

SCALING LIMITS OF MULTI-PERIOD DISTRIBUTIONALLY ROBUST OPTIMIZATION PROBLEMS

MAX NENDEL, ARIEL NEUFELD, KYUNGHYUN PARK, AND ALESSANDRO SGARABOTTOLO

ABSTRACT. We examine the scaling limit of multi-period distributionally robust optimization (DRO) problems via a semigroup approach. Each period involves a worst-case maximization over distributions in a Wasserstein ball around the transition probability of a reference process with radius proportional to the length of the period, and the multi-period DRO problem arises through its sequential composition. We show that the scaling limit of the multi-period DRO, as the length of each period tends to zero, is a strongly continuous monotone semigroup on C_b . Furthermore, we show that its infinitesimal generator is equal to the generator associated with the non-robust scaling limit plus an additional perturbation term induced by the Wasserstein uncertainty. As an application, we show that when the reference process follows an Itô process, the viscosity solution of the associated nonlinear PDE coincides with the value of continuous-time robust optimization problems under parametric uncertainty.

1. INTRODUCTION

When implementing an optimization method based on a stochastic model, it is likely that a margin for potential inaccuracies in the corresponding probabilities or parameters exists. In the economic literature this phenomenon is often referred to as Knightian uncertainty, see, e.g., [21, 33, 43, 44], or model misspecification, see, e.g., [20, 49]. Robust optimization, which presents a worst-case approach, accommodates this issue by solving min-max or max-min problems defined over two sets: the set of admissible controls (to be optimized) and the set of plausible probability distributions (representing model uncertainty). In this context, there exists a wide range of mathematical formulations, which can be broadly classified depending on whether the uncertainty stems from probabilistic imprecision, i.e., distributional uncertainty, see, e.g., [9–12, 42, 59, 67, 69, 92], or from model parameters, see, e.g., [24, 29, 56, 64–66, 68, 74–76, 83].

A primary interest across these frameworks is to obtain a dynamic equation—typically a partial differential equation (PDE)—that characterizes the value function of the robust problem and allows the computation of optimal strategies. In this context, the dynamic programming principle or semigroup property play a central role. The recent works [27, 28, 31, 32, 40, 62] have extended classical semigroup theory to dynamic optimization problems, see also [35, Chapter II] for a broad discussion on the use of semigroups in the context of viscosity solutions and optimal control. Moreover, in a series of works [13–16], Chernoff-type approximation results for nonlinear expectations have been developed, which allow to recover dynamic consistency starting from families

Date: November 26, 2025.

Key words: Distributionally robust optimization, nonlinear semigroup, infinitesimal generator, viscosity solution, Hamilton-Jacobi-Bellman-Isaacs equation, g -expectation, model uncertainty.

MSC2020 Subject Classification: Primary 90C17; 47H20; Secondary 90C30; 90C31; 60J35.

Funding: M. Nendel is financially supported by the Natural Sciences and Engineering Research Council of Canada via Discovery Grant no. RGPIN-2025-04219. A. Neufeld gratefully acknowledges support by the MOE AcRF Tier 2 Grant MOE-T2EP20222-0013. K. Park and A. Sgarabottolo gratefully acknowledge support by the Presidential Postdoctoral Fellowship of Nanyang Technological University and by the National Research Foundation of Korea (Grant DOI: RS-2025-02633175).

of operators that need not be dynamically consistent. The key idea of a Chernoff approximation is to start from a family of one-step operators, e.g., $(I(t))_{t \geq 0}$ with $I(t) : C_b \rightarrow C_b$, which satisfy suitable properties, and show that these properties are preserved under iterations of the type

$$I(t_1 - t_0) \circ \cdots \circ I(t_k - t_{k-1}), \quad \text{for } 0 = t_0 < \cdots < t_k = t,$$

where $\{t_0, \dots, t_k\}$ is a partition of a finite time interval $[0, t]$. The dynamic consistency that is naturally obtained on the multi-period level can then be transferred to continuous time when taking the limit over increasingly finer partitions of finite time intervals. While such results are obtained under rather abstract conditions and can be applied to a variety of frameworks, they require convexity of the one-step operators $(I(t))_{t \geq 0}$. Hence, they are not suitable for addressing robust optimization problems, where, in general, the additional infimum breaks convexity. Some exceptions arise when the control enters in the one-step operators in a convenient way, preserving the convexity of the operators themselves, see, e.g., [17]. However, a direct application of the previous results in a general robust optimization framework is not possible.

The papers [4, 40], which consider transition semigroups for classes of time-homogeneous Markov processes under distributional uncertainty, form the starting point of this article. In these works, the authors study a sequential composition of nonlinear expectation operators under distributional uncertainty, where the nonlinear expectation operators are modeled as a perturbation of the transition probabilities of the Markov process within a proximity in Wasserstein distance. In this setting, it is shown that, as the time horizon is divided into progressively finer periods with a proportional scaling of the level of uncertainty, the sequential composition converges to a semigroup in the mixed topology, see (2.1) in Section 2.1. Furthermore, by examining the infinitesimal generator of the semigroup, it is shown that the semigroup gives rise to viscosity solutions to a nonlinear PDE.

The goal of this article is to extend the results of [40] by proposing a semigroup approach for a class of distributionally robust optimization (DRO) problems, thus providing an analytic bridge between Wasserstein distributionally robust optimization problems, see, e.g., [11, 12, 42, 59, 67, 69, 92], and continuous-time robust optimization problems, see, e.g., [50, 57, 72, 73, 80, 81, 86]. While certain proof techniques in this article resemble those used in [4, 40] for nonlinear expectations involving merely a supremum (or infimum) over the uncertainty set, the min-max structure of the robust optimization introduces major technical challenges. At the core of our analysis lies an explicit description of the nonlinear infinitesimal generator, which is closely tied to the sensitivity of DRO problems, a topic that has recently gained a significant amount of attention. We refer to [3, 41, 47, 63] for sensitivity analyses for Wasserstein distributionally robust optimization problems in one period, [8, 54, 79] in a multi-period setting, and [6, 7, 23, 51, 52, 54] in continuous time.

We proceed with a rigorous description of our results. We consider semigroups on the space C_b of all bounded continuous functions $\mathbb{R}^d \rightarrow \mathbb{R}$. For $p \geq 1$, let \mathcal{P}_p be the set of all probability measures on the Borel σ -algebra $\mathcal{B}(\mathbb{R}^d)$ with finite moment of order p . Throughout, we fix a nonempty set of actions A . For each action $a \in A$, let $(\mu_t^a)_{t \geq 0} \subset \mathcal{P}_p$ be a family of probability measures and let $(\psi_t^a)_{t \geq 0}$ be a family of continuous maps $\mathbb{R}^d \rightarrow \mathbb{R}^d$. For any $a \in A$ and $t \geq 0$, we define a linear operator $T^a(t)$ on C_b by setting

$$(1.1) \quad (T^a(t)f)(x) := \int_{\mathbb{R}^d} f(\psi_t^a(x) + y_t) \mu_t^a(dy_t) \quad \text{for all } x \in \mathbb{R}^d \text{ and } \in C_b.$$

Moreover, we assume that, for each $a \in A$, the family $(T^a(t))_{t \geq 0}$ defined in (1.1) is the transition semigroup of a time-homogeneous controlled Markov process, and we refer to it as the *reference semigroup* associated with the action a . Natural examples include Brownian motions with drift and Ornstein-Uhlenbeck processes, see Example 2.3.

To account for distributional uncertainty in the reference measure μ^a , we consider the set

$$(1.2) \quad \mathcal{B}_t^a(m) := \{\nu \in \mathcal{P}_p \mid \mathcal{W}_p(\mu_t^a, \nu) \leq tm\} \quad \text{for all } t \geq 0 \text{ and } a \in A,$$

where \mathcal{W}_p denotes the p -Wasserstein distance and $m \geq 0$ can be seen as a confidence parameter describing the degree of uncertainty.

We then define a family of one-period operators $(I(t))_{t \geq 0}$ on C_b by

$$(1.3) \quad (I(t)f)(x) := \inf_{a \in A} \sup_{\nu \in \mathcal{B}_t^a(m)} \int_{\mathbb{R}^d} f(\psi_t^a(x) + z)\nu(dz) \quad \text{for all } t \geq 0, f \in C_b, \text{ and } x \in \mathbb{R}^d.$$

Thus, $(I(t)f)(x)$ represents the value of a distributionally robust optimization problem for the uncertainty set $\mathcal{B}_t^a(m)$ in (1.2).

To obtain a dynamically consistent family from $(I(t))_{t \geq 0}$, we consider the set of partitions $P_t := \{\{t_0, \dots, t_k\} \mid 0 = t_0 < \dots < t_k = t, k \in \mathbb{N}\}$ for $t \geq 0$. We then consider the sequential composition operator $\mathcal{I}(\pi)$ over a partition $\pi = \{t_0, \dots, t_k\} \in P_t$ for $t \geq 0$, given by

$$(1.4) \quad (\mathcal{I}(\pi)f)(x) := (I(t_1 - t_0) \cdots I(t_k - t_{k-1})f)(x) \quad \text{for all } f \in C_b \text{ and } x \in \mathbb{R}^d,$$

which represents the multi-period version of the distributionally robust optimization problem given in (1.3). Here and throughout, for $s, t \geq 0$, we use the notation $I(s)I(t) := I(s) \circ I(t)$ for the composition of the operators $I(s)$ and $I(t)$, and similarly for the composition of all operators in the paper.

With these notations and definitions in place, Theorem 2.7 shows that if the reference components $(\mu_t^a, \psi_t^a)_{t \geq 0}$ appearing in (1.1) satisfy certain integrability and regularity conditions, see Assumption 2.2, then the family of operators $(\mathcal{S}(t))_{t \geq 0}$, defined by

$$(1.5) \quad (\mathcal{S}(t)f)(x) := \inf_{\pi \in P_t} (\mathcal{I}(\pi)f)(x) \quad \text{for all } t \geq 0, f \in C_b, \text{ and } x \in \mathbb{R}^d,$$

is a strongly continuous monotone semigroup on C_b , see Definition 2.6. This result can be seen as an analogue of the scaling limit derived in [40] with an additional optimization: as the partition in P_t becomes increasingly finer, the family of sequential composition operators in (1.4) converges to a semigroup $(\mathcal{S}(t))_{t \geq 0}$ in the mixed topology.

Furthermore, given the semigroup $(\mathcal{S}(t))_{t \geq 0}$ in (1.5), Theorem 2.10 shows that, under the same assumptions on $(\mu_t^a, \psi_t^a)_{t \geq 0}$, the infinitesimal generator \mathcal{L} of $(\mathcal{S}(t))_{t \geq 0}$, see Definition 2.8, satisfies

$$(1.6) \quad (\mathcal{L}f)(x) = \inf_{a \in A} (\mathcal{L}^a f)(x) + m\|\nabla f(x)\| \quad \text{for all } x \in \mathbb{R}^d$$

if the function f is sufficiently regular, where, for each $a \in A$, \mathcal{L}^a is the infinitesimal generator of the reference semigroup T^a in (1.1). In other words, $\mathcal{L}f$ equals the Bellman operator $\inf_{a \in A} \mathcal{L}^a f$ corresponding to an optimal control problem over the action set A plus an additional perturbation term $m\|\nabla f(\cdot)\|$, induced by the Wasserstein uncertainty in (1.2).

Lastly, as an application of our main results, Theorem 3.3 shows that the map $v: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}$, defined by $v(t, x) := (\mathcal{S}(t)u_0)(x)$ for $(t, x) \in [0, \infty) \times \mathbb{R}^d$, is a viscosity solution to the PDE with initial condition $v(0, x) = u_0(x)$, see (3.1) and Definition 3.1, which is governed by the nonlinear integro-differential operator \mathcal{L} . We point out that the infinitesimal generator is of (additively separated) inf-sup-type, so that the aforementioned PDE corresponds to the Isaacs equation related to an abstract inf-sup optimal control problem. Theorem 3.3 then allows us to show that, for specific choices of the reference components $(\mu^a, \psi^a)_{a \in A}$, the viscosity solution v coincides with values of a g -expectation, see Corollary 3.7, or with the value function of a continuous-time robust optimization problem under drift uncertainty, see Corollary 3.9. This resembles a collapse of nonparametric uncertainty to parametric uncertainty in the scaling limit, which has also been observed in [4, 40] without an additional optimization.

The rest of the paper is structured as follows. In Section 2, we introduce the setup and state the first two main results, Theorem 2.7 and Theorem 2.10, concerning continuity properties, dynamic consistency, and the infinitesimal behavior of the scaling limit \mathcal{S} of the multi-period DRO problems. Section 3 contains the third main result, Theorem 3.3, which shows that the semigroup \mathcal{S} gives rise to viscosity solutions to a possibly nonlocal Hamilton-Jacobi-Bellman-Isaacs (HJBI) equation, and applications in the context of g -expectations and robust optimization with drift uncertainty. Section 4 discusses the related ‘best-case’ optimization, which plays a crucial role in the proofs of the main results. Sections 5, 6, and 7 contain the proofs of Sections 2.3, 2.4, and 3, respectively. Appendix A contains a series of auxiliary statements.

2. CONSTRUCTION OF THE SCALING LIMIT OF MULTI-PERIOD DROS

2.1. Notation and preliminaries. Throughout this paper, let $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ be the Euclidean norm and inner product on \mathbb{R}^d , respectively. If a subset $K \subset \mathbb{R}^d$ is compact, we write $K \Subset \mathbb{R}^d$.

We denote by $C_b = C_b(\mathbb{R}^d)$ the set of all continuous functions $f: \mathbb{R}^d \rightarrow \mathbb{R}$ with

$$\|f\|_\infty := \sup_{x \in \mathbb{R}^d} |f(x)| < \infty.$$

We endow the space C_b with the *mixed topology* τ_M between the topology induced by the supremum norm $\|\cdot\|_\infty$ and the topology τ_C of uniform convergence on compacts, i.e., the strongest locally convex topology on C_b that coincides with τ_C on $\|\cdot\|_\infty$ -bounded subsets of C_b , cf. [37, 90]. We point out that, in this particular setup, the mixed topology τ_M is the Mackey topology of the dual pair (C_b, ca) , where ca denotes the space of all countably additive signed Borel measures on \mathbb{R}^d of finite variation. Moreover, τ_M belongs to the class of *strict topologies*, cf. [18, 82, 89]. We also refer to [45, 46, 55] for surveys on semigroups in mixed or strict topologies and their relation to Markov processes.

