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: 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 non-stochastic setting, is widely recognised within many fields such as statistics, operations research, finance, economic theory, and other related fields, 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. [11], Denk et al. [14], Fadina et al. [18], Hu and Peng [21], Kühn [24], Neufeld and Nutz [27], and Peng [30]. Nonlinear semigroups are intimately related to BSDEs and stochastic optimal control, cf. Cheridito et al. [10], El Karoui et al. [17], Kazi-Tani et al. [22], and Soner et al. [35].

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

Date: May 12, 2021.
We thank Daniel Bartl, Jonas Blessing, and Stephan Eckstein for helpful comments and discussions. Financial support through the German Research Foundation via CRC 1283 is gratefully acknowledged by the third author.

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 [25], Gao and Kleywegt [19], Pflug and Wozobal [31], Zhao and Guan [40], 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 Markov process (Ξ_t^x) of the form

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

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 $\varphi\colon [0,\infty) \to [0,\infty]$ 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

$$\big(I(t)u\big)(x) := \sup \int_X u\big(\psi_t(x) + z\big)\, \nu(\mathrm{d}z),$$

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 $I(\pi) = I(t_1) \circ I(t_2 - t_1) \circ \cdots \circ I(t - t_n)$ for a partition $\pi = \{0 < t_1 < \cdots < t_n < t\}$. A crucial step in the construction is the monotonicity $I(\pi) \geq I(\pi')$ for partitions $\pi \subset \pi'$ of $[0,t]$, see Lemma 3.9. 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 3.14. Our main result is Theorem 3.13, 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

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

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 a 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

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

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 4.5 for further details. The results and possible applications are illustrated in various examples in Section 5, where we also depict sensitivity bounds for the resulting semigroup.

The present construction is related to the one of the Nisio semigroup [28], see Section 4.6, where the family $(I(t))_{t\geq 0}$ is replaced by a supremum over a class of linear semigroups, cf. Denk et al. [14] and Nendel and Röckner [26]. 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 [9]. 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 [32, 33]. 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 [16] and Yang [39]. For the broad topic of model uncertainty in Markovian systems, we refer to Wiesemann et al. [38] and Dentcheva and Ruszczyński [15] in the context of Markov decision processes, Rudolf and Schweizer [34] for applications to Markov chain Monte Carlo algorithms, and De Cooman [13], Hartfiel [20], Krak et al. [23], and Škulj [37] for Markov chains with interval probabilities.

The paper is organized as follows. In Section 2, we introduce the setup, all standing assumptions, and basic properties. Wasserstein robust semigroups are discussed in Section 3. Therein, we present the worst-case operator, as well as the iterated optimization scheme. The main result is Theorem 3.13, which provides the limiting convex semigroup and its infinitesimal generator. In Section 4, we provide a detailed discussion 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 $K\colon X \to X$, i.e., K is continuous, linear, and the set $\{Kx\colon \|x\| \leq r\}$ is compact for all $r \geq 0$. We further assume that $\{Kx\colon x \in X\}$ 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 $\mathrm{Lip_b} = \mathrm{Lip_b^{\mathit{K}}}(X)$ the space of all bounded functions $u\colon X \to \mathbb{R}$ satisfying

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

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

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

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

By $\mathrm{C_b} = \mathrm{C_b^{\mathit{K}}}(X)$, we denote the space of all $u\colon X \to \mathbb{R}$, for which there exists a sequence $(u_n)_{n\in\mathbb{N}} \subset \mathrm{Lip_b}$ with $u_n \to u$ as $n \to \infty$ in the sense of (2.2).[2] By construction, $\mathrm{C_b}$ is a subspace of the space of all bounded functions $u\colon X \to \mathbb{R}$, 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 $\int_X \|y\|^p \,\mu(\mathrm{d}y) < \infty$. 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 $x_1, x_2 \in X$. We endow $\mathcal{P}_p(X)$ with the p-Wasserstein distance $\mathcal{W}_p$, which is defined as

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

where $\mathrm{Cpl}(\mu, \nu)$ 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,

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

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

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 [36].

Remark 2.1.

a) Using Minkowski's inequality, $\operatorname{Cpl}(\mu,\nu)\subset\mathcal{P}_p(X\times X)$ for all $\mu,\nu\in\mathcal{P}_p(X)$.
b) Another consequence of Minkowski's inequality is that, for every Lipschitz continuous function $u\colon X\to\mathbb{R}$ and $\mu,\nu\in\mathcal{P}_p(X)$,
$$\left|\left(\int_X |u(y)|^p\,\mu(\mathrm{d}y)\right)^{1/p}-\left(\int_X |u(z)|^p\,\nu(\mathrm{d}z)\right)^{1/p}\right|\leq \|u\|_{\mathrm{Lip}}\mathcal{W}_p(\mu,\nu).$$
In particular, the map $\mathcal{P}_p(X)\to\mathbb{R},\ \mu\mapsto\int_X|u(y)|^p\,\mu(\mathrm{d}y)$ is continuous.
c) For $\mu\in\mathcal{P}_p(X)$, we have
$$\mathcal{W}_p(\mu,\delta_0)=\left(\int_X\|y\|^p\,\mu(\mathrm{d}y)\right)^{1/p},$$
where δ_0 denotes the Dirac measure with barycenter 0.

On X, we consider a time-homogeneous Markov process with transition probabilities $(p_t)_{t\geq0}$ of the form
$$p_t(x,B):=\mu_t\big(\{y\in X\colon \psi_t(x)+y\in B\}\big)\quad\text{for all } t\geq0,\ x\in X,\ \text{and } B\in\mathcal{B}(X),$$
where $(\mu_t)_{t\geq0}\subset\mathcal{P}_p(X)$ is a family of probability measures and $(\psi_t)_{t\geq0}$ is a family of continuous maps $\psi_t\colon X\to X$. In particular, we assume that the transition probabilities satisfy the Chapman-Kolmogorov equations, i.e.,
$$\int_X u(\psi_{t+s}(x)+y_{t+s})\,\mu_{t+s}(\mathrm{d}y_{t+s})=\int_X\int_X u\Big(\psi_s\big(\psi_t(x)+y_t\big)+y_s\Big)\,\mu_s(\mathrm{d}y_s)\,\mu_t(\mathrm{d}y_t)\tag{2.3}$$
for all $s,t\geq0$, $u\in\mathrm{C_b}$, and $x\in X$. In other words, the Markov process belonging to the family of transition kernels $(p_t)_{t\geq0}$ is of the form
$$\Xi_t^x=\psi_t(x)+Y_t\quad\text{with}\quad Y_t\sim\mu_t\quad\text{for all } t\geq0 \text{ and } x\in X.$$

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

(A1) For all $t\geq0$ and $x_1,x_2\in X$,
$$\|\psi_t(x_1)-\psi_t(x_2)\|\leq\|x_1-x_2\|.\tag{2.4}$$
(A2) We assume that
$$K\psi_t(x)=\psi_t(Kx)\quad\text{for all } t\geq0 \text{ and } x\in X.$$
Moreover,
$$\lim_{h\downarrow0}\sup_{\|x\|\leq r}\|\psi_h(Kx)-Kx\|=0\quad\text{for all } r\geq0.$$
(A3) We assume that $\mathcal{W}_p(\mu_h,\delta_0)\to0$ as $h\downarrow0$ or, in other words,
$$\lim_{h\downarrow0}\int_X\|y\|^p\,\mu_h(\mathrm{d}y)=0.$$

We briefly discuss our assumptions in the following remark.

Remark 2.3.

a) Note that our global Assumption (A2) implies that, for all $u \in \mathrm{C_b}$ and all $r \geq 0$,
$$\sup_{\|x\|\leq r} \big|u\big(\psi_h(x)\big) - u(x)\big| \to 0 \quad \text{as } h \downarrow 0.$$
Indeed, for $u \in \mathrm{Lip_b}$,
$$\sup_{\|x\|\leq r} \big|u\big(\psi_h(x)\big) - u(x)\big| \leq \|u\|_{\mathrm{Lip}} \sup_{\|x\|\leq r} \|\psi_h(Kx) - Kx\| \to 0 \quad \text{as } h \downarrow 0.$$
For general $u \in \mathrm{C_b}$, the statement can be obtained by the following argument that will be used in a similar form on various occasions: Let $u \in \mathrm{C_b}$. Then, there exists a sequence $(u_n)_{n\in\mathbb{N}} \in \mathrm{Lip_b}$ with $u_n \to 0$. By our global Assumptions (A1) and (A2), there exists some $h_0 > 0$ such that
$$M_r := \sup_{h\in[0,h_0]} \sup_{\|x\|\leq r} \|\psi_h(x)\| \leq \sup_{h\in[0,h_0]} \|\psi_h(0)\| + r < \infty \quad \text{for all } r \geq 0.$$
Therefore, for arbitrary $\varepsilon > 0$,
$$\begin{aligned}\sup_{\|x\|\leq r} \big|u\big(\psi_h(x)\big) - u(x)\big| &\leq \sup_{\|x\|\leq r} \big|u\big(\psi_h(x)\big) - u(x)\big| + \sup_{\|\xi\|\leq M_r} |u(\xi) - u_n(\xi)| \\ &< \sup_{\|x\|\leq r} \big|u\big(\psi_h(x)\big) - u(x)\big| + \varepsilon\end{aligned}$$
for $n \in \mathbb{N}$ sufficiently large. Letting $h \downarrow 0$ and $\varepsilon \downarrow 0$, it follows that
$$\lim_{h\downarrow 0} \sup_{\|x\|\leq r} \big|u\big(\psi_h(x)\big) - u(x)\big| \to 0 \quad \text{for all } r \geq 0.$$

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) Condition (2.4) can be weakened to
$$\|\psi_t(x_1) - \psi_t(x_2)\| \leq e^{Lt}\|x_1 - x_2\| \quad \text{for all } t \geq 0 \text{ and } x_1, x_2 \in X$$
with an arbitrary constant $L \geq 0$, see Section 4.2.

d) Consider the family $T = \big(T(t)\big)_{t\geq 0}$, given by
$$\big(T(t)u\big)(x) := \int_X u(\xi)\, p_t(x, \mathrm{d}\xi) = \int_X u\big(\psi_t(x) + y\big)\, \mu_t(\mathrm{d}y) \tag{2.5}$$
for all $t \geq 0$, $u \in \mathrm{C_b}$, and $x \in X$. Due to the Chapman-Kolmogorov equations (2.3), it follows that T is a semigroup on $\mathrm{C_b}$. 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 3.12, below.

