to BSDEs and stochastic optimal control, cf. Cheridito et al. [11], El Karoui et al. [18], Kazi-Tani et al. [24], and Soner et al. [38].

Date: March 8, 2022.

We thank Daniel Bartl, Jonas Blessing, and Stephan Eckstein for valuable comments and discussions, as well as two anonymous referees for their helpful suggestions and remarks. Financial support through the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1283/2 2021 – 317210226 is gratefully acknowledged.

On the other hand, nonparametric uncertainty refers to the impossibility of precisely capturing certain aspects of a model, such as independence or, more generally, the joint distributions of certain random variables – phenomena that are in general not implementable using a finite number of parameters. One way to tackle such situations is to consider perturbations of a reference model using a certain notion of proximity, e.g., in terms of Wasserstein distances, cf. Bartl et al. [5], Blanchet and Murthy [8], Mohajerin Esfahani and Kuhn [27], Gao and Kleywegt [21], Pflug and Wozabal [33], Zhao and Guan [43], as well as Bartl et al. [6] for a dynamic setting.

The starting point of the present article is the work of Bartl et al. [6], where perturbations of certain Lévy processes in \mathbb{R}^d were considered. Our focus is on more general classes of stochastic processes with values in infinite-dimensional state spaces and their transition semigroups under nonparametric uncertainty. For this purpose, we start with a reference model which is a time-homogeneous Markov process (Ξ_t^x) of the form

$$\Xi_t^x := \psi_t(x) + Y_t \quad \text{for } t \geq 0 \text{ and } x \in X, \quad \text{with } Y_t \sim \mu_t, \quad (1.1)$$

taking values in a separable Banach space X . Here, $(\mu_t)_{t \geq 0}$ is a family of probability measures satisfying a certain moment condition. Although at first glance slightly restrictive, this class of Markov processes is analytically tractable, can easily be simulated, and includes the following prominent dynamics, see Section 5:

- suitably integrable infinite-dimensional Lévy processes,
- infinite-dimensional Ornstein-Uhlenbeck processes,
- geometric Brownian motions,
- deterministic flows arising from solutions to Banach-space-valued differential equations.

The multiplicative structure of the geometric Brownian motion can for instance be captured by the dynamics $\Xi_t^x = \psi_t(x) + Y_t := xY_t$, where the “addition” is understood as the multiplication on $(0, \infty)$, see Section 4.1 for further details. Given a Markov process of the form (1.1), we take nonparametric uncertainty of the respective transition probabilities into account by considering the worst-case scenario among perturbations within a proximity in Wasserstein distance. This is expressed by using a penalisation function $\varphi: [0, \infty) \rightarrow [0, \infty]$ applied to the Wasserstein distance to the family $(\mu_t)_{t \geq 0}$. A canonical example would be the indicator function $\varphi = \infty \cdot \mathbb{1}_{(a, \infty)}$ with $a \geq 0$, which allows for all transition probabilities in a Wasserstein neighborhood around the reference process. Then, the worst-case expectation at time $t \geq 0$ of a suitable function u in state x is given by

$$(I(t)u)(x) := \sup \int_X u(\psi_t(x) + z) \nu(dz),$$

where the supremum is taken over all ν with Wasserstein distance $\mathcal{W}(\mu_t, \nu) \leq at$ and the level of uncertainty at is proportional to the time horizon t . The family $(I(t))_{t \geq 0}$ can be seen as solutions to static optimization problems, which, in general, do not fulfill the Dynamic Programming Principle or the Chapman-Kolmogorov equations. In order to come up with a dynamically consistent family (semigroup), we split the optimization on $[0, t]$ into a composition of static optimization problems $I(\pi) = I(t_1) \circ I(t_2 - t_1) \circ \dots \circ I(t - t_n)$ for a partition $\pi = \{0 < t_1 < \dots < t_n < t\}$. A crucial step in the construction is the monotonicity $I(\pi) \geq I(\pi')$ for partitions $\pi \subset \pi'$ of $[0, t]$, see Lemma 3.10. This allows to define $S(t)$ as a limit of $I(\pi)$ over progressively finer partitions. We emphasise that the operator $S(t)$ does *not* depend on the choice of the approximating partitions, see Remark 3.15. Our main result is Theorem 3.14,

which states that the family $(S(t))_{t \geq 0}$ forms a strongly continuous convex semigroup on a suitable function space. Moreover, the infinitesimal behaviour of the semigroup, described in terms of its generator A , relates to the generator B of the reference Markov process (Ξ_t^x) via

$$(Au)(x) = (Bu)(x) + \varphi^*(\|u'(x)\|),$$

where φ^* is the convex conjugate of the penalisation function φ and u' denotes the Fréchet derivative of the function u . In particular, under a proper definition of proximity (e.g., for a geometric Brownian motion, we consider a logarithmic version of the Wasserstein distance), the uncertainty in the generator does not depend on the order of the Wasserstein distance. Our results in Section 3 extend and complement the work [6] as follows. While the analysis in [6] is performed on the space C_0 endowed with the supremum norm, we are working here with locally uniform convergence on a suitable space of (linearly) bounded continuous functions. Moreover, our construction allows for state-dependent dynamics such as geometric Brownian motions or Ornstein-Uhlenbeck processes. On a technical level, we make use of a suitable compact operator K in order to enforce the compactness of balls. In finite dimensions, the operator K can be neglected. Yet, in infinite-dimensional situations, it enables us to treat, for example, Ornstein-Uhlenbeck processes with an unbounded generator in the drift term. A minor yet subtle issue is the a priori chosen dyadic decomposition (partition) of the time interval $[0, t]$ for the definition of the operators $(S(t))_{t \geq 0}$ in [6]. In the present work, we also use dyadic partitions for the definition of the nonlinear semigroup. However, in a second step, we show that the latter does *not* depend on this particular choice of approximating partitions.

In Section 4, we discuss possible extensions of the setup and derive sensitivity bounds for the convex semigroup $(S(t))_{t \geq 0}$ relative to

$$(T(t)u)(x) = \mathbb{E}[u(\Xi_t^x)] = \int_X u(\psi_t(x) + y) \mu_t(dy).$$

Note that $(T(t))_{t \geq 0}$ is the transition semigroup corresponding to the reference Markov process. Due to our assumption on the additive separability of state and randomness, the expectation $\mathbb{E}[u(\Xi_t^x)]$ can be computed using a Monte-Carlo simulation for the law μ_t (independent of x) and simply adding the deterministic expression $\psi_t(x)$ to the simulated random variable Y_t , see Section 4.4 for further details. The results and possible applications are illustrated in various examples in Section 5.

The present construction is related to the one of the Nisio semigroup [30], see Section 4.5, where the family $(I(t))_{t \geq 0}$ is replaced by a supremum over a class of linear semigroups, cf. Denk et al. [15] and Nendel and Röckner [28]. In contrast to the Nisio semigroup, where $I(\pi)$ is increasing over refining partitions, the Wasserstein robust semigroup is obtained as a decreasing limit. Although the two approaches are based on a different notion of uncertainty (parametric vs. nonparametric) and a different monotonicity over refining partitions (increasing vs. decreasing), in standard situations, they result in the same convex semigroup $(S(t))_{t \geq 0}$ – the Nisio semigroup related to parametric drift uncertainty. An alternative construction, which does not rely on monotonicity arguments, was recently proposed by Blessing and Kupper [10]. In the multiperiod optimization problem $I(\pi)$, we consider the Wasserstein distance on the level of transition kernels, which is related to the nested distance introduced by Pflug and Pichler [34, 35]. For further results on adapted Wasserstein distances, we refer to Backhoff-Veraguas et al. [2], Backhoff-Veraguas et al. [3], and the references therein. Additionally, Markov processes under Wasserstein perturbations are studied in Eckstein [17] and Yang [42].

For the broad topic of model uncertainty in Markovian systems, we refer to Wiese-
mann et al. [41] and Dentcheva and Ruszczyński [16] in the context of Markov decision
processes, Rudolf and Schweizer [37] for applications to Markov chain Monte Carlo al-
gorithms, and De Cooman [14], Hartfiel [22], Krak et al. [25], and Škulj [40] for Markov
chains with interval probabilities.

The paper is organized as follows. In Section 2, we introduce the setup, all stand-
ing assumptions, and basic properties. Wasserstein robust semigroups are discussed in
Section 3. Therein, we present the worst-case operator, as well as the iterated opti-
mization scheme. The main result is Theorem 3.14, which provides the limiting convex
semigroup and its infinitesimal generator. In Section 4, we provide a detailed discus-
sion on the setup and our standing assumptions, present implications for sensitivity
bounds, and discuss the relation to nonlinear PDEs and parametric drift uncertainty.
The paper concludes with several examples in Section 5.

2. SETUP AND BASIC PROPERTIES

Throughout, let $(X, \|\cdot\|)$ be a separable Banach space with Borel σ -algebra $\mathcal{B}(X)$.
We consider a fixed compact linear operator $K: X \rightarrow X$, i.e., K is linear and the set
 $\{Kx: \|x\| \leq r\}$ is compact for all $r \geq 0$. We further assume that $\{Kx: x \in X\}$ is dense
in X and that $\|Kx\| \leq \|x\|$ for all $x \in X$. Note that, for a finite-dimensional space
 X , the operator K can simply be chosen to be the identity. In an infinite-dimensional
setting, K could be the resolvent of a generator of a compact C_0 -semigroup on X . We
refer to Section 5 for further details. We denote by $\text{Lip} = \text{Lip}^K(X)$ the space of all
functions $u: X \rightarrow \mathbb{R}$ satisfying

$$|u(x_1) - u(x_2)| \leq L\|K(x_1 - x_2)\| \quad \text{for all } x_1, x_2 \in X \text{ and some } L \geq 0. \quad (2.1)$$

For $u \in \text{Lip}$, we denote the smallest constant $L \geq 0$ such that (2.1) is satisfied by $\|u\|_{\text{Lip}}$.
Moreover, let Lip_b denote the subspace of all $u \in \text{Lip}$ with $\|u\|_\infty := \sup_{x \in X} |u(x)| < \infty$.
Note that $u \in \text{Lip}_b$ if and only if there exists a bounded Lipschitz continuous function
 $u_0: X \rightarrow \mathbb{R}$ with $u(x) = u_0(Kx)$ for all $x \in X$.¹ Since K is continuous, Lip_b is a
subspace of the space of all bounded Lipschitz continuous functions $X \rightarrow \mathbb{R}$.

For a sequence $(u_n)_{n \in \mathbb{N}}$ of bounded continuous functions $X \rightarrow \mathbb{R}$ and $u: X \rightarrow \mathbb{R}$, we
write $u_n \rightarrow u$ as $n \rightarrow \infty$ if

$$\sup_{n \in \mathbb{N}} \|u_n\|_\infty < \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \sup_{\|x\| \leq r} |u(x) - u_n(x)| = 0 \quad \text{for all } r \geq 0. \quad (2.2)$$

By $C_b = C_b^K(X)$, we denote the space of all $u: X \rightarrow \mathbb{R}$, for which there exists a sequence
 $(u_n)_{n \in \mathbb{N}} \subset \text{Lip}_b$ with $u_n \rightarrow u$ as $n \rightarrow \infty$ in the sense of (2.2).² By construction, C_b
is a subspace of the space of all bounded functions $u: X \rightarrow \mathbb{R}$, which are uniformly
continuous on bounded sets.

Let $p \in (1, \infty)$, and consider the set $\mathcal{P}_p(X)$ of all probability measures μ on $\mathcal{B}(X)$
with finite p -th moment $\int_X \|y\|^p \mu(dy) < \infty$. In a similar way, we consider $\mathcal{P}_p(X \times X)$
with X being replaced by $X \times X$ together with the norm $\|(x_1, x_2)\| := \|x_1\| + \|x_2\|$ for

¹For the nontrivial direction, define $u_0(x) := \sup_{y \in X} (u(y) - L\|x - Ky\|)$, for $x \in X$, with $L \geq \|u\|_{\text{Lip}}$.
The boundedness of u_0 follows from the density of the image $\{Kx: x \in X\}$.

²For a sequence $(u_n)_{n \in \mathbb{N}} \subset C_b$ with $u_n \rightarrow u$ as $n \rightarrow \infty$, it follows that $u \in C_b$. This can be seen by
approximating u_n with the inf-convolution $u_{n,k}: X \rightarrow \mathbb{R}$, $x \mapsto \inf_{y \in X} (u_n(y) + k\|Kx - Ky\|)$.

$x_1, x_2 \in X$. We endow $\mathcal{P}_p(X)$ with the p -Wasserstein distance \mathcal{W}_p , which is defined as

$$\mathcal{W}_p(\mu, \nu) := \left(\inf_{\pi \in \text{Cpl}(\mu, \nu)} \int_{X \times X} \|y - z\|^p \pi(\mathrm{d}y, \mathrm{d}z) \right)^{\frac{1}{p}} \quad \text{for } \mu, \nu \in \mathcal{P}_p(X),$$

where $\text{Cpl}(\mu, \nu)$ denotes the set of all probability measures on $\mathcal{B}(X \times X)$ with first marginal μ and second marginal ν , respectively. Then, \mathcal{W}_p defines a metric on $\mathcal{P}_p(X)$ and, in a similar way, the Wasserstein metric \mathcal{W}_p can be defined on $\mathcal{P}_p(X \times X)$. Throughout, we endow $\mathcal{P}_p(X \times X)$ with the topology generated by \mathcal{W}_p .

In the following remark, we collect some basic properties related to the Wasserstein distance \mathcal{W}_p that we frequently use in this work. For a detailed discussion on transport distances and their properties, we refer to Ambrosio et al. [1] or Villani [39].

Remark 2.1.

- a) Using Minkowski's inequality, $\text{Cpl}(\mu, \nu) \subset \mathcal{P}_p(X \times X)$ for all $\mu, \nu \in \mathcal{P}_p(X)$.
- b) Let $u: X \rightarrow \mathbb{R}$ and $L \geq 0$ with $|u(y) - u(z)| \leq L\|y - z\|$ for all $y, z \in X$. Then, for all $\mu, \nu \in \mathcal{P}_p(X)$ and $\pi \in \text{Cpl}(\mu, \nu)$, Minkowski's inequality implies that

$$\begin{aligned} \left| \left(\int_X |u(y)|^p \mu(\mathrm{d}y) \right)^{\frac{1}{p}} - \left(\int_X |u(z)|^p \nu(\mathrm{d}z) \right)^{\frac{1}{p}} \right| &\leq \left(\int_{X \times X} |u(y) - u(z)|^p \pi(\mathrm{d}y, \mathrm{d}z) \right)^{\frac{1}{p}} \\ &\leq L \left(\int_{X \times X} \|y - z\|^p \pi(\mathrm{d}y, \mathrm{d}z) \right)^{\frac{1}{p}}. \end{aligned}$$

In particular, the map $\mathcal{P}_p(X) \rightarrow \mathbb{R}$, $\mu \mapsto \int_X |u(y)|^p \mu(\mathrm{d}y)$ is continuous.

- c) For $\mu \in \mathcal{P}_p(X)$, we have

$$\mathcal{W}_p(\mu, \delta_0) = \left(\int_X \|y\|^p \mu(\mathrm{d}y) \right)^{\frac{1}{p}},$$

where δ_0 denotes the Dirac measure with barycenter 0.