While the explicit definition of the mixed topology is rather involved, convergence of sequences in the mixed topology can be characterized as follows. A sequence $(f_n)_{n \in \mathbb{N}} \subset C_b$ converges to $f \in C_b$ in the mixed topology if and only if

$$(2.1) \quad \sup_{n \in \mathbb{N}} \|f_n\|_\infty < \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \|f_n - f\|_{\infty, K} = 0 \quad \text{for every } K \Subset \mathbb{R}^d,$$

where $\|f\|_{\infty, K} := \sup_{x \in K} |f(x)|$ for $f \in C_b$ and $K \Subset \mathbb{R}^d$. In this case, we write $f_n \xrightarrow{\tau_M} f$.

We point out that the mixed topology is not metrizable and continuity is not equivalent to sequential continuity. However, for convex and concave monotone operators defined on C_b and taking values in a locally convex vector lattice, continuity in the mixed topology is equivalent to sequential continuity, cf. [61, Corollary 2.9], so that, for our purposes, it is enough to restrict our attention to the convergence of sequences in the mixed topology and sequential continuity of operators defined on C_b .

In our analysis, we also consider the subspace Lip_b of all bounded Lipschitz continuous functions $f: \mathbb{R}^d \rightarrow \mathbb{R}$. For $f \in Lip_b$, we write $\|f\|_{Lip}$ for the (smallest) Lipschitz constant of f , i.e.,

$$\|f\|_{Lip} := \sup_{\substack{x, y \in \mathbb{R}^d \\ x \neq y}} \frac{|f(x) - f(y)|}{\|x - y\|}.$$

Moreover, for $k \in \mathbb{N}$, C_b^k denotes the space of all k -times continuously differentiable functions $f \in C_b$ with bounded derivatives up to order k and C_c^k denotes the space of all $f \in C_b^k$ with compact support $\text{supp}(f)$. As usual, we set $C_c^\infty := \bigcap_{k \in \mathbb{N}} C_c^k$, and we use the notation ∇f for the gradient of f and $\nabla^2 f$ for its Hessian.

Finally, throughout this article, we fix some $p \in (1, \infty)$ and write $\mathcal{P}_p := \mathcal{P}_p(\mathbb{R}^d)$ for the set of probability measures μ on the Borel σ -algebra $\mathcal{B}(\mathbb{R}^d)$ with finite moment of order p , i.e., $\int_{\mathbb{R}^d} \|y\|^p \mu(dy) < \infty$. On \mathcal{P}_p , we consider the p -Wasserstein distance \mathcal{W}_p , defined by

$$\mathcal{W}_p(\mu, \nu) := \left(\inf_{\gamma \in \text{Cpl}(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|z - y\|^p \gamma(dy, dz) \right)^{1/p},$$

where $\text{Cpl}(\mu, \nu)$ is the set of all couplings between μ and ν , i.e., probability measures on the product σ -algebra of $\mathcal{B}(\mathbb{R}^d)$ on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal μ and second marginal ν . We refer to [1, 88] for a detailed discussion of the Wasserstein distance in the broader context of optimal transport.

2.2. The controlled reference Markov process and its transition semigroup. Throughout, we consider a fixed nonempty *action set* A . Following the framework in [40], we consider a time-homogeneous controlled Markov process along with its transition semigroup, which is assumed to be described by the following two components.

Setting 2.1. For any $a \in A$,

- (i) $\mu^a := (\mu_t^a)_{t \geq 0} \subset \mathcal{P}_p$ is a family of probability measures with finite p -moment,
- (ii) $\psi^a := (\psi_t^a)_{t \geq 0}$ is a family of continuous maps $\mathbb{R}^d \rightarrow \mathbb{R}^d$.

We assume that the families μ^a and ψ^a satisfy the *Chapman-Kolmogorov equations*

$$(2.2) \quad \int_{\mathbb{R}^d} f(\psi_{s+t}^a(x) + y_{s+t}) \mu_{s+t}^a(dy_{s+t}) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f\left(\psi_s^a(\psi_t^a(x) + y_t) + y_s\right) \mu_s^a(dy_s) \mu_t^a(dy_t)$$

for all $s, t \geq 0$ and $f \in C_b$.

For each $a \in A$, the families μ^a and ψ^a correspond to a time-homogeneous Markov process $(X_t^{:,a})_{t \geq 0}$ of the form

$$(2.3) \quad X_t^{x,a} := \psi_t^a(x) + Y_t^a \quad \text{with} \quad Y_t^a \sim \mu_t^a \quad \text{for all } x \in \mathbb{R}^d \text{ and } t \geq 0,$$

whose transition kernels $p_t^a : \mathbb{R}^d \rightarrow \mathcal{P}_p$, $x \mapsto p_t^a(x; \cdot)$ are given by

$$(2.4) \quad p_t^a(x; B) := \mu_t^a(\{y \in \mathbb{R}^d \mid \psi_t^a(x) + y \in B\})$$

for all $t \geq 0$, $x \in \mathbb{R}^d$, and $B \in \mathcal{B}(\mathbb{R}^d)$.

For any $a \in A$ and $t \geq 0$, we define a linear operator $T^a(t)$ on C_b via

$$(2.5) \quad (T^a(t)f)(x) := \int_{\mathbb{R}^d} f(x_t) p_t^a(x; dx_t) = \int_{\mathbb{R}^d} f(\psi_t^a(x) + y_t) \mu_t^a(dy_t)$$

for all $f \in C_b$ and $x \in \mathbb{R}^d$, where the last equality follows from the particular shape of the transition kernel p_t^a in (2.4).

By (2.2), for each $a \in A$, the family $T^a := (T^a(t))_{t \geq 0}$ in (2.5) satisfies the semigroup property

$$T^a(s+t)f = T^a(t)T^a(s)f \quad \text{for all } s, t \geq 0 \text{ and } f \in C_b.$$

That is, T^a is the transition semigroup of the time-homogeneous Markov process $(X_t^{:,a})_{t \geq 0}$ in (2.3), and we refer to it as the *reference semigroup* associated to the action $a \in A$.

For our analysis, we need to impose several conditions for the behavior of the reference semigroups $(T^a)_{a \in A}$, more precisely, of its components $(\mu^a)_{a \in A}$ and $(\psi^a)_{a \in A}$, which we collect in the following assumption.

Assumption 2.2. The following conditions hold:

- (i) For all $(x, a) \in \mathbb{R}^d \times A$, $\psi_0^a(x) = x$. Moreover, there exists a constant $C > 0$ such that, for all $x \in \mathbb{R}^d$,

$$\sup_{t>0} \sup_{a \in A} \frac{\|\psi_t^a(x) - x\|}{t} \leq C(1 + \|x\|).$$

- (ii) There exists a constant $c \geq 0$ such that, for every $t \geq 0$, $(\psi_t^a)_{a \in A}$ are Lipschitz continuous with their Lipschitz constants uniformly bounded by e^{ct} , i.e., for all $t \geq 0$ and $x_1, x_2 \in \mathbb{R}^d$,

$$\sup_{a \in A} \|\psi_t^a(x_1) - \psi_t^a(x_2)\| \leq e^{ct} \|x_1 - x_2\|.$$

- (iii) For every $\chi \in C_c^\infty$, it holds that

$$\limsup_{t \downarrow 0} \sup_{a \in A} \left| \int_{\mathbb{R}^d} \frac{\chi(y) - \chi(0)}{t} \mu_t^a(dy) \right| < \infty.$$

- (iv) For all $a \in A$, $\mu_0^a = \delta_0$ and $\lim_{t \downarrow 0} \sup_{a \in A} \int_{\mathbb{R}^d} \|y\|^p \mu_t^a(dy) = 0$. Moreover, for every $\varepsilon > 0$, there exists some $M > 0$ such that

$$\sup_{a \in A} \sup_{t > 0} \frac{\mu_t^a(\{y \in \mathbb{R}^d \mid \|y\| > M\})}{t} < \varepsilon.$$

Before we start our analysis, we provide two examples for $(\mu^a)_{a \in A}$ and $(\psi^a)_{a \in A}$ satisfying Assumption 2.2 and the Chapman-Kolmogorov equation (2.2).

Example 2.3. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ be a filtered probability space and let $(W_t)_{t \geq 0}$ be a d -dimensional Brownian motion on that filtered probability space. Moreover, let A be a nonempty set.

- a) (Brownian motion with drift) For all $t \geq 0$, $a \in A$, and $x \in \mathbb{R}^d$, define

$$(2.6) \quad \psi_t^a(x) := x + b(a)t,$$

$$(2.7) \quad Y_t^a = \sigma(a)W_t, \quad Y_0^a = 0,$$

with bounded functions $b: A \rightarrow \mathbb{R}^d$ and $\sigma: A \rightarrow \mathbb{R}^{d \times d}$.

- b) (Ornstein-Uhlenbeck process) For every $t \geq 0$, $a \in A$, and $x \in \mathbb{R}^d$, define

$$(2.8) \quad \psi_t^a(x) := e^{-\theta(a)t}x + \int_0^t e^{-\theta(a)s} \kappa(a) ds,$$

$$(2.9) \quad Y_t^a = \int_0^t e^{-\theta(a)s} \sigma(a) dW_s,$$

where $\theta: A \rightarrow \mathbb{R}^{d \times d}$, $\kappa: A \rightarrow \mathbb{R}^d$, and $\sigma: A \rightarrow \mathbb{R}^{d \times d}$ are bounded functions and $\theta(a)$ is symmetric and positive semi-definite for all $a \in A$.

Then, both previously described cases satisfy Assumption 2.2 and the Chapman-Kolmogorov equations (2.2).

Indeed, Assumption 2.2 (i) and (ii) immediately follow from the explicit form of $(\psi^a)_{a \in A}$ together with the boundedness of the functions b , κ , and θ and the fact that $\theta(a)$ is symmetric and positive semi-definite for all $a \in A$. Since both cases lead to Markov processes $(X_t^{\cdot, a})_{t \geq 0}$ for each $a \in A$, the Chapman-Kolmogorov equations (2.2) are satisfied.

For Assumption 2.2 (iii) and (iv), observe that

$$(2.10) \quad Y_t^a \sim \mathcal{N}(0, \sigma(a)\sigma(a)^\top t) \quad \text{or}$$

$$(2.11) \quad Y_t^a \sim \mathcal{N}\left(0, \int_0^t e^{-\theta(a)(t-s)} \sigma(a)\sigma(a)^\top (e^{-\theta(a)(t-s)})^\top ds\right)$$

in a) or b), respectively, where \mathcal{N} refers to a d -dimensional normal distribution. Recall that, for a normally distributed random vector $Y \sim \mu := \mathcal{N}(0, \sigma\sigma^\top)$ with $\sigma \in \mathbb{R}^{d \times d}$, $\chi \in C_c^\infty$ and $t \geq 0$, by Taylor's theorem,

$$\int_{\mathbb{R}^d} \frac{\chi(y) - \chi(0)}{t} \mu(dy) \leq \frac{\|\sigma\|^2}{2t} \|\nabla^2 \chi\|_\infty$$

and, by Markov's inequality,

$$\frac{\mu(\{y \in \mathbb{R}^d \mid \|y\| > M\})}{t} \leq \frac{\|\sigma\|^2}{M^2 t}.$$

Hence, using the boundedness of σ and the fact that $\theta(a)$ is symmetric and positive semi-definite for all $a \in A$, properties (iii) and (iv) follow in both cases, since $\mu_0^a = \delta_0$ for all $a \in A$ and $\lim_{t \downarrow 0} \sup_{a \in A} \int_{\mathbb{R}^d} \|y\|^p \mu_t^a(dy) = 0$ due to the fact that μ_t^a is a normal distribution of the form (2.10) or (2.11).

2.3. Single-period distributionally robust optimization. In this section, we introduce a family of operators that represent a class of single-period distributionally robust optimization (DRO) problems.

Recall the family $(\mu^a)_{a \in A}$ given in Setting 2.1 (i). Throughout, we fix a radius $m \geq 0$, which corresponds to a *degree of uncertainty* that is scaled proportionally to time. Then, for all $a \in A$ and $t \geq 0$, we denote by

$$(2.12) \quad \mathcal{B}_t^a(m) := \{\nu \in \mathcal{P}_p \mid \mathcal{W}_p(\mu_t^a, \nu) \leq tm\}$$

the set of all probability measures that lie within a tm -radius of the reference distribution $\mu_t^a \in \mathcal{P}_p$ w.r.t. the p -Wasserstein distance \mathcal{W}_p .

We define a family of operators $I := (I(t))_{t \geq 0}$ on C_b by setting

$$(2.13) \quad (I(t)f)(x) := \inf_{a \in A} \sup_{\nu \in \mathcal{B}_t^a(m)} \int_{\mathbb{R}^d} f(\psi_t^a(x) + z) \nu(dz)$$

for all $t \geq 0$, $f \in C_b$, and $x \in \mathbb{R}^d$, so that $I(t)f$ corresponds to a static DRO problem for the ambiguity sets $(\mathcal{B}_t^a(m))_{a \in A}$ given in (2.12). Moreover, we consider the family of operators $T := (T(t))_{t \geq 0}$, given by

$$(T(t)f)(x) := \inf_{a \in A} (T^a(t)f)(x) \quad \text{for all } t \geq 0, f \in C_b, \text{ and } x \in \mathbb{R}^d.$$

If $m = 0$, we have $I(t) = T(t)$ for all $t \geq 0$, i.e., the family $(I(t))_{t \geq 0}$ corresponds to a minimization of the reference semigroups $(T^a(t))_{t \geq 0}$ given in (2.5) over $a \in A$.

The following proposition forms the starting point of our theoretical analysis. Its proof can be found in Section 5. For similar statements without the additional minimization, we refer to [4, 40].

Proposition 2.4. *Suppose that Assumption 2.2 holds.*

- (i) *For every $t \geq 0$, the operator $I(t) : C_b \rightarrow C_b$ is well-defined, monotone, and a contraction, i.e., $I(t)f \geq I(t)g$ for all $f, g \in C_b$ with $f \geq g$ and*

$$\|I(t)f - I(t)g\|_\infty \leq \|f - g\|_\infty \quad \text{for all } f, g \in C_b.$$

- (ii) *For all $t \geq 0$ and $f \in \text{Lip}_b$, we have $\|I(t)f\|_{\text{Lip}} \leq e^{ct} \|f\|_{\text{Lip}}$, where $c \geq 0$ is defined in Assumption 2.2 (ii).*

- (iii) *For all $s, t \geq 0$, $f \in C_b$ and $x \in \mathbb{R}^d$,*

$$(I(s)I(t)f)(x) \leq (I(s+t)f)(x).$$

In particular, choosing $m = 0$, these properties also hold for the family T instead of I .