Throughout, we consider a convex lower semicontinuous function $\varphi\colon [0,\infty) \to [0,\infty]$ with $\varphi(0) = 0$ and $\varphi(v) \neq 0$ for some $v > 0$. Note that φ is nondecreasing and continuous on $\mathrm{dom}(\varphi) := \{v \in [0,\infty)\colon \varphi(v) < \infty\}$. Since φ is nondecreasing, either

$\mathrm{dom}(\varphi) = [0,\infty)$ or $\mathrm{dom}(\varphi) = [0,a]$ for some $a \geq 0$. We further assume that the map $[0,\infty) \to [0,\infty],\ v \mapsto \varphi\big(v^{1/p}\big)$ is convex. Since $\varphi \not\equiv 0$, this implies that

$$\liminf_{v\to\infty} \frac{\varphi(v)}{v^p} > 0. \tag{2.6}$$

In particular, the conjugate

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

defines a continuous function $[0,\infty) \to [0,\infty)$ with $\varphi^*(0) = 0$. Typical examples for φ are $[0,\infty) \to [0,\infty),\ v \mapsto v^p$ or $\varphi_a := \infty \cdot \mathbb{1}_{(a,\infty)}$ 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 (2.5). We start with a preliminary version by considering the following family of operators:

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

$$\big(I(t)u\big)(x) = \sup_{\nu\in\mathcal{P}_p(X)} \bigg(\int_X u\big(\psi_t(x) + z\big)\,\nu(\mathrm{d}z) - \varphi_t\big(\mathcal{W}_p(\mu_t,\nu)\big)\bigg), \tag{3.1}$$

where $\varphi_t(v) := t\varphi\big(\frac{v}{t}\big)$, for $v \geq 0$ and $t > 0$, and $\varphi_0 := \infty \cdot \mathbb{1}_{(0,\infty)}$, i.e., $\varphi_0(0) = 0$ and $\varphi_0(v) = \infty$ for all $v > 0$.

Remark 3.2.

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

$$\|I(t)u_1 - I(t)u_2\|_\infty \leq \|u_1 - u_2\|_\infty. \tag{3.2}$$

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

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

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

b) Let $t \geq 0$ and $u \in \mathrm{Lip_b}$. Then,

$$\|I(t)u\|_{\mathrm{Lip}} \leq \|u\|_{\mathrm{Lip}}.$$

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

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

We start with the following observations.

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

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

For every $L \geq 0$, there exists a constant $a \geq 0$, depending only on φ and L, such that *a) for every* $u \in \mathrm{Lip_b}$ *with* $\|u\|_{\mathrm{Lip}} \le L$,

$$\big(I(t)u\big)(x) = \sup_{\mathcal{W}_p(\mu_t,\nu)\le at} \left(\int_X u\big(\psi_t(x)+z\big)\,\nu(\mathrm{d}z) - \varphi_t\big(\mathcal{W}_p(\mu_t,\nu)\big)\right) \tag{3.4}$$

for all $t \ge 0$ *and* $x \in X$,

b) for all $u_1, u_2 \in \mathrm{Lip_b}$ *with* $\|u_i\|_{\mathrm{Lip}} \le L$ *for* $i = 1,2$,

$$I(t)u_1 - I(t)u_2 \le I_a(t)(u_1-u_2) \quad \text{for all } t \ge 0. \tag{3.5}$$

Proof. Let $L \ge 0$ and

$$J := \big\{v \in [0,\infty)\colon \varphi(v) \le 1 + Lv\big\}.$$

By (2.6), $J = [0,a]$ for some $a \ge 0$, depending only on φ and L. Now, let $u \in \mathrm{Lip_b}$ with $\|u\|_{\mathrm{Lip}} \le L$, $t \ge 0$, and $x \in X$. For $t = 0$, Equality (3.4) is trivial. Therefore, assume that $t > 0$. Then, there exists some $\nu \in \mathcal{P}_p(X)$ with

$$\int_X u\big(\psi_t(x)+y\big)\,\mu_t(\mathrm{d}y) \le \big(I(t)u\big)(x) \le t + \int_X u\big(\psi_t(x)+z\big)\,\nu(\mathrm{d}z) - \varphi_t\big(\mathcal{W}_p(\mu_t,\nu)\big),$$

which leads to the inequality

$$\begin{aligned}\varphi\left(\frac{\mathcal{W}_p(\mu_t,\nu)}{t}\right) &\le 1 + \frac{1}{t}\int_X\int_X \Big(u\big(\psi_t(x)+z\big) - u\big(\psi_t(x)+y\big)\Big)\,\nu(\mathrm{d}z)\,\mu_t(\mathrm{d}y)\\ &\le 1 + L\frac{\mathcal{W}_p(\mu_t,\nu)}{t}.\end{aligned}$$

Therefore, in (3.1), it is sufficient to take the supremum over the set

$$\Big\{\nu \in \mathcal{P}_p(X)\colon \tfrac{\mathcal{W}_p(\mu_t,\nu)}{t} \in J\Big\} = \Big\{\nu \in \mathcal{P}_p(X)\colon \tfrac{\mathcal{W}_p(\mu_t,\nu)}{t} \in [0,a]\Big\}.$$

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

$$\big(I(t)u_1\big)(x) - \big(I(t)u_2\big)(x) \le \sup_{\mathcal{W}_p(\mu_t,\nu)\le at} \int_X \Big(u_1\big(\psi_t(x)+z\big) - u_2\big(\psi_t(x)+z\big)\Big)\,\nu(\mathrm{d}z).$$

The proof is complete. ☐

Note that the definition of the operator $I_a(t)$ by virtue of (3.3) is exactly the same as (3.1) with $\varphi = \varphi_a = \infty \cdot \mathbb{1}_{(a,\infty)}$.

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

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

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

Proof. By (2.6), there exists some constant $M > 0$, depending only on φ, such that

$$v^p \le M\big(1 + \varphi(v)\big) \quad \text{for all } v \ge 0.$$

Let $C \ge 0$, $u \in \mathrm{C_b}$ with $\|u\|_\infty \le C$, $t \in [0,1]$, 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

$$\int_X u\big(\psi_t(x)+y\big)\,\mu_t(\mathrm{d}y) \le \big(I(t)u\big)(x) \le 1 + \int_X u\big(\psi_t(x)+z\big)\,\nu(\mathrm{d}z) - \varphi_t\big(\mathcal{W}_p(\mu_t,\nu)\big),$$

which leads to the inequality

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

Therefore, in (3.1) it is sufficient to take the supremum over the set

$$\Big\{\nu \in \mathcal{P}_p(X) \colon \varphi\Big(\tfrac{\mathcal{W}_p(\mu_t,\nu)}{t}\Big) \leq \tfrac{1+2C}{t}\Big\} \subset \Big\{\nu \in \mathcal{P}_p(X) \colon \mathcal{W}_p(\mu_t,\nu) \leq bt^\alpha\Big\}$$

with $b := \big(2M(1+C)\big)^{1/p}$. □

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

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

Proof. By (3.2), we have

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

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

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

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

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

Then, by Assumption (A1),

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

Let $\varepsilon > 0$, $r \geq 0$, and $M > 0$ with $\frac{C(1+bh_0^\alpha)}{M} < \varepsilon$. Then, using Markov's inequality,

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

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

Lemma 3.6. *Let $h_0 > 0$ as in Lemma 3.5. Then, the operator $I(h)\colon \mathrm{C_b} \to \mathrm{C_b}$ is well-defined, convex, and monotone for all $h \in [0,h_0]$. Moreover, for all $u \in \mathrm{C_b}$,*

$$I(h)u \to u \quad \textit{as } h \downarrow 0.$$

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

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

which, by taking the supremum over all $x \in X$, leads to the inequality

$$\big\|I(h)u - T(h)u\big\|_\infty \leq h\varphi^*\big(\|u\|_{\mathrm{Lip}}\big). \tag{3.7}$$