- d) For $\mu, \nu \in \mathcal{P}_p$, by the triangular inequality and Hölder's inequality,

$$\int_X \|z\| \nu(\mathrm{d}z) \leq \mathcal{W}_p(\mu, \nu) + \int_X \|y\| \mu(\mathrm{d}y).$$

On X , we consider a time-homogeneous Markov process with transition probabilities $(p_t)_{t \geq 0}$ of the form

$$p_t(x, B) := \mu_t(\{y \in X : \psi_t(x) + y \in B\}) \quad \text{for all } t \geq 0, x \in X, \text{ and } B \in \mathcal{B}(X),$$

where $(\mu_t)_{t \geq 0} \subset \mathcal{P}_p(X)$ is a family of probability measures and $(\psi_t)_{t \geq 0}$ is a family of continuous maps $\psi_t: X \rightarrow X$. In particular, we assume that the transition probabilities satisfy the Chapman-Kolmogorov equations, i.e.,

$$\int_X u(\psi_{t+s}(x) + y_{t+s}) \mu_{t+s}(\mathrm{d}y_{t+s}) = \int_X \int_X u(\psi_s(\psi_t(x) + y_t) + y_s) \mu_s(\mathrm{d}y_s) \mu_t(\mathrm{d}y_t) \quad (2.3)$$

for all $s, t \geq 0$, $u \in C_b$, and $x \in X$. In other words, the Markov process belonging to the family of transition kernels $(p_t)_{t \geq 0}$ is of the form

$$\Xi_t^x = \psi_t(x) + Y_t \quad \text{with } Y_t \sim \mu_t \quad \text{for all } t \geq 0 \text{ and } x \in X. \quad (2.4)$$

Assumption 2.2. Throughout, we work under the following assumptions:

- (A1) There exists a constant $c \geq 0$ such that, for all $t \geq 0$ and $x_1, x_2 \in X$,

$$\|\psi_t(x_1) - \psi_t(x_2)\| \leq e^{ct} \|x_1 - x_2\|. \quad (2.5)$$

(A2) We assume that

$$K\psi_t(x) = \psi_t(Kx) \quad \text{for all } t \geq 0 \text{ and } x \in X.$$

Moreover, $\psi_0(x) = x$ for all $x \in X$, and

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq r} \|\psi_h(Kx) - Kx\| = 0 \quad \text{for all } r \geq 0.$$

(A3) We assume that $\mathcal{W}_p(\mu_h, \delta_0) \rightarrow 0$ as $h \downarrow 0$ or, in other words,

$$\lim_{h \downarrow 0} \int_X \|y\|^p \mu_h(dy) = 0.$$

We briefly discuss our assumptions in the following remark.

Remark 2.3.

a) Note that our global Assumption (A2) implies that, for all $u \in C_b$ and all $r \geq 0$,

$$\sup_{\|x\| \leq r} |u(\psi_h(x)) - u(x)| \rightarrow 0 \quad \text{as } h \downarrow 0.$$

Indeed, for $u \in \text{Lip}_b$,

$$\sup_{\|x\| \leq r} |u(\psi_h(x)) - u(x)| \leq \|u\|_{\text{Lip}} \sup_{\|x\| \leq r} \|\psi_h(Kx) - Kx\| \rightarrow 0 \quad \text{as } h \downarrow 0.$$

For general $u \in C_b$, the statement can be obtained by the following argument that will be used in a similar form on various occasions. Let $u \in C_b$. Then, there exists a sequence $(u_n)_{n \in \mathbb{N}} \in \text{Lip}_b$ with $u_n \rightarrow u$. By our global Assumptions (A1) and (A2), there exists some $h_0 > 0$ such that

$$M_r := \sup_{h \in [0, h_0]} \sup_{\|x\| \leq r} \|\psi_h(x)\| \leq \sup_{h \in [0, h_0]} \|\psi_h(0)\| + e^{ch_0} r < \infty \quad \text{for all } r \geq 0. \quad (2.6)$$

Therefore, for arbitrary $\varepsilon > 0$,

$$\begin{aligned} \sup_{\|x\| \leq r} |u(\psi_h(x)) - u(x)| &\leq \sup_{\|x\| \leq r} |u_n(\psi_h(x)) - u_n(x)| + 2 \sup_{\|\xi\| \leq M_r} |u(\xi) - u_n(\xi)| \\ &< \sup_{\|x\| \leq r} |u_n(\psi_h(x)) - u_n(x)| + \varepsilon \end{aligned}$$

for $n \in \mathbb{N}$ sufficiently large. Letting, first, $h \downarrow 0$ and then $\varepsilon \downarrow 0$, it follows that

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq r} |u(\psi_h(x)) - u(x)| = 0 \quad \text{for all } r \geq 0.$$

b) We assume that the family $(p_t)_{t \geq 0}$ of transition kernels is such that the related Markov process is of the form

$$\Xi_t^x = \psi_t(x) + Y_t.$$

However, also more general dynamics can be considered. For example, the additive operation $+$ could be the multiplication on the positive half line $(0, \infty)$, leading to geometric dynamics

$$\Xi_t^x = xY_t.$$

For more details, we refer to Section 4.1, in particular, Example 4.2.

c) Consider the family $T = (T(t))_{t \geq 0}$, given by

$$(T(t)u)(x) := \int_X u(\xi) p_t(x, d\xi) = \int_X u(\psi_t(x) + y) \mu_t(dy) \quad (2.7)$$

for all $t \geq 0$, $u \in C_b$, and $x \in X$. Due to the Chapman-Kolmogorov equations (2.3), it follows that T is a semigroup on C_b . More precisely, it is the transition semigroup of the Markov process (Ξ_t^x) . Due to our global assumptions, the semigroup T is strongly continuous on C_b , see Remark 3.13, below.

Throughout, we consider a convex lower semicontinuous function $\varphi: [0, \infty) \rightarrow [0, \infty]$ with $\varphi(0) = 0$ and $\varphi(v) \neq 0$ for some $v > 0$. Note that φ is nondecreasing and continuous on $\text{dom}(\varphi) := \{v \in [0, \infty): \varphi(v) < \infty\}$. Since φ is nondecreasing, either $\text{dom}(\varphi) = [0, \infty)$ or $\text{dom}(\varphi) = [0, a]$ for some $a \geq 0$. We further assume that the map $[0, \infty) \rightarrow [0, \infty]$, $v \mapsto \varphi(v^{1/p})$ is convex. Since $\varphi \not\equiv 0$, this implies that

$$\liminf_{v \rightarrow \infty} \frac{\varphi(v)}{v^p} > 0. \quad (2.8)$$

In particular, the conjugate

$$\varphi^*(w) := \sup_{v \in [0, \infty)} (vw - \varphi(v)), \quad \text{for } w \geq 0,$$

defines a continuous function $[0, \infty) \rightarrow [0, \infty)$ with $\varphi^*(0) = 0$. Typical examples for φ are $[0, \infty) \rightarrow [0, \infty)$, $v \mapsto v^p$ or $\varphi_a := \infty \cdot \mathbf{1}_{(a, \infty)}$, i.e., $\varphi_a(v) = 0$, for $v \in [0, a]$, and $\varphi_a(v) = \infty$, for $v > a$, with $a \geq 0$.

3. WASSERSTEIN ROBUST SEMIGROUPS

In this section, we aim to study a distributionally robust version of the semigroup T given in Equation (2.7). We start with a preliminary version by considering the following family of operators:

Definition 3.1. For $t \geq 0$, $u \in C_b$, and $x \in X$, we define

$$(I(t)u)(x) = \sup_{\nu \in \mathcal{P}_p(X)} \left(\int_X u(\psi_t(x) + z) \nu(dz) - \varphi_t(\mathcal{W}_p(\mu_t, \nu)) \right), \quad (3.1)$$

where $\varphi_t(v) := t\varphi(e^{-ct} \frac{v}{t})$, for $v \geq 0$ and $t > 0$, and $\varphi_0 := \infty \cdot \mathbf{1}_{(0, \infty)}$.

Remark 3.2.

a) Let $t \geq 0$ and $u_1, u_2 \in C_b$. Then,

$$\|I(t)u_1 - I(t)u_2\|_\infty \leq \|u_1 - u_2\|_\infty. \quad (3.2)$$

Moreover, $I(t)0 = 0$, which implies that $\|I(t)u\|_\infty \leq \|u\|_\infty$ for all $u \in C_b$. In fact,

$$\begin{aligned} |(I(t)u_1)(x) - (I(t)u_2)(x)| &\leq \sup_{\nu \in \mathcal{P}_p(X)} \int_X |u_1(\psi_t(x) + z) - u_2(\psi_t(x) + z)| \nu(dz) \\ &\leq \|u_1 - u_2\|_\infty \quad \text{for all } x \in X. \end{aligned}$$

By taking the supremum over all $x \in X$, the inequality (3.2) follows.

b) Let $t \geq 0$ and $u \in \text{Lip}_b$. Then,

$$\|I(t)u\|_{\text{Lip}} \leq e^{ct} \|u\|_{\text{Lip}}.$$

In particular, $I(t)u \in \text{Lip}_b$ for all $u \in \text{Lip}_b$. Indeed, due to our global assumptions (A1) and (A2),

$$\begin{aligned} |(I(t)u)(x_1) - (I(t)u)(x_2)| &\leq \sup_{\nu \in \mathcal{P}_p(X)} \int_X |u(\psi_t(x_1) + z) - u(\psi_t(x_2) + z)| \nu(dz) \\ &\leq e^{ct} \|u\|_{\text{Lip}} \|K(x_1 - x_2)\| \quad \text{for all } x_1, x_2 \in X. \end{aligned}$$

We start with the following observations.

Lemma 3.3. *For all $a \geq 0$, $t \geq 0$, $u \in \text{Lip}$, and $x \in X$, let*

$$(I_a(t)u)(x) := \sup_{\mathcal{W}_p(\mu_t, \nu) \leq at} \int_X u(\psi_t(x) + z) \nu(dz). \quad (3.3)$$

For every $L \geq 0$, there exists a constant $a \geq 0$, depending only on φ , c , and L such that

a) for every $u \in \text{Lip}$ with $\|u\|_{\text{Lip}} \leq L$,

$$(I(t)u)(x) = \sup_{\mathcal{W}_p(\mu_t, \nu) \leq at} \left(\int_X u(\psi_t(x) + z) \nu(dz) - \varphi_t(\mathcal{W}_p(\mu_t, \nu)) \right) \quad (3.4)$$

for all $t \in [0, 1]$ and $x \in X$,

b) for all $u_1, u_2 \in \text{Lip}$ with $\|u_i\|_{\text{Lip}} \leq L$ for $i = 1, 2$,

$$I(t)u_1 - I(t)u_2 \leq I_a(t)(u_1 - u_2) \quad \text{for all } t \in [0, 1]. \quad (3.5)$$

Proof. Let $L \geq 0$ and

$$J := \{v \in [0, \infty) : \varphi(e^{-c}v) \leq 1 + Lv\}.$$

By Equation (2.8), $J = [0, a]$ for some $a \geq 0$, depending only on φ , c , and L . Now, let $u \in \text{Lip}$ with $\|u\|_{\text{Lip}} \leq L$, $t \geq 0$, and $x \in X$. For $t = 0$, the equality stated in (3.4) is trivial. Therefore, assume that $t \in (0, 1]$. Then, there exists some $\nu \in \mathcal{P}_p(X)$ with

$$\int_X u(\psi_t(x) + y) \mu_t(dy) \leq (I(t)u)(x) \leq t + \int_X u(\psi_t(x) + z) \nu(dz) - \varphi_t(\mathcal{W}_p(\mu_t, \nu)),$$

which leads to the inequality

$$\begin{aligned} \varphi\left(e^{-ct} \frac{\mathcal{W}_p(\mu_t, \nu)}{t}\right) &\leq 1 + \frac{1}{t} \int_X \int_X \left(u(\psi_t(x) + z) - u(\psi_t(x) + y)\right) \nu(dz) \mu_t(dy) \\ &\leq 1 + L \frac{\mathcal{W}_p(\mu_t, \nu)}{t}. \end{aligned}$$

Therefore, since $t \in (0, 1]$, it is sufficient to take the supremum in (3.1) over the set

$$\left\{ \nu \in \mathcal{P}_p(X) : \frac{\mathcal{W}_p(\mu_t, \nu)}{t} \in J \right\} = \left\{ \nu \in \mathcal{P}_p(X) : \frac{\mathcal{W}_p(\mu_t, \nu)}{t} \in [0, a] \right\}.$$

The proof of part a) is complete. In order to prove part b), let $u_1, u_2 \in \text{Lip}$ with $\|u_i\|_{\text{Lip}} \leq L$ for $i = 1, 2$, $t \in [0, 1]$, and $x \in X$. Then, by part a),

$$(I(t)u_1)(x) - (I(t)u_2)(x) \leq \sup_{\mathcal{W}_p(\mu_t, \nu) \leq at} \int_X \left(u_1(\psi_t(x) + z) - u_2(\psi_t(x) + z)\right) \nu(dz).$$

The proof is complete. \square

Note that, for $u \in \text{Lip}_b$, the definition of $I_a(t)u$ by virtue of (3.3) is exactly the same as (3.1) with $\varphi = \varphi_a = \infty \cdot \mathbf{1}_{(a, \infty)}$.

Lemma 3.4. *For every $C \geq 0$, there exists a constant $b \geq 0$, depending only on φ , c , p , and C such that*

$$I(t)u = \sup_{\mathcal{W}_p(\mu_t, \nu) \leq bt^\alpha} \left(\int_X u(\psi_t(x) + z) \nu(dz) - \varphi_t(\mathcal{W}_p(\mu_t, \nu)) \right)$$

for all $u \in C_b$ with $\|u\|_\infty \leq C$, $t \in [0, 1]$, and $x \in X$, where $\alpha := \frac{p-1}{p}$.

Proof. By (2.8), there exists some constant $M > 0$, depending only on φ and p , such that

$$v^p \leq M(1 + \varphi(v)) \quad \text{for all } v \geq 0.$$

Let $C \geq 0$, $u \in C_b$ with $\|u\|_\infty \leq C$, $t \in [0, 1]$, and $x \in X$. For $t = 0$, the statement is trivial. Therefore, assume that $t > 0$. Then, there exists some $\nu \in \mathcal{P}_p(X)$ with

$$\int_X u(\psi_t(x) + y) \mu_t(dy) \leq (I(t)u)(x) \leq 1 + \int_X u(\psi_t(x) + z) \nu(dz) - \varphi_t(\mathcal{W}_p(\mu_t, \nu)),$$

which leads to the inequality

$$\varphi_t(\mathcal{W}_p(\mu_t, \nu)) \leq 1 + 2\|u\|_\infty \leq 1 + 2C.$$

Therefore, since $t \in [0, 1]$, it is sufficient to take the supremum in (3.1) over the set

$$\left\{ \nu \in \mathcal{P}_p(X) : \varphi \left(e^{-c \frac{\mathcal{W}_p(\mu_t, \nu)}{t}} \right) \leq \frac{1+2C}{t} \right\} \subset \left\{ \nu \in \mathcal{P}_p(X) : \mathcal{W}_p(\mu_t, \nu) \leq bt^\alpha \right\}$$

with $b := e^c(2M(1+C))^{1/p}$. □

Lemma 3.5. *For $i = 1, 2$, let $(u_n^i)_{n \in \mathbb{N}} \subset C_b$ with $(u_n^1 - u_n^2) \rightarrow 0$ as $n \rightarrow \infty$. Then, there exists some $h_0 > 0$ such that $(I(h)u_n^1 - I(h)u_n^2) \rightarrow 0$ uniformly in $h \in [0, h_0]$ as $n \rightarrow \infty$, i.e., $\sup_{n \in \mathbb{N}} \sup_{h \in [0, h_0]} \|I(h)u_n^1 - I(h)u_n^2\|_\infty < \infty$ and*

$$\lim_{n \rightarrow \infty} \sup_{h \in [0, h_0]} \sup_{\|x\| \leq r} |(I(h)u_n^1)(x) - (I(h)u_n^2)(x)| = 0 \quad \text{for all } r \geq 0. \quad (3.6)$$

Proof. By Equation (3.2), we have

$$\|I(t)u_n^1 - I(t)u_n^2\|_\infty \leq \|u_n^1 - u_n^2\|_\infty \quad \text{for all } t \geq 0 \text{ and } n \in \mathbb{N}.$$

Therefore, it remains to verify (3.6). Let $u_n := u_n^1 - u_n^2$ and $\alpha := \frac{p-1}{p}$. By Lemma 3.4 with $C := \sup_{n \in \mathbb{N}} \|u_n\|_\infty$, there exists some $b \geq 0$ such that

$$|(I(t)u_n^1)(x) - (I(t)u_n^2)(x)| \leq \sup_{\mathcal{W}_p(\mu_t, \nu) \leq bt^\alpha} \int_X |u_n(\psi_t(x) + z)| \nu(dz)$$

for all $t \in [0, 1]$ and $x \in X$. By our global Assumptions (A2) and (A3), there exists some $h_0 \in (0, 1]$ such that

$$\sup_{h \in [0, h_0]} \|\psi_h(0)\| < \infty \quad \text{and} \quad \sup_{h \in [0, h_0]} \int_X \|y\|^p \mu_h(dy) \leq 1.$$