We now turn our focus on the asymptotic relation between the families I and T as $t \rightarrow 0$, which can be understood as a sensitivity w.r.t. the degree of uncertainty, cf. [3, 63] and [6, 8] in a dynamic context. Observe that, by [88], for all $f \in C_b^1$, $t > 0$, $x \in \mathbb{R}^d$, $a \in A$, and $\varepsilon > 0$, there exists $\nu_t^{a,x} \in \mathcal{B}_t^a(m)$ and an optimal coupling $\gamma_t^{a,x} \in \mathcal{P}_p(\mathbb{R}^d \times \mathbb{R}^d)$ between μ_t^a and $\nu_t^{a,x}$ such that

$$\begin{aligned} \frac{(I^a(t)f)(x) - (T^a(t)f)(x)}{t} &\leq \frac{1}{t} \left(\int_{\mathbb{R}^d} f(\psi_t^a(x) + z) \nu_t^{a,x}(dz) - (T^a(t)f)(x) \right) + \varepsilon \\ &= \frac{1}{t} \int_{\mathbb{R}^d \times \mathbb{R}^d} (f(\psi_t^a(x) + z) - f(\psi_t^a(x) + y)) \gamma_t^{a,x}(dy, dz) + \varepsilon \\ &\leq \|\nabla f\|_\infty \frac{1}{t} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|z - y\| \gamma_t^{a,x}(dy, dz) + \varepsilon \\ &\leq \|\nabla f\|_\infty \frac{\mathcal{W}_p(\mu_t^a, \nu_t^{a,x})}{t} + \varepsilon \leq m \|\nabla f\|_\infty + \varepsilon, \end{aligned}$$

where

$$(2.14) \quad (I^a(t)f)(x) := \sup_{\nu \in \mathcal{B}_t^a(m)} \int_{\mathbb{R}^d} f(\psi_t^a(x) + z) \nu(dz)$$

for all $a \in A$, $t \geq 0$, $f \in C_b$, and $x \in \mathbb{R}^d$. Letting $\varepsilon \rightarrow 0$, we thus obtain the following a priori estimate

$$(2.15) \quad \frac{(I^a(t)f)(x) - (T^a(t)f)(x)}{t} \leq m \|\nabla f\|_\infty,$$

which is further strengthened in the following proposition.

Proposition 2.5. *Suppose that Assumption 2.2 holds. For every $m \geq 0$ and every $f \in C_b^1$,*

$$(2.16) \quad \lim_{t \downarrow 0} \frac{I(t)f - T(t)f}{t} = m \|\nabla f(\cdot)\|,$$

where the limit is to be understood in the mixed topology.¹

The proof of Proposition 2.5 is contained in Section 5.

2.4. Multi-period DRO and its scaling limit. In this section, we consider a multi-period version of the DRO problem described in (2.13). By ‘multi-period’, we mean that the decision maker *iteratively* optimizes their worst-case objective over a finite time horizon. We then consider the scaling limit of this multi-period DRO problem as the number of decisions within the finite time horizon tends to infinity.

To that end, we consider the set

$$P := \{\pi \subset [0, \infty) \mid 0 \in \pi, \#\pi < \infty\}$$

of all finite partitions of closed intervals on the positive half line that contain zero. Observe that, by definition, every $\pi \in P$ is either the set containing only zero or of the form $\pi = \{t_0, \dots, t_k\}$ with $0 = t_0 < \dots < t_k$ and $k \in \mathbb{N}$. In the latter case, we define

$$\text{mesh } \pi := \max_{i=1, \dots, k} (t_i - t_{i-1})$$

as the *maximal mesh size* and, for the sake of completeness, we set $\text{mesh}\{0\} := 0$.

Then, for $\pi = \{t_0, \dots, t_k\} \in P$ with $0 = t_0 < \dots < t_k$ and $k \in \mathbb{N}$, we define

$$(2.17) \quad \mathcal{I}(\pi)f := (I(t_1 - t_0) \cdots I(t_k - t_{k-1}))f \quad \text{for all } f \in C_b$$

¹Here and throughout $\|\nabla f(\cdot)\|$ refers to the map $\mathbb{R}^d \rightarrow \mathbb{R}^d$, $x \mapsto \|\nabla f(x)\|$.

as the composition of the operators I given in (2.13) over the partition π , which represents a multi-period DRO problem over the time interval $[0, t_k]$, see Figure 1. Again, for the sake of completeness, we define $\mathcal{I}(\{0\})f := f$.

Since $I(t)$ maps C_b to C_b for all $t \geq 0$, see Proposition 2.4 (i), the multi-period operator $\mathcal{I}(\pi)$ in (2.17) also maps C_b to C_b for all $\pi \in P$.

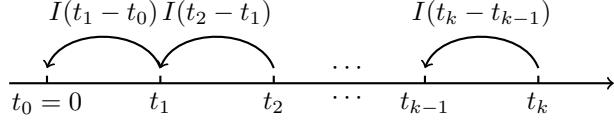


FIGURE 1. Multi-period DRO problem as iteration of the one-step DRO operator I .

By definition of \mathcal{I} in (2.17) and the monotonicity property of I in Proposition 2.4 (iii), for all $\pi^1, \pi^2 \in P$ with $\pi^2 \subset \pi^1$, i.e., π^1 is finer than π^2 , it holds

$$\mathcal{I}(\pi^1)f \leq \mathcal{I}(\pi^2)f \quad \text{for all } f \in C_b.$$

This allows us to define

$$(2.18) \quad (\mathcal{S}(t)f)(x) := \inf_{\pi \in P_t} (\mathcal{I}(\pi)f)(x)$$

for all $t \geq 0$, $f \in C_b$, and $x \in \mathbb{R}^d$, where

$$P_t := \{\pi \in P \mid \max \pi = t\}$$

is the set of all finite partitions of the interval $[0, t]$ containing 0 and $t \geq 0$.

In a first step, we show that the family $\mathcal{S} := (\mathcal{S}(t))_{t \geq 0}$, given in (2.18), is a strongly continuous monotone semigroup according to the following definition.

Definition 2.6. A family $S := (S(t))_{t \geq 0}$ of continuous operators $S(t): C_b \rightarrow C_b$ is a *strongly continuous monotone semigroup* on C_b if the following conditions are satisfied:

- (i) $S(0)f = f$ for every $f \in C_b$.
- (ii) (Monotonicity) $S(t)f \leq S(t)g$ for all $t \geq 0$ and all $f, g \in C_b$ with $f \leq g$;
- (iii) (Semigroup property) $S(t)S(s)f = S(t+s)f$ for all $t, s \geq 0$ and $f \in C_b$;
- (iv) (Strong continuity) the map $t \mapsto S(t)f$ is continuous for all $f \in C_b$.

The following theorem is the first main result of this paper.

Theorem 2.7. Suppose that Assumption 2.2 holds. Then, the family \mathcal{S} in (2.18) is a strongly continuous monotone semigroup on C_b . Moreover,

$$(2.19) \quad \mathcal{S}(t)f = \inf_{\substack{\pi \in P_t \\ \text{mesh } \pi \leq h}} \mathcal{I}(\pi)f \quad \text{for all } h > 0.$$

The proof of Theorem 2.7 can be found in Section 6. In the following, we briefly describe the key points of the proof. Since the properties (i) and (ii) in Definition 2.6 immediately follow from the definition of \mathcal{S} , the proof of Theorem 2.7 mainly focuses on verifying the semigroup property and the strong continuity of \mathcal{S} . Proving the semigroup property of \mathcal{S} hinges on the definition of an auxiliary operator $\hat{\mathcal{S}}: C_b \rightarrow C_b$, see (6.3), which represents the multi-period distributionally robust optimization over (almost) dyadic partitions. Leveraging the monotonicity of the operator I with respect to refining partitions, see Proposition 2.4 (iii), and the continuity properties of the best-case maximization problem, which is studied in detail in Section 4, it is possible to show the semigroup property for the operator $\hat{\mathcal{S}}$ and its strong continuity. In a second step, it is then

shown that $\hat{\mathcal{S}} = \mathcal{S}$. The restriction to partitions $\pi \in P_t$ with mesh $\pi \leq h$ in (2.19), for arbitrary $h > 0$, then follows from the definition of the operator $\hat{\mathcal{S}}$ and its representation as a multi-period optimization over arbitrarily fine dyadic partitions, see (6.4).

We next aim to characterize the infinitesimal generator of the strongly continuous monotone semigroup \mathcal{S} , which is defined as follows.

Definition 2.8. Let S be a strongly continuous monotone semigroup on C_b . Then, the *infinitesimal generator* $L: D(L) \subset C_b \rightarrow C_b$ of S is defined by

$$D(L) := \left\{ f \in C_b \mid \lim_{t \downarrow 0} \frac{S(t)f - f}{t} \in C_b \text{ exists w.r.t. the mixed topology} \right\},$$

$$Lf := \lim_{t \downarrow 0} \frac{S(t)f - f}{t} \quad \text{for } f \in D(L).$$

Remark 2.9. We consider the following infinitesimal generators in the sense of Definition 2.8.

- (i) Observe that, by Assumption 2.2 and Theorem 2.7 (for the special case $m = 0$ and $A = \{a\}$), the reference semigroup T^a in (2.5) is a strongly continuous monotone semigroup on C_b for every $a \in A$. We denote its infinitesimal generator by $\mathcal{L}^a: D(\mathcal{L}^a) \subset C_b \rightarrow C_b$ for all $a \in A$.
- (ii) We denote the infinitesimal generator of the strongly continuous monotone semigroup \mathcal{S} in (2.18), see Theorem 2.7, by $\mathcal{L}: D(\mathcal{L}) \subset C_b \rightarrow C_b$.

The following theorem, which is the main result of this section, provides an explicit description of the infinitesimal generator \mathcal{L} of the strongly continuous monotone semigroup \mathcal{S} .

Theorem 2.10. Suppose that Assumption 2.2 holds, and let $(\mathcal{L}^a)_{a \in A}$ and \mathcal{L} be the infinitesimal generators defined in Remark 2.9. Then, for every $f \in \cap_{a \in A} D(\mathcal{L}^a) \cap C_b^1$ with

$$(2.20) \quad \sup_{a \in A} \|\mathcal{L}^a f\|_\infty < \infty \quad \text{and}$$

$$(2.21) \quad \lim_{t \downarrow 0} \sup_{a \in A} \sup_{x \in K} \left| \frac{T^a(t)f(x) - f(x)}{t} - (\mathcal{L}^a f)(x) \right| = 0 \quad \text{for all } K \Subset \mathbb{R}^d,$$

it holds that $f \in D(\mathcal{L})$ and

$$\mathcal{L}f = \inf_{a \in A} \mathcal{L}^a f + m\|\nabla f(\cdot)\|.$$

The proof of Theorem 2.10 is presented in Section 6. The main challenge is to show that the generator of the semigroup actually coincides with the derivative $\lim_{t \downarrow 0} \frac{I(t)f - f}{t}$ of the one-period operators at zero for functions f as in Theorem 2.10. The result is then a direct consequence of Proposition 2.5, splitting the generator of the DRO into the sum of the generator of the non-robust dynamics and a component arising from uncertainty according to

$$\frac{I(t)f - f}{t} = \frac{T(t)f - f}{t} + \frac{I(t)f - T(t)f}{t}.$$

3. CONNECTION TO NONLINEAR EXPECTATIONS AND ROBUST OPTIMIZATION UNDER DRIFT UNCERTAINTY.

In this section, we connect the strongly continuous monotone semigroup \mathcal{S} , obtained as the scaling limit of the multi-period DRO problems, see Theorem 2.7, with nonlinear PDEs of Hamilton-Jacobi-Bellman-Isaacs-type that arise from nonlinear expectations and robust optimization problems under drift uncertainty.

3.1. Nonlinear PDEs arising from the semigroup \mathcal{S} and their viscosity solutions. In order to establish the connection between the semigroup \mathcal{S} and a nonlinear PDE, the infinitesimal generator \mathcal{L} of \mathcal{S} , see Remark 2.9 (ii), plays a crucial role. For given $u_0 \in C_b$, we consider the nonlinear PDE

$$(3.1) \quad \begin{cases} \partial_t v(t, x) = \mathcal{L}v(t, x) & \text{for all } (t, x) \in (0, \infty) \times \mathbb{R}^d, \\ v(0, x) = u_0(x) & \text{for all } x \in \mathbb{R}^d. \end{cases}$$

In what follows, we introduce the notion of a viscosity solution to (3.1), as considered in [39].

Definition 3.1. Let $u_0 \in C_b$ and $D \subseteq D(\mathcal{L})$ be nonempty. We say that $v: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a D -viscosity subsolution or supersolution to (3.1) if $v \in C_b([0, \infty) \times \mathbb{R}^d)$, $v(0, x) = u_0(x)$ for all $x \in \mathbb{R}^d$ and, for all $(t, x) \in (0, \infty) \times \mathbb{R}^d$ and $\varphi \in C_b^1((0, \infty) \times \mathbb{R}^d)$ with $\varphi(t, \cdot) \in D$, $v(t, x) = \varphi(t, x)$, and $v \leq \varphi$ or $v \geq \varphi$, it holds

$$\partial_t \varphi(t, x) \leq \mathcal{L}\varphi(t, x) \quad \text{or} \quad \partial_t \varphi(t, x) \geq \mathcal{L}\varphi(t, x), \quad \text{respectively.}^2$$

We say v a D -viscosity solution to (3.1) if it is both a D -viscosity sub- and supersolution.

If the set D contains C_b^2 as a subset, then the notion of D -viscosity solutions in Definition 3.1 is stronger than classical notions of viscosity solutions presented, for example, in [25, 26, 35, 65] and slightly weaker than the notion of a D - M viscosity solution, considered in [39]. For a detailed discussion, we refer to [39, Remark 2.11].

In order to establish the existence of a D -viscosity solution to the PDE in (3.1), we consider the following property for operators on C_b , introduced in [39].

Definition 3.2. Let $K := (K(t))_{t \geq 0}$ be a family of operators $C_b \rightarrow C_b$. A family $S := (S(t))_{t \geq 0}$ of operators $C_b \rightarrow C_b$ is called K -convex if for all $\lambda \in [0, 1]$, $f, g \in C_b$, and $t \geq 0$,

$$S(t)(\lambda f + (1 - \lambda)g) \leq \lambda S(t)f + (1 - \lambda)K(t)g.$$

Theorem 3.3. Suppose that Assumption 2.2 holds. Let \mathcal{S} be the strongly continuous monotone semigroup on C_b defined in (2.18), \mathcal{L} be its infinitesimal generator, see Remark 2.9 (ii), and $D \subseteq D(\mathcal{L})$ be nonempty. Then, for every $u_0 \in C_b$, the function $v: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}$, defined by

$$(3.2) \quad v(t, x) := (\mathcal{S}(t)u_0)(x) \quad \text{for all } x \in \mathbb{R}^d,$$

is a D -viscosity solution to the PDE (3.1).