By Remark 2.3 a),

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

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

$$\sup_{\|x\|\leq r} \big|\big(T(h)u\big)(x) - u(x)\big| \leq \|u\|_{\mathrm{Lip}} \bigg(\int_X \|y\|^p \, \mu_h(\mathrm{d}y) \bigg)^{1/p} + \sup_{\|x\|\leq r} \big|u\big(\psi_h(x)\big) - u(x)\big| \to 0 \quad \text{as } h \downarrow 0,$$

which, together with (3.7), implies that $I(h)u \to u$ as $h \downarrow 0$. For general $u \in \mathrm{C_b}$, the strong continuity follows by approximating with $(u_n)_{n\in\mathbb{N}} \subset \mathrm{Lip_b}$ and using the uniform in $h \in [0, h_0]$ convergence $I(h)u_n \to I(h)u$ as $n \to \infty$, see Lemma 3.5. □

Alternatively, the family of operators $I = \big(I(t)\big)_{t\geq 0}$ 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 3.7. Let

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

Then, by definition, $\mathcal{W}_p(\mu, \nu) = \inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \Delta_p \pi$ for all $\mu, \nu \in \mathcal{P}_p(X)$. By Remark 2.1 b), $|\Delta_p \pi - \Delta_p \pi'| \leq \mathcal{W}_p(\pi, \pi')$, i.e., the map $\mathcal{P}_p(X \times X) \to \mathbb{R}$, $\pi \mapsto \Delta_p \pi$ 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 $t \geq 0$, $u \in \mathrm{C_b}$, and $x \in X$,

$$\big(I(t)u\big)(x) = \sup_{\substack{\pi \in \mathcal{P}_p(X\times X)\\ \pi_1 = \mu_t}} \bigg(\int_X u\big(\psi_t(x) + z\big) \, \pi_2(\mathrm{d}z) - \varphi_t\big(\Delta_p \pi\big) \bigg). \tag{3.9}$$

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 [36, Theorem 4.1], there exists an optimal coupling $\pi^* \in \mathrm{Cpl}(\mu, \nu)$ with $\mathcal{W}_p(\mu_t, \nu) = \Delta_p \pi^*$. By definition of $\mathrm{Cpl}(\mu_t, \nu)$, it follows that $\pi_1^* = \mu_t$ and $\pi_2^* = \nu$. Therefore,

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

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

$$\big(I(t)u\big)(x) \leq \int_X u\big(\psi_t(x) + z\big) \, \pi_2^x(\mathrm{d}z) - \varphi_t\big(\Delta_p \pi^x\big) + \varepsilon,$$

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

Proof. Let $D_t := \{\pi \in \mathcal{P}_p(X \times X) \colon \pi_1 = \mu_t \text{ and } \varphi_t(\Delta_p \pi) < \infty\}$. Then, by Remark 3.7, the map

$$f \colon X \times D_t \to \mathbb{R}, \quad (x,\pi) \mapsto \int_{X\times X} u\big(\psi_t(x) + z\big)\, \pi_1(\mathrm{d}z) - \varphi_t\big(\Delta_p \pi\big)$$

is continuous, and

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

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

Lemma 3.9. *Let* $s,t \geq 0$ *and* $u \in \mathrm{C_b}$. *Then,*

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

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

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

where $\nu_t := \pi_{t,2} \in \mathcal{P}_p(X)$ is the second marginal of π_t. By Lemma 3.8, there exists a measurable map $\pi_s \colon X \to \mathcal{P}_p(X \times X)$, $z \mapsto \pi_s^z$ such that, for all $z \in X$, the first marginal $\pi_{s,1}^z$ of π_s^z is μ_s and

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

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

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

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

$$\begin{aligned}\int_{X\times X} g(y_{t+s}, z_{t+s})\, \pi_{t+s}(\mathrm{d}y_{t+s}, \mathrm{d}z_{t+s}) = \int_{X\times X}\int_{X\times X} g\Big(\psi_s\big(\psi_t(x) + y_t\big) + y_s - \psi_{t+s}(x), \\ \psi_s\big(\psi_t(x) + z_t\big) + z_s - \psi_{t+s}(x)\Big)\, \pi_s^{z_t}(\mathrm{d}y_s, \mathrm{d}z_s)\, \pi_t(\mathrm{d}y_t, \mathrm{d}z_t)\end{aligned} \tag{3.11}$$

for all bounded continuous functions $g \colon X \times X \to \mathbb{R}$. 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 $z \mapsto \pi^z$, 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

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

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

$$\varphi_{t+s}\big(\Delta_p \pi_{t+s}\big) \leq \varphi_t\big(\Delta_p \pi_t\big) + \int_X \varphi_s\big(\Delta_p \pi_s^{z_t}\big)\, \nu_t(\mathrm{d}z_t). \tag{3.12}$$

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

$$\Delta_p \pi_{t+s} \leq \Delta_p \pi_t + \bigg(\int_X \int_{X\times X} \|y_s - z_s\|^p \, \pi_s^{z_t}(\mathrm{d}y_s, \mathrm{d}z_s)\, \nu_t(\mathrm{d}z_t) \bigg)^{1/p}.$$

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

$$\begin{aligned}
\Delta_p \pi_{t+s} &= \bigg(\int_{X\times X} \|y_{t+s} - z_{t+s}\|^p \, \pi_{t+s}(\mathrm{d}y_{t+s}, \mathrm{d}z_{t+s}) \bigg)^{1/p} \\
&= \bigg(\int_{X\times X} \big\|\psi_{t+s}(x) + y_{t+s} - \big(\psi_{t+s}(x) + z_{t+s}\big)\big\|^p \, \pi_{t+s}(\mathrm{d}y_{t+s}, \mathrm{d}z_{t+s}) \bigg)^{1/p} \\
&\leq \bigg(\int_{X\times X} \int_{X\times X} \big\|\psi_s\big(\psi_t(x) + y_t\big) - \psi_s\big(\psi_t(x) + z_t\big)\big\|^p \, \pi_t(\mathrm{d}y_t, \mathrm{d}z_t) \bigg)^{1/p} \\
&\qquad + \bigg(\int_X \int_{X\times X} \|y_s - z_s\|^p \, \pi_s^{z_t}(\mathrm{d}y_s, \mathrm{d}z_s)\, \nu_t(\mathrm{d}z_t) \bigg)^{1/p} \\
&\leq \Delta_p \pi_t + \bigg(\int_X \int_{X\times X} \|y_s - z_s\|^p \, \pi_s^{z_t}(\mathrm{d}y_s, \mathrm{d}z_s)\, \nu_t(\mathrm{d}z_t) \bigg)^{1/p}.
\end{aligned}$$

Note that the function $z \mapsto \int_{X\times X} \|y_s - z_s\|^p \, \pi_s^z(\mathrm{d}y_s, \mathrm{d}z_s)$ is in fact measurable, since the map $z \mapsto \pi_s^z$ is measurable and the map $\pi \mapsto \int_{X\times X} \|y - z\|^p \, \pi(\mathrm{d}y, \mathrm{d}z)$ is continuous, by Remark 2.1 b), and thus measurable. Finally, we show that

$$\varphi_{t+s}\big(\Delta_p \pi_{t+s}\big) \leq \varphi_t\big(\Delta_p \pi_t\big) + \int_X \varphi_s\big(\Delta_p \pi_s^{z_t}\big)\, \nu_t(\mathrm{d}z_t).$$

Since we assumed $v \mapsto \varphi\big(v^{1/p}\big)$ to be convex, the map $v \mapsto \varphi_s\big(v^{1/p}\big)$ is convex as well, which, by Jensen's inequality, implies that

$$\varphi_s\Bigg(\bigg(\int_X \int_{X\times X} \|y_s - z_s\|^p \, \pi_s^{z_t}(\mathrm{d}y_s, \mathrm{d}z_s)\, \nu_t(\mathrm{d}z_t) \bigg)^{1/p} \Bigg) \leq \int_X \varphi_s\big(\Delta_p \pi_s^{z_t}\big)\, \nu_t(\mathrm{d}z_t).$$

The convexity of φ yields that

$$\varphi_{t+s}\big(\Delta_p \pi_{t+s}\big) = (t+s)\varphi\bigg(\frac{\Delta_p \pi_{t+s}}{t+s}\bigg) \leq \varphi_t\big(\Delta_p \pi_t\big) + \int_X \varphi_s\big(\Delta_p \pi_s^{z_t}\big)\, \nu_t(\mathrm{d}z_t).$$

The proof is complete. $\square$

By C_b^1, we denote the space of all $u \in \mathrm{Lip_b}$ with Fréchet derivative $u' \colon X \to X'$ satisfying $x \mapsto \|u'(x)\| \in \mathrm{C_b}$. 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 3.10. *Let $u \in \mathrm{C}_\mathrm{b}^1$. Then,*

$$\frac{\big(I(h)u\big)(x) - \big(T(h)u\big)(x)}{h} \to \varphi^*\big(\|u'(x)\|\big) \quad \textit{as } h \downarrow 0. \tag{3.13}$$