Let $\varepsilon > 0$, $r \geq 0$, M_r be given by Equation (2.6), and $M > 0$ with $\frac{C(1+bh_0^\alpha)}{M} < \varepsilon$. Then, using Markov's inequality and Remark 2.1 d),

$$\begin{aligned} & \sup_{h \in [0, h_0]} \sup_{\|x\| \leq r} \sup_{\mathcal{W}_p(\mu_h, \nu) \leq bh^\alpha} \int_X |u_n(\psi_h(x) + z)| \nu(dz) \\ & \leq \sup_{\|\xi\| \leq M+M_r} |u_n(\xi)| + \sup_{h \in [0, h_0]} \sup_{\mathcal{W}_p(\mu_h, \nu) \leq bh^\alpha} \frac{C}{M} \int_X \|z\| \nu(dz) \\ & \leq \sup_{\|\xi\| \leq M+M_r} |u_n(\xi)| + \frac{C(1+bh_0^\alpha)}{M} < \varepsilon \end{aligned}$$

for $n \in \mathbb{N}$ sufficiently large. \square

Remark 3.6. For $u \in \text{Lip}$,

$$\|I(h)u - T(h)u\|_\infty \leq h\varphi^*(e^{ch}\|u\|_{\text{Lip}}). \quad (3.7)$$

In fact, for all $x \in X$, by Remark 2.1 b),

$$(I(h)u)(x) - (T(h)u)(x) \leq \sup_{\nu \in \mathcal{P}_p} \|u\|_{\text{Lip}} \mathcal{W}_p(\mu_h, \nu) - \varphi_h(\mathcal{W}_p(\mu_h, \nu)) = h\varphi^*(e^{ch}\|u\|_{\text{Lip}}).$$

Lemma 3.7. Let $h_0 > 0$ as in Lemma 3.5. Then, the operator $I(h): \mathbf{C}_b \rightarrow \mathbf{C}_b$ is well-defined, convex, and monotone for all $h \in [0, h_0]$. Moreover, for all $u \in \mathbf{C}_b$,

$$I(h)u \rightarrow u \quad \text{as } h \downarrow 0.$$

Proof. By Remark 3.2, $I(t)u \in \text{Lip}_b$ for all $t \geq 0$ and $u \in \text{Lip}_b$. By Lemma 3.5, it follows that $I(h): \mathbf{C}_b \rightarrow \mathbf{C}_b$ is well-defined for all $h \in [0, h_0]$, since, by definition of \mathbf{C}_b , every $u \in \mathbf{C}_b$ can be approximated by a sequence $(u_n)_{n \in \mathbb{N}} \subset \text{Lip}_b$ in the sense that $u_n \rightarrow u$ as $n \rightarrow \infty$. The convexity and monotonicity of the operator $I(h)$, for $h \in [0, h_0]$, are an immediate consequence of its definition. Let $u \in \text{Lip}_b$ and $r \geq 0$. By Remark 2.3 a),

$$\sup_{\|x\| \leq r} |u(\psi_h(x)) - u(x)| \rightarrow 0 \quad \text{as } h \downarrow 0.$$

By virtue of Assumption (A3), we may conclude that

$$\begin{aligned} \sup_{\|x\| \leq r} |(T(h)u)(x) - u(x)| & \leq \|u\|_{\text{Lip}} \left(\int_X \|y\|^p \mu_h(dy) \right)^{1/p} \\ & + \sup_{\|x\| \leq r} |u(\psi_h(x)) - u(x)| \rightarrow 0 \quad \text{as } h \downarrow 0, \end{aligned}$$

which, together with (3.7), implies that $I(h)u \rightarrow u$ as $h \downarrow 0$. For general $u \in \mathbf{C}_b$, the strong continuity follows by approximating with $(u_n)_{n \in \mathbb{N}} \subset \text{Lip}_b$ and using the convergence $I(h)u_n \rightarrow I(h)u$, uniformly in $h \in [0, h_0]$, as $n \rightarrow \infty$, see Lemma 3.5. \square

Alternatively, the family of operators $I = (I(t))_{t \geq 0}$ can also be defined via couplings with first marginals given in terms of the family of laws $(\mu_t)_{t \geq 0}$, as we point out in the following remark.

Remark 3.8. Let

$$\Delta_p \pi := \left(\int_{X \times X} \|y - z\|^p \pi(dy, dz) \right)^{1/p} \quad \text{for all } \pi \in \mathcal{P}_p(X \times X). \quad (3.8)$$

Then, by definition, $\mathcal{W}_p(\mu, \nu) = \inf_{\pi \in \text{Cpl}(\mu, \nu)} \Delta_p \pi$ for all $\mu, \nu \in \mathcal{P}_p(X)$. Moreover, by Remark 2.1 b),

$$|\Delta_p \pi - \Delta_p \pi'| \leq \mathcal{W}_p(\pi, \pi') \quad \text{for all } \pi' \in \mathcal{P}_p(X \times X),$$

i.e., the map $\mathcal{P}_p(X \times X) \rightarrow \mathbb{R}$, $\pi \mapsto \Delta_p \pi$ is 1-Lipschitz. For $\pi \in \mathcal{P}_p(X \times X)$ and $i = 1, 2$, we denote by π_i the i -th marginal of π . Using these notations, we find that, for all $t \geq 0$, $u \in C_b$, and $x \in X$,

$$(I(t)u)(x) = \sup_{\substack{\pi \in \mathcal{P}_p(X \times X) \\ \pi_1 = \mu_t}} \left(\int_X u(\psi_t(x) + z) \pi_2(dz) - \varphi_t(\Delta_p \pi) \right). \quad (3.9)$$

In fact, since φ is nondecreasing, \geq holds in (3.9). In order to establish the other inequality, let $\nu \in \mathcal{P}_p(X)$. By [39, Theorem 4.1], there exists an optimal coupling $\pi^* \in \text{Cpl}(\mu, \nu)$ with $\mathcal{W}_p(\mu_t, \nu) = \Delta_p \pi^*$. By definition of $\text{Cpl}(\mu_t, \nu)$, it follows that $\pi_1^* = \mu_t$ and $\pi_2^* = \nu$. Therefore,

$$\begin{aligned} \int_X u(\psi_t(x) + z) \nu(dz) - \varphi_t(\mathcal{W}_p(\mu_t, \nu)) &= \int_X u(\psi_t(x) + z) \pi_2^*(dz) - \varphi_t(\Delta_p \pi^*) \\ &\leq \sup_{\substack{\pi \in \mathcal{P}_p(X \times X) \\ \pi_1 = \mu_t}} \left(\int_X u(\psi_t(x) + z) \pi_2(dz) - \varphi_t(\Delta_p \pi) \right). \end{aligned}$$

Taking the supremum over all $\nu \in \mathcal{P}_p(X)$ yields the desired claim.

The following measurable selection result forms the basis for several proofs in this section.

Lemma 3.9. *Let $t \geq 0$, $u \in C_b$, and $\varepsilon > 0$. Then, there exists a measurable map $\pi: X \rightarrow \mathcal{P}_p(X \times X)$, $x \mapsto \pi^x$ such that, for each $x \in X$, the first marginal π_1^x of π^x equals μ_t and*

$$(I(t)u)(x) \leq \int_X u(\psi_t(x) + z) \pi_2^x(dz) - \varphi_t(\Delta_p \pi^x) + \varepsilon,$$

where $\pi_2^x \in \mathcal{P}_p(X)$ denotes the second marginal of π^x .

Proof. Let $D_t := \{\pi \in \mathcal{P}_p(X \times X) : \pi_1 = \mu_t \text{ and } \varphi_t(\Delta_p \pi) < \infty\}$. Then, by Remark 3.8, the map

$$f: X \times D_t \rightarrow \mathbb{R}, \quad (x, \pi) \mapsto \int_{X \times X} u(\psi_t(x) + z) \pi_1(dz) - \varphi_t(\Delta_p \pi)$$

is continuous, and

$$(I(t)u)(x) = \sup_{\pi \in D_t} f(x, \pi) \quad \text{for all } x \in X.$$

The statement now follows from [7, Proposition 7.34]. \square

Lemma 3.10. *Let $s, t \geq 0$ and $u \in C_b$. Then,*

$$(I(t)I(s)u)(x) \leq (I(t+s)u)(x) \quad \text{for all } x \in X.$$

Proof. Let $\varepsilon > 0$, $x \in X$, and $\pi_t \in \mathcal{P}_p(X \times X)$ with first marginal $\pi_{t,1} = \mu_t$ and

$$(I(t)I(s)u)(x) \leq \int_X (I(s)u)(\psi_t(x) + z_t) \nu_t(dz_t) - \varphi_t(\Delta_p \pi_t) + \frac{\varepsilon}{2},$$

where $\nu_t := \pi_{t,2} \in \mathcal{P}_p(X)$ is the second marginal of π_t . By Lemma 3.9, there exists a measurable map $\pi_s: X \rightarrow \mathcal{P}_p(X \times X)$, $z \mapsto \pi_s^z$ such that, for all $z \in X$, the first marginal $\pi_{s,1}^z$ of π_s^z is μ_s and

$$(I(s)u)(\psi_t(x) + z) \leq \int_X u(\psi_s(\psi_t(x) + z) + z_s) \nu_s^z(dz_s) - \varphi_s(\Delta_p \pi_s^z) + \frac{\varepsilon}{2}$$

with $\nu_s^z := \pi_{s,2}^z \in \mathcal{P}_p(X)$ being the second marginal of π_s^z . It follows that

$$\begin{aligned} (I(t)I(s)u)(x) &\leq \int_X \int_X u(\psi_s(\psi_t(x) + z_t) + z_s) \nu_s^{z_t}(dz_s) \nu_t(dz_t) \\ &\quad - \int_X \varphi_s(\Delta_p \pi_s^{z_t}) \nu_t(dz_t) - \varphi_t(\Delta_p \pi_t) + \varepsilon, \end{aligned} \quad (3.10)$$

where the measurability of the map $z \mapsto \varphi_s(\Delta_p \pi_s^z)$ follows from the measurability of the map $z \mapsto \pi_s^z$ and the lower semicontinuity of the map $\pi \mapsto \varphi_s(\Delta_p \pi)$. We now define a new coupling π_{t+s} via

$$\begin{aligned} \int_{X \times X} g(y_{t+s}, z_{t+s}) \pi_{t+s}(dy_{t+s}, dz_{t+s}) &= \int_{X \times X} \int_{X \times X} g(\psi_s(\psi_t(x) + y_t) + y_s - \psi_{t+s}(x), \\ &\quad \psi_s(\psi_t(x) + z_t) + z_s - \psi_{t+s}(x)) \pi_s^{z_t}(dy_s, dz_s) \pi_t(dy_t, dz_t) \end{aligned} \quad (3.11)$$

for all bounded continuous functions $g: X \times X \rightarrow \mathbb{R}$. Note that the outer integral on the right-hand side of (3.11) is well-defined since the inner integral depending on y_t and z_t is jointly measurable in (y_t, z_t) . The latter follows from the measurability of the map $z \mapsto \pi_s^z$, the choice of g as a bounded continuous function, and the (Lipschitz) continuity of ψ_s . Then, the consistency condition (2.3) and the definition of π_{t+s} in (3.11) with $g(y, z) = u(\psi_{t+s}(x) + y)$, for $y, z \in X$, imply that the first marginal of π_{t+s} equals to μ_{t+s} , i.e., $\pi_{1,t+s} = \mu_{t+s}$. In particular, for $g(y, z) = u(\psi_{t+s}(x) + z)$ with $y, z \in X$ in (3.11), we have

$$\begin{aligned} (I(t+s)u)(x) &\geq \int_X u(\psi_{t+s}(x) + z_{t+s}) \nu_{t+s}(dz_{t+s}) - \varphi_{t+s}(\Delta_p \pi_{t+s}) \\ &= \int_X \int_X u(\psi_s(\psi_t(x) + z_t) + z_s) \nu_s^{z_t}(dz_s) \nu_t(dz_t) - \varphi_{t+s}(\Delta_p \pi_{t+s}), \end{aligned}$$

where $\nu_{t+s} := \pi_{2,t+s}$ denotes the second marginal of π_{t+s} . Note that π_{t+s} is an element of $\mathcal{P}_p(X \times X)$. This follows from $\pi_{1,t+s} = \mu_{t+s}$ together with the estimate

$$\varphi_{t+s}(\Delta_p \pi_{t+s}) \leq \varphi_t(\Delta_p \pi_t) + \int_X \varphi_s(\Delta_p \pi_s^{z_t}) \nu_t(dz_t). \quad (3.12)$$

It remains to prove (3.12), since, in view of (3.10), the estimate (3.12) implies $I(t+s)u \geq I(t)I(s)u$ after taking the limit $\varepsilon \downarrow 0$. We start by showing that

$$\Delta_p \pi_{t+s} \leq \Delta_p \pi_t + \left(\int_X \int_{X \times X} \|y_s - z_s\|^p \pi_s^{z_t}(dy_s, dz_s) \nu_t(dz_t) \right)^{1/p}.$$

Using (3.11), Minkowski's inequality, and Assumption (A1),

$$\begin{aligned}
\Delta_p \pi_{t+s} &= \left(\int_{X \times X} \|y_{t+s} - z_{t+s}\|^p \pi_{t+s}(dy_{t+s}, dz_{t+s}) \right)^{1/p} \\
&= \left(\int_{X \times X} \|\psi_{t+s}(x) + y_{t+s} - (\psi_{t+s}(x) + z_{t+s})\|^p \pi_{t+s}(dy_{t+s}, dz_{t+s}) \right)^{1/p} \\
&\leq \left(\int_{X \times X} \int_{X \times X} \|\psi_s(\psi_t(x) + y_t) - \psi_s(\psi_t(x) + z_t)\|^p \pi_t(dy_t, dz_t) \right)^{1/p} \\
&\quad + \left(\int_X \int_{X \times X} \|y_s - z_s\|^p \pi_s^{z_t}(dy_s, dz_s) \nu_t(dz_t) \right)^{1/p} \\
&\leq e^{cs} \Delta_p \pi_t + \left(\int_X \int_{X \times X} \|y_s - z_s\|^p \pi_s^{z_t}(dy_s, dz_s) \nu_t(dz_t) \right)^{1/p}.
\end{aligned}$$

Note that the function $z \mapsto \int_{X \times X} \|y_s - z_s\|^p \pi_s^z(dy_s, dz_s)$ is in fact measurable, since the map $z \mapsto \pi_s^z$ is measurable and the map $\pi \mapsto \int_{X \times X} \|y - z\|^p \pi(dy, dz)$ is continuous, by Remark 2.1 b), and thus measurable. Finally, we show that (3.12) holds. Since we assumed $v \mapsto \varphi(v^{1/p})$ to be convex, the map $v \mapsto \varphi_s(v^{1/p})$ is convex as well, which, by Jensen's inequality, implies that

$$\varphi_s \left(\left(\int_X \int_{X \times X} \|y_s - z_s\|^p \pi_s^{z_t}(dy_s, dz_s) \nu_t(dz_t) \right)^{1/p} \right) \leq \int_X \varphi_s(\Delta_p \pi_s^{z_t}) \nu_t(dz_t).$$

The convexity of φ yields that

$$\varphi_{t+s}(\Delta_p \pi_{t+s}) = (t+s) \varphi \left(e^{-c(t+s)} \frac{\Delta_p \pi_{t+s}}{t+s} \right) \leq \varphi_t(\Delta_p \pi_t) + \int_X \varphi_s(\Delta_p \pi_s^{z_t}) \nu_t(dz_t).$$

The proof is complete. \square

By C_b^1 , we denote the space of all $u \in \text{Lip}_b$ with Fréchet derivative $u': X \rightarrow X'$ satisfying $x \mapsto \|u'(x)\| \in C_b$. Here and throughout, the dual space X' of X is endowed with the operator norm

$$\|x'\| := \sup_{\|x\| \leq 1} |x'x| \quad \text{for all } x' \in X'.$$

Lemma 3.11. *Let $u \in C_b^1$. Then,*

$$\frac{I(h)u - T(h)u}{h} \rightarrow \varphi^*(\|u'(\cdot)\|) \quad \text{as } h \downarrow 0. \quad (3.13)$$