The proof of Theorem 3.3, which is contained in Section 7, is based on the construction of a suitable family $K := (K(t))_{t \geq 0}$ of operators $C_b \rightarrow C_b$ in order to apply [39, Theorem 3.1] to establish the existence of a D -viscosity solution to the PDE (3.1).

Remark 3.4. Let $D \subset \bigcap_{a \in A} D(\mathcal{L}^a) \cap C_b^1$ be a non-empty subset of functions satisfying (2.20) and (2.21) for all $K \in \mathbb{R}^d$. Then, by Theorem 2.10 and Theorem 3.3, the function $v: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ defined in (3.2) is a D -viscosity solution to

$$\begin{cases} \partial_t v(t, x) = \inf_{a \in A} \mathcal{L}^a v(t, x) + m \|\nabla v(t, x)\| & \text{for all } (t, x) \in (0, \infty) \times \mathbb{R}^d, \\ v(0, x) = u_0(x) & \text{for all } x \in \mathbb{R}^d. \end{cases}$$

²We write $C_b([0, \infty) \times \mathbb{R}^d)$ for the set of all continuous bounded functions on $[0, \infty) \times \mathbb{R}^d$. Moreover, we write $C_b^1((0, \infty) \times \mathbb{R}^d)$ for the set of all bounded functions on $(0, \infty) \times \mathbb{R}^d$ having bounded continuous derivative.

3.2. Connection to g -expectation. The result from the preceding section allows us to establish a connection between the semigroup \mathcal{S} and the g -expectation framework, see [24, 74], which we discuss in the sequel.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, $W := (W_t)_{t \geq 0}$ be a d -dimensional standard Brownian motion with its Brownian standard filtration $\mathbb{F} := (\mathcal{F}_t)_{t \geq 0}$, i.e., the natural filtration generated by W augmented by all \mathbb{P} -null sets, and $T > 0$ be a finite time horizon. For any probability measure \mathbb{Q} on (Ω, \mathcal{F}_T) , we write $\mathbb{E}^{\mathbb{Q}}[\cdot]$ for the expectation under \mathbb{Q} .

Definition 3.5. We define the g -expectation $\mathcal{E}: L^2(\mathcal{F}_T) \ni \xi \mapsto \mathcal{E}^g[\xi] \in \mathbb{R}$ by³

$$\mathcal{E}[\xi] := Y_0,$$

where $(Y, Z) \in \mathbb{S}^2(\mathbb{R}) \times \mathbb{L}^2(\mathbb{R}^d)$ is the unique solution to the backward stochastic differential equation (BSDE), see [70, Theorem 3.1],

$$Y_t = \xi + \int_t^T m \|Z_s\| ds - \int_t^T Z_s^\top dW_s.$$

Moreover, its conditional g -expectation with respect to \mathcal{F}_t is defined by

$$\mathcal{E}[\xi | \mathcal{F}_t] := Y_t \quad \text{for all } t \in [0, T].$$

Remark 3.6. By [34, Proposition 3.6], the g -expectation in Definition 3.5 satisfies the following representation for every $\xi \in L^2(\mathcal{F}_T)$ and $t \in [0, T]$:

$$(3.3) \quad \mathcal{E}[\xi | \mathcal{F}_t] = \underset{\vartheta \in \mathcal{C}(m)}{\text{ess sup}} \mathbb{E}^{\mathbb{P}^\vartheta} [\xi | \mathcal{F}_t],$$

where $\mathcal{C}(m)$ is the set of all \mathbb{R}^d -valued, \mathbb{F} -progressively measurable processes $\vartheta = (\vartheta_t)_{t \in [0, T]}$ with $|\vartheta_t| \leq m$ $\mathbb{P} \otimes dt$ -a.e., and for each $\vartheta \in \mathcal{C}(m)$ the probability measure \mathbb{P}^ϑ on (Ω, \mathcal{F}_T) is defined by

$$\frac{d\mathbb{P}^\vartheta}{d\mathbb{P}|_{\mathcal{F}_T}} := \exp \left(-\frac{1}{2} \int_0^T \|\vartheta_s\|^2 ds + \int_0^T \vartheta_s^\top dW_s \right).$$

By the variational representation in (3.3), the conditional g -expectation $(\mathcal{E}[\cdot | \mathcal{F}_t])_{t \in [0, T]}$ is often considered to be a dynamic risk measure, see, e.g., [50, 57, 72, 73, 80, 81, 86].

We now consider the following g -expectation value function defined w.r.t. a given $h \in C_b$. Set

$$(3.4) \quad V(t, x) := \mathcal{E}[h(X_T^{t,x}) | \mathcal{F}_t], \quad \text{for } (t, x) \in [0, T] \times \mathbb{R}^d,$$

where $X_T^{t,x} := x + b(T-t) + (W_T - W_t)$ with $b \in \mathbb{R}^d$.

The following corollary establishes the connection between the strongly continuous monotone semigroup \mathcal{S} with the g -expectation value in (3.4). The proof can be found in Section 7.

Corollary 3.7. Let $(\psi^a)_{a \in A}$ and $(Y^a)_{a \in A}$ be given by (2.6) and (2.7) in Example 2.3 a), respectively, with $b(a) := b \in \mathbb{R}^d$ and $\sigma(a) := I_d$, where I_d is the identity matrix in $\mathbb{R}^{d \times d}$. Then, the following hold:

- (i) The inclusion $C_b^2 \subseteq D(\mathcal{L}^a)$ holds for all $a \in A$, and every $f \in C_b^2$ satisfies (2.20) and (2.21).

³We write $L^2(\mathcal{F}_T)$ for the set of all real-valued, \mathcal{F}_T -measurable random variables ξ with $\mathbb{E}^{\mathbb{P}}[|\xi|^2] < \infty$, $\mathbb{L}^2(\mathbb{R}^d)$ for the set of all \mathbb{R}^d -valued, \mathbb{F} -predictable processes $Z = (Z_t)_{t \in [0, T]}$ with $\mathbb{E}^{\mathbb{P}}[\int_0^{T_1} \|Z_t\|^2 dt] < \infty$, and $\mathbb{S}^2(\mathbb{R})$ for the set of all real-valued, \mathbb{F} -progressively measurable càdlàg (i.e., right-continuous with left limits) processes $Y = (Y_t)_{t \in [0, T]}$ with $\mathbb{E}^{\mathbb{P}}[\sup_{t \in [0, T]} |Y_t|^2] < \infty$.

- (ii) The value function V in (3.4) coincides with the function $v: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, defined by
- $$(3.5) \quad v(t, x) := (\mathcal{S}(T-t)h)(x) \quad \text{for all } (t, x) \in [0, T] \times \mathbb{R}^d,$$

and $V = v$ is the unique viscosity solution⁴ to the following PDE:

$$(3.6) \quad \begin{cases} \partial_t v(t, x) + \frac{1}{2} \Delta v(t, x) + \langle b, \nabla v(t, x) \rangle + m \|\nabla v(t, x)\| = 0 & \text{for all } (t, x) \in [0, T] \times \mathbb{R}^d, \\ v(T, x) = h(x) & \text{for all } x \in \mathbb{R}^d. \end{cases}$$

Remark 3.8. In view of Remark 3.6 and Corollary 3.7, the function v in (3.5) can also be seen as the unique viscosity solution of a nonlinear Kolmogorov equation under drift uncertainty, see, e.g., [7, 56, 64, 65], or, alternatively, as the solution to a Hamilton–Jacobi–Bellman (HJB) equation arising from stochastic control problems with drift control, see, e.g., [35, 77, 91]. For similar observations as well as related settings involving drift and volatility uncertainty (or control), cf. [5, 6, 30, 66, 71, 78, 87], corresponding to second-order backward stochastic differential equations, cf. [22, 58, 84, 85], within the framework of nonlinear semigroups, we refer to [15–17, 32, 40, 62].

3.3. Connection to robust optimization under drift uncertainty. In this section, we aim to establish the connection between the semigroup \mathcal{S} and robust optimization problems under drift uncertainty, which will be formulated within the framework of stochastic differential games, see, e.g., [19, 36, 38, 48, 86]. To that end, we consider a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a d -dimensional standard Brownian Motion $(W_t)_{t \geq 0}$ together with its natural filtration $\mathbb{F} := (\mathcal{F}_t)_{t \geq 0}$ augmented by all \mathbb{P} -null sets.

Assume there exists a decision maker who aims to minimize a terminal cost by choosing an optimal control (so that they act as the minimizer), while the environment systematically acts in opposition to the decision maker's interest (so that it plays the role of the maximizer in adversarial way).

Fix $m \geq 0$, a nonempty set of actions $A \subseteq \mathbb{R}^d$, and a time-horizon $T > 0$. We then consider the following set of admissible controls for the environment

$$(3.7) \quad \Lambda_t := \{\vartheta := (\vartheta_s)_{s \in [t, T]} \mid \vartheta \text{ is } \mathbb{F}\text{-progressively measurable and } \|\vartheta_t\| \leq m \text{ } \mathbb{P}\text{-a.s.}\}$$

and the set of admissible controls set for the decision maker

$$\mathcal{A}_t := \{\alpha = (\alpha_s)_{s \in [t, T]} \mid \alpha \text{ is } \mathbb{F}\text{-progressively measurable and } \alpha_t \in A \text{ } \mathbb{P}\text{-a.s.}\}.$$

We now consider the robust optimization problem of the decision maker w.r.t. a given function $h \in C_b$. Set

$$(3.8) \quad V(t, x) := \inf_{\alpha \in \mathcal{A}_t} \sup_{\vartheta \in \Lambda_t} \mathbb{E}^\mathbb{P} \left[h(X_T^{t, x; \alpha, \vartheta}) \right] \quad \text{for all } (t, x) \in [0, T] \times \mathbb{R}^d,$$

where, for all $(t, x) \in [0, T] \times \mathbb{R}^d$ and all $(\alpha, \vartheta) \in \mathcal{A}_t \times \Lambda_t$, we consider a controlled Brownian Motion $X^{t, x; \alpha, \vartheta} := (X_s^{t, x; \alpha, \vartheta})_{s \in [t, T]}$ starting at time t in x , given by

$$(3.9) \quad X_s^{t, x; \alpha, \vartheta} := x + \int_t^s (b(\alpha_u) + \vartheta_u) du + \int_t^s \sigma(\alpha_u) dW_u \quad \text{for } s \in [t, T],$$

with bounded functions $b: A \rightarrow \mathbb{R}^d$ and $\sigma: A \rightarrow \mathbb{R}^{d \times d}$.

The following corollary establishes the connection between the strongly continuous monotone semigroup \mathcal{S} with the value function v of the robust optimal control problem in (3.8). The proof is contained in Section 7.

Corollary 3.9. Let $(\psi^a)_{a \in A}$ and $(Y^a)_{a \in A}$ be given by (2.6) and (2.7) in Example 2.3 a), respectively, with bounded functions $b: A \rightarrow \mathbb{R}^d$, and $\sigma: A \rightarrow \mathbb{R}^{d \times d}$. Then, the following hold:

⁴ In the classical sense of [25, 26, 35].

- (i) $C_b^2 \subseteq \cap_{a \in A} D(\mathcal{L}^a)$ and every $f \in C_b^3$ satisfies (2.20) and (2.21) in Theorem 2.10.
- (ii) The value function V given in (3.8) coincides with the function $v: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, defined by

$$(3.10) \quad v(t, x) := (\mathcal{S}(T-t)h)(x) \quad \text{for all } (t, x) \in [0, T] \times \mathbb{R}^d,$$

and $V = v$ is the unique viscosity solution to the following PDE:

$$(3.11) \quad \begin{cases} \partial_t v(t, x) + \mathcal{L}v(t, x) = 0 & \text{for all } (t, x) \in [0, T] \times \mathbb{R}^d, \\ v(T, x) = h(x) & \text{for all } x \in \mathbb{R}^d, \end{cases}$$

where the infinitesimal generator \mathcal{L} is given by

$$\mathcal{L}v(t, x) = \inf_{a \in A} \left\{ \frac{1}{2} \operatorname{tr} (\sigma(a)\sigma(a)^\top \nabla^2 v(t, x)) + \langle b(a), \nabla v(t, x) \rangle \right\} + m \|\nabla v(t, x)\|.$$

Remark 3.10. While the robust optimization problem in (3.8) is given as a min–max problem, the resulting PDE in (3.11) takes the form of a HJB equation involving the minimization over controls, i.e., $\inf_{a \in A}$, rather than a full Isaacs-type equation. This reduction stems from the specific structure of the uncertainty in (3.7) and the controlled process in (3.9): the uncertainty appears only in the drift term and is additively separated from the control process. As a consequence, the associated Hamilton-Jacobi-Bellman-Isaacs (HJBI) equation for the problem (3.8), see (7.3) in Section 7, reduces to an HJB equation with an additive nonlinear first-order term in this setup.

4. BEST-CASE MAXIMIZATION

In this section, we discuss the ‘best-case’ maximization problem, where the infimum over the action set A is replaced by a supremum in the definition of the family of operators I . The properties of the resulting operators are essential to prove our main results Theorem 2.7 and Theorem 2.10. We consider the family $J = (J(t))_{t \geq 0}$ of operators, defined by

$$(4.1) \quad (J(t)f)(x) := \sup_{a \in A} \sup_{\nu \in \mathcal{B}_t^a(m)} \int_{\mathbb{R}^d} f(\psi_t^a(x) + z) \nu(dz)$$

for all $t \geq 0$, $f \in C_b$, and $x \in \mathbb{R}^d$, so that $J(t)$ corresponds to the ‘best-case’ maximization problem for the ambiguity set $(\mathcal{B}_t^a(m))_{a \in A}$ for $t \geq 0$. Then, similarly to what has been done for the family I , for $\pi = \{t_0, \dots, t_k\} \in \mathcal{P}$ with $0 = t_0 < \dots < t_k$ and $k \in \mathbb{N}$, we consider the composition

$$(4.2) \quad \mathcal{J}(\pi)f := (J(t_1 - t_0) \cdots J(t_k - t_{k-1}))f \quad \text{for all } f \in C_b.$$

Remark 4.1. Note that, for all $t \geq 0$, $f \in C_b$, $\|J(t)f\|_\infty \leq \|f\|_\infty$ and, for all $t \geq 0$, $f, g \in C_b$, and $x \in \mathbb{R}^d$, it holds

$$|(J(t)f)(x) - (J(t)g)(x)| \leq \sup_{a \in A} \sup_{\nu \in \mathcal{B}_t^a(m)} \int_{\mathbb{R}^d} |f(\psi_t^a(x) + z) - g(\psi_t^a(x) + z)| \nu(dz) \leq \|f - g\|_\infty.$$

Therefore, the operator $J(t)$ is a contraction for each $t \geq 0$, i.e.,

$$(4.3) \quad \|J(t)f - J(t)g\|_\infty \leq \|f - g\|_\infty \quad \text{for all } t \geq 0 \text{ and } f, g \in C_b.$$

A direct computation shows that $J(t)$ is subadditive and positive homogeneous for each $t \geq 0$, i.e.,

$$(J(t)(f+g))(x) \leq (J(t)f)(x) + (J(t)g)(x) \quad \text{and} \quad (J(t)(\lambda f))(x) = \lambda(J(t)f)(x)$$

for all $f, g \in C_b$, $\lambda \geq 0$, and $x \in \mathbb{R}^d$.