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

Proof. Let $r\geq 0$. We start by showing that, for $h>0$ sufficiently small,

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

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

$$\begin{aligned}\frac{\big(I(h)u\big)(x)-\big(T(h)u\big)(x)}{h}&\geq \frac{1}{h}\bigg(\int_X u\big(\psi_h(x)+z\big)\,\nu_{h,\theta}(\mathrm{d}z)-\varphi_h(h\|\theta\|)-\big(T(h)u\big)(x)\bigg)\\&=\int_X\frac{u\big(\psi_h(x)+y+h\theta\big)-u\big(\psi_h(x)+y\big)}{h}\,\mu_h(\mathrm{d}y)-\varphi(\|\theta\|)\end{aligned} \tag{3.15}$$

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

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

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

$$u'(x)\theta\leq\frac{u\big(\psi_h(x)+y+h\theta\big)-u\big(\psi_h(x)+y\big)}{h}+\frac{\varepsilon}{3} \tag{3.16}$$

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

$$\bigg|\int_{\{\|y\|>\delta\}}\frac{u\big(\psi_h(x)+y+h\theta\big)-u\big(\psi_h(x)+y\big)}{h}\,\mu_h(\mathrm{d}y)\bigg|\leq\frac{a\|u'\|_\infty}{\delta}\mathcal{W}_p(\mu_h,\delta_0)\leq\frac{\varepsilon}{3} \tag{3.17}$$

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

$$u'(x)\theta\leq\int_X\frac{u\big(\psi_h(x)+y+h\theta\big)-u\big(\psi_h(x)+y\big)}{h}\,\mu_h(\mathrm{d}y)+\varepsilon$$

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,

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

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

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

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 $h>0$ 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

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

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

$$\begin{aligned}\frac{\big(I(h)u\big)(x)-\big(T(h)u\big)(x)}{h} &\leq \frac{1}{h}\int_{X\times X} g_{h,x}(y,z)\|y-z\|\,\pi_h^x(\mathrm{d}y,\mathrm{d}z)-\varphi\Big(\tfrac{\mathcal{W}_p(\mu_h,\nu_h^x)}{h}\Big)+\frac{\varepsilon}{2}\\ &\leq \bigg(\int_{X\times X}\big|g_{h,x}(y,z)\big|^q\,\pi_h^x(\mathrm{d}y,\mathrm{d}z)\bigg)^{1/q}\frac{\mathcal{W}_p(\mu_h,\nu_h^x)}{h}-\varphi\Big(\tfrac{\mathcal{W}_p(\mu_h,\nu_h^x)}{h}\Big)+\frac{\varepsilon}{2}\\ &\leq \varphi^*\Bigg(\bigg(\int_{X\times X}\big|g_{h,x}(y,z)\big|^q\,\pi_h^x(\mathrm{d}y,\mathrm{d}z)\bigg)^{1/q}\Bigg)+\frac{\varepsilon}{2},\end{aligned}$$

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 $x\in X$,

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

Since φ^* is nondecreasing, (3.18) follows as soon as we are able to show that

$$\sup_{\|x\|\leq r}\Bigg(\bigg(\int_{X\times X}\big|g_{h,x}(y,z)\big|^q\,\pi_h^x(\mathrm{d}y,\mathrm{d}z)\bigg)^{1/q}-\|u'(x)\|\Bigg)\leq\delta$$

for $h>0$ sufficiently small. 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

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

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, we may conclude that

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

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

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

The proof is complete. □

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

Definition 3.11. We say that a family $S=\big(S(t)\big)_{t\geq 0}$ is a *strongly continuous convex transition semigroup* (on $\mathrm{C_b}$) if the following conditions hold:

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

(iv) $S(t)u \to u$ as $t \downarrow 0$ for all $u \in \mathrm{C_b}$.

For a strongly continuous convex transition semigroup S, we define the *generator* $A\colon D(A) \subset \mathrm{C_b} \to \mathrm{C_b}$ of S by

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

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

Remark 3.12. Choosing $\varphi := \infty \cdot \mathbb{1}_{(0,\infty)}$, the results of this section imply that the family $T = \big(T(t)\big)_{t \geq 0}$, given by

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

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 d). Throughout the remainder of this section, we denote by $B\colon D(B) \subset \mathrm{C_b} \to \mathrm{C_b}$ the generator of T.

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

$$I^n(t)u := \big(\underbrace{I(2^{-n}) \cdots I(2^{-n})}_{k\text{-times}}\big) I\big(t - k2^{-n}\big)u = I(2^{-n})^k I\big(t - k2^{-n}\big)u, \tag{3.20}$$

where $k \in \mathbb{N}_0$ denotes the largest natural number with $k2^{-n} \leq t$. Then, by Lemma 3.9,

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

We define

$$S(t)u := \inf_{n \in \mathbb{N}_0} I^n(t)u \quad \text{for all } t \geq 0 \text{ and } u \in \mathrm{C_b}. \tag{3.21}$$

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

Theorem 3.13. *The family S is a strongly continuous convex transition semigroup. Let A denote the generator of S. Then, $D(B) \cap \mathrm{C_b^1} \subset D(A)$ with*

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

Proof. Let $t \geq 0$. Then, for $n \in \mathbb{N}$ sufficiently large, $I^n(t)u \in \mathrm{C_b}$ for all $u \in \mathrm{C_b}$ since, by definition, $I^n(t)$ is a finite composition of operators $I(h)\colon \mathrm{C_b} \to \mathrm{C_b}$ with $h \in (0, h_0]$, cf. Lemma 3.6. Hence, for $n \in \mathbb{N}$ sufficiently large, the operator $I^n(t)\colon \mathrm{C_b} \to \mathrm{C_b}$ is well-defined, convex, and monotone with $I^n(t)m = m$ for all $m \in \mathbb{R}$ and $\|I^n(t)u\|_{\mathrm{Lip}} \leq \|u\|_{\mathrm{Lip}}$ for all $u \in \mathrm{Lip_b}$ as all these properties carry over from $I(h)$, for $h > 0$, to $I^n(t)$. Since $\|I^n(t)u\|_{\mathrm{Lip}} \leq \|u\|_{\mathrm{Lip}}$ for all $u \in \mathrm{Lip_b}$ and

$$T(t)u \leq S(t)u \leq I^n(t)u \quad \text{for all } n \in \mathbb{N} \text{ and } u \in \mathrm{C_b}, \tag{3.22}$$

it follows that $S(t)u \in \mathrm{Lip_b}$ with $\|S(t)u\|_{\mathrm{Lip}} \leq \|u\|_{\mathrm{Lip}}$ for all $u \in \mathrm{Lip_b}$. Next, we verify property (iii) in Definition 3.11. For this, we even show a slightly stronger property. For $i = 1, 2$, let $(u_k^i)_{k\in\mathbb{N}} \subset \mathrm{C_b}$ with $(u_k^1 - u_k^2) \to 0$ as $k \to \infty$, i.e.,

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

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

$$\lim_{k\to\infty} \sup_{t\in[0,s]} \sup_{\|x\|\leq r} \big|\big(S(t)u_k^1\big)(x) - \big(S(t)u_k^2\big)(x)\big| = 0. \tag{3.23}$$

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

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

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 $u \in \mathrm{C_b}$ with a sequence $(u_k)_{n\in\mathbb{N}} \subset \mathrm{Lip_b}$, it follows that $S(t)\colon \mathrm{C_b} \to \mathrm{C_b}$ is well-defined. Since all other properties stated in (i) of Definition 3.11 directly carry over from $I^n(t)$ to the limit, $S(t)$ satisfies these properties. Moreover, (3.22) together with Lemma 3.6 and Remark 3.12 implies that $S(t)u \to u$ as $t \downarrow 0$ for all $u \in \mathrm{C_b}$.
In order to verify the semigroup property (ii), we fix $s, t \geq 0$ and $u \in \mathrm{Lip_b}$. Consider the set $\mathcal{D} := \{k2^{-n} \colon k, n \in \mathbb{N}_0\}$ 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,

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

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 $k \in \mathbb{N}$,

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

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

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

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

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

Now, let $u \in D(B) \cap \mathrm{C_b^1}$ and $g(x) := \big(Bu\big)(x) + \varphi^*\big(\|u'(x)\|\big)$ for all $x \in X$. By definition of the generator B and Lemma 3.10, we find that $g \in \mathrm{C_b}$. Let $a \geq 0$ and $I_a(t)$, for $t \geq 0$, be given as in Lemma 3.3 with $L := \|u\|_{\mathrm{Lip}}$. Let $t > 0$ and $0 = t_0 < \ldots < t_m = t$

[3]The 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 $z_1 = 0$ and $z_2 = \frac{u_k^1-u_k^2}{\lambda}$.

be a partition of the interval $[0,t]$ with $m \in \mathbb{N}$ and $h_k := t_k - t_{k-1}$ for $k \in \{1,\ldots,m\}$. Then, by Lemma 3.3,

$$\frac{I(h_1)\cdots I(h_{k-1})u - I(h_1)\cdots I(h_k)u}{h_k} + g \le I_a(t_{k-1})\Big(\frac{u - I(h_k)u}{h_k}\Big) + g.$$