In particular, the map $X \rightarrow \mathbb{R}$, $x \mapsto \varphi^(\|u'(x)\|)$ is an element of C_b .*

Proof. By Remark 3.6,

$$\sup_{h \in [0,1]} \left\| \frac{I(h)u - T(h)u}{h} \right\|_\infty < \infty,$$

since $u \in C_b^1 \subset \text{Lip}_b$. Let $r \geq 0$. We start by showing that, for $h > 0$ sufficiently small,

$$\sup_{\|x\| \leq r} \left(\varphi^*(\|u'(x)\|) - \frac{(I(h)u)(x) - (T(h)u)(x)}{h} \right) \leq \varepsilon. \quad (3.14)$$

For $h \geq 0$ and $\theta \in X$, let $\nu_{h,\theta} := \mu_h * \delta_{h\theta}$, where $*$ denotes the convolution of two probability measures. Clearly, $\mathcal{W}_p(\mu_h, \nu_{h,\theta}) \leq h\|\theta\|$, and therefore

$$\begin{aligned} \frac{(I(h)u)(x) - (T(h)u)(x)}{h} &\geq \frac{1}{h} \left(\int_X u(\psi_h(x) + z) \nu_{h,\theta}(dz) - \varphi_h(h\|\theta\|) - (T(h)u)(x) \right) \\ &= \int_X \frac{u(\psi_h(x) + y + h\theta) - u(\psi_h(x) + y)}{h} \mu_h(dy) - \varphi(e^{-ch}\|\theta\|) \end{aligned} \quad (3.15)$$

for all $x, \theta \in X$ and $h \geq 0$. Since u' is bounded, our assumptions on φ imply that there exists some constant $a \geq 0$ such that

$$\varphi^*(\|u'(x)\|) = \sup_{v \in [0, a]} \{v\|u'(x)\| - \varphi(v)\} = \sup_{\|\theta\| \leq a} \{u'(x)\theta - \varphi(\|\theta\|)\} \quad \text{for all } x \in X.$$

Fix $\varepsilon > 0$. Using Taylor's theorem, the uniform continuity of u' on bounded subsets of X , and Remark 2.3 a), there exists some $\delta > 0$ such that

$$u'(x)\theta \leq \frac{u(\psi_h(x) + y + h\theta) - u(\psi_h(x) + y)}{h} + \frac{\varepsilon}{2} \quad (3.16)$$

for all $x, y, \theta \in X$ with $\|x\| \leq r$, $\|y\| \leq \delta$, and $\|\theta\| \leq a$, and $h > 0$ sufficiently small. Moreover, since u' is bounded, it follows from Markov's inequality that

$$\left| \int_{\{\|y\| > \delta\}} \frac{u(\psi_h(x) + y + h\theta) - u(\psi_h(x) + y)}{h} \mu_h(dy) \right| \leq \frac{a\|u'\|_\infty}{\delta} \mathcal{W}_p(\mu_h, \delta_0) < \frac{\varepsilon}{2} \quad (3.17)$$

for all $x, \theta \in X$ with $\|x\| \leq r$ and $\|\theta\| \leq a$, and $h > 0$ sufficiently small. Therefore, since $\mu_h(\{\|y\| \leq \delta\}) \rightarrow 1$ as $h \downarrow 0$ by Assumption (A3), a combination of (3.15), (3.16), and (3.17) together with $e^{-ch} \leq 1$, for all $h \geq 0$, yields that

$$u'(x)\theta - \varphi(\|\theta\|) \leq \int_X \frac{u(\psi_h(x) + y + h\theta) - u(\psi_h(x) + y)}{h} \mu_h(dy) - \varphi(e^{-ch}\|\theta\|) + \varepsilon$$

for all $x, \theta \in X$ with $\|x\| \leq r$ and $\|\theta\| \leq a$, and $h > 0$ sufficiently small. Taking the supremum over all $x, \theta \in X$ with $\|x\| \leq r$ and $\|\theta\| \leq a$, (3.14) follows.

It remains to show that, for $h > 0$ sufficiently small,

$$\sup_{\|x\| \leq r} \left(\frac{(I(h)u)(x) - (T(h)u)(x)}{h} - \varphi^*(\|u'(x)\|) \right) \leq \varepsilon. \quad (3.18)$$

To that end, for $x \in X$ and $h \geq 0$, we consider the auxiliary function $g_{h,x} : X \times X \rightarrow \mathbb{R}$, defined by

$$g_{h,x}(y, z) := \frac{|u(\psi_h(x) + y) - u(\psi_h(x) + z)|}{\|y - z\|} \quad \text{for all } y, z \in X.$$

Let $\varepsilon > 0$, $b \geq 0$ as in Lemma 3.4 with $C := \|u\|_\infty$, and $\alpha := \frac{p-1}{p}$. Then, by Lemma 3.4, for all $h \in [0, 1]$ and $x \in X$, there exists some $\nu_h^x \in \mathcal{P}_p(X)$ with $\mathcal{W}_p(\mu_h, \nu_h^x) \leq bh^\alpha$ and

$$(I(h)u)(x) \leq \frac{\varepsilon h}{2} + \int_X u(\psi_h(x) + z) \nu_h^x(dz) - \varphi_h(\mathcal{W}_p(\mu_h, \nu_h^x)). \quad (3.19)$$

For each $h \in [0, 1]$ and $x \in X$, let $\pi_h^x \in \mathcal{P}_p(X \times X)$ be an optimal coupling between μ_h and ν_h^x . Then,

$$\begin{aligned} & \frac{(I(h)u)(x) - (T(h)u)(x)}{h} \\ & \leq \frac{1}{h} \int_{X \times X} g_{h,x}(y, z) \|y - z\| \pi_h^x(dy, dz) - \frac{\varphi_h(\mathcal{W}_p(\mu_h, \nu_h^x))}{h} + \frac{\varepsilon}{2} \\ & \leq \left(\int_{X \times X} |g_{h,x}(y, z)|^q \pi_h^x(dy, dz) \right)^{1/q} \frac{\mathcal{W}_p(\mu_h, \nu_h^x)}{h} - \varphi \left(e^{-ch} \frac{\mathcal{W}_p(\mu_h, \nu_h^x)}{h} \right) + \frac{\varepsilon}{2} \\ & \leq \varphi^* \left(e^{ch} \left(\int_{X \times X} |g_{h,x}(y, z)|^q \pi_h^x(dy, dz) \right)^{1/q} \right) + \frac{\varepsilon}{2} \end{aligned}$$

for all $h \in [0, 1]$ and $x \in X$, where $q = \frac{p}{p-1}$ is the conjugate exponent of p . Since φ^* is convex and continuous, thus uniformly continuous on bounded intervals, and u' is bounded, there exists some $\delta > 0$ such that, for all $x \in X$,

$$\varphi^*((1 + \delta)(\|u'(x)\| + \delta)) \leq \varphi^*(\|u'(x)\|) + \frac{\varepsilon}{2}.$$

Since φ^* is nondecreasing and $e^{ch} \leq 1 + \delta$, for $h > 0$ sufficiently small, (3.18) follows as soon as we are able to show that

$$\sup_{\|x\| \leq r} \left(\left(\int_{X \times X} |g_{h,x}(y, z)|^q \pi_h^x(dy, dz) \right)^{1/q} - \|u'(x)\| \right) \leq \delta$$

for $h > 0$ sufficiently small. Observe that

$$\begin{aligned} g_{h,x}(y, z) &= \frac{1}{\|y - z\|} \left| \int_0^1 u'(\psi_h(x) + sy + (1-s)z)(y - z) ds \right| \\ &\leq \int_0^1 \|u'(\psi_h(x) + sy + (1-s)z)\| ds. \end{aligned} \quad (3.20)$$

Using (3.20), the uniform continuity of u' on bounded subsets of X , and Remark 2.3 a), there exists some $\delta' > 0$ such that

$$g_{h,x}(y, z) \leq \|u'(x)\| + \frac{\delta}{2}$$

for all $x, y, z \in X$ with $\|x\| \leq r$ and $\|y\| + \|z\| \leq \delta'$, and $h > 0$ sufficiently small. Using Minkowski's inequality and Markov's inequality together with (3.20), we may conclude that

$$\begin{aligned} \left(\int_{X \times X} |g_{h,x}(y, z)|^q \pi_h^x(dy, dz) \right)^{1/q} &\leq \|u'(x)\| + \frac{\delta}{2} + \left(\frac{\|u'\|_\infty}{\delta'} \mathcal{W}_p(\pi_h^x, \delta_0) \right)^{1/q} \\ &\leq \|u'(x)\| + \delta \end{aligned}$$

for all $x \in X$ with $\|x\| \leq r$ and $h > 0$ sufficiently small, since, by Minkowski's inequality and our global assumption (A3),

$$\begin{aligned} \sup_{x \in X} \mathcal{W}_p(\pi_h^x, \delta_0) &\leq \left(\int_X \|y\|^p \mu_h(dy) \right)^{1/p} + \sup_{x \in X} \mathcal{W}_p(\mu_h, \nu_h^x) + \mathcal{W}_p(\mu_h, \delta_0) \\ &\leq 2 \left(\int_X \|y\|^p \mu_h(dy) \right)^{1/p} + bh^\alpha \rightarrow 0 \quad \text{as } h \downarrow 0. \end{aligned}$$

The proof is complete. \square

Before we start with the construction of the distributionally robust version of the family $(T(t))_{t \geq 0}$, cf. Remark 2.3 c), we define the fundamental object for the rest of our study.

Definition 3.12. We say that a family $S = (S(t))_{t \geq 0}$ is a *strongly continuous convex transition semigroup* (on C_b) if the following conditions hold:

- (i) For all $t \geq 0$, $S(t): C_b \rightarrow C_b$ is convex and monotone with $S(t)m = m$, for every constant (function) $m \in \mathbb{R}$, and $\|S(t)u\|_{\text{Lip}} \leq \|u\|_{\text{Lip}}$ for all $u \in \text{Lip}_b$.
- (ii) For all $s, t \geq 0$, $S(t)S(s) = S(t+s)$.
- (iii) $S(t)u_n \rightarrow S(t)u$ uniformly in $t \in [0, s]$ as $n \rightarrow \infty$ for all $s \geq 0$, $(u_n)_{n \in \mathbb{N}} \subset C_b$, and $u \in C_b$ with $u_n \rightarrow u$ as $n \rightarrow \infty$.
- (iv) $S(t)u \rightarrow u$ as $t \downarrow 0$ for all $u \in C_b$.

For a strongly continuous convex transition semigroup S , we define the *generator* $A: D(A) \subset C_b \rightarrow C_b$ of S by

$$D(A) := \left\{ u \in C_b : \frac{S(h)u - u}{h} \rightarrow g \in C_b \text{ as } h \downarrow 0 \right\} \quad \text{and} \quad Au := g$$

for $u \in D(A)$ and $g \in C_b$ with $\frac{S(h)u - u}{h} \rightarrow g$ as $h \downarrow 0$.

Remark 3.13. Choosing $\varphi := \infty \cdot \mathbb{1}_{(0, \infty)}$, the results of this section imply that the family $T = (T(t))_{t \geq 0}$, given by

$$(T(t)u)(x) = \int_X u(\psi_t(x) + y) \mu_t(dy) \quad \text{for all } t \geq 0, u \in C_b, \text{ and } x \in X,$$

is a strongly continuous convex (even linear) transition semigroup. Note that, for all $t \geq 0$, the operator $T(t) = I(t)$ is linear, which shows that the properties (i), (ii), and (iii) of a strongly continuous convex transition semigroup are satisfied. The semigroup property (ii) follows from (2.3), see also Remark 2.3 c). Throughout the remainder of this section, we denote by $B: D(B) \subset C_b \rightarrow C_b$ the generator of T .

For $n \in \mathbb{N}_0$ and $t \geq 0$, we define $I^n(t): C_b \rightarrow C_b$ via the following construction. For $u \in C_b$, we define

$$I^n(t)u := \underbrace{(I(2^{-n}) \cdots I(2^{-n}))}_{k\text{-times}} I(t - k2^{-n})u = I(2^{-n})^k I(t - k2^{-n})u, \quad (3.21)$$

where $k \in \mathbb{N}_0$ denotes the largest natural number with $k2^{-n} \leq t$. Then, by Lemma 3.10,

$$I^{n+1}(t)u \leq I^n(t)u \quad \text{for all } n \in \mathbb{N}_0, t \geq 0, \text{ and } u \in C_b.$$

We define

$$S(t)u := \inf_{n \in \mathbb{N}_0} I^n(t)u \quad \text{for all } t \geq 0 \text{ and } u \in C_b. \quad (3.22)$$

Note that $S(t)u = \lim_{n \rightarrow \infty} I^n(t)u$, since $I^{n+1}(t)u \leq I^n(t)u$, for all $n \in \mathbb{N}$, $t \geq 0$, and $u \in C_b$. The previous results allow us to state the following main result of this section.

Theorem 3.14. *The family S is a strongly continuous convex transition semigroup. Let A denote the generator of S . Then, $D(B) \cap C_b^1 \subset D(A)$ with*

$$(Au)(x) = (Bu)(x) + \varphi^*(\|u'(x)\|) \quad \text{for all } u \in D(B) \cap C_b^1 \text{ and } x \in X.$$

Proof. Let $t \geq 0$. Then, for $n \in \mathbb{N}$ sufficiently large, $I^n(t)u \in C_b$ for all $u \in C_b$ since, by definition, $I^n(t)$ is a finite composition of operators $I(h): C_b \rightarrow C_b$ with $h \in (0, h_0]$, cf. Lemma 3.7. Hence, for $n \in \mathbb{N}$ sufficiently large, the operator $I^n(t): C_b \rightarrow C_b$ is

well-defined, convex, and monotone with $I^n(t)m = m$ for all $m \in \mathbb{R}$ and $\|I^n(t)u\|_{\text{Lip}} \leq e^{ct}\|u\|_{\text{Lip}}$ for all $u \in \text{Lip}_b$ as all these properties carry over from $I(h)$, for $h > 0$, to $I^n(t)$. Since $\|I^n(t)u\|_{\text{Lip}} \leq e^{ct}\|u\|_{\text{Lip}}$ for all $u \in \text{Lip}_b$ and

$$T(t)u \leq S(t)u \leq I^n(t)u \quad \text{for all } n \in \mathbb{N} \text{ and } u \in C_b, \quad (3.23)$$

it follows that $S(t)u \in \text{Lip}_b$ with $\|S(t)u\|_{\text{Lip}} \leq e^{ct}\|u\|_{\text{Lip}}$ for all $u \in \text{Lip}_b$. Next, we verify property (iii) in Definition 3.12. For this, we even show a slightly stronger property. For $i = 1, 2$, let $(u_k^i)_{k \in \mathbb{N}} \subset C_b$ with $(u_k^1 - u_k^2) \rightarrow 0$ as $k \rightarrow \infty$, i.e.,

$$\sup_{n \in \mathbb{N}} \|u_k^1 - u_k^2\|_\infty < \infty \quad \text{and} \quad \lim_{k \rightarrow \infty} \sup_{\|x\| \leq r} |u_k^1(x) - u_k^2(x)| \rightarrow 0 \quad \text{for all } r \geq 0.$$

We prove that, for $s \geq 0$ and $r \geq 0$,

$$\lim_{k \rightarrow \infty} \sup_{t \in [0, s]} \sup_{\|x\| \leq r} |(S(t)u_k^1)(x) - (S(t)u_k^2)(x)| = 0. \quad (3.24)$$

That is, $(S(t)u_k^1 - S(t)u_k^2) \rightarrow 0$ uniformly in $t \in [0, s]$ as $k \rightarrow \infty$ for all $s \geq 0$. To that end, observe that, for $t \geq 0$, $\lambda \in (0, 1)$, and $n \in \mathbb{N}$,³

$$\begin{aligned} S(t)u_k^1 - S(t)u_k^2 &\leq \lambda \left(S(t) \left(u_k^2 + \frac{u_k^1 - u_k^2}{\lambda} \right) - S(t)u_k^2 \right) \leq \lambda \left(I^n(t) \left(u_k^2 + \frac{u_k^1 - u_k^2}{\lambda} \right) - S(t)u_k^2 \right) \\ &\leq \lambda \left(I^n(t) \left(u_k^2 + \frac{u_k^1 - u_k^2}{\lambda} \right) - I^n(t)u_k^2 \right) + 2\lambda \|u_k^2\|_\infty. \end{aligned}$$

The statement now follows from a symmetry argument together with Lemma 3.5 and an appropriate choice of $\lambda \in (0, 1)$ (sufficiently small), $n \in \mathbb{N}$ (sufficiently large in order to achieve $2^{-n} \leq h_0$ in Lemma 3.5), and $k \in \mathbb{N}$ (sufficiently large). Approximating $u \in C_b$ with a sequence $(u_k)_{k \in \mathbb{N}} \subset \text{Lip}_b$, it follows that $S(t): C_b \rightarrow C_b$ is well-defined. Since all other properties stated in (i) of Definition 3.12 directly carry over from $I^n(t)$ to the limit, $S(t)$ satisfies these properties. Moreover, (3.23) together with Lemma 3.7 and Remark 3.13 implies that $S(t)u \rightarrow u$ as $t \downarrow 0$ for all $u \in C_b$.