We then have the following continuity results, which play a crucial role in the subsequent discussion.

Lemma 4.2. Suppose that Assumption 2.2 holds. Then, for all $\varepsilon > 0$, $t, r \geq 0$, and $K \Subset \mathbb{R}^d$, there exist $h > 0$, $c' \geq 0$, and $K' \Subset \mathbb{R}^d$ such that, for all partitions $\pi \in P$ with $\max \pi \leq t$ and $\text{mesh } \pi \leq h$ and all $f, g \in C_b$ with $\|f\|_\infty \leq r$ and $\|g\|_\infty \leq r$,

$$\|\mathcal{J}(\pi)f - \mathcal{J}(\pi)g\|_{\infty, K} \leq c'\|f - g\|_{\infty, K'} + \varepsilon.$$

Proof. Note that, for all $\pi \in P$, $\mathcal{J}(\pi)$ is monotone, subadditive, and positive homogeneous, since it is the composition of the monotone, subadditive, and positive homogeneous operators J , see Remark 4.1. Moreover, $\mathcal{J}(\pi)0 = 0$ and $\mathcal{J}(\pi)$ satisfies condition (A.8) in Appendix A.3 since, by (4.3), $\|J(h)f\|_\infty \leq \|f\|_\infty$ for every $h \geq 0$ and $f \in C_b$, and this estimate can be iterated to compositions of J .

We will show the statement by employing Lemma A.4 with $\mathcal{K} = \{\pi \in P \mid \max \pi \leq t, \text{mesh } \pi \leq h\}$, for $h > 0$ small enough, and $\mathfrak{I}_\pi = \mathcal{J}(\pi)$ for $\pi \in \mathcal{K}$. To that end, we fix $K \Subset \mathbb{R}^d$ and construct a function $\zeta \in C_c^\infty$ satisfying the conditions of Lemma A.4.

Consider a partition $\pi = \{t_0, t_1, \dots, t_k\} \in P$. Using (4.3) and the translation invariance of $J(s)$, for all $s \geq 0$, we can write, for all $x \in K$ and all $\zeta: \mathbb{R}^d \rightarrow \mathbb{R}$ with $\zeta(x) = 1$ for $x \in K$,

$$(4.4) \quad \begin{aligned} (\mathcal{J}(\pi)(1 - \zeta))(x) &\leq \|\mathcal{J}(\pi)(1 - \zeta) - (1 - \zeta)\|_\infty = \|\mathcal{J}(\pi)(-\zeta) - (-\zeta)\|_\infty \\ &\leq \sum_{i=1}^k \|J(t_i - t_{i-1})(-\zeta) - (-\zeta)\|_\infty. \end{aligned}$$

Using (2.15), for all $s \geq 0$ and $\zeta \in C_c^\infty$, we can write

$$\begin{aligned} \|J(s)(-\zeta) - (-\zeta)\|_\infty &\leq \sup_{a \in A} \|I^a(s)(-\zeta) - T^a(s)(-\zeta)\|_\infty + \sup_{a \in A} \|T^a(s)(-\zeta) - (-\zeta)\|_\infty \\ &\leq s \sup_{x \in \mathbb{R}^d} \left(\sup_{a \in A} \frac{(I^a(s)(-\zeta))(x) - (T^a(s)(-\zeta))(x)}{s} \right) \\ &\quad + \sup_{a \in A} \|T^a(s)(-\zeta) - (-\zeta)\|_\infty \\ &\leq sm\|\nabla \zeta\|_\infty + \sup_{a \in A} \|T^a(s)(-\zeta) - (-\zeta)\|_\infty. \end{aligned}$$

Moreover, by Lemma A.2, for all $\varepsilon > 0$, there exist $C' > 0$ such that, for sufficiently small $h > 0$,

$$\sup_{a \in A} \|T^a(h)(-\zeta) - (-\zeta)\|_\infty \leq h \left[\frac{\varepsilon}{2t} + C' \left(\sup_{y \in \mathbb{R}^d} (1 + \|y\|) \|\nabla \zeta(y)\| + \|\nabla^2 \zeta\|_\infty \right) \right].$$

Since K is compact, there exists $R > 0$ such that $K \subset \{x \in \mathbb{R}^d \mid \|x\| \leq R\}$. Hence, by Lemma A.5, for each $\delta > 0$, there exists a function $\zeta \in C_c^\infty$ with $0 \leq \zeta \leq 1$ and $\zeta(x) = 1$ for all $x \in K$, which satisfies

$$(4.5) \quad \sup_{y \in \mathbb{R}^d} (1 + \|y\|) \|\nabla \zeta(y)\| + \|\nabla^2 \zeta\|_\infty \leq \delta.$$

For this function ζ , it then holds

$$\|J(h)(-\zeta) - (-\zeta)\|_\infty \leq h \left(\frac{\varepsilon}{2t} + (m + C')\delta \right) \quad \text{for all } h > 0 \text{ sufficiently small.}$$

Selecting $\zeta \in C_c^\infty$ that satisfies (4.5) with $\delta = \varepsilon/(2t(m + C'))$, together with a sufficiently small $h > 0$, we can plug the previous estimate in (4.4) and get

$$(\mathcal{J}(\pi)(r(1 - \zeta)))(x) \leq t \left(\frac{\varepsilon}{2t} + (m + C') \frac{\varepsilon}{2t(m + C')} \right) \leq \varepsilon$$

for all $\pi \in P$ with $\max \pi \leq t$ and $\text{mesh } \pi \leq h$, and all $x \in K$. The function ζ selected above satisfies condition (i)-(iii) in Lemma A.4 and the result follows. \square

Lemma 4.3. Suppose that Assumption 2.2 holds. Then, for all $\varepsilon > 0$, $r \geq 0$, and $K \in \mathbb{R}^d$, there exist $h > 0$, $c' \geq 0$, and $K' \in \mathbb{R}^d$ such that, for all partitions $\pi \in P$ with $\max \pi \leq h$ and all $f, g \in C_b$ with $\|f\|_\infty \leq r$ and $\|g\|_\infty \leq r$,

$$\|\mathcal{J}(\pi)f - g\|_{\infty, K} \leq c' \|f - g\|_{\infty, K'} + \varepsilon.$$

Proof. Fix $\varepsilon > 0$ and $K \in \mathbb{R}^d$. We first show that $\|\mathcal{J}(\pi)g - g\|_{\infty, K}$ is small for $\pi \in P$ with $\max \pi$ small enough. To that end, observe that, for every function $\zeta \in C_c^2$, there exist $\alpha > 0$ and $h > 0$ such that

$$(4.6) \quad \sup_{\pi \in P_t} \|\mathcal{J}(\pi)\zeta - \zeta\|_\infty \leq \alpha t \quad \text{for all } t \in [0, h].$$

As a matter of fact, thanks to lemma A.2, there exists $\beta > 0$ and $h > 0$ such that

$$(4.7) \quad \sup_{a \in A} \sup_{x \in \mathbb{R}^d} |(T^a(t)\zeta)(x) - \zeta(x)| \leq \beta t \quad \text{for all } t \in [0, h].$$

Moreover, using (2.15), we also obtain

$$(4.8) \quad \sup_{a \in A} \sup_{x \in \mathbb{R}^d} |(I^a(t)\zeta)(x) - (T^a(t)\zeta)(x)| \leq m \|\nabla \zeta\|_\infty t \quad \text{for all } t \geq 0.$$

Therefore, putting together (4.7) and (4.8), there exists $\alpha > 0$ such that

$$(4.9) \quad |(J(t)\zeta)(x) - \zeta(x)| \leq \sup_{a \in A} |(I^a(t)\zeta)(x) - (T^a(t)\zeta)(x)| + \sup_{a \in A} |(T^a(t)\zeta)(x) - \zeta(x)| \leq \alpha t$$

for all $t \in [0, h]$ and all $x \in \mathbb{R}^d$. Now, (4.6) is obtained by iterating (4.9) and using (4.3) as in (4.4).

To prove the claim, we approximate $g \in C_b$ with a sequence of functions $(g_n)_{n \in \mathbb{N}} \in C_c^2$ such that $g_n \xrightarrow{\tau, M} g$ as $n \rightarrow \infty$, see Lemma A.1. We can then find $n \in \mathbb{N}$ and $h > 0$ such that

$$\begin{aligned} \|\mathcal{J}(\pi)f - g\|_{\infty, K} &\leq \|\mathcal{J}(\pi)f - \mathcal{J}(\pi)g\|_{\infty, K} + \|\mathcal{J}(\pi)g - \mathcal{J}(\pi)g_n\|_{\infty, K} \\ &\quad + \|\mathcal{J}(\pi)g_n - g_n\|_{\infty, K} + \|g_n - g\|_{\infty, K} \\ &\leq c' \|f - g\|_{\infty, K'} + \varepsilon \end{aligned}$$

for all partitions $\pi \in P$ with $\max \pi \leq h$. To that end, we first use Lemma 4.2 and obtain $h > 0$, $c' \geq 0$, and $K' \in \mathbb{R}^d$ such that, for all $\pi \in P$ with $\max \pi \leq h$,

$$\|\mathcal{J}(\pi)f - \mathcal{J}(\pi)g\|_{\infty, K} \leq c' \|f - g\|_{\infty, K'} + \varepsilon/4$$

and

$$\|\mathcal{J}(\pi)g - \mathcal{J}(\pi)g_n\|_{\infty, K} + \|g - g_n\|_{\infty, K} \leq (c' + 1) \|g - g_n\|_{\infty, K'} + \varepsilon/4 \leq \varepsilon/2,$$

where, in the last inequality, we have chosen $n \in \mathbb{N}$ large enough and used that $g_n \xrightarrow{\tau, M} g$. Finally, after potentially choosing a smaller $h > 0$, we can use (4.6) to write

$$\|\mathcal{J}(\pi)g_n - g_n\|_{\infty, K} \leq \varepsilon/4,$$

and the claim is proved. \square

For the proof of Theorem 3.3, we also consider the operator

$$(4.10) \quad K(t)f := \inf_{h>0} \sup_{\substack{\pi \in P_t \\ \text{mesh } \pi \leq h}} \mathcal{J}(\pi)f = \lim_{h \downarrow 0} \sup_{\substack{\pi \in P_t \\ \text{mesh } \pi \leq h}} \mathcal{J}(\pi)f \quad \text{for all } f \in C_b \text{ and } t \geq 0.$$

We have the following auxiliary result.

Lemma 4.4. For all $t \geq 0$, $K(t) : C_b \rightarrow C_b$ is well defined and

$$K(t_n)f_n \xrightarrow{\tau, M} f$$

for all $f \in C_b$ and all sequences $(t_n)_{n \in \mathbb{N}} \subset [0, \infty)$ with $t_n \rightarrow t$ and $(f_n)_{n \in \mathbb{N}} \subset C_b$ with $f_n \xrightarrow{\tau, M} f$.

Proof. Let $f \in \text{Lip}_b$ and $t \geq 0$. Then, for all $x_1, x_2 \in \mathbb{R}^d$,

$$\begin{aligned} \sup_{a \in A} \sup_{\nu \in \mathcal{B}_t^a(m)} \int_{\mathbb{R}^d} |f(\psi_t^a(x_1) + z) - f(\psi_t^a(x_2) + z)| \nu(dz) &\leq \sup_{a \in A} \|f\|_{\text{Lip}} \|\psi_t^a(x_1) - \psi_t^a(x_2)\| \\ (4.11) \quad &\leq e^{ct} \|f\|_{\text{Lip}} \|x_1 - x_2\|, \end{aligned}$$

where the last inequality follows from Assumption 2.2 (ii). Using the triangle inequality, we find that $J(t)f \in \text{Lip}_b$ with $\|J(t)f\|_{\text{Lip}} \leq e^{ct} \|f\|_{\text{Lip}}$. Iterating this estimate and using the fact that $\|K(t)g\|_\infty \leq \|g\|_\infty$ for all $g \in C_b$, yield that $K(t): \text{Lip}_b \rightarrow \text{Lip}_b$ is well-defined. Now, let $f \in C_b$ and $(f_n)_{n \in \mathbb{N}} \subset \text{Lip}_b$ with $f_n \xrightarrow{\tau_M} f$, see Lemma A.1. Then, by Lemma 4.2, it follows that

$$K(t)f_n \xrightarrow{\tau_M} K(t)f.$$

Hence, $K(t) \in C_b$ as uniform limit on compacts of a sequence of continuous functions, which is uniformly bounded. As a direct consequence of Lemma 4.3, we obtain that

$$K(t_n)f_n \xrightarrow{\tau_M} f$$

for all $f \in C_b$ and all sequences $(t_n)_{n \in \mathbb{N}} \subset [0, \infty)$ with $t_n \rightarrow t$ and $(f_n)_{n \in \mathbb{N}} \subset C_b$ with $f_n \xrightarrow{\tau_M} f$. The proof is complete. \square

5. PROOFS OF RESULTS IN SECTION 2.3

Proof of Proposition 2.4. We start by showing that $I(t)$ is well defined as an operator $\text{Lip}_b \rightarrow \text{Lip}_b$ for all $t \geq 0$, and we later extend such operator to C_b using Lemma 4.2. To that end, let $t \geq 0$ and $f \in \text{Lip}_b$. Then, by (4.11), for every $x_1, x_2 \in \mathbb{R}^d$,

$$\begin{aligned} |(I(t)f)(x_1) - (I(t)f)(x_2)| &\leq \sup_{a \in A} \sup_{\nu \in \mathcal{B}_t^a(m)} \int_{\mathbb{R}^d} |f(\psi_t^a(x_1) + z) - f(\psi_t^a(x_2) + z)| \nu(dz) \\ &\leq e^{ct} \|f\|_{\text{Lip}} \|x_1 - x_2\|, \end{aligned}$$

so that $I(t)f \in \text{Lip}_b$ with $\|I(t)f\|_{\text{Lip}} \leq e^{ct} \|f\|_{\text{Lip}}$ and (ii) is proved.