We thus obtain that

$$\begin{aligned}\frac{u - I(h_1)\cdots I(h_m)u}{t} + g &= \sum_{k=1}^m \frac{h_k}{t}\Big(\frac{I(h_1)\cdots I(h_{k-1})u - I(h_1)\cdots I(h_k)u}{h_k} + g\Big)\\ &\le \sum_{k=1}^m \frac{h_k}{t}\bigg(I_a(t_{k-1})\Big(\frac{u - I(h_k)u}{h_k}\Big) + g\bigg).\end{aligned}$$

Using the convergence results for the family of operators $I_a = \big(I_a(t)\big)_{t\ge 0}$ from Lemma 3.5 and Lemma 3.6 together with Lemma 3.10 and the observation

$$\frac{S(t)u - u}{t} - g \le \frac{I(t)u - u}{t} - g,$$

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

Remark 3.14. For $t > 0$, let P_t denote the set of all partitions $\pi = \{t_0,\ldots,t_m\}$ with $0 = t_0 < \ldots < t_m = t$ and $m \in \mathbb{N}$. For $t > 0$, $\pi \in P_t$, and $u \in \mathrm{C_b}$, we define

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

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

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

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

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

4. Extensions and Remarks

4.1. **On the particular form of the transition kernels.** The structural assumption on the transition kernels

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

at first seems rather restrictive. However, it contains a large class of examples, as we will discuss in this subsection. For this purpose, let X be a nonempty set, and assume that there exists a bijective function $V\colon X \to M$, where M is a separable Banach space. Let $\psi_t\colon X \to X$ be a map with

$$\big\|V\big(\psi_t(x_1)\big) - V\big(\psi_t(x_2)\big)\big\| \le \|V(x_1) - V(x_2)\| \quad \text{for all } t \ge 0 \text{ and } x_1, x_2 \in X.$$

Then, for $x, y \in X$ and $\lambda \in \mathbb{R}$, the operations

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

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

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

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

$$\begin{aligned} u'(x)z &= \lim_{h\downarrow 0} \frac{u\big(x +_V (h \cdot_V z)\big) - u(x)}{h} = \lim_{h\downarrow 0} \frac{u\Big(V^{-1}\big(V(x) + hV(z)\big)\Big) - u(x)}{h} \\ &= \Big(D\big(u \circ V^{-1}\big)\big(V(x)\big)\Big)V(z) \quad \text{for all } z \in X. \end{aligned} \tag{4.1}$$

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

$$u'(x)z = \big(Du(x)\big)\big(DV(x)\big)^{-1}V(z), \tag{4.2}$$

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 $\big(DV(x)\big)^{-1}$ is a bounded linear operator from M to X_0. In this case,

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

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 topological dual space of M.

Choosing $V = \log$, our setup covers, for example, the case of geometric dynamics, and the term in the penalisation closely relates to the first order term appearing in the generator of a geometric Brownian motion.

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

$$\Xi_t^x = xY_t,$$

where Y_t corresponds, for example, to a geometric Brownian motion starting in 1, i.e., $Y_t = \exp\big(\big(\mu - \frac{\sigma^2}{2}\big) + \sigma W_t\big)$ with $\mu \in \mathbb{R}$, $\sigma \geq 0$, and a standard Brownian motion $(W_t)_{t\geq 0}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In this case, for any continuously differentiable function $u \in \mathrm{C_b}$, the derivative u' is given by

$$u'(x) = \lim_{h\downarrow 0} \frac{u\big(x +_V (h \cdot_V e)\big) - u(x)}{h} = \lim_{h\downarrow 0} \frac{u(xh) - u(x)}{h} = x\frac{\mathrm{d}}{\mathrm{d}x}u(x),$$

where e is the Euler constant and $\frac{\mathrm{d}}{\mathrm{d}x}u$ denotes the usual derivative of u. Note that the equality $u'(x) = x\frac{\mathrm{d}}{\mathrm{d}x}u(x)$ also follows from (4.2) with $z = e$, since $\big(\frac{\mathrm{d}}{\mathrm{d}x}\log(x)\big)^{-1} = x$ for all $x \in (0, \infty)$. In particular, $\|u'(x)\| = \big|x\frac{\mathrm{d}}{\mathrm{d}x}u(x)\big|$ for all $x \in (0, \infty)$.

4.2. **Relaxation of Assumption (A1).** Assumption (A1) states that, for every $t \geq 0$, the map $\psi_t\colon X \to X$ is Lipschitz continuous with Lipschitz constant 1. However, using a simple observation, one can weaken it to

(A1') There exists some $L \geq 0$, such that

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

In this case, one modifies the scaling in the penalisation φ_t and considers, for $t > 0$,

$$\bar{\varphi}_t(v) := t\varphi\big(e^{-Lt}\tfrac{v}{t}\big), \quad \text{for } v \geq 0,$$

instead of φ_t. Therefore, the definition of the operator $I(t)$ changes to

$$\big(\bar{I}(t)u\big)(x) := \bigg(\sup_{\nu \in \mathcal{P}_p} \int_X u\big(\psi_t(x) + z\big)\,\nu(\mathrm{d}z) - \bar{\varphi}_t\big(\mathcal{W}_p(\mu_t, \nu)\big)\bigg).$$

Note that the operator $\bar{I}(t)$ constitutes to the definition of the semigroup S only for infinitesimally small times $t \geq 0$. Since φ is lower semicontinuous and $\varphi^*(w) < \infty$ for all $w \geq 0$,

$$I(t)u \leq \bar{I}(t)u \leq I(t)u + \varepsilon$$

for all $\varepsilon > 0$ and sufficiently small $t \geq 0$ depending only on ε, φ, and the norm $\|u\|_\infty$ of $u \in \mathrm{C_b}$. That is, the infinitesimal behaviour of $I(t)$ is not affected by the additional constant e^{-Lt} for small times $t \geq 0$. Hence, the semigroup resulting from the operators $\bar{I}(t)$ is the same as the one resulting from the construction using $I(t)$, for $t \geq 0$. The modification using (A1') instead of (A1) influences the construction of the semigroup and its properties only at the following points.

- The Lipschitz constant evolves according to

$$\|\bar{I}(t)u\|_{\mathrm{Lip}} \leq e^{Lt}\|u\|_{\mathrm{Lip}} \quad \text{for } u \in \mathrm{Lip_b} \text{ and } t \geq 0. \tag{4.3}$$

  This leads to the modified estimate $\|S(t)u\|_{\mathrm{Lip}} \leq e^{Lt}\|u\|_{\mathrm{Lip}}$ for all $u \in \mathrm{Lip_b}$ and $t \geq 0$.
- The estimate $S(t)u \leq \bar{I}(t)u$ still holds for all $t \geq 0$ and $u \in \mathrm{C_b}$. However, the operators $S(t)$ and $I(t)$ can only be compared for small times $t \geq 0$. More precisely, for all $\varepsilon > 0$ and sufficiently small $t \geq 0$ depending on ε, φ, and the norm $\|u\|_\infty$ of $u \in \mathrm{C_b}$,

$$S(t)u \leq I(t)u + \varepsilon. \tag{4.4}$$

- The key estimate (3.12) in the proof of Lemma 3.9: Assuming (A1') instead of (A1) in Lemma 3.9, using the same reasoning as in the proof, we obtain the modified estimate

$$\Delta_p\pi_{t+s} \leq e^{Ls}\Delta_p\pi_t + \bigg(\int_X \int_{X\times X} \|y_s - z_s\|^p \pi_s^{z_t}(\mathrm{d}y_s, \mathrm{d}z_s)\,\nu_t(\mathrm{d}z_t)\bigg)^{1/p}. \tag{4.5}$$

  Considering $\bar{\varphi}_t$ instead of φ_t, Equation (4.5) leads to the following modified version of the key estimate (3.12)

$$\bar{\varphi}_{t+s}\big(\Delta_p\pi_{t+s}\big) \leq \bar{\varphi}_t\big(\Delta_p\pi_t\big) + \int_X \bar{\varphi}_s\big(\Delta_p\pi_s^{z_t}\big)\,\nu_t(\mathrm{d}z_t),$$

  which finally ensures the inequality $\bar{I}(t)\bar{I}(s)u \leq \bar{I}(t+s)u$ for all $u \in \mathrm{C_b}$.

- The estimates using the penalisation in Lemma 3.10: observe that the convex conjugate $\big(\bar\varphi_t\big)^*$ of $\bar\varphi_t$ is given by
$$\big(\bar\varphi_t\big)^*(w) = t\varphi^*(e^{Lt}w) \quad \text{for all } t>0 \text{ and } w\geq 0.$$
Using the continuity of φ^*, it follows that
$$\frac{1}{h}\big(\bar\varphi_h\big)^*(w) = \varphi^*(e^{Lh}w) \to \varphi^*(w) \quad \text{as } h\downarrow 0.$$
We thus obtain the same generator for the semigroup related to the penalisation $\bar\varphi_t$ as for the one related to φ_t.

Concluding, modifying (A1) to (A1') only affects the evolution of the Lipschitz constant under the semigroup S, see Equation (4.3), and the comparison between $S(t)$ and $I(t)$ for $t\geq 0$, see Equation (4.4). All statements from Section 3 related to asymptotic properties as $t\downarrow 0$ are not affected by the relaxation (A1').