In order to verify the semigroup property (ii), we fix $s, t \geq 0$ and $u \in \text{Lip}_b$. Consider the set $\mathcal{D} := \{k2^{-n} : k, n \in \mathbb{N}_0\}$ of all dyadic numbers. We first show the semigroup property in the case, where t is a dyadic number, i.e., $t \in \mathcal{D}$. Then, by definition of S ,

$$S(t+s)u = \lim_{n \rightarrow \infty} I^n(t+s)u = \lim_{n \rightarrow \infty} I^n(t)I^n(s)u,$$

where, in the second equality, we used the fact that t is a dyadic number. Due to the monotonicity of S , $I^n(t)I^n(s)u \geq S(t)S(s)u$ for all $n \in \mathbb{N}$. On the other hand, for fixed $k \in \mathbb{N}$,

$$\lim_{n \rightarrow \infty} I^n(t)I^n(s)u \leq \lim_{n \rightarrow \infty} I^n(t)I^k(s)u = S(t)I^k(s)u.$$

Now, since $I^k(s)u$, for $k \in \mathbb{N}$, is a decreasing sequence of monotone functions in Lip_b , it follows that $I^k(s)u \rightarrow S(s)u$. In fact, since $I^k(s)u \in \text{Lip}_b$, there exist bounded and Lipschitz continuous functions $u^k: X \rightarrow \mathbb{R}$ such that $(I^k(s)u)(x) = u^k(Kx)$ for all $x \in X$ and $k \in \mathbb{N}$. Again, since $S(t) \in \text{Lip}_b$, there exists a bounded and Lipschitz continuous function $u_0: X \rightarrow \mathbb{R}$. By Dini's lemma, it follows that

$$\lim_{k \rightarrow \infty} \sup_{\|x\| \leq r} |I^k(s)u(x) - (S(s)u)(x)| = \lim_{k \rightarrow \infty} \sup_{\|x\| \leq r} |u^k(Kx) - u_0(Kx)| = 0$$

for all $r \geq 0$, where we used the fact that K is a compact operator. We have therefore shown that $S(t+s)u = S(t)S(s)u$, when t is a dyadic number and $u \in \text{Lip}_b$. For the

³The first inequality of the following estimate follows from the convexity of the mapping $z \mapsto S(t)(u_k^2 + z) - S(t)u_k^2$ for $z_1 = \frac{u_k^1 - u_k^2}{\lambda}$ and $z_2 = 0$.

general case, we approximate $t \geq 0$ with dyadic numbers $(t_n)_{n \in \mathbb{N}} \subset \mathcal{D}$ and $u \in C_b$ with functions $(u_n)_{n \in \mathbb{N}} \subset \text{Lip}_b$, and obtain, using the properties (iii) and (iv),

$$S(t+s)u = \lim_{n \rightarrow \infty} S(t_n + s)u_n = \lim_{n \rightarrow \infty} S(t_n)S(s)u_n = S(t)S(s)u.$$

Now, let $u \in D(B) \cap C_b^1$ and $g(x) := (Bu)(x) + \varphi^*(\|u'(x)\|)$ for all $x \in X$. By definition of the generator B and Lemma 3.11, we find that $g \in C_b$. Let $a \geq 0$ and $I_a(t)$, for $t \geq 0$, be given as in Lemma 3.3 with $L := \|u\|_{\text{Lip}}$. Let $h \in [0, 1]$ and $0 = t_0 < \dots < t_m = h$ be a partition of the interval $[0, h]$ with $m \in \mathbb{N}$ and $h_k := t_k - t_{k-1}$ for $k \in \{1, \dots, m\}$. Then, by Lemma 3.3 and Lemma 3.10 for the family of operators $(I_a(t))_{t \geq 0}$,

$$\begin{aligned} \frac{I(h_1) \cdots I(h_{k-1})u - I(h_1) \cdots I(h_k)u}{h_k} + g &\leq I_a(h_1) \cdots I_a(h_{k-1}) \left(\frac{u - I(h_k)u}{h_k} \right) + g \\ &\leq I_a(t_{k-1}) \left(\frac{u - I(h_k)u}{h_k} \right) + g. \end{aligned}$$

We thus obtain that

$$\begin{aligned} \frac{u - I(h_1) \cdots I(h_m)u}{h} + g &= \sum_{k=1}^m \frac{h_k}{h} \left(\frac{I(h_1) \cdots I(h_{k-1})u - I(h_1) \cdots I(h_k)u}{h_k} + g \right) \\ &\leq \sum_{k=1}^m \frac{h_k}{h} \left(I_a(t_{k-1}) \left(\frac{u - I(h_k)u}{h_k} \right) + g \right). \end{aligned}$$

Taking the supremum over all finite partitions of the interval $[0, h]$, it follows that

$$\frac{u - S(h)u}{h} + g \leq \sup_{t, h' \in (0, h]} \frac{h'}{h} \left(I_a(t) \left(\frac{u - I(h')u}{h'} \right) + g \right).$$

Using the convergence results for the family of operators $(I_a(t))_{t \geq 0}$ from Lemma 3.5 and Lemma 3.7 together with Lemma 3.11, we observe that the right-hand side converges to zero as $h \downarrow 0$. This together with the observation

$$\frac{S(h)u - u}{h} - g \leq \frac{I(h)u - u}{h} - g,$$

yields that $u \in D(A)$ with $Au = g$. □

Remark 3.15. For $t > 0$, let P_t denote the set of all partitions $\pi = \{t_0, \dots, t_m\}$ with $0 = t_0 < \dots < t_m = t$ and $m \in \mathbb{N}$. For $t > 0$, $\pi \in P_t$, and $u \in C_b$, we define

$$I(\pi)u := I(t_1 - t_0) \cdots I(t_m - t_{m-1})u.$$

Then, for $t > 0$ and $u \in C_b$,

$$S(t)u = \inf_{\pi \in P_t} I(\pi)u.$$

Indeed, by definition of $S(t)$, the inequality $S(t)u \geq \inf_{\pi \in P_t} I(\pi)u$ holds true. In order to establish the other inequality, let $\pi \in P_t$. Then, using the semigroup property of S ,

$$I(\pi)u \geq S(t_1 - t_0) \cdots S(t_m - t_{m-1})u = S(t)u \quad \text{for all } u \in C_b.$$

4. EXTENSIONS AND REMARKS

4.1. On the particular form of the transition kernels. In this section, we discuss the limitations of our setup in terms of the structural assumption (2.4), i.e.,

$$\Xi_t^x = \psi_t(x) + Y_t,$$

on the reference Markov process (Ξ_t^x) . In the following, we explore how this class behaves regarding

- a) deterministic dynamics,
- b) stochastic dynamics arising from affine processes,
- c) the structure of the underlying Banach space and vector operations.

a) Deterministic dynamics. In this part, we consider the case of deterministic dynamics, i.e., $Y_t \equiv 0$ for all $t \geq 0$ in (2.4). Then, the consistency condition (2.3), simplifies to the flow property of the family $(\psi_t)_{t \geq 0}$, i.e.,

$$\psi_t \circ \psi_s = \psi_{t+s} \quad \text{for all } s, t \geq 0.$$

Since $\psi_0(x) = x$ for all $x \in X$, this means that $(\psi_t)_{t \geq 0}$ is a so-called semiflow, and the reference semigroup is the so-called Koopman semigroup related to $(\psi_t)_{t \geq 0}$. In particular, the dynamics comprise all solutions to stationary ordinary differential equations with a global Lipschitz condition as well as solutions to linear partial differential equations as we discuss in Example 5.1 and Example 5.2, respectively.

b) Stochastic dynamics arising from affine processes. In stochastic modeling, random evolutions are oftentimes implemented using stochastic differential equations (SDEs). In this case, structural assumptions are imposed on the coefficients appearing in the SDE. A widely used assumption is the one that the Itô diffusion is an affine process, i.e., up to technicalities, the coefficients appearing in the SDE are affine linear functions depending on the current state of the process, cf. Filipović [20]. In this section, we investigate how the class of Markov processes that admit a decomposition of the form (2.4) relates to the class of affine processes.

Assume that (Ξ_t^x) is an \mathbb{R}^d -valued stochastic process, which is a solution to the SDE

$$d\Xi_t^x = b(\Xi_t^x) dt + \varrho(\Xi_t^x) dW_t, \quad \Xi_0^x = x \in \mathbb{R}^d,$$

where $(W_t)_{t \geq 0}$ is a d -dimensional Brownian motion on a suitably filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ satisfying the usual assumptions, $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is continuous, and $\varrho: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ is continuous and takes values in the set of all symmetric positive semidefinite matrices. Assume that (Ξ_t^x) is an affine process in the sense that b and $a := \varrho^T \varrho$ are affine functions, i.e.,

$$a(x) = \alpha_0 + \sum_{i=1}^d x_i \alpha_i \quad \text{and} \quad b(x) = \beta_0 + \sum_{i=1}^d x_i \beta_i = \beta_0 + \beta x$$

with some matrices $\alpha_0, \dots, \alpha_d \in \mathbb{R}^{d \times d}$, $\beta_0, \dots, \beta_d \in \mathbb{R}^d$ and $\beta = (\beta_1, \dots, \beta_d)^T$, and that, for all $u \in i\mathbb{R}^d$, the Riccati equations

$$\partial_t \Phi(t, u) = \frac{1}{2} \Psi(t, u)^T \alpha_0 \Psi(t, u) + \beta_0^T \Psi(t, u), \quad \Phi(0, u) = 0, \quad (4.1)$$

$$\partial_t \Psi_i(t, u) = \frac{1}{2} \Psi(t, u)^T \alpha_i \Psi(t, u) + \beta_i^T \Psi(t, u), \quad \Psi(0, u) = u, \quad (4.2)$$

have a solution (Φ, Ψ) with $\Psi = (\Psi_1, \dots, \Psi_d)$ and $\operatorname{Re}(\Phi(t, u) + \Psi(t, u)^T x) \leq 0$. Then, the characteristic function φ_{Ξ} of (Ξ_t^x) is of the form

$$\varphi_{\Xi}(t, x, u) := \mathbb{E}(e^{u^T \Xi_t^x}) = e^{\Phi(t, u) + \Psi(t, u)^T x} \quad \text{for all } t \geq 0, x \in \mathbb{R}^d, \text{ and } u \in i\mathbb{R}^d.$$

In this case, our structural assumption that (Ξ_t^x) is of the form (2.4), implies that

$$e^{\Phi(t, u) + \Psi(t, u)^T x} = \mathbb{E}(e^{u^T \Xi_t^x}) = e^{u^T \psi_t(x)} \mathbb{E}(e^{u^T Y_t}) \quad (4.3)$$

for all $t \geq 0$, $x \in \mathbb{R}^d$, and $u \in i\mathbb{R}^d$. For $x = 0$, this yields that $e^{\Phi(t, u)} = e^{u^T \psi_t(0)} \mathbb{E}(e^{u^T Y_t})$ for all $t \geq 0$ and $u \in i\mathbb{R}^d$, which together with (4.3) leads to

$$\Psi(t, u)^T x = u^T (\psi_t(x) - \psi_t(0)) \quad \text{for all } t \geq 0, x \in \mathbb{R}^d, \text{ and } u \in i\mathbb{R}^d.$$

Hence, for all $t \geq 0$, the maps $u \mapsto \Psi(t, u)$ and $x \mapsto \psi_t(x) - \psi_t(0)$ are linear in u and x , respectively. Since (Φ, Ψ) solve the Riccati equations (4.1) and (4.2), it follows that $\alpha_i = 0$ for $i = 1, \dots, d$ and

$$\Psi(t, u) = e^{t\beta} u \quad \text{for all } t \geq 0 \text{ and } u \in i\mathbb{R}^d.$$

As a consequence,

$$d\Xi_t^x = (\beta_0 + \beta X_t^x) dt + \sqrt{\alpha_0} dW_t, \quad \Xi_0^x = x \in \mathbb{R}^d,$$

and

$$\psi_t(x) = e^{t\beta} x + \int_0^t e^{\beta s} \beta_0 ds \quad \text{for all } t \geq 0 \text{ and } x \in \mathbb{R}^d.$$

We conclude that every affine process (Ξ_t^x) , which satisfies (2.4), is, up to a (e.g., logarithmic) transformation, see part c), an Ornstein-Uhlenbeck process.

c) Structure of the underlying Banach space and vector operations. In this part, we show how different vector operations on an a priori arbitrary nonempty set X can be used to incorporate more general versions of the additive operation $+$ in (2.4), for instance a multiplication. For this purpose, let X be a nonempty set, and assume that there exists a bijective function $V: X \rightarrow M$, where M is a separable Banach space. Then, for $x, y \in X$ and $\lambda \in \mathbb{R}$, the operations

$$x +_V y := V^{-1}(V(x) + V(y)), \quad \lambda \cdot_V x := V^{-1}(\lambda V(x)), \quad \text{and} \quad \|x\|_V := \|V(x)\|$$

define an addition, a scalar multiplication, and a norm on X , respectively. By definition, the map $V: X \rightarrow M$ is an isometric isomorphism making X a separable Banach space, as well. In this setup, one may consider flows of the form

$$\Xi_t^x = \psi_t(x) +_V Y_t.$$

For a function $u \in C_b$ with $u \circ V^{-1}: M \rightarrow \mathbb{R}$ differentiable and $x \in X$, the derivative $u'(x): X \rightarrow \mathbb{R}$ (w.r.t. the addition $+_V$ and scalar multiplication \cdot_V) is then given by

$$\begin{aligned} u'(x)z &= \lim_{h \downarrow 0} \frac{u(x +_V (h \cdot_V z)) - u(x)}{h} = \lim_{h \downarrow 0} \frac{u(V^{-1}(V(x) + hV(z))) - u(x)}{h} \\ &= \left(D(u \circ V^{-1})(V(x)) \right) V(z) \quad \text{for all } z \in X. \end{aligned} \quad (4.4)$$

Remark 4.1. Assume that X is an open subset of some Banach space X_0 and that $V: X \rightarrow M$ is a C^1 -diffeomorphism from X (as an open subset of X_0) to M . Then, Equation (4.4) together with the chain rule implies that, for any continuously differentiable function $u: X \rightarrow \mathbb{R}$ and $x, z \in X$,

$$u'(x)z = (Du(x))(DV(x))^{-1}V(z), \quad (4.5)$$

where Du and DV denote the derivatives of u and V on X as a subset of X_0 , respectively. That is, for every $x \in X$, $Du(x)$ is an element of the topological dual space of X_0 and $(DV(x))^{-1}$ is a bounded linear operator from M to X_0 . In this case,

$$\|u'(x)\| = \sup_{\substack{z \in X \\ V(z) \neq 0}} \frac{|u'(x)z|}{\|V(z)\|} \leq \left\| (Du(x))(DV(x))^{-1} \right\|,$$

where the norm appearing on the left-hand side is the operator norm w.r.t. $\|\cdot\|_V$ and the norm on the right-hand side is the operator norm on the dual space of M .

In the following example, we discuss the previously introduced setup for logarithmic transformations.