We can then use Lemma 4.2 to extend $I(t)$ to a map $C_b \rightarrow C_b$ for all $t \geq 0$, since, for $f \in C_b$, $(f_n)_{n \in \mathbb{N}} \subset \text{Lip}_b$ with $f_n \xrightarrow{\tau_M} f$, see Lemma A.1, and $K \in \mathbb{R}^d$, it holds

$$\|I(t)f - I(t)f_n\|_{\infty, K} \leq \|J(t)|f - f_n|\|_{\infty, K} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Therefore, $I(t)f$ is continuous, being a uniform limit on compacts of Lipschitz continuous functions. The monotonicity of $I(t)$ follows directly from the definition and, in order to show that it is a contraction, it is enough to write

$$|(I(t)f)(x) - (I(t)g)(x)| \leq \sup_{a \in A} \sup_{\nu \in \mathcal{B}_t^a(m)} \int_{\mathbb{R}^d} |f(\psi_t^a(x) + z) - g(\psi_t^a(x) + z)| \nu(dz) \leq \|f - g\|_\infty.$$

Therefore, (i) is proved.

We now prove (iii). Let $s, t \geq 0$ and $f \in C_b$ be given. Note that, given $a \in A$, the operator I^a , defined in (2.14), corresponds to a specific instance of the operator I studied in [40] (obtained by choosing $(\mu_t^a)_{t \geq 0}$ as reference family of measures and $\varphi(v) = \infty \cdot \mathbf{1}_{(m, \infty)}(v)$, for $v \geq 0$, as penalty term). Moreover, Assumption 2.2 is stronger than [40, Assumption 2.2]. Therefore, we can use [40, Lemma 3.10] to get that, for every $x \in \mathbb{R}^d$,

$$(I^a(t)I^a(s)f)(x) \leq (I^a(t+s)f)(x).$$

This ensures that, for every $x \in \mathbb{R}^d$,

$$(I(t+s)f)(x) = (\inf_{a \in A} I^a(t+s)f)(x) \geq (\inf_{a \in A} I^a(t)I^a(s)f)(x)$$

$$\begin{aligned} &\geq \left(\inf_{a \in A} I^a(t) \inf_{a' \in A} I^{a'}(s) f \right)(x) \\ &= (I(t)I(s)f)(x), \end{aligned}$$

as claimed. \square

Proof of Proposition 2.5. The proof closely follows the arguments in the proof of [40, Lemma 3.11]. Let $f \in C_b^1$. First, recall the definition of I^a from (2.14) and observe that

$$\frac{(I(t)f)(x) - (T(t)f)(x)}{t} \leq \sup_{a \in A} \frac{(I^a(t)f)(x) - (T^a(t)f)(x)}{t}.$$

Using (2.15), we obtain, for all $t \geq 0$ and $x \in \mathbb{R}^d$,

$$\sup_{a \in A} \frac{(I^a(t)f)(x) - (T^a(t)f)(x)}{t} \leq m \|\nabla f\|_\infty.$$

Now, let $K \Subset \mathbb{R}^d$ and $\varepsilon > 0$. We prove that, for $t > 0$ sufficiently small,

$$(5.1) \quad \sup_{x \in K} \left(\sup_{a \in A} \frac{(I^a(t)f)(x) - (T^a(t)f)(x)}{t} - m \|\nabla f(x)\| \right) \leq \varepsilon.$$

Using the Fundamental Theorem of Calculus, we estimate

$$\begin{aligned} \sup_{a \in A} \frac{(I^a(t)f)(x) - (T^a(t)f)(x)}{t} &= \sup_{a \in A} \frac{1}{t} \int_{\mathbb{R}^d \times \mathbb{R}^d} (f(\psi_t^a(x) + z) - f(\psi_t^a(x) + y)) \gamma_t^{a,x}(dy, dz) \\ &= \sup_{a \in A} \frac{1}{t} \int_0^1 \int_{\mathbb{R}^d \times \mathbb{R}^d} \nabla f(\psi_t^a(x) + y + s(z-y))^\top (z-y) \gamma_t^{a,x}(dy, dz) ds \\ &\leq m \|\nabla f(x)\| + \sup_{a \in A} \frac{1}{t} \int_0^1 \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\nabla f(\psi_t^a(x) + y + s(z-y)) - \nabla f(x)\| \|z-y\| \gamma_t^{a,x}(dy, dz) ds, \end{aligned}$$

and it remains to bound the last term by ε uniformly for $x \in K$. Using the continuity of ∇f together with Assumption 2.2 (i), there exists some $\delta > 0$ such that

$$\begin{aligned} \sup_{x \in K} \sup_{a \in A} \frac{1}{t} \int_0^1 \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\nabla f(\psi_t^a(x) + y + s(z-y)) - \nabla f(x)\| \|z-y\| \gamma_t^{a,x}(dy, dz) ds \\ &\leq \frac{\varepsilon}{2} + 2 \|\nabla f\|_\infty \sup_{x \in K} \sup_{a \in A} \frac{1}{t} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|z-y\| (\mathbf{1}_{\{\|y-z\|>\delta\}} + \mathbf{1}_{\{\|y\|>\delta\}}) \gamma_t^{a,x}(dy, dz) \\ &\leq \frac{\varepsilon}{2} + 2 \|\nabla f\|_\infty \sup_{x \in K} \sup_{a \in A} \left(\frac{1}{\delta^{p-1}} \frac{\mathcal{W}_p(\mu_t^a, \nu_t^{a,x})^p}{t} + \left(\frac{1}{\delta} \int_{\mathbb{R}^d} \|y\| \mu_t^a(dy) \right)^{\frac{1}{q}} \frac{\mathcal{W}_p(\mu_t^a, \nu_t^{a,x})}{t} \right) \\ &\leq \frac{\varepsilon}{2} + 2m \|\nabla f\|_\infty \left(\frac{t^{p-1}}{\delta^{p-1}} + \sup_{a \in A} \left(\frac{1}{\delta} \int_{\mathbb{R}^d} \|y\| \mu_t^a(dy) \right)^{\frac{1}{q}} \right) \leq \varepsilon, \end{aligned}$$

where, in the second step, we used Markov's inequality twice. Moreover, using the observation

$$\begin{aligned} \frac{(I(t)f)(x) - (T(t)f)(x)}{t} &\geq \inf_{a \in A} \frac{(I^a(t)f)(x) - T^a(t)f)(x)}{t} \\ &\geq - \sup_{a \in A} \frac{(I^a(t)(-f))(x) - (T^a(t)(-f))(x)}{t}, \end{aligned}$$

the claim follows from (2.15) and (5.1). \square

6. PROOFS OF RESULTS IN SECTION 2.4

In the proof of Theorem 2.7, we work with an auxiliary operator $(\hat{\mathcal{S}}(t))_{t \geq 0}$, which we define by using dyadic partitions. To that end, we denote by \mathcal{D} the set of dyadic numbers, i.e.,

$$(6.1) \quad \mathcal{D} := \{k2^{-n} : k, n \in \mathbb{N}\}.$$

Next, we define

$$(6.2) \quad \pi_t^n := \{0, 2^{-n}, 2 \cdot 2^{-n}, \dots, k_t^n 2^{-n}, t\} \quad \text{for } t > 0 \text{ and } n \in \mathbb{N},$$

where $k_t^n := \max\{k \in \mathbb{N} \cup \{0\} : k2^{-n} < t\}$ for $t > 0$ and $n \in \mathbb{N}$, and we use the convention $\pi_0^n := \{0\}$ for all $n \in \mathbb{N}$. We will refer to $(\pi_t^n)_{n \in \mathbb{N}}$ as the dyadic partitions of $[0, t]$ for $t \geq 0$.

We can then define the operator $\hat{\mathcal{S}} := (\hat{\mathcal{S}}(t))_{t \geq 0}$

$$(6.3) \quad (\hat{\mathcal{S}}(t)f)(x) := \inf_{n \in \mathbb{N}} (\mathcal{I}(\pi_t^n)f)(x) = \lim_{n \rightarrow \infty} (\mathcal{I}(\pi_t^n)f)(x), \quad \text{for } t \geq 0, f \in C_b, \text{ and } x \in \mathbb{R}^d.$$

where the infimum is well-defined since $((\mathcal{I}(\pi_t^n)f)(x))_{n \in \mathbb{N}}$ is a decreasing sequence which is bounded from below by $-\|f\|_\infty > -\infty$ and the equality between the infimum and the limit follows from the monotonicity of \mathcal{I} over refining partitions, see Proposition 2.4 (iii). As a consequence, we shall write, for all $n_0 \in \mathbb{N}$,

$$(6.4) \quad (\hat{\mathcal{S}}(t)f)(x) = \inf_{n \geq n_0} (\mathcal{I}(\pi_t^n)f)(x) \quad \text{for } t \geq 0, f \in C_b, \text{ and } x \in \mathbb{R}^d.$$

With the definition of the previous auxiliary operator at hand, we can state the proof of Theorem 2.7.

Proof of Theorem 2.7. We organize the proof in six steps.

Step 1: For all $t \geq 0$, $\hat{\mathcal{S}}(t)$ is a map $\text{Lip}_b \rightarrow \text{Lip}_b$. Iterating the estimate in Proposition 2.4 (ii), we get $\|\mathcal{I}(\pi_t^n)f\|_{\text{Lip}} \leq e^{ct} \|f\|_{\text{Lip}}$ for all $n \in \mathbb{N}$ and $t \geq 0$, which in turns implies

$$\begin{aligned} |(\hat{\mathcal{S}}(t)f)(x) - (\hat{\mathcal{S}}(t)f)(y)| &\leq \sup_{n \in \mathbb{N}} |(\mathcal{I}(\pi_t^n)f)(x) - (\mathcal{I}(\pi_t^n)f)(y)| \\ &\leq e^{ct} \|f\|_{\text{Lip}} \|x - y\| \end{aligned}$$

for all $x, y \in \mathbb{R}^d$.

Step 2: For all $t \geq 0$, $\hat{\mathcal{S}}(t)$ is a sequentially continuous map $C_b \rightarrow C_b$ (w.r.t. the mixed topology) and $\mathcal{I}(\pi_t^n)f \xrightarrow{\tau_M} \hat{\mathcal{S}}(t)f$ as $n \rightarrow \infty$. Let $f \in C_b$ and $(f_k)_{k \in \mathbb{N}} \subset \text{Lip}_b$ such that $f_k \xrightarrow{\tau_M} f$ as $k \rightarrow \infty$ (see Lemma A.1). For all $K \Subset \mathbb{R}^d$ and $\varepsilon > 0$, using Lemma 4.2, we obtain

$$\|\hat{\mathcal{S}}(t)f_k - \hat{\mathcal{S}}(t)f\|_{\infty, K} \leq \sup_{n \geq n_0} \|\mathcal{J}(\pi_t^n)f_k - \mathcal{J}(\pi_t^n)f\|_{\infty, K} \leq \varepsilon$$

for $n_0, k \in \mathbb{N}$ sufficiently large. This implies

$$(6.5) \quad \|\hat{\mathcal{S}}(t)f_k - \hat{\mathcal{S}}(t)f\|_{\infty, K} \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

Therefore, $\hat{\mathcal{S}}(t)f \in C_b$ for all $t \geq 0$ and $f \in C_b$, being uniform limit over compacts of (Lipschitz) continuous functions. The sequential continuity of $\hat{\mathcal{S}}(t)$ is again obtained from equation (6.5) since, for all $(f_k)_{k \in \mathbb{N}} \subset C_b$, $f \in C_b$ with $f_k \xrightarrow{\tau_M} f$ as $k \rightarrow \infty$, one has

$$(6.6) \quad \lim_{k \rightarrow \infty} \hat{\mathcal{S}}(t)f_k = \hat{\mathcal{S}}(t)f \quad \text{for all } t \geq 0.$$

Finally, using that the sequence $(\mathcal{I}(\pi_t^n)f)(x)$ is monotonically decreasing for all $x \in \mathbb{R}^d$ and $\hat{\mathcal{S}}(t)f \in C_b$, Dini's Theorem implies that the convergence $(\mathcal{I}(\pi_t^n)f)(x) \rightarrow (\hat{\mathcal{S}}(t)f)(x)$ is uniform

over compacts, i.e.,

$$(6.7) \quad \sup_{x \in K} |(\mathcal{I}(\pi_t^n)f)(x) - (\hat{\mathcal{S}}(t)f)(x)| \rightarrow 0 \text{ as } n \rightarrow \infty,$$

for all $f \in C_b$, $t \geq 0$, and $K \Subset \mathbb{R}^d$. Since $\|\mathcal{I}(\pi_t^n)f\|_\infty \leq \|f\|_\infty$ for all $n \in \mathbb{N}$, we conclude that $\mathcal{I}(\pi_t^n)f \xrightarrow{\tau_M} \hat{\mathcal{S}}(t)f$ as $n \rightarrow \infty$.

Step 3: $\hat{\mathcal{S}}$ is left-continuous in time. We now show that, for every $f \in C_b$, $t > 0$, and $K \Subset \mathbb{R}^d$ it holds

$$(6.8) \quad \sup_{x \in K} |(\hat{\mathcal{S}}(s)f)(x) - (\hat{\mathcal{S}}(t)f)(x)| \rightarrow 0 \quad \text{as } s \uparrow t.$$

Fix $f \in C_b$, $t > 0$, $K \Subset \mathbb{R}^d$, and $\varepsilon > 0$. By (6.7), there exists $n_t \in \mathbb{N}$ such that, for every $x \in K$,

$$(\hat{\mathcal{S}}(t)f)(x) \geq (\mathcal{I}(\pi_t^{n_t})f)(x) - \varepsilon/3.$$

Let $s > 0$ be such that $t - s < 2^{-n_t}$. Then, for every $x \in K$, it holds

$$\begin{aligned} (6.9) \quad (\hat{\mathcal{S}}(t)f)(x) - (\hat{\mathcal{S}}(s)f)(x) &\geq (\mathcal{I}(\pi_t^{n_t})f)(x) - (\mathcal{I}(\pi_s^{n_t})f)(x) - \varepsilon/3 \\ &\geq (\mathcal{I}(\pi_s^{n_t})I(t-s)f)(x) - (\mathcal{I}(\pi_s^{n_t})f)(x) - \varepsilon/3 \\ &\geq -|(\mathcal{J}(\pi_s^{n_t})(I(t-s)f - f))(x)| - \varepsilon/3, \end{aligned}$$

where the second inequality follows from Proposition 2.4 (iii).