4.3. **Viscosity solutions.** Note that Theorem 3.13 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
$$\big(Au\big)(x) = \big(Bu\big)(x) + \varphi^*\big(\|u'(x)\|\big) \quad \text{for } u\in D(B)\cap \mathrm{C}_\mathrm{b}^1 \text{ and } x\in X,$$
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
$$\partial_t u = Au \quad \text{for } t\geq 0. \tag{4.6}$$
Using a standard procedure from Denk et al. [14], Röckner and Nendel [26], 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 $u_0\in \mathrm{C_b}$, is a viscosity solution to (4.6) using the following notion of a viscosity solution.

Definition 4.3. We say that a function $u\colon[0,\infty)\to \mathrm{C_b}$ is a *viscosity subsolution* to the abstract differential equation (4.6) if $u\colon[0,\infty)\to\mathrm{C_b}$ is continuous and
$$\big(\psi'(t)\big)(x) \leq \big(A\psi(t)\big)(x)$$
for all $t>0$, $x\in M$, and every differentiable function $\psi\colon(0,\infty)\to\mathrm{C_b}$ satisfying $\psi(t)\in D(B)\cap\mathrm{C}_\mathrm{b}^1$, $\big(\psi(t)\big)(x) = \big(u(t)\big)(x)$, and $\psi(s)\geq u(s)$ for all $s>0$. A function $u\colon[0,\infty)\to\mathrm{C_b}$ 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 $u\colon[0,\infty)\to\mathrm{C_b}$ 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 $u_0\in\mathrm{C_b}$. Then, the function $u\colon[0,\infty)\to\mathrm{C_b}$, $t\mapsto S(t)u_0$ 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 $\mathrm{C_b}$) almost literally the same as in Denk et al. [14, Proposition 5.11], Röckner and Nendel [26, Theorem 4.5], or Bartl et al. [6, Theorem 2.12].

4.4. **Unbounded functions and additional continuity properties.** Although we consider the space $\mathrm{C_b}$ resulting as a closure from bounded Lipschitz continuous functions, the construction of the semigroup and many statements carry over to arbitrary Lipschitz continuous functions. We denote by Lip the space of all functions $u\colon X\to\mathbb{R}$ satisfying (2.1). For such functions $\|\cdot\|_{\mathrm{Lip}}$ is defined as before. The key ingredient in order to establish the semigroup property of $\big(S(t)\big)_{t\geq 0}$ 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 $u\in\mathrm{Lip}$ and $t\geq 0$. We consider $T(t)u$, $I(t)$, and $S(t)$ given by (2.5), (3.1), and (3.21), respectively. Then,

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

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

$$\big|\big(T(t)u\big)(x)\big|\leq C\bigg(1+\|\psi_t(x)\|+\int_X\|y\|\,\mu_t(\mathrm{d}y)\bigg)<\infty.$$

Moreover, by equation (3.7) in the proof of Lemma 3.6, we have seen that

$$I(t)u\leq T(t)u+t\varphi^*\big(\|u\|_{\mathrm{Lip}}\big)$$

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

$$\big(T(t)u\big)(x)\leq\big(S(t)u\big)(x)\leq\big(I(t)u\big)(x)\leq\big(T(t)u\big)(x)+t\varphi^*\big(\|u\|_{\mathrm{Lip}}\big), \tag{4.8}$$

which shows that the functions $I(t)u$ and $S(t)u$ are well-defined. Moreover, we have seen that

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

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

$$\sup_{\|x\|\leq r}\big|\big(T(t)u\big)(x)-u(x)\big|\leq\|u\|_{\mathrm{Lip}}\bigg(\sup_{\|x\|\leq r}\|\psi_t(x)-x\|+\int_X\|y\|\,\mu_t(\mathrm{d}y)\bigg)\to 0$$

as $t\downarrow 0$. This observation together with (4.8) again implies the following version of strong continuity, which is in line with the next definition: for all $r\geq 0$,

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

We introduce the following convergence on Lip.

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

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

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

Proof. Let $a\geq 0$ be as in Lemma 3.3 with $L:=\sup_{k\in\mathbb{N}}\|u_k\|_{\mathrm{Lip}}$. Then,

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

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 $k\to\infty$ uniformly in $[0,h_0]$ with $h_0>0$ sufficiently small. Let $h_0>0$ sufficiently small such that

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

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

$$M_r := \sup_{h\in[0,h_0]}\sup_{\|x\|\le r}\|\psi_h(x)\| \le \sup_{h\in[0,h_0]}\|\psi_h(0)\| + r < \infty \quad \text{for all } r\ge 0. \tag{4.9}$$

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

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

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

We conclude with the following remark on the semigroup property for S on Lip.

Remark 4.8. By the previous proposition,

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

Indeed, for all $k\in\mathbb{N}$, let $\varrho_k\colon\mathbb{R}\to\mathbb{R}$ be 1-Lipschitz with $0\le\varrho_k\le 1$, $\varrho_k(w)=1$ for all $w\in\mathbb{R}$ with $|w|\le k$, and $\varrho_k(w)=0$ for all $w\in\mathbb{R}$ with $|w|\ge k+1$. Let $u\in\mathrm{Lip}$ and $u_k(x):=u(x)\varrho_k(\|Kx\|)$ for all $x\in X$ and $k\in\mathbb{N}$. Since $u_k\rightrightarrows u$ as $k\to\infty$, the previous proposition implies that

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

4.5. **Sensitivity analysis and numerical computation of the robust semigroup.** The sensitivity of functionals depending on the distribution μ of a random source X with respect to "small" nonparametric 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 $(p_t)_{t\ge0}$. In order to compute a realisation of Ξ_t^x for some $t\ge0$ 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 $T(t)u\le S(t)u\le I(t)u$ for all $u\in\mathrm{Lip}$. Combined with (3.7), we have seen that

$$S(t)u - T(t)u \le I(t)u - T(t)u \le t\varphi^*\big(\|u\|_{\mathrm{Lip}}\big) \quad \text{for all } u\in\mathrm{Lip}. \tag{4.10}$$

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 3.13,

$$\left\| \frac{S(t)u - T(t)u}{t} - \varphi^*\big(\|u'(x)\|\big) \right\|_\infty \to 0 \quad \text{as } t \downarrow 0, \tag{4.11}$$

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

Consider the case, where $\varphi = \infty \cdot \mathbb{1}_{(a,\infty)}$ with some $a \geq 0$. The fact that $v(t,x) = \big(S(t)u\big)(x)$, for $t \geq 0$, $x \in X$, and $u \in \mathrm{C}_\mathrm{b}^1$, defines a viscosity solution to the nonlinear PDE

$$\partial_t v(t,x) = Bv(t,x) + a\|D_x v(t,x)\| \quad \text{for } t \geq 0 \text{ and } x \in X, \tag{4.12}$$

where B is the generator of T and D_x is the Fréchet derivative in the space variable, provides tools for the numerical computation of $S(t)u$ also for large times $t \geq 0$.

Implicitly, we have also derived sensitivity bounds for the influence of Wasserstein perturbations around the reference semigroup T in terms of the penalisation φ^*. Note that, for $a = 0$, $S = T$. For $a > 0$, the operator I is given by

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

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

$$\sup_{x \in X} \left\| \frac{\big(I(t)u\big)(x) - \big(T(t)u\big)(x)}{t} - a\|u'(x)\| \right\| \to 0 \quad \text{as } t \downarrow 0, \tag{4.13}$$

for $u \in \mathrm{C}_\mathrm{b}^1$, and provides a uniform (in $x \in X$) 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 simplified version of (4.13):

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

4.6. **Relation to parametric uncertainty.** 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 [25], Zhao and Guan [40], 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. [28], 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 $t \geq 0$, $u \in \mathrm{C}_\mathrm{b}$, and $x \in X$, we consider

$$\big(E(t)u\big)(x) := \sup_{\theta \in X} \left(\int_X u\big(x + y + t\theta\big)\,\mu_t(\mathrm{d}y) - t\varphi\big(\|\theta\|\big) \right). \tag{4.14}$$

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

$$T(t)u \le E(t)u \le I(t)u \quad \text{for all } t \ge 0. \tag{4.15}$$

Moreover, one readily verifies that $E(t)\colon \mathrm{Lip_b} \to \mathrm{Lip_b}$ with

$$\|E(t)u\|_{\mathrm{Lip}} \le \|u\|_{\mathrm{Lip}} \quad \text{and} \quad \|E(t)u\|_\infty \le \|u\|_\infty \quad \text{for all } u \in \mathrm{Lip_b}.$$

Similar as in the proof of Lemma 3.3, one sees that also $E(h)\colon \mathrm{C_b} \to \mathrm{C_b}$ is well-defined and 1-Lipschitz for sufficiently small $h \ge 0$. For $u \in \mathrm{C_b}$ and $\pi = \{t_0, \ldots, t_m\} \in P_t$ with $m \in \mathbb{N}$ and $0 = t_0 < \ldots < t_m = t$ (see Remark 3.14), we define

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

Let $u \in \mathrm{C_b}$. Then, (4.15) implies that

$$E(\pi)u \le I(\pi)u. \tag{4.16}$$

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

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

for all $n \in \mathbb{N}$. Hence, using (4.16),

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

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

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

This shows that

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

for all $t \ge 0$ and $u \in \mathrm{C_b}$. Proceeding as in Nendel and Röckner [26], one sees that $N = \big(N(t)\big)_{t\ge0}$ is a strongly continuous convex transition semigroup, where the strong continuity (property (iii) in Definition 3.11) follows from $T(t)u \le N(t)u \le S(t)u$ for all $u \in \mathrm{C_b}$.