Example 4.2. Consider the particular choices $X = (0, \infty)$, $M = \mathbb{R}$, and $V(x) = \log x$. Then,

$$x +_V y = xy, \quad \lambda \cdot_V x = x^\lambda, \quad \text{and} \quad \|x\|_V = |\log x|.$$

This leads to dynamics of the form

$$\Xi_t^x = \psi_t(x)Y_t,$$

see Example 5.5 for the case of a geometric Brownian motion. Then, for any continuously differentiable function $u \in C_b$, the derivative u' is given by

$$u'(x) := u'(x)e = \lim_{h \downarrow 0} \frac{u(x +_V (h \cdot_V e)) - u(x)}{h} = \lim_{h \downarrow 0} \frac{u(xe^h) - u(x)}{h} = x \frac{d}{dx} u(x),$$

where e is the Euler constant and $\frac{d}{dx}u$ denotes the usual derivative of u . Note that the equality $u'(x) = x \frac{d}{dx}u(x)$ also follows from (4.5) with $z = e$, since $(\frac{d}{dx} \log(x))^{-1} = x$ for all $x \in (0, \infty)$. In particular, $\|u'(x)\| = |x \frac{d}{dx}u(x)|$ for all $x \in (0, \infty)$.

4.2. Viscosity solutions. Note that Theorem 3.14 shows that the family S is a semigroup, which is, at least formally, closely related to a dynamic programming principle. Moreover, we have shown that the local behaviour at time $t = 0$, given in terms of the generator A , is given by

$$(Au)(x) = (Bu)(x) + \varphi^*(\|u'(x)\|) \quad \text{for } u \in D(B) \cap C_b^1 \text{ and } x \in X,$$

where B is the generator of the reference semigroup T related to the transition kernels $(p_t)_{t \geq 0}$. In view of these results, it is a natural question, whether the semigroup S gives rise to viscosity solutions to the abstract HJB-type differential equation

$$\partial_t u = Au \quad \text{for } t \geq 0. \quad (4.6)$$

Using a standard procedure from Denk et al. [15], Nendel and Röckner [28], or Bartl et al. [6], which can be almost literally adapted to our setup, one can show that the map $t \mapsto S(t)u_0$, for fixed $u_0 \in C_b$, is a viscosity solution to (4.6) using the following notion of a viscosity solution.

Definition 4.3. We say that a function $u: [0, \infty) \rightarrow C_b$ is a *viscosity subsolution* to the abstract differential equation (4.6) if $u: [0, \infty) \rightarrow C_b$ is continuous and

$$(\psi'(t))(x) \leq (A\psi(t))(x)$$

for all $t > 0$, $x \in M$, and every differentiable function $\psi: (0, \infty) \rightarrow C_b$ satisfying $\psi(t) \in D(B) \cap C_b^1$, $(\psi(t))(x) = (u(t))(x)$, and $\psi(s) \geq u(s)$ for all $s > 0$. A function $u: [0, \infty) \rightarrow C_b$ is called a *viscosity supersolution* to (4.6) if $-u$ is a viscosity subsolution to (4.6) with A replaced by $-A(-\cdot)$. A function $u: [0, \infty) \rightarrow C_b$ is called a *viscosity solution* to (4.6) if it is both a viscosity subsolution and a viscosity supersolution to (4.6).

We obtain the following proposition.

Proposition 4.4. *Let $u_0 \in C_b$. Then, the function $u: [0, \infty) \rightarrow C_b$, $t \mapsto S(t)u_0$ is a viscosity solution to the abstract differential equation (4.6) with $u(0) = u_0$.*

The proof is up to the considered function space (here C_b) almost literally the same as in Denk et al. [15, Proposition 5.11], Nendel and Röckner [28, Theorem 4.5], or Bartl et al. [6, Theorem 2.12].

4.3. Unbounded functions and additional continuity properties. Although we consider the space C_b resulting as a closure from bounded and suitably Lipschitz continuous functions, the construction of the semigroup and many statements carry over to the space Lip of all functions $u: X \rightarrow \mathbb{R}$ satisfying (2.1). For such functions, $\|\cdot\|_{\text{Lip}}$ is defined as before. The key ingredient in order to establish the semigroup property of $(S(t))_{t \geq 0}$ on Lip is an additional continuity property of $S(t)$ for all $t \geq 0$. We start with the following observations.

Remark 4.5. Let $u \in \text{Lip}$ and $t \geq 0$. We consider $T(t)u$, $I(t)u$, and $S(t)u$ given by (2.7), (3.1), and (3.22), respectively. Then,

$$|u(x)| \leq C(1 + \|x\|) \quad \text{for all } x \in X, \quad (4.7)$$

with $C := \max\{|u(0)|, \|u\|_{\text{Lip}}\}$, which, by Assumption (A3), implies that

$$|(T(t)u)(x)| \leq C \left(1 + \|\psi_t(x)\| + \int_X \|y\| \mu_t(dy) \right) < \infty.$$

Moreover, by Remark 3.6, it follows that

$$I(t)u \leq T(t)u + t\varphi^*(e^{ct}\|u\|_{\text{Lip}})$$

for all $u \in \text{Lip}$. We thus end up with the estimate

$$T(t)u \leq S(t)u \leq I(t)u \leq T(t)u + t\varphi^*(e^{ct}\|u\|_{\text{Lip}}), \quad (4.8)$$

which shows that the functions $I(t)u$ and $S(t)u$ are well-defined. Moreover, by similar arguments as in Remark 3.2 b) and Theorem 3.14, we obtain

$$\|I(t)u\|_{\text{Lip}} \leq e^{ct}\|u\|_{\text{Lip}} \quad \text{and} \quad \|S(t)u\|_{\text{Lip}} \leq e^{ct}\|u\|_{\text{Lip}},$$

which shows that $I(t)u$ and $S(t)u$ are elements of Lip . Note that, for all $r \geq 0$,

$$\sup_{\|x\| \leq r} |(T(t)u)(x) - u(x)| \leq \|u\|_{\text{Lip}} \left(\sup_{\|x\| \leq r} \|\psi_t(x) - x\| + \int_X \|y\| \mu_t(dy) \right) \rightarrow 0$$

as $t \downarrow 0$. This observation together with (4.8) implies for all $r \geq 0$,

$$\sup_{\|x\| \leq r} |(S(t)u)(x) - u(x)| \rightarrow 0 \quad \text{as } t \downarrow 0.$$

This version of strong continuity motivates the following convergence on Lip.

Notation 4.6. For $(u_k)_{k \in \mathbb{N}}$ in Lip and $u \in \text{Lip}$, we write $u_k \rightrightarrows u$ as $n \rightarrow \infty$ if

$$\sup_{k \in \mathbb{N}} \|u_k\|_{\text{Lip}} < \infty, \quad \text{and} \quad \lim_{k \rightarrow \infty} \sup_{\|x\| \leq r} |u(x) - u_k(x)| = 0 \quad \text{for all } r \geq 0.$$

Proposition 4.7. *Let $s \geq 0$, $(u_k)_{k \in \mathbb{N}}$ in Lip, and $u \in \text{Lip}$ with $u_k \rightrightarrows u$ as $k \rightarrow \infty$. Then, $S(t)u_k \rightrightarrows S(t)u$ uniformly in $t \in [0, s]$ as $k \rightarrow \infty$. Moreover,*

$$S(t+s)u = S(t)S(s)u \quad \text{for all } s, t \geq 0 \text{ and } u \in \text{Lip}.$$

Proof. Let $s \geq 0$. Similar to Lemma 3.3 and the argumentation in Theorem 3.14, since $L := \sup_{k \in \mathbb{N}} \|u_k\|_{\text{Lip}} < \infty$, there exists $a \geq 0$ such that

$$|S(t)u_k - S(t)u| \leq I_a^n(t)|u_k - u| \quad \text{for all } t \in [0, s] \text{ and } k, n \in \mathbb{N}.$$

Therefore, it is sufficient to consider the case, where $u = 0$ and $u_k \geq 0$ for all $k \in \mathbb{N}$, and it remains to show that $I_a(h)u_k \rightrightarrows 0$ as $k \rightarrow \infty$ uniformly in $[0, h_0]$ with $h_0 > 0$ sufficiently small. Let $h_0 > 0$ sufficiently small such that

$$\sup_{h \in [0, h_0]} \|\psi_h(0)\| < \infty \quad \text{and} \quad \sup_{h \in [0, h_0]} \int_X \|y\|^p \mu_h(dy) \leq 1,$$

where we used our global assumptions (A2) and (A3). Then, by Assumption (A1),

$$M_r := \sup_{h \in [0, h_0]} \sup_{\|x\| \leq r} \|\psi_h(x)\| \leq \sup_{h \in [0, h_0]} \|\psi_h(0)\| + e^{ch_0} r < \infty \quad \text{for all } r \geq 0. \quad (4.9)$$

Let $\varepsilon > 0$ and $M > 0$ with $LM^{1-p}(1+ah_0)^p < \varepsilon$. Then,

$$\begin{aligned} 0 &\leq \sup_{h \in [0, h_0]} \sup_{\|x\| \leq r} (I_a(h)u_k)(x) \\ &\leq \sup_{\|\xi\| \leq M+M_r} |u_k(\xi)| + \sup_{h \in [0, h_0]} \sup_{\|x\| \leq r} \sup_{\mathcal{W}_p(\mu_h, \nu) \leq ah} \int_{\{\|z\| > M\}} (u_k(\psi_h(x)) + L\|z\|) \nu(dz) \\ &\leq 2 \sup_{\|\xi\| \leq M+M_r} |u_k(\xi)| + \sup_{h \in [0, h_0]} \sup_{\mathcal{W}_p(\mu_h, \nu) \leq ah} LM^{1-p} \int_{\{\|z\| > M\}} \|z\|^p \nu(dz) \\ &\leq 2 \sup_{\|\xi\| \leq M+M_r} |u_k(\xi)| + LM^{1-p} \sup_{h \in [0, h_0]} (\mathcal{W}_p(\mu_h, \delta_0) + ah)^p \\ &\leq 2 \sup_{\|\xi\| \leq M+M_r} |u_k(\xi)| + LM^{1-p}(1+ah_0)^p < \varepsilon \end{aligned}$$

for $k \in \mathbb{N}$ sufficiently large.

Finally, we show the semigroup property. To do so, for all $k \in \mathbb{N}$, let $\varrho_k: \mathbb{R} \rightarrow \mathbb{R}$ be 1-Lipschitz with $0 \leq \varrho_k \leq 1$, $\varrho_k(w) = 1$ for all $w \in \mathbb{R}$ with $|w| \leq k$, and $\varrho_k(w) = 0$ for all $w \in \mathbb{R}$ with $|w| \geq k+1$. For $u \in \text{Lip}$, let $u_k(x) := u(x)\varrho_k(\|Kx\|)$ for all $x \in X$ and $k \in \mathbb{N}$, so that $(u_k)_{k \in \mathbb{N}} \subset \text{Lip}_b$ and $u_k \rightrightarrows u$ as $k \rightarrow \infty$. Hence, it follows from the first part and Theorem 3.14 that

$$S(t+s)u = \lim_{k \rightarrow \infty} S(t+s)u_k = \lim_{k \rightarrow \infty} S(t)S(s)u_k = S(t)S(s)u. \quad \square$$

4.4. Sensitivity analysis for the robust semigroup. The sensitivity of functionals depending on the distribution μ of a random source X with respect to “small” non-parametric perturbations of the distribution μ has recently received a lot of attention. In this context, sensitivity is typically understood as a suitable derivative w.r.t. the degree of uncertainty expressed in terms of a transport distance, typically, a Wasserstein distance, cf. Bartl et al. [4]. In this subsection, we address this issue in the context of

Wasserstein perturbed semigroups. We show that the generator of the semigroup can, in certain situations, be understood as a sensitivity estimate for infinitesimally small amounts of model uncertainty. On the other hand, we address the numerical implementation of the semigroup S using, e.g., Monte-Carlo methods together with the obtained sensitivity bounds.

Note that the simulation of the semigroup T is of a very simple nature due to our structural assumption on the kernels $(p_t)_{t \geq 0}$. In order to compute a realisation of Ξ_t^x for some $t \geq 0$ and *all* $x \in X$, it is sufficient to perform *one* Monte-Carlo simulation for the random variable Y_t with law μ_t independent of $x \in X$. In a second step, one simulates Ξ_t^x by computing the sum of the deterministic value $\psi_t(x)$ and the simulated random variable Y_t for all $x \in X$.

Recall that $T(t)u \leq S(t)u \leq I(t)u$ for all $u \in \text{Lip}$. Combined with (3.7), we have seen that

$$S(t)u - T(t)u \leq I(t)u - T(t)u \leq t\varphi^*(e^{ct}\|u\|_{\text{Lip}}) \quad \text{for all } u \in \text{Lip}. \quad (4.10)$$

Although this estimate is very rough at points where the function u is (almost) constant, it is very attractive due to its simple nature and delivers a reliable estimate for S and I in terms of T that scales linearly in time. Using the modified version of (4.10) given in Theorem 3.14, for all $r \geq 0$,

$$\sup_{\|x\| \leq r} \left| \frac{(S(h)u)(x) - (T(h)u)(x)}{h} - \varphi^*(\|u'(x)\|) \right| \rightarrow 0 \quad \text{as } h \downarrow 0, \quad (4.11)$$

we also take into account the local behaviour of the function $u \in C_b^1$. Note that this is a tighter but asymptotic bound for infinitesimally small times $h \geq 0$ that still scales linearly in time.

Implicitly, we have also derived sensitivity bounds for the influence of Wasserstein perturbations around the reference semigroup T in terms of the penalisation φ^* . Consider the case, where $\varphi = \infty \cdot \mathbb{1}_{(a, \infty)}$ with some $a \geq 0$. Note that, for $a = 0$, $S = T$. For $a > 0$, the operator I is given by

$$(I(t)u)(x) = \sup_{\mathcal{W}_p(\mu_t, \nu) \leq at} \int_X u(\psi_t(x) + z) \nu(dz) \quad \text{for } t \geq 0, u \in \text{Lip}, \text{ and } x \in X.$$

In this case, the derivative at $t = 0$, (3.13), reformulates to

$$\sup_{\|x\| \leq r} \left| \frac{(I(h)u)(x) - (T(h)u)(x)}{t} - a\|u'(x)\| \right| \rightarrow 0 \quad \text{as } h \downarrow 0, \quad (4.12)$$

for $u \in C_b^1$, and provides a sensitivity estimate for the influence of nonparametric model uncertainty in terms of Wasserstein balls around the baseline model given in terms of the kernels $(p_t)_{t \geq 0}$. Moreover, (4.10) leads to the following version of (4.12):

$$\|I(t)u - T(t)u\|_\infty \leq ta\|u\|_{\text{Lip}} \quad \text{for all } t \geq 0 \text{ and } u \in \text{Lip}.$$

4.5. Relation to parametric uncertainty and dynamic risk measures. Closely related to the topic of sensitivity analysis that we discussed in the previous subsection, a lot of interest has recently been paid to reducing optimisation procedures under nonparametric uncertainty to low-dimensional optimisation problems, cf. Bartl et al. [5], Mohajerin Esfahani and Kuhn [27], Zhao and Guan [43], and Blanchet and Murthy [8]. In this subsection, we show how, in special yet relevant cases, the nonparametric uncertainty in the semigroup S reduces to a simple form of parametric uncertainty. This

leads to a low-dimensional optimisation scheme instead of the infinite-dimensional optimisation problem related to $I(t)$. More precisely, we discuss the relation between the Wasserstein perturbed semigroup S and the Nisio semigroup, cf. [30], w.r.t. drift uncertainty of the related stochastic process (Ξ_t^x) .