Since $\|I(t-s)f\|_\infty \leq \|f\|_\infty$, we can apply Lemma 4.2 (if necessary, taking n_t larger than previously stated) to obtain that there exist some $K' \Subset \mathbb{R}^d$ and $c' > 0$ such that

$$(6.10) \quad \|\mathcal{J}(\pi_s^{n_t})(I(t-s)f - f)\|_{\infty, K'} \leq c' \|I(t-s)f - f\|_{\infty, K'} + \varepsilon/3 \quad \text{for all } s \in [0, t].$$

Then by Lemma 2.4 (iv), we can choose $\delta \in (0, 2^{-n_t})$ such that $c' \|I(t-s)f - f\|_{\infty, K'} \leq \varepsilon/3$ whenever $s > 0$ satisfies $t - s < \delta$. Therefore, combining (6.9) with (6.10), we find that, for all $s > 0$ with $t - s < \delta$ and $x \in K$,

$$\begin{aligned} (\hat{\mathcal{S}}(t)f)(x) - (\hat{\mathcal{S}}(s)f)(x) &\geq -\|\mathcal{J}(\pi_s^{n_t})(I(t-s)f - f)\|_{\infty, K} - \varepsilon/3 \\ &\geq -c' \|I(t-s)f - f\|_{\infty, K'} - \frac{2}{3}\varepsilon \geq -\varepsilon. \end{aligned}$$

In order to obtain the opposite inequality, we proceed in a similar way, starting from the set $[0, t) \cap \mathcal{D}$, where \mathcal{D} is the set of dyadic numbers, see (6.1)). For every $s \in [0, t) \cap \mathcal{D}$, let $n_s \in \mathbb{N}$ be such that, for every $x \in K$,

$$(\hat{\mathcal{S}}(s)f)(x) \geq (\mathcal{I}(\pi_s^{n_s})f)(x) - \varepsilon/3,$$

and we repeat the same estimates as above to obtain some (sufficiently small) $\delta > 0$ such that, for every $s \in (t - \delta, t) \cap \mathcal{D}$ and every $x \in K$,

$$\begin{aligned} (\hat{\mathcal{S}}(t)f)(x) - (\hat{\mathcal{S}}(s)f)(x) &\leq (\mathcal{I}(\pi_s^{n_s})I(t-s)f)(x) - (\mathcal{I}(\pi_s^{n_s})f)(x) + \varepsilon/3 \\ &\leq \|\mathcal{J}(\pi_s^{n_s})(I(t-s)f - f)\|_{\infty, K} + \varepsilon/3 \\ &\leq c' \|I(t-s)f - f\|_{\infty, K'} + \frac{2}{3}\varepsilon \leq \varepsilon. \end{aligned}$$

Finally, to extend the last inequality to every s sufficiently close to t , we can choose $\delta_t > 0$ such that, for every $u \in (t - \delta_t, t) \cap \mathcal{D}$,

$$\|\hat{\mathcal{S}}(t)f - \hat{\mathcal{S}}(u)f\|_{\infty, K} \leq \varepsilon/2.$$

Analogously, for every $s \in (t - \delta_t, t)$, we can choose $\delta_s > 0$ such that, for every $u \in (s - \delta_s, s) \cap \mathcal{D}$,

$$\|\hat{\mathcal{S}}(s)f - \hat{\mathcal{S}}(u)f\|_{\infty, K} \leq \varepsilon/2.$$

Hence, choosing $\delta = \delta_t$ and then taking any $u \in (t - \delta, t) \cap (s - \delta_s, s) \cap \mathcal{D}$, for each $s \in (t - \delta, t)$, we have that

$$\|\hat{\mathcal{S}}(t)f - \hat{\mathcal{S}}(s)f\|_{\infty, K} \leq \|\hat{\mathcal{S}}(t)f - \hat{\mathcal{S}}(u)f\|_{\infty, K} + \|\hat{\mathcal{S}}(s)f - \hat{\mathcal{S}}(u)f\|_{\infty, K} \leq \varepsilon,$$

and the left-continuity of the map $t \mapsto \hat{\mathcal{S}}(t)f$ follows.

Step 4: Semigroup property of $\hat{\mathcal{S}}$. In order to prove the semigroup property, i.e., condition (iii) in Definition 2.6, we follow the strategy presented in [40]. Fix $f \in C_b$ and, at first, $t \in \mathcal{D}$. Then, for every $x \in \mathbb{R}^d$ and $s \geq 0$,

$$\begin{aligned} (\hat{\mathcal{S}}(t+s)f)(x) &= \lim_{n \rightarrow \infty} (\mathcal{I}(\pi_{t+s}^n)f)(x) = \lim_{n \rightarrow \infty} (\mathcal{I}(\pi_t^n)\mathcal{I}(\pi_s^n)f)(x) \\ &\geq (\hat{\mathcal{S}}(t)\hat{\mathcal{S}}(s)f)(x), \end{aligned}$$

where the second equality holds because $t \in \mathcal{D}$, and the inequality holds by definition of $\hat{\mathcal{S}}$.

To obtain the opposite inequality, fix $\hat{n} \in \mathbb{N}$. Then, for every $x \in \mathbb{R}^d$ and $s \geq 0$,

$$\begin{aligned} (\hat{\mathcal{S}}(t+s)f)(x) &= \lim_{n \rightarrow \infty} (\mathcal{I}(\pi_{t+s}^n)f)(x) = \lim_{n \rightarrow \infty} (\mathcal{I}(\pi_t^n)\mathcal{I}(\pi_s^n)f)(x) \\ &\leq \lim_{n \rightarrow \infty} (\mathcal{I}(\pi_t^n)\mathcal{I}(\pi_s^{\hat{n}})f)(x) = (\hat{\mathcal{S}}(t)\mathcal{I}(\pi_s^{\hat{n}})f)(x), \end{aligned}$$

where the inequality holds since π_s^n becomes finer than $\pi_s^{\hat{n}}$ as $n \rightarrow \infty$ and \mathcal{I} is monotonically decreasing over refining partitions.

Moreover, by Step 2, we know that $\mathcal{I}(\pi_s^{\hat{n}})f \xrightarrow{\tau_{\mathcal{M}}} \hat{\mathcal{S}}(s)f$ as $\hat{n} \rightarrow \infty$. Therefore, using the continuity in the mixed topology of $\hat{\mathcal{S}}(t)$, we obtain, for every $x \in \mathbb{R}^d$ and $s \geq 0$,

$$(\hat{\mathcal{S}}(t+s)f)(x) \leq \lim_{\hat{n} \rightarrow \infty} (\hat{\mathcal{S}}(t)\mathcal{I}(\pi_s^{\hat{n}})f)(x) = (\hat{\mathcal{S}}(t)\hat{\mathcal{S}}(s)f)(x).$$

To extend the semigroup property to arbitrary $t > 0$, we consider an approximating sequence $(t_n)_{n \in \mathbb{N}} \subset \mathcal{D}$ such that $t_n \uparrow t$ as $n \rightarrow \infty$. Then it holds

$$\hat{\mathcal{S}}(t+s)f = \lim_{n \rightarrow \infty} \hat{\mathcal{S}}(t_n+s)f = \lim_{n \rightarrow \infty} \hat{\mathcal{S}}(t_n)\hat{\mathcal{S}}(s)f = \hat{\mathcal{S}}(t)\hat{\mathcal{S}}(s)f,$$

where the first and last equality follow from the time continuity of $\hat{\mathcal{S}}$ from the left, and all the limits above are to be understood w.r.t. the mixed topology.

Step 5: $\hat{\mathcal{S}}$ is strongly continuous. By Step 3, it is enough to show that

$$(6.11) \quad \hat{\mathcal{S}}(t)f \xrightarrow{\tau_{\mathcal{M}}} f \quad \text{as } t \downarrow 0 \text{ for all } f \in C_b.$$

As a matter of fact, for $s \downarrow t$, we can then use the semigroup property and the sequential continuity of $\hat{\mathcal{S}}(t)$, for fixed $t > 0$, to write

$$\hat{\mathcal{S}}(s)f = \hat{\mathcal{S}}(t)\hat{\mathcal{S}}(s-t)f \xrightarrow{\tau_{\mathcal{M}}} \hat{\mathcal{S}}(t)f \quad \text{as } s \downarrow t.$$

We therefore prove (6.11). Note that, by (6.4) and Lemma 4.3, for each compact $K \Subset \mathbb{R}^d$, $f \in C_b$, and $\varepsilon > 0$, there exists $h > 0$ such that

$$\sup_{x \in K} (\hat{\mathcal{S}}(t)f)(x) - f(x) \leq \sup_{x \in K} \sup_{n \in \mathbb{N}} (\mathcal{J}(\pi_t^n)f(x) - f(x)) \leq \varepsilon$$

for all $t \in [0, h]$. On the other hand, again by (6.4) and Lemma 4.3, there exists $h > 0$ such that

$$\sup_{x \in K} (f(x) - (\hat{\mathcal{S}}(t)f)(x)) \leq \sup_{x \in K} \sup_{n \in \mathbb{N}} (f(x) + (\mathcal{J}(\pi_t^n)(-f))(x)) \leq \varepsilon$$

for all $t \in [0, h]$. Since $\varepsilon > 0$ is arbitrarily small, putting the two estimates together we obtain (6.11).

Step 6: $\hat{\mathcal{S}} = \mathcal{S}$. We show that $\hat{\mathcal{S}}(t)f = \mathcal{S}(t)f$ for all $t \geq 0$ and $f \in C_b$. From Proposition 2.4 (iii), it follows that, for every $t \geq 0$, $\mathcal{S}(t)f \leq \hat{\mathcal{S}}(t)f$, since the infimum in the definition of $\mathcal{S}(t)f$ is taken over more partitions. Moreover, for all $\pi = \{t_0 = 0, t_1, \dots, t_k = t\} \in P_t$

$$\begin{aligned}\mathcal{I}(\pi)f &= I(t_1 - t_0)I(t_2 - t_1) \cdots I(t_n - t_{k-1})f \\ &\geq \hat{\mathcal{S}}(t_1 - t_0)\hat{\mathcal{S}}(t_2 - t_1) \cdots \hat{\mathcal{S}}(t_n - t_{k-1})f = \hat{\mathcal{S}}(t)f,\end{aligned}$$

where the last equality comes from the semigroup property of $\hat{\mathcal{S}}$. By taking the infimum over all $\pi \in P_t$, we get $\mathcal{S}(t)f \geq \hat{\mathcal{S}}(t)f$.

It is immediate to show that \mathcal{S} satisfies properties (i) and (ii) in Definition 2.6. Moreover, we have shown in Step 4 and Step 5, respectively, that properties (iii) and (iv) are satisfied by $\hat{\mathcal{S}}$ and therefore by \mathcal{S} . Therefore, \mathcal{S} is a strongly continuous monotone semigroup on C_b . To obtain (2.19), note that, using (6.4), we can restrict the definition of $\hat{\mathcal{S}}$ to arbitrary fine partitions. Therefore,

$$\mathcal{S}(t)f \leq \inf_{\substack{\pi \in P_t \\ \text{mesh } \pi \leq h}} \mathcal{I}(\pi)f \leq \inf_{n \geq n_0} \mathcal{I}(\pi_t^n)f = \hat{\mathcal{S}}(t)f \quad \text{for all } h > 0,$$

where $n_0 \in \mathbb{N}$ with $n_0 \geq \frac{1}{h}$. The proof is complete. \square

Proof of Theorem 2.10. We prove the statement in two steps.

Step 1: Generator of the one-period DRO. Let $f \in \cap_{a \in A} D(\mathcal{L}^a) \cap C_b^1$. We start proving that

$$(6.12) \quad \lim_{t \downarrow 0} \frac{I(t)f - f}{t} = \inf_{a \in A} \mathcal{L}^a f + m\|\nabla f(\cdot)\|,$$

where the limit is to be understood w.r.t. the mixed topology.

To that end, we first show that $\|\frac{I(t)f - f}{t}\|_\infty < \infty$ for all $t > 0$. Indeed, using (2.15) we get, for every $t > 0$,

$$\begin{aligned}\left\| \frac{I(t)f - f}{t} \right\|_\infty &\leq \left\| \frac{I(t)f - T(t)f}{t} \right\|_\infty + \left\| \frac{T(t)f - f}{t} \right\|_\infty \\ &\leq m\|\nabla f\|_\infty + \sup_{a \in A} \left\| \frac{T^a(t)f - f}{t} \right\|_\infty.\end{aligned}$$

Since, for each $a \in A$, T^a is a linear semigroup on C_b , we can apply [53, Lemma 4.1.14] to obtain, for every $t > 0$,

$$\begin{aligned}(6.13) \quad \sup_{a \in A} \left\| \frac{T^a(t)f - f}{t} \right\|_\infty &= \sup_{a \in A} \left\| \frac{1}{t} \int_0^t T^a(s) \mathcal{L}^a f \, ds \right\|_\infty \leq \sup_{a \in A} \frac{1}{t} \int_0^t \|T^a(s) \mathcal{L}^a f\|_\infty \, ds \\ &\leq \sup_{a \in A} \|\mathcal{L}^a f\|_\infty < \infty,\end{aligned}$$

where the second inequality holds because $\|T^a(t)g\|_\infty \leq \|g\|_\infty$ for all $a \in A$, $t \geq 0$, and $g \in C_b$, and the last inequality follows from condition (2.20). Moreover, for every $K \Subset \mathbb{R}^d$, we write

$$\begin{aligned}&\left\| \frac{I(t)f - f}{t} - \left(\inf_{a \in A} \mathcal{L}^a f + m\|\nabla f(\cdot)\| \right) \right\|_{\infty, K} \\ &\leq \left\| \frac{I(t)f - T(t)f}{t} - m\|\nabla f(\cdot)\| \right\|_{\infty, K} + \left\| \frac{T(t)f - f}{t} - \inf_{a \in A} \mathcal{L}^a f \right\|_{\infty, K} \\ &\leq \left\| \frac{I(t)f - T(t)f}{t} - m\|\nabla f(\cdot)\| \right\|_{\infty, K} + \sup_{a \in A} \left\| \frac{T^a(t)f - f}{t} - \mathcal{L}^a f \right\|_{\infty, K},\end{aligned}$$

where the first and the second term in the last line vanish as $t \downarrow 0$ because of Lemma 2.5 and condition (2.21), respectively. Therefore, the limit (6.12) holds, as claimed.