Let $u \in D(B) \cap \mathrm{C_b^1}$ and $\varepsilon > 0$. In the proof of Lemma 3.10, we have implicitly proved that, for sufficiently small $h > 0$,

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

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

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

showing that the generator of S and N coincide on $D(B) \cap \mathrm{C_b^1}$. Similar as in Section 4.3, one sees that the map $t \mapsto N(t)u_0$, for $u_0 \in \mathrm{C_b}$, 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. [12], that

$$S(t)u = N(t)u \quad \text{for all } t \ge 0 \text{ and } u \in \mathrm{C_b}. \tag{4.17}$$

Approximating $u \in \mathrm{Lip}$ with $(u_n)_{n\in\mathbb{N}}$ in $\mathrm{C_b}$ as in Section 4.4, it follows that (4.17) holds for all $u \in \mathrm{Lip}$. We therefore see that the nonparametric uncertainty captured by $I(t)$ collapses to pure drift uncertainty as $t \downarrow 0$.

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 $F\colon X \to X$ with Lipschitz constant $L \geq 0$, we consider the initial value problem

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

For $x \in X$, we define $\big(\psi_t(x)\big)_{t\in\mathbb{R}}$ 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 given, e.g., in terms of a suitable (stochastic integral w.r.t. 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,

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

which shows that Assumption (A1') in Section 4.2 is satisfied. Another application of Gronwall's lemma shows that

$$\|\psi_h(x) - x\| \leq he^{Lh}\|F(x)\| \leq he^{Lh}\big(\|F(0)\| + L\|x\|\big) \tag{5.3}$$

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

$$\sup_{\|x\|\leq r} \|\psi_h(x) - x\| \leq he^{Lh}\big(\|F(0)\| + Lr\big) \to 0 \quad \text{as } h \downarrow 0.$$

From (5.2) and (5.3), it follows that

$$\begin{aligned} \|x\| = \big\|\psi_{-h}\big(\psi_h(x)\big) - \psi_{-h}\big(\psi_h(0)\big)\big\| &\leq e^{Lh}\|\psi_h(x) - \psi_h(0)\| \\ &\leq e^{Lh}\|\psi_h(x)\| + he^{2Lh}\|F(0)\|, \end{aligned}$$

which implies that $\|\psi_h(x)\| \geq e^{-Lh}\|x\| - he^{Lh}\|F(0)\|$ for all $h \geq 0$ and $x \in X$. Hence, for every $h_0 \geq 0$,

$$\inf_{h\in[0,h_0]} \|\psi_h(x)\| \to \infty \quad \text{as } \|x\| \to \infty,$$

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

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

Then, the semigroup $S = \big(S(t)\big)_{t\geq 0}$ 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

$$\partial_t u(t,x) = \big(D_x u(t,x)\big)\big(F(x)\big) + \varphi^*\big(\|D_x u(t,x)\|\big), \tag{5.4}$$

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 $G\colon D(G) \subset X \to X$ be the generator of a C_0-semigroup $P = \big(P(t)\big)_{t\geq 0}$ on X with compact resolvent.[4] 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. [29, 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

$$\psi_t(x) := P(t)x + \int_0^t P(s)m \, \mathrm{d}s$$

be the unique solution to the abstract Cauchy problem

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

Since P is a C_0-semigroup on X, there exist $M \geq 0$ 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. 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 identity[5] $GKx = \lambda Kx - x$ together with $\|Kx\| \leq \frac{1}{\lambda - \omega} = 1$ for all $x \in X$ yields

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

as $h \downarrow 0$ for all $r \geq 0$. As in the previous example, we choose $\mu_t = \delta_0$, so that the assumptions (A1'), (A2), and (A3) are satisfied. The nonlinear transition semigroup yields a solution to the nonlinear PDE

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

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

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

[5]Note that $Kx \in D(G)$.

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

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

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 $u \in \mathrm{C_b}$, the operator $I(t)$ is given by

$$(I(t)u)(x) = \sup_{\nu\in\mathcal{P}_p(X)} \bigg(\int_X u(x+z)\,\nu(\mathrm{d}z) - \varphi_t\big(\mathcal{W}_p(\mu_t,\nu)\big)\bigg).$$

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

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

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 $\varphi = \infty \cdot \mathbb{1}_{(a,\infty)}$ with $a \geq 0$. In the context of Mathematical Finance, this corresponds to a Bachelier model with a Wasserstein perturbation in the law of the underlying Brownian motion. In this case, the PDE (5.6) simplifies to

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

We see that this is the same PDE that appears in the context of a Bachelier Model with an uncertain drift within the interval $[-a,a]$, cf. Coquet et al. [11]. For $t \geq 0$ and $u_0 \in \mathrm{Lip}$, the functions $S(t)u_0$ and $-S(t)(-u_0)$ can be interpreted as upper and lower price bounds (depending on the current price $x \in \mathbb{R}$) of the contingent claim $u_0(\Xi_t^x)$ under Wasserstein uncertainty, respectively. The discussion in Section 4.6 shows that the consideration of nonparametric uncertainty actually reduces to parametric uncertainty in terms of pure drift uncertainty. If $u_0(x) = (x-k)^+$ is a call option with strike price $k \in \mathbb{R}$, by (4.17), the upper and lower bounds are given by

$$\big(S(t)u_0\big)(x) = \mathbb{E}\big(u_0(x+at+W_t)\big) \quad \text{and} \quad -\big(S(t)(-u_0)\big)(x) = \mathbb{E}\big(u_0(x-at+W_t)\big) \tag{5.8}$$

for all $t \geq 0$ and $x \in \mathbb{R}$, where $(W_t)_{t\geq 0}$ is a Brownian motion with standard deviation $\sigma > 0$ on some probability space. The price bounds in (5.8) can, e.g., be computed using a simple Monte-Carlo simulation, see Section 4.5. The price bounds are depicted in Figure 1.

5.4. **Ornstein-Uhlenbeck processes.** We consider a separable Hilbert space X. Let $G\colon D(G) \subset X \to X$ be the generator of a C_0-semigroup $P = \big(P(t)\big)_{t\geq 0}$ 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 $C\colon X \to X$ on some filtered probability space, satisfying the usual assumptions. For $t \geq 0$ and $x \in X$, let

$$\psi_t(x) := P(t)x + \int_0^t P(s)m\,\mathrm{d}s$$

be the unique solution to the abstract Cauchy problem

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

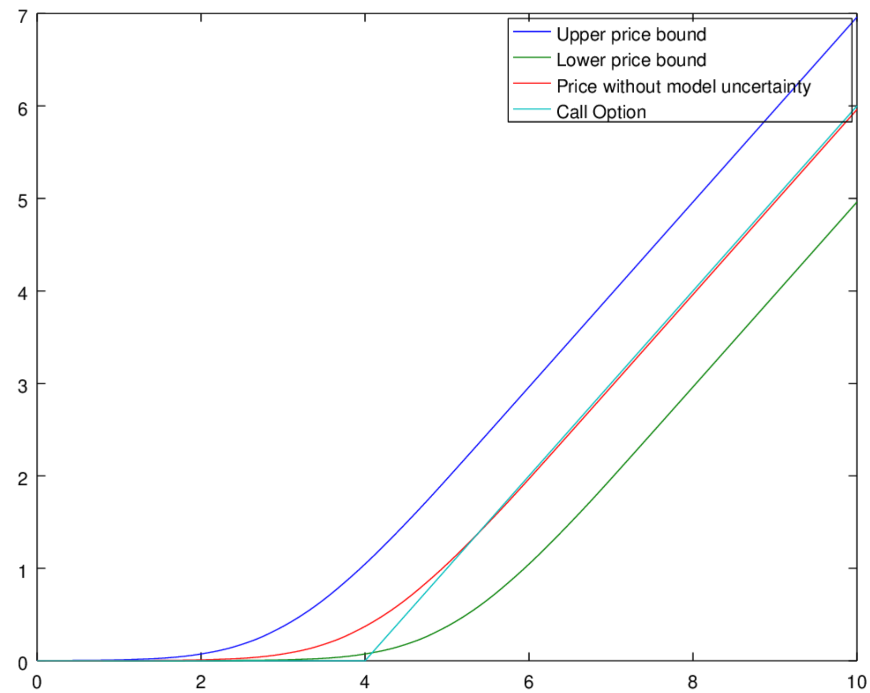

FIGURE 1. Upper and lower price bounds for a call option $u_0 = (x-k)^+$ in a Bachelier model with $\sigma = 1$, $k = 4$, $t = 1$, $a = 1$, and $x \in [0, 10]$ in blue and green, respectively. In red and cyan, the price without model uncertainty and the function u_0 are depicted, respectively.