Throughout this subsection, we work under the assumption that $\psi_t(x) = x$ for all $t \geq 0$ and $x \in X$. For $t \geq 0$, $u \in C_b$, and $x \in X$, we consider

$$(E(t)u)(x) := \sup_{\theta \in X} \left(\int_X u(x + y + t\theta) \mu_t(dy) - t\varphi(e^{-ct}\|\theta\|) \right). \quad (4.13)$$

Since $\mathcal{W}_p(\mu_t, \mu_t * \delta_{t\theta}) \leq t\|\theta\|$ for all $t \geq 0$ and $\theta \in X$, it follows that

$$T(t)u \leq E(t)u \leq I(t)u \quad \text{for all } t \geq 0. \quad (4.14)$$

Moreover, one readily verifies that $E(t): \text{Lip}_b \rightarrow \text{Lip}_b$ with

$$\|E(t)u\|_{\text{Lip}} \leq e^{ct}\|u\|_{\text{Lip}} \quad \text{and} \quad \|E(t)u\|_{\infty} \leq \|u\|_{\infty} \quad \text{for all } u \in \text{Lip}_b.$$

Similar as in the proof of Lemma 3.3, one sees that also $E(h): C_b \rightarrow C_b$ is well-defined and 1-Lipschitz for sufficiently small $h \geq 0$. For $u \in C_b$ and $\pi = \{t_0, \dots, t_m\} \in P_t$ with $m \in \mathbb{N}$ and $0 = t_0 < \dots < t_m = t$ (see Remark 3.15), we define

$$E(\pi)u := E(t_1 - t_0) \cdots E(t_m - t_{m-1})u.$$

Let $u \in C_b$. Then, (4.14) implies that

$$E(\pi)u \leq I(\pi)u. \quad (4.15)$$

For all $n \in \mathbb{N}$, we consider the partition $\pi_n := \{j2^{-n}: j = 0, \dots, k\} \cup \{t\} \in P_t$ with $k \in \mathbb{N}_0$ being the largest natural number such that $k2^{-n} \leq t$. Then,

$$E(t)u \leq E(\pi_n)u \leq E(\pi_{n+1})u \quad \text{and} \quad I(\pi_n)u = I^n(t)u$$

for all $n \in \mathbb{N}$. Hence, using (4.15),

$$E(t)u \leq \sup_{n \in \mathbb{N}} E(\pi_n)u = \lim_{n \rightarrow \infty} E(\pi_n)u \leq \lim_{n \rightarrow \infty} I^n(t)u = S(t)u.$$

Now, let $\pi = \{t_0, \dots, t_m\} \in P_t$ with $m \in \mathbb{N}$ and $0 = t_0 < \dots < t_m$. Then,

$$E(\pi)u = E(t_1 - t_0) \cdots E(t_m - t_{m-1})u \leq S(t_1 - t_0) \cdots S(t_m - t_{m-1})u = S(t)u$$

This shows that

$$N(t)u := \sup_{\pi \in P_t} E(\pi)u \leq S(t)u$$

for all $t \geq 0$ and $u \in C_b$. Proceeding as in Nendel and Röckner [28], one sees that $N = (N(t))_{t \geq 0}$ is a strongly continuous convex transition semigroup, where the strong continuity (property (iii) in Definition 3.12) follows from $T(t)u \leq N(t)u \leq S(t)u$ for all $u \in C_b$.

Let $u \in D(B) \cap C_b^1$ and $\varepsilon > 0$. It follows from the proof of Lemma 3.11 that, for sufficiently small $h > 0$,

$$\varphi^*(\|u'(x)\|) - \varepsilon \leq \frac{(E(h)u)(x) - (T(h)u)(x)}{h}.$$

Since $E(t)u \leq N(t)u \leq S(t)u \leq I(t)u$, for $t \geq 0$, it follows that

$$\left\| \frac{N(t)u - u}{u} - Au \right\|_{\infty} \rightarrow 0 \quad \text{as } t \downarrow 0,$$

showing that the generator of S and N coincide on $D(B) \cap C_b^1$. Similar as in Section 4.2, one sees that the map $t \mapsto N(t)u_0$, for $u_0 \in C_b$, is a viscosity solution to the abstract PDE

$$\partial_t u = Au \quad \text{with} \quad u(0) = u_0.$$

In certain cases, e.g., if $X = \mathbb{R}$ and $B = \frac{\sigma^2}{2} \partial_{xx}$ (see Section 5.3) or if $X = (0, \infty)$ and $B = \frac{\sigma^2 x^2}{2} \partial_{xx}$ (see Section 5.5), it follows from standard comparison results for viscosity solutions to HJB equations, cf. Crandall et al. [13], that

$$S(t)u = N(t)u \quad \text{for all } t \geq 0 \text{ and } u \in C_b. \quad (4.16)$$

Approximating $u \in \text{Lip}$ with $(u_n)_{n \in \mathbb{N}}$ in C_b as in Section 4.3, it follows that (4.18) holds for all $u \in \text{Lip}$. We therefore see that the nonparametric uncertainty captured by $I(t)$ collapses to pure drift uncertainty as $t \downarrow 0$.

Remark 4.8. We discuss the link to dynamic risk measures. For $u \in C_b$, the value $(I(t)u)(x)$ in (3.1) can be viewed as the conditional risk in terms of a capital requirement of the position $u(\Xi_t^x)$ under a convex risk measure conditioned on $\{\Xi_0^x = x\}$.⁴ In this context, for $t = k2^{-n}$, we interpret $I^n(t)u = I(2^{-n})^k u$ in (3.21) as the k -times composition of the conditional risk measures $I(2^{-n})$, which is a time-consistent dynamic risk measure at the discrete time points $\{0, 2^{-n}, \dots, k2^{-n}\}$. By making the partition progressively finer, we obtain, in the limit, a dynamic risk measure in continuous time. Here, $(S(t)u)(x)$ is the dynamic risk of the position $u(\Xi_t^x)$ conditioned on $\{\Xi_0^x = x\}$, and the time-consistency is ensured by the semigroup property. As a result of the Lemma 3.10, we obtain that finer partitions lead to smaller capital requirements, i.e., $(I^{n+1}u)(x) \leq (I^n u)(x)$ for all $n \in \mathbb{N}$. The previous discussion leads to the following insight: although we start with a family of risk measures $(I^n(t))_{n \in \mathbb{N}}$ that bear nonparametric uncertainty, in the limit, we obtain a dynamic risk measure $S(t)$ that incorporates only parametric drift-uncertainty. In the particular case of a Brownian filtration, the limiting dynamic risk measures are represented as g -expectations, see Rosazza Gianin [36].

Moreover, due to Lemma 3.10, we obtain that the limiting dynamic risk measure is dominated by the one-step risk measure, i.e., $S(t)u \leq I(t)u$ for all $u \in C_b$ and $t \geq 0$. In particular, the inequality in (4.10) allows to simulate bounds for dynamic risk measures (g -expectations) in terms of the reference dynamics and the convex conjugate φ^* . In the case $X = \mathbb{R}$ with $\psi_t(x) = x$ for all $x \in \mathbb{R}$, $\mu_t = \mathcal{N}(0, t)$, and $\varphi(v) = \frac{v^2}{2}$, we obtain the entropic risk measure

$$(S(t)u)(x) = \log \int_{\mathbb{R}} \exp(u(x+y)) \mu_t(dy),$$

which is dominated by the respective one-step conditional risk measure $I(t)u$, see also Blessing et al. [9, Corollary 5.12] for a discussion on the corresponding Talagrand T_2 inequality.

4.6. Martingale constraints. Motivated by risk-neutral pricing, we consider the case where $\int_X y \mu_t(dy) = 0$ for all $t \geq 0$. In line with Bartl et al. [4, Section 3.1], one can add a martingale constraint to the operator $I(t)$, for $t \geq 0$, by considering

$$(I_{\text{Mart}}(t)u)(x) := \sup_{\nu \in \mathcal{P}_p^0(X)} \left(\int_X u(\psi_t(x) + z) \nu(dz) - \varphi_t(\mathcal{W}_p(\mu_t, \nu)) \right)$$

⁴Instead of $I(t)$, one could even consider the entropic version $(\tilde{I}(t)u)(x) := (\log I(t)e^u)(x)$, where the position $u(\Xi_t^x)$ is weighted with the exponential loss function.

for all $t \geq 0$, $u \in C_b$, and $x \in X$, where $\mathcal{P}_p^0(X)$ denotes the set of all $\nu \in \mathcal{P}_p(X)$ with $\int_X z \nu(dz) = 0$. The proof of Lemma 3.11 shows that, for $u \in C_b^1$,

$$\frac{I_{\text{Mart}}(h)u - T(h)u}{h} \rightarrow 0 \quad \text{as } h \downarrow 0.$$

In fact, for all $h \in [0, 1]$ and $x \in X$, let $\nu_h^x \in \mathcal{P}_p^0(X)$ with $\mathcal{W}_p(\mu_h, \nu_h^x) \leq bh^\alpha$ and

$$(I_{\text{Mart}}(h)u)(x) \leq \varepsilon h + \int_X u(\psi_h(x) + z) \nu_h^x(dz) - \varphi_h(\mathcal{W}_p(\mu_h, \nu_h^x)). \quad (4.17)$$

For each $h \in [0, 1]$ and $x \in X$, let $\pi_h^x \in \mathcal{P}_p(X \times X)$ be an optimal coupling between μ_h and ν_h^x . Then, using the fundamental theorem of infinitesimal calculus and the fact that $\nu_h^x \in \mathcal{P}_p^0(X)$, we obtain that

$$\begin{aligned} & \frac{(I_{\text{Mart}}(h)u)(x) - (T(h)u)(x)}{h} \\ & \leq \frac{1}{h} \int_{X \times X} \int_0^1 u'(\psi_h(x) + (1-s)y + sz)(z-y) ds \pi_h^x(dy, dz) - \frac{\varphi_h(\mathcal{W}_p(\mu_h, \nu_h^x))}{h} + \varepsilon \\ & = \frac{1}{h} \int_{X \times X} \int_0^1 \left(u'(\psi_h(x) + (1-s)y + sz) - u'(\psi_h(x)) \right) (z-y) ds \pi_h^x(dy, dz) \\ & \quad - \varphi \left(e^{-ch} \frac{\mathcal{W}_p(\mu_h, \nu_h^x)}{h} \right) + \varepsilon \\ & \leq \left(\int_{X \times X} \int_0^1 \|u'(\psi_h(x) + (1-s)y + sz) - u'(\psi_h(x))\|^q ds \pi_h^x(dy, dz) \right)^{1/q} \\ & \quad \times \frac{\mathcal{W}_p(\mu_h, \nu_h^x)}{h} - \varphi \left(e^{-ch} \frac{\mathcal{W}_p(\mu_h, \nu_h^x)}{h} \right) + \varepsilon \\ & \leq \varphi^* \left(e^{ch} \left(\int_{X \times X} \int_0^1 \|u'(\psi_h(x) + (1-s)y + sz) - u'(\psi_h(x))\|^q ds \pi_h^x(dy, dz) \right)^{1/q} \right) \\ & \quad + \varepsilon \end{aligned}$$

for all $h \in [0, 1]$ and $x \in X$, where $q = \frac{p}{p-1}$ is the conjugate exponent of p . Proceeding as in the proof of Lemma 3.11, we find that

$$\sup_{\|x\| \leq r} \int_{X \times X} \int_0^1 \|u'(\psi_h(x) + (1-s)y + sz) - u'(\psi_h(x))\|^q ds \pi_h^x(dy, dz) \rightarrow 0 \quad \text{as } h \downarrow 0.$$

In particular,

$$\sup_{\|x\| \leq r} \frac{(I_{\text{Mart}}(h)u)(x) - (T(h)u)(x)}{h} \rightarrow 0 \quad \text{as } h \downarrow 0.$$

Similar as in Section 3, the family $(I_{\text{Mart}}(t))_{t \geq 0}$ gives rise to a semigroup $(S_{\text{Mart}}(t))_{t \geq 0}$. With an analogous argumentation as in the proof of Theorem 3.14, one sees that

$$\frac{S_{\text{Mart}}(h)u - u}{h} \rightarrow Bu \quad \text{as } h \downarrow 0.$$

Using comparison results as in the previous section, we find that, in certain cases, e.g., if $X = \mathbb{R}$ and $B = \frac{\sigma^2}{2} \partial_{xx}$ or if $X = (0, \infty)$ and $B = \frac{\sigma^2 x^2}{2} \partial_{xx}$,

$$S_{\text{Mart}}(t)u = T(t)u \quad \text{for all } t \geq 0 \text{ and } u \in C_b. \quad (4.18)$$

In particular, under our assumption that the level of uncertainty is proportional to the time interval, we see that S_{Mart} as a limit $I_{\text{Mart}}(\pi)$ over progressively finer time partitions π coincides with the transition semigroup T of the reference Markov process.

5. EXAMPLES

5.1. Koopman semigroups. Let $X = \mathbb{R}^d$ with $d \in \mathbb{N}$. In this example, we consider the case of a deterministic drift that might be susceptible to an uncertain random noise. More precisely, for a fixed Lipschitz continuous function $F: X \rightarrow X$ with Lipschitz constant $c \geq 0$, we consider the initial value problem

$$\begin{aligned} \partial_t x(t) &= F(x(t)) \quad \text{for all } t \in \mathbb{R}, \\ x(0) &= x \in X. \end{aligned} \tag{5.1}$$

For $x \in X$, we define $(\psi_t(x))_{t \geq 0}$ as the unique solution to the above initial value problem. As a reference model, we choose the deterministic dynamics

$$\Xi_t^x = \psi_t(x) \quad \text{for } t \geq 0 \text{ and } x \in X.$$

We assume that a random noise Z with uncertain distribution $\nu \in \mathcal{P}_p(X)$ enters in an additive way leading to the uncertain stochastic dynamics

$$\psi_t(x) + Z \quad \text{for } t \geq 0 \text{ and } x \in X,$$

where Z can be seen as a generalisation of a known random source, such as a Brownian motion. In the setup of our previous discussion, we thus consider the flow $\mu_t = \delta_0$, for $t \geq 0$, where δ_0 is the Dirac measure with barycenter 0. Assumption (A3) is therefore trivially satisfied for all $p \geq 1$. The consistency condition (2.3) is exactly the flow property of solutions to the ODE (5.1). Using Gronwall's lemma,

$$\|\psi_t(x_1) - \psi_t(x_2)\| \leq e^{ct} \|x_1 - x_2\| \quad \text{for all } t \geq 0 \text{ and } x_1, x_2 \in X, \tag{5.2}$$

which shows that Assumption (A1) is satisfied. Another application of Gronwall's lemma shows that

$$\|\psi_h(x) - x\| \leq h e^{ch} \|F(x)\| \leq h e^{ch} (\|F(0)\| + c\|x\|) \tag{5.3}$$

for all $h \geq 0$ and $x \in X$. In particular, for every $r \geq 0$,

$$\sup_{\|x\| \leq r} \|\psi_h(x) - x\| \leq h e^{ch} (\|F(0)\| + cr) \rightarrow 0 \quad \text{as } h \downarrow 0.$$

Moreover, the estimate in (5.3) together with the inverse triangular inequality implies that

$$\|x\| - \|\psi_h(x)\| \leq \|\psi_h(x) - x\| \leq h e^{ch} (\|F(0)\| + c\|x\|).$$

Hence, for $h_0 > 0$ with $ch_0 e^{ch_0} < 1$,

$$\inf_{h \in [0, h_0]} \|\psi_h(x)\| \geq \|x\| (1 - ch_0 e^{ch_0}) - h_0 e^{ch_0} \|F(0)\| \rightarrow \infty \quad \text{as } \|x\| \rightarrow \infty,$$

which shows that Assumption (A2) is satisfied as well. For $t \geq 0$, $x \in X$, and $u \in C_b$, we have

$$(I(t)u)(x) := \sup_{\nu \in \mathcal{P}_p(X)} \left(\int_X u(\psi_t(x) + z) \nu(dz) - \varphi_t \left(\left(\int_X \|z\|^p \nu(dz) \right)^{1/p} \right) \right).$$

Then, the semigroup $S = (S(t))_{t \geq 0}$ can be seen as transition kernels of solutions to the ODE (5.1) with an additive robust noise. The related Kolmogorov equation is a nonlinear PDE, and given by

$$\partial_t u(t, x) = (D_x u(t, x))(F(x)) + \varphi^*(\|D_x u(t, x)\|), \quad (5.4)$$

where D_x denotes the Fréchet derivative in the space variable. Observe that, in contrast to an additive noise in terms of a Brownian motion, no second-order terms appear in Equation (5.4).