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

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

In this case, μ_t is the law of the stochastic convolution $\int_0^t P(t-s)\,\mathrm{d}W_s$. By definition, the process (Ξ_t^x) is a mild solution to the SPDE

$$\mathrm{d}\Xi_t^x = (G\Xi_t^x + m)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0^x = x.$$

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

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

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

$$\partial_t u(t,x) = \big\langle Gx + m, D_x u(t,x)\big\rangle + \frac{1}{2}\mathrm{tr}\big(CD_{xx}u(t,x)\big) + \varphi^*\big(\|D_x u(t,x)\|\big). \tag{5.9}$$

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, (5.9) shows how the two forms of noise, aleatoric and epistemic, enter the equation. The aleatoric (random) noise leads to the term $\frac{1}{2}\operatorname{tr}\big(CD_{xx}u(t,x)\big)$ and the epistemic (uncertain) noise leads to the term $\varphi^*\big(\|D_x u(t,x)\|\big)$. For a detailed discussion of (5.9) in the context of parametric uncertainty, we refer to Nendel and Röckner [26].

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

$$Y_t:=\exp\Big(\big(\alpha-\tfrac{\sigma^2}{2}\big)t+\sigma W_t\Big),$$

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

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

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

$$\big(I(t)u\big)(x)=\sup_{\nu\in\mathcal{P}_p(X)}\bigg(\int_X u(xy)\,\nu(\mathrm{d}z)-\varphi_t\big(\mathcal{W}_p(\mu_t,\nu)\big)\bigg),$$

where the considered metric $\mathcal{W}_p$ is a logarithmic Wasserstein distance. In the context of Mathematical Finance, this corresponds to the Black-Scholes model with an additional uncertain perturbation in the law of the underlying geometric Brownian motion. Similar to the discussion in Section 5.3, the functions $S(t)u_0$ and $-S(t)(-u_0)$ can be interpreted as upper and lower price bounds in an uncertain Black-Scholes model, respectively. The related semigroup $\big(S(t)\big)_{t\geq 0}$ can be computed by solving the nonlinear PDE

$$\partial_t u(t,x)=\alpha\partial_x u(t,x)+\tfrac{\sigma^2}{2}\partial_{xx}u(t,x)+\varphi^*\big(|x\partial_x u(t,x)|\big). \tag{5.10}$$

Choosing $\varphi:=\infty\cdot\mathbb{1}_{(a,\infty)}$ with $a\geq 0$, (5.10) simplifies to

$$\partial_t u(t,x)=\tfrac{\sigma^2}{2}\partial_{xx}u(t,x)+\sup_{\beta=\alpha\pm a}\beta x\partial_x u(t,x). \tag{5.11}$$

If $u_0(x)=(x-k)^+$ is a call option with strike price $k\geq 0$, by (4.17), the upper and lower bounds can be computed via

$$\big(S(t)u_0\big)(x)=\mathbb{E}\big(u_0(xe^{at}Y_t)\big) \quad \text{and} \quad -\big(S(t)(-u_0)\big)(x)=\mathbb{E}\big(u_0(xe^{-at}Y_t)\big), \tag{5.12}$$

for $t\geq 0$ and $x\in X$, by numerically simulating the random variable Y_t. We depict the bounds in Figure 2.

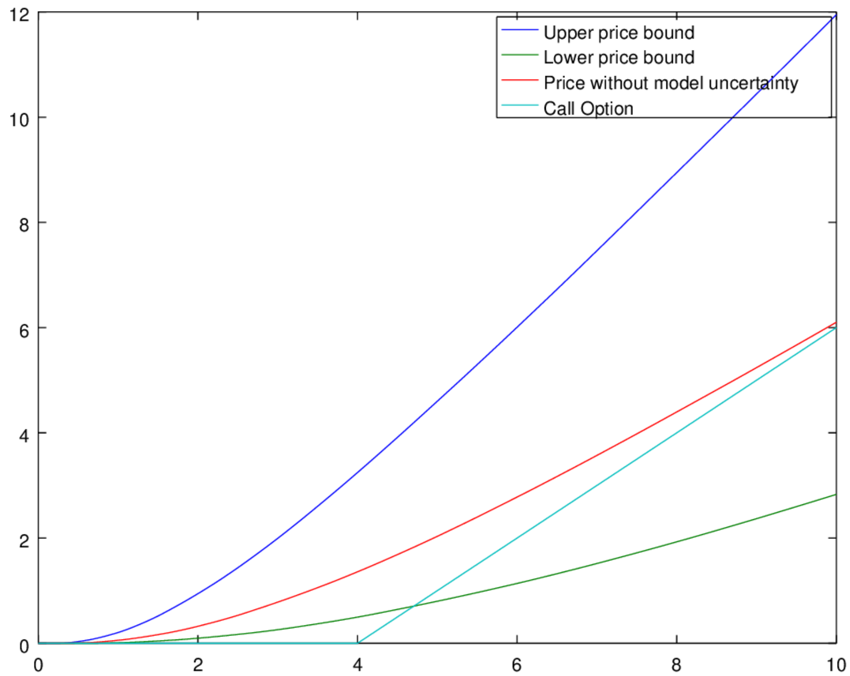

Figure 2. Upper and lower price bounds for a call option $u_0 = (x-k)^+$ in a Black-Scholes model with $\alpha = 0$, $\sigma = 1$, $k = 4$ $t = 1$, $a = 0.5$, and $x \in [0, 10]$ in blue and green, respectively. In red and cyan, the price without model uncertainty and the function u_0 are depicted, respectively.

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. Robust uncertainty sensitivity analysis. *Preprint*, arXiv:2006:12022, 2020.

[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. Forthcoming in *Electronic Communications in Probability*.

[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 and M. Kupper. Nonlinear semigroups built on generating families and their Lipschitz sets. *Preprint*, arXiv:2011.13664, 2020.

[10] 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.

[11] 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.

[12] 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.
[13] G. de Cooman, F. Hermans, and E. Quaeghebeur. Imprecise Markov chains and their limit behavior. *Probab. Engrg. Inform. Sci.*, 23(4):597–635, 2009.
[14] R. Denk, M. Kupper, and M. Nendel. A semigroup approach to nonlinear Lévy processes. *Stochastic Processes and their Applications*, 130(3):1616–1642, 2020.
[15] D. Dentcheva and A. Ruszczyński. Time-coherent risk measures for continuous-time Markov chains. *SIAM J. Financial Math.*, 9(2):690–715, 2018.
[16] S. Eckstein. Extended Laplace principle for empirical measures of a Markov chain. *Adv. in Appl. Probab.*, 51(1):136–167, 2019.
[17] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Math. Finance*, 7(1):1–71, 1997.
[18] T. Fadina, A. Neufeld, and T. Schmidt. Affine processes under parameter uncertainty. *Probab. Uncertain. Quant. Risk*, 4:Paper No. 1, 35, 2019.
[19] R. Gao and A. J. Kleywegt. Distributionally Robust Stochastic Optimization with Wasserstein Distance. *Preprint*, arXiv:1604.02199, 2016.
[20] D. J. Hartfiel. *Markov set-chains*, volume 1695 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1998.
[21] M. Hu and S. Peng. *G*-Lévy Processes under Sublinear Expectations. *Preprint*, arXiv:0911.3533, 2009.
[22] 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.
[23] T. Krak, J. De Bock, and A. Siebes. Imprecise continuous-time Markov chains. *Internat. J. Approx. Reason.*, 88:452–528, 2017.
[24] 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.
[25] 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.
[26] M. Nendel and M. Röckner. Upper envelopes of families of Feller semigroups and viscosity solutions to a class of nonlinear Cauchy problems. *Preprint*, arXiv:1906.04430, 2019.
[27] A. Neufeld and M. Nutz. Nonlinear Lévy processes and their characteristics. *Trans. Amer. Math. Soc.*, 369(1):69–95, 2017.
[28] M. Nisio. On a non-linear semi-group attached to stochastic optimal control. *Publ. Res. Inst. Math. Sci.*, 12(2):513–537, 1976/77.
[29] A. Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983.
[30] 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.
[31] G. Pflug and D. Wozabal. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442, 2007.
[32] G. C. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM J. Optim.*, 22(1):1–23, 2012.
[33] G. C. Pflug and A. Pichler. *Multistage stochastic optimization*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2014.
[34] D. Rudolf and N. Schweizer. Perturbation theory for Markov chains via Wasserstein distance. *Bernoulli*, 24(4A):2610–2639, 2018.
[35] H. M. Soner, N. Touzi, and J. Zhang. Wellposedness of second order backward SDEs. *Probab. Theory Related Fields*, 153(1-2):149–190, 2012.
[36] C. Villani. *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media, 2008.
[37] D. Škulj. Discrete time Markov chains with interval probabilities. *Internat. J. Approx. Reason.*, 50(8):1314–1329, 2009.
[38] W. Wiesemann, D. Kuhn, and B. Rustem. Robust Markov decision processes. *Math. Oper. Res.*, 38(1):153–183, 2013.
[39] I. Yang. A convex optimization approach to distributionally robust Markov decision processes with Wasserstein distance. *IEEE Control Syst. Lett.*, 1(1):164–169, 2017.

[40] 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
E-mail address: sven.fuhrmann@uni-konstanz.de

Department of Mathematics and Statistics, University of Konstanz
E-mail address: kupper@uni-konstanz.de

Center for Mathematical Economics, Bielefeld University
E-mail address: max.nendel@uni-bielefeld.de