5.2. Semiflows arising from linear dynamics. Let X be a separable Banach space, $m \in X$, and $G: D(G) \subset X \rightarrow X$ be the generator of a C_0 -semigroup $P = (P(t))_{t \geq 0}$ on X with compact resolvent.⁵ A typical example for such a generator G is given by the Dirichlet Laplacian in $L^q(\Omega)$, where $\Omega \subset \mathbb{R}^d$ is a bounded domain and $1 \leq q < \infty$, or any generator of a compact C_0 -semigroup, cf. [31, Section 2.3]. For the Dirichlet Laplacian, the compactness of the resolvent follows from the Rellich-Kondrachov theorem. For $t \geq 0$ and $x \in X$, let

$$\psi_t(x) := P(t)x$$

be the unique solution to the abstract Cauchy problem

$$\partial_t x(t) = Gx(t) \quad \text{for } t \geq 0, \quad x(0) = x.$$

Since P is a C_0 -semigroup on X , there exist constants $M \geq 0$ and $\omega \geq 0$ such that $\|P(t)x\| \leq Me^{\omega t}\|x\|$ for all $x \in X$ and $t \geq 0$. By passing to the equivalent norm

$$\|x\|_P := \sup_{t \geq 0} e^{-\omega t} \|P(t)x\| \quad \text{for } x \in X,$$

we may, w.l.o.g., assume that $M = 1$. Since $\psi_t(x_1) - \psi_t(x_2) = P(t)(x_1 - x_2)$ for all $x_1, x_2 \in X$, it follows that Assumption (A1) is satisfied with $c := \max\{\omega, 0\}$. Choosing $K := (\lambda - G)^{-1}$ with $\lambda := 1 + \omega$, it follows that $K\psi_t(x) = \psi_t(Kx)$ for all $x \in X$ and $t \geq 0$, since the resolvent commutes with the semigroup. Moreover, the resolvent identity⁶ $GKx = \lambda Kx - x$ together with $\|Kx\| \leq \frac{1}{\lambda - \omega} = 1$ for all $x \in X$ yields

$$\sup_{\|x\| \leq r} \|\psi_h(Kx) - Kx\| \leq he^{\omega h} \sup_{\|x\| \leq r} \|GKx\| \leq he^{\omega h} (2 + \omega)r \rightarrow 0$$

as $h \downarrow 0$ for all $r \geq 0$. We choose $Y_t := \int_0^t P(s)m \, ds$ for $t \geq 0$, so that the assumptions (A1), (A2), and (A3) are satisfied. Then, the nonlinear transition semigroup yields a solution to the nonlinear PDE

$$\partial_t u(t, x) = (D_x u(t, x))(Gx + m) + \varphi^*(\|D_x u(t, x)\|). \quad (5.5)$$

5.3. Lévy processes. In this example, we consider the case, where (Ξ_t^x) is a Lévy process taking values in a separable Hilbert space X . For a detailed discussion in the finite-dimensional case, we refer to Bartl et al. [6]. Throughout this example, we consider $\psi_t(x) = x$ for all $t \geq 0$ and $x \in X$. Since ψ_t is the identity for all $t \geq 0$, the compact operator K can be chosen arbitrarily. Furthermore, let $\mu = (\mu_t)_{t \geq 0}$ in $\mathcal{P}_p(X)$ be a family of infinitely divisible distributions with

$$\lim_{h \downarrow 0} \int_X \|y\|^p \mu_h(dy) = 0,$$

⁵Recall that a generator G of a C_0 -semigroup on X has compact resolvent if $(\lambda - G)^{-1}: X \rightarrow X$ is a compact operator for some (or equivalently all) $\lambda > \omega$, where $\omega \in \mathbb{R}$ is the growth bound of the related semigroup.

⁶Note that $Kx \in D(G)$.

so that Assumption (A3) holds. A typical example for μ is the distribution of a Brownian motion with trace class covariance operator. Since ψ_t is the identity for all $t \geq 0$, the Assumptions (A1) and (A2) are trivially satisfied. In this case, for $t \geq 0$, $x \in X$, and $u \in C_b$, the operator $I(t)$ is given by

$$(I(t)u)(x) = \sup_{\nu \in \mathcal{P}_p(X)} \left(\int_X u(x+z) \nu(dz) - \varphi_t(\mathcal{W}_p(\mu_t, \nu)) \right).$$

Let B_μ be the generator of the Lévy process related to the family μ . Then, the nonlinear semigroup $(S(t))_{t \geq 0}$ can be computed by solving the PDE

$$\partial_t u(t, x) = B_\mu u(t, x) + \varphi^*(\|D_x u(t, x)\|). \quad (5.6)$$

As an illustration, we consider the case, where $H = \mathbb{R}$, μ is the law of a Brownian motion (starting in 0) with standard deviation $\sigma > 0$, and $\varphi = \infty \cdot \mathbf{1}_{(a, \infty)}$ with $a \geq 0$. In this case, $B_\mu = \frac{\sigma^2}{2} \partial_{xx}$, $\varphi^* = a|\cdot|$, and the PDE (5.6) simplifies to

$$\partial_t u(t, x) = \frac{\sigma^2}{2} \partial_{xx} u(t, x) + a|\partial_x u(t, x)|. \quad (5.7)$$

5.4. Ornstein-Uhlenbeck processes. We consider a separable Hilbert space X . Let $G: D(G) \subset X \rightarrow X$ be the generator of a C_0 -semigroup $P = (P(t))_{t \geq 0}$ on X with compact resolvent, and $m \in X$. We modify the approach in Example 5.2 by adding a noise in terms of a Brownian motion $(W_t)_{t \geq 0}$ taking values in X with a trace class covariance operator $C: X \rightarrow X$ on some filtered probability space, satisfying the usual assumptions. For $t \geq 0$ and $x \in X$, let $\psi_t(x) := P(t)x$ be the unique solution to the abstract Cauchy problem

$$\partial_t x(t) = Gx(t), \quad \text{for } t \geq 0, \quad x(0) = x.$$

Choosing $K := (\lambda - G)^{-1}$ with $\lambda > 0$ sufficiently large, we have seen in Example 5.2 that the Assumptions (A1) and (A2) are satisfied. Then, we consider the Ornstein-Uhlenbeck process (Ξ_t^x) given by

$$\Xi_t^x = \psi_t(x) + \int_0^t P(s)m ds + \int_0^t P(t-s) dW_s, \quad \text{for } t \geq 0 \text{ and } x \in X.$$

In this case, μ_t is the law of the stochastic convolution $\int_0^t P(s)m ds + \int_0^t P(t-s) dW_s$. By definition, the process (Ξ_t^x) is a mild solution to the SPDE

$$d\Xi_t^x = (G\Xi_t^x + m) dt + dW_t, \quad X_0^x = x.$$

By the Burkholder-Davis-Gundy inequality, Assumption (A3) is satisfied. Here, the operator $I(t)$ is given by

$$(I(t)u)(x) = \sup_{\nu \in \mathcal{P}_p(X)} \left(\int_X u(\psi_t(x) + z) \nu(dz) - \varphi_t(\mathcal{W}_p(\mu_t, \nu)) \right)$$

for $t \geq 0$, $u \in C_b$, and $x \in X$. The related semigroup $(S(t))_{t \geq 0}$ can be computed by solving the semilinear PDE

$$\partial_t u(t, x) = \langle Gx + m, D_x u(t, x) \rangle + \frac{1}{2} \text{tr}(CD_{xx}u(t, x)) + \varphi^*(\|D_x u(t, x)\|). \quad (5.8)$$

We point out that, in comparison to (5.5), the resulting PDE incorporates second-order terms. This is due to the consideration of the a priori random noise in terms of a Brownian motion. Roughly speaking, (5.8) shows how the two forms of noise, aleatoric and epistemic, enter the equation. The aleatoric (random) noise leads to the term

$\frac{1}{2}\text{tr}(CD_{xx}u(t, x))$ and the epistemic (uncertain) noise leads to the term $\varphi^*(\|D_x u(t, x)\|)$. For a detailed discussion of (5.8) in the context of parametric uncertainty, we refer to Nendel and Röckner [28].

5.5. Geometric Brownian motion. We follow Example 4.2, and consider the case, where $X = (0, \infty)$, endowed with the multiplication as an “additive” operation and the norm $\|x\| := |\log(x)|$ for all $x \in (0, \infty)$. Let $\psi_t(x) := x$, for $x \in (0, \infty)$, and μ_t be given as the law of the random variable

$$Y_t := \exp\left(\left(\alpha - \frac{\sigma^2}{2}\right)t + \sigma W_t\right),$$

where $\alpha \in \mathbb{R}$, $\sigma \geq 0$ and $(W_t)_{t \geq 0}$ is a standard Brownian motion on some probability space. We thus consider dynamics (Ξ_t^x) of the form

$$\Xi_t^x = xY_t \quad \text{for } t \geq 0 \text{ and } x \in (0, \infty).$$

Trivially, the Assumptions (A1) and (A2) are satisfied since ψ_t is the identity for all $t \geq 0$. Assumption (A3) is satisfied since

$$\mathbb{E}(\|Y_h\|^p)^{1/p} = \mathbb{E}(|\log(Y_h)|^p)^{1/p} \leq h|\alpha - \frac{\sigma^2}{2}| + \sigma\mathbb{E}(|W_h|^p)^{1/p} \rightarrow 0 \quad \text{as } h \downarrow 0.$$

Here, for $t \geq 0$, $x \in X$, and $u \in C_b$, the operator $I(t)$ is given by

$$(I(t)u)(x) = \sup_{\nu \in \mathcal{P}_p(X)} \left(\int_X u(xy) \nu(dz) - \varphi_t(\mathcal{W}_p(\mu_t, \nu)) \right),$$

where the considered metric \mathcal{W}_p is a logarithmic Wasserstein distance.

REFERENCES

- [1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
- [2] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. All adapted topologies are equal. *Probab. Theory Related Fields*, 178(3-4):1125–1172, 2020.
- [3] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and J. Wiesel. Estimating processes in adapted Wasserstein distance. Forthcoming in *Annals of Applied Probability*, 2021.
- [4] D. Bartl, S. Drapeau, J. Obłój, and J. Wiesel. Sensitivity analysis of Wasserstein distributionally robust optimization problems. *Proc. A.*, 477(2256):Paper No. 20210176, 19, 2021.
- [5] D. Bartl, S. Drapeau, and L. Tangpi. Computational aspects of robust optimized certainty equivalents and option pricing. *Math. Finance*, 30(1):287–309, 2020.
- [6] D. Bartl, S. Eckstein, and M. Kupper. Limits of random walks with distributionally robust transition probabilities. *Electron. Commun. Probab.*, 26:Paper No. 28, 13, 2021.
- [7] D. P. Bertsekas and S. Shreve. *Stochastic Optimal Control: The Discrete Time Case*. 2004.
- [8] J. Blanchet and K. Murthy. Quantifying distributional model risk via optimal transport. *Math. Oper. Res.*, 44(2):565–600, 2019.
- [9] J. Blessing, R. Denk, M. Kupper, and M. Nendel. Convex monotone semigroups and their generators with respect to γ -convergence. *arXiv preprint arXiv:2202.08653*, 2022.
- [10] J. Blessing and M. Kupper. Nonlinear semigroups built on generating families and their Lipschitz sets. Forthcoming in *Potential Analysis*, 2022.
- [11] P. Cheridito, H. M. Soner, N. Touzi, and N. Victoir. Second-order backward stochastic differential equations and fully nonlinear parabolic PDEs. *Comm. Pure Appl. Math.*, 60(7):1081–1110, 2007.
- [12] F. Coquet, Y. Hu, J. Mémin, and S. Peng. Filtration-consistent nonlinear expectations and related g -expectations. *Probab. Theory Related Fields*, 123(1):1–27, 2002.
- [13] M. G. Crandall, H. Ishii, and P.-L. Lions. User’s guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992.
- [14] G. de Cooman, F. Hermans, and E. Quaeghebeur. Imprecise Markov chains and their limit behavior. *Probab. Engrg. Inform. Sci.*, 23(4):597–635, 2009.
- [15] R. Denk, M. Kupper, and M. Nendel. A semigroup approach to nonlinear Lévy processes. *Stochastic Processes and their Applications*, 130(3):1616–1642, 2020.

- [16] D. Dentcheva and A. Ruszczyński. Time-coherent risk measures for continuous-time Markov chains. *SIAM J. Financial Math.*, 9(2):690–715, 2018.
- [17] S. Eckstein. Extended Laplace principle for empirical measures of a Markov chain. *Adv. in Appl. Probab.*, 51(1):136–167, 2019.
- [18] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Math. Finance*, 7(1):1–71, 1997.
- [19] T. Fadina, A. Neufeld, and T. Schmidt. Affine processes under parameter uncertainty. *Probab. Uncertain. Quant. Risk*, 4:Paper No. 1, 35, 2019.
- [20] D. Filipović. *Term-structure models*. Springer Finance. Springer-Verlag, Berlin, 2009. A graduate course.
- [21] R. Gao and A. J. Kleywegt. Distributionally Robust Stochastic Optimization with Wasserstein Distance. *Preprint*, arXiv:1604.02199, 2016.
- [22] D. J. Hartfiel. *Markov set-chains*, volume 1695 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1998.
- [23] M. Hu and S. Peng. G -Lévy Processes under Sublinear Expectations. *Preprint*, arXiv:0911.3533, 2009.
- [24] N. Kazi-Tani, D. Possamaï, and C. Zhou. Second order BSDEs with jumps: existence and probabilistic representation for fully-nonlinear PIDEs. *Electron. J. Probab.*, 20:1–31, 2015.
- [25] T. Krak, J. De Bock, and A. Siebes. Imprecise continuous-time Markov chains. *Internat. J. Approx. Reason.*, 88:452–528, 2017.
- [26] F. Kühn. Viscosity solutions to Hamilton-Jacobi-Bellman equations associated with sublinear Lévy(-type) processes. *ALEA Lat. Am. J. Probab. Math. Stat.*, 16(1):531–559, 2019.
- [27] P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Math. Program.*, 171(1-2, Ser. A):115–166, 2018.
- [28] M. Nendel and M. Röckner. Upper envelopes of families of Feller semigroups and viscosity solutions to a class of nonlinear Cauchy problems. *SIAM J. Control Optim.*, 59(6):4400–4428, 2021.
- [29] A. Neufeld and M. Nutz. Nonlinear Lévy processes and their characteristics. *Trans. Amer. Math. Soc.*, 369(1):69–95, 2017.
- [30] M. Nisio. On a non-linear semi-group attached to stochastic optimal control. *Publ. Res. Inst. Math. Sci.*, 12(2):513–537, 1976/77.
- [31] A. Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983.
- [32] S. Peng. G -expectation, G -Brownian motion and related stochastic calculus of Itô type. In *Stochastic analysis and applications*, volume 2 of *Abel Symp.*, pages 541–567. Springer, Berlin, 2007.
- [33] G. Pflug and D. Wozabal. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442, 2007.
- [34] G. C. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM J. Optim.*, 22(1):1–23, 2012.
- [35] G. C. Pflug and A. Pichler. *Multistage stochastic optimization*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2014.
- [36] E. Rosazza Gianin. Risk measures via g -expectations. *Insurance Math. Econom.*, 39(1):19–34, 2006.
- [37] D. Rudolf and N. Schweizer. Perturbation theory for Markov chains via Wasserstein distance. *Bernoulli*, 24(4A):2610–2639, 2018.
- [38] H. M. Soner, N. Touzi, and J. Zhang. Wellposedness of second order backward SDEs. *Probab. Theory Related Fields*, 153(1-2):149–190, 2012.
- [39] C. Villani. *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media, 2008.
- [40] D. Škulj. Discrete time Markov chains with interval probabilities. *Internat. J. Approx. Reason.*, 50(8):1314–1329, 2009.
- [41] W. Wiesemann, D. Kuhn, and B. Rustem. Robust Markov decision processes. *Math. Oper. Res.*, 38(1):153–183, 2013.
- [42] I. Yang. A convex optimization approach to distributionally robust Markov decision processes with Wasserstein distance. *IEEE Control Syst. Lett.*, 1(1):164–169, 2017.
- [43] C. Zhao and Y. Guan. Data-driven risk-averse stochastic optimization with Wasserstein metric. *Oper. Res. Lett.*, 46(2):262–267, 2018.

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF KONSTANZ
Email address: sven.fuhrmann@uni-konstanz.de

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF KONSTANZ
Email address: kupper@uni-konstanz.de

CENTER FOR MATHEMATICAL ECONOMICS, BIELEFELD UNIVERSITY
Email address: max.nendel@uni-bielefeld.de