Step 2: Extension to the multi-period DRO. By Step 1, for any $\varepsilon > 0$, $f \in \cap_{a \in A} D(\mathcal{L}^a) \cap C_b^1$ and $K \Subset \mathbb{R}^d$, and any sufficiently small $t > 0$, it holds

$$\begin{aligned} & \sup_{x \in K} \left(\frac{(\mathcal{S}(t)f)(x) - f(x)}{t} - \left(\inf_{a \in A} \mathcal{L}^a f(x) + m \|\nabla f(x)\| \right) \right) \\ & \leq \sup_{x \in K} \left(\frac{(I(t)f)(x) - f(x)}{t} - \left(\inf_{a \in A} \mathcal{L}^a f(x) + m \|\nabla f(x)\| \right) \right) < \varepsilon. \end{aligned}$$

To show the other inequality, we write

$$\begin{aligned} \frac{(\mathcal{S}(t)f)(x) - f(x)}{t} &= \inf_{\pi \in P_t} \frac{(\mathcal{I}(\pi)f)(x) - f(x)}{t} \\ &= \inf_{\pi \in P_t} \sum_{i=0}^{k-1} \frac{(\mathcal{I}(\{t_0, \dots, t_i\})I(t_{i+1} - t_i)f)(x) - (\mathcal{I}(\{t_0, \dots, t_i\})f)(x)}{t} \\ &\geq - \sup_{\pi \in P_t} \sum_{i=0}^{k-1} \frac{t_{i+1} - t_i}{t} \mathcal{J}(\{t_0, \dots, t_i\}) \left(\frac{f - I(t_{i+1} - t_i)f}{t_{i+1} - t_i} \right)(x) \\ &\geq - \sup_{\max \pi \leq t} \sup_{h \in (0, t]} \mathcal{J}(\pi) \left(\frac{f - I(h)f}{h} \right)(x), \end{aligned}$$

where, in each step, $\pi = \{t_0, \dots, t_k\}$ with $0 = t_0 < \dots < t_k = t$. Therefore, combining (6.12) with Lemma 4.3, we get that, for every $f \in \cap_{a \in A} D(\mathcal{L}^a) \cap C_b^1$ and sufficiently small $t > 0$,

$$\begin{aligned} & \sup_{x \in K} \left(- \frac{(\mathcal{S}(t)f)(x) - f(x)}{t} + \inf_{a \in A} (\mathcal{L}^a f)(x) + m \|\nabla f(x)\| \right) \\ & \leq \sup_{x \in K} \sup_{\max \pi \leq t} \sup_{h \in (0, t]} \left(\mathcal{J}(\pi) \left(\frac{f - I(h)f}{h} \right)(x) + \inf_{a \in A} (\mathcal{L}^a f)(x) + m \|\nabla f(x)\| \right) < \varepsilon, \end{aligned}$$

which concludes the proof. \square

7. PROOFS OF RESULTS IN SECTION 3

Proof of Theorem 3.3. Let the family $K := (K(t))_{t \geq 0}$ of operators $C_b \rightarrow C_b$ be given by (4.10), see Lemma 4.4. Then, by (2.19),

$$\begin{aligned} \mathcal{S}(t)(\lambda f + (1 - \lambda g)) &= \lim_{h \downarrow 0} \sup_{\substack{\pi \in P_t \\ \text{mesh } \pi \leq h}} \mathcal{I}(\pi)(\lambda f + (1 - \lambda g)) \leq \lim_{h \downarrow 0} \sup_{\substack{\pi \in P_t \\ \text{mesh } \pi \leq h}} (\lambda \mathcal{I}(\pi)f + (1 - \lambda) \mathcal{J}(\pi)g) \\ &\leq \lambda \mathcal{S}(t)f + (1 - \lambda)K(t)g \end{aligned}$$

for all $\lambda \in [0, 1]$, $f, g \in C_b$, and $t \geq 0$. Hence, by Theorem 2.7 and Theorem 2.10, it follows that $\mathcal{S} = (\mathcal{S}(t))_{t \geq 0}$ is a K -convex monotone semigroup with generator \mathcal{L} and $S(t)0 = 0$ for all $t \geq 0$. Moreover, by Lemma 4.4, K is strongly right-continuous in the sense of [39, Definition 2.7], so that we can apply [39, Theorem 3.1] with $M := \{\delta_x \mid x \in \mathbb{R}^d\}$, where δ_x denotes the Dirac measure with barycenter $x \in \mathbb{R}^d$, to conclude that the function $v: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}$, defined in (3.2), is a D -viscosity solution to the PDE (3.1). \square

Proof of Corollary 3.7. We start proving (i). To that end, let $f \in C_b^2$. Then, using Itô's formula, for every $t > 0$ and $x \in \mathbb{R}^d$,

$$(7.1) \quad \frac{(T^a(t)f)(x) - f(x)}{t} = \frac{1}{t} \mathbb{E}^\mathbb{P} \left[f(X_t^{x,a}) - f(x) \right] = \frac{1}{t} \mathbb{E}^\mathbb{P} \left[\int_0^t (Lf)(X_s^{x,a}) ds \right],$$

where

$$(7.2) \quad (Lf)(x) := \frac{1}{2} \Delta f(x) + \langle b, \nabla f(x) \rangle,$$

independent of $a \in A$. Since $Lf \in C_b$ and $X_s^{x,a} - x = bt + W_t$ for all $x \in \mathbb{R}^d$ and $a \in A$, part (i) follows.

We now prove (ii). By part (i), Theorem 2.10, and Theorem 3.3, the function v given in (3.5) is a C_b^2 -viscosity solution to (3.6). Moreover, since $h \in C_b$ and the underlying process $(X_s^{t,x})_{s \in [t,T]}$ is Markovian, the conditional g -expectation value function V given in (3.4) corresponds to the unique solution to the following (Markovian) forward-backward SDE: for every $(t, x) \in [0, T] \times \mathbb{R}^d$,

$$\begin{aligned} Y_s^{t,x} &= h(X_T^{t,x}) + \int_t^T m \|Z_s^{t,x}\| ds - \int_t^T Z_s^{t,x} dW_s, \\ X_s^{t,x} &= x + b(s-t) + (W_s - W_t), \quad \text{for } s \in [t, T], \end{aligned}$$

with $(Y^{t,x}, Z^{t,x}) \in \mathbb{S}^2(\mathbb{R}) \times \mathbb{L}^2(\mathbb{R}^d)$, see Definition 3.5.

Therefore, we can apply [2, Theorem 3.4] to ensure that $V(t, x) = Y_t^{t,x}$ is a viscosity solution to the PDE (3.6) in the classical sense of [2]. Then, by the comparison principle for viscosity solutions to a general class of nonlinear PDEs that includes (3.6), see, e.g., [65, Proposition 5.5, Section 5], [2, Theorem 3.5], we conclude that both two functions V and v coincide, as claimed. \square

Proof of Corollary 3.9. Since $b: A \rightarrow \mathbb{R}^d$, and $\sigma: A \rightarrow \mathbb{R}^{d \times d}$ are bounded, one can use similar arguments as in the proof of Corollary 3.7 (i) in order to prove part (i) and that, for every $f \in C_b^2$

$$\begin{aligned} (\mathcal{L}f)(x) &= \inf_{a \in A} \mathcal{L}^a f(x) + m \|\nabla f(x)\| \\ &= \inf_{a \in A} \left\{ \frac{1}{2} \operatorname{tr} (\sigma(a) \sigma(a)^\top \nabla^2 f(x)) + \langle b(a), \nabla f(x) \rangle \right\} + m \|\nabla f(x)\|, \quad \text{for all } x \in \mathbb{R}^d. \end{aligned}$$

For part (ii), we apply Theorem 2.10 and Theorem 3.3 with $D := C_b^2$ to obtain that the function v , given in (3.10), is a D -viscosity solution to (3.11). Moreover, it follows from [36, Theorem 2.6] that the robust optimization problem V in (3.8) is the unique viscosity solution to the following HJBI equation:

$$(7.3) \quad \begin{cases} \partial_t V(t, x) + \inf_{a \in A} \sup_{\|\theta\| \leq m} \{\mathcal{L}^{a,\theta} V(t, x)\} = 0 & \text{for all } (t, x) \in [0, T] \times \mathbb{R}^d, \\ V(T, x) = h(x) & \text{for all } x \in \mathbb{R}^d, \end{cases}$$

where

$$\mathcal{L}^{a,\theta} V(t, x) = \frac{1}{2} \operatorname{tr} (\sigma(a) \sigma(a)^\top \nabla^2 V(t, x)) + \langle b(a) + \theta, \nabla V(t, x) \rangle \quad \text{for } a \in A \text{ and } \|\theta\| \leq m.$$

In particular, since $\sup_{\|\theta\| \leq m} \langle \theta, w \rangle = m \|w\|$ for every $w \in \mathbb{R}^d$, the HJBI equation in (7.3) coincides with the PDE (3.11). Therefore, by the uniqueness result from [36, Theorem 2.6], the robust optimization problem V coincides with the function v given in (3.10), as claimed. \square

APPENDIX A. SUPPLEMENTARY STATEMENTS

A.1. Approximation of continuous functions. We often approximate functions in C_b (in the mixed topology) with functions with stronger regularity, e.g., functions in Lip_b or C_c^∞ . In fact any function $f \in C_b$ can be approximated in the mixed topology by a sequence of functions in C_c^∞ . This follows directly from the fact that \mathbb{R}^d is a σ -compact space.

Lemma A.1. *Let $f \in C_b$. Then there exists a sequence $(f_k)_{k \in \mathbb{N}} \subset C_c^\infty$ such that $f_k \xrightarrow{\tau_M} f$ as $k \rightarrow \infty$.*

Proof. Let $K_k := \{x \in \mathbb{R}^d \mid \|x\| \leq k\}$ for all $k \in \mathbb{N}_0$. Then, by the Stone-Weierstrass theorem, for every $k \in \mathbb{N}$, there exists a smooth function $g_k: \mathbb{R}^d \rightarrow \mathbb{R}$ with $\|g_k - f\|_{\infty, K_k} \leq \frac{1}{k}$. Moreover, we

consider a family of cut-off functions $(\chi_k)_{k \in \mathbb{N}} \subset C_c^\infty$ with $0 \leq \chi_k \leq 1$, $\chi_k(x) = 1$ for all $x \in K_{k-1}$, and $\chi_k(x) = 0$ for all $x \in \mathbb{R}^d \setminus K_k$. Then, $f_k := g_k \chi_k \in C_c^\infty$ with

$$\|f_k\|_\infty \leq \|g_k\|_{\infty, K_k} \leq \|f\|_{\infty, K_k} + \frac{1}{k} \leq \|f\|_\infty + 1$$

for all $k \in \mathbb{N}$ and, for all $\varepsilon > 0$ and $K \Subset \mathbb{R}^d$, there exists some $k_0 \in \mathbb{N}$ with $k_0 \geq \frac{1}{\varepsilon}$ and $K \subset K_{k_0-1}$, so that

$$\|f_k - f\|_{\infty, K} \leq \|g_k - f\|_{\infty, K_{k-1}} \leq \|g_k - f\|_{\infty, K_k} \leq \frac{1}{k} \leq \varepsilon \quad \text{for all } k \in \mathbb{N} \text{ with } k \geq k_0.$$

We have therefore shown that $C_c^\infty \ni f_k \xrightarrow{\tau_{\mathcal{M}}} f$ as claimed. \square

A.2. A priori estimates on the linear semigroups. In this section, we report an a priori estimate for the linear semigroups $(T^a(t))_{t \geq 0}$, which is central for the proof of Lemma 4.2.

Lemma A.2. *Suppose that Assumption 2.2 holds. Then, for every $\varepsilon > 0$, there exist a constant $C' \geq 0$ such that, for all $f \in C_c^2$ and $t > 0$ sufficiently small,*

$$\sup_{a \in A} \sup_{x \in \mathbb{R}^d} \left| \frac{(T^a(t)f)(x) - f(x)}{t} \right| \leq \varepsilon \|f\|_\infty + C \left(\sup_{y \in \mathbb{R}^d} (1 + \|y\|) \|\nabla f(y)\| + \|\nabla^2 f\|_\infty \right).$$

Before stating the proof of Lemma A.2, we need a simple estimate on the deterministic drift of the controlled dynamics. The following lemma enables us to ban $\psi_t^a(x)$ from any arbitrarily large neighborhood of 0, uniformly in $a \in A$, by taking x out of a larger neighborhood and choosing t small enough. The proof is a straightforward modification of [60, Lemma 5.4 c)] to our setup.

Lemma A.3. *Suppose that Assumption 2.2 (i), (ii) hold. Then, for every $R' > R \geq 0$, there exists $t_0 > 0$ such that, for every $x \in \mathbb{R}^d$ with $\|x\| \geq R'$ and every $t \in [0, t_0]$,*

$$\inf_{a \in A} \|\psi_t^a(x)\| \geq R.$$

Proof. Let $R' > R \geq 0$. Then, we set $t_0 := \frac{R'-R}{C(1+R')}$ with $C > 0$ appearing in Assumption 2.2 (i). Then, for every $x \in \mathbb{R}^d$ with $\|x\| \geq R'$ and every $t \in [0, t_0]$,

$$\sup_{a \in A} \|\psi_t^a(x) - x\| \leq t_0 C (1 + \|x\|).$$

Hence, by the inverse triangle inequality together with the fact that $R' > R$, we conclude that, for all $x \in \mathbb{R}^d$ with $\|x\| \geq R'$ and all $t \in [0, t_0]$,

$$\begin{aligned} \inf_{a \in A} \|\psi_t^a(x)\| &\geq \|x\| - \sup_{a \in A} \|\psi_t^a(x) - x\| \geq \|x\| - t_0 C (1 + \|x\|) = (1 + \|x\|)(1 - t_0 C) - 1 \\ &= \frac{1 + \|x\|}{1 + R'} (1 + R) - 1 \geq R, \end{aligned}$$

as claimed. \square

Proof of Lemma A.2. Fix $\varepsilon, r > 0$ and let $f \in C_c^2$ be such that $\text{supp}(f) \subset \{x \in \mathbb{R}^d \mid \|x\| \leq r\}$. If $f = 0$, the statement is trivial. Therefore assume that $f \neq 0$. By Assumption 2.2 (iv), there exists some constant $M > 0$ such that, for every $t > 0$,

$$\sup_{a \in A} \frac{1}{t} \mu_t^a \{y \in \mathbb{R}^d \mid \|y\| > M\} \leq \frac{\varepsilon}{4}.$$

This ensures that, for all $t > 0$ and $a \in A$,

$$(A.1) \quad \left| \frac{T^a(t)f(x) - f(x)}{t} \right| \leq \left| \int_{\|y\| \leq M} \frac{f(\psi_t^a(x) + y) - f(x)}{t} \mu_t^a(dy) \right| + \frac{\varepsilon}{2} \|f\|_\infty.$$

Note that, by Lemma A.3, there exists $R > r$ such that $\|\psi_t^a(x) + y\| > r$ for all $x, y \in \mathbb{R}^d$ with $\|x\| > R$ and $\|y\| \leq M$, all $a \in A$, and all $t > 0$ sufficiently small. Hence,

$$(A.2) \quad f(\psi_t^a(x) + y) - f(x) = 0$$

for all $x, y \in \mathbb{R}^d$ with $\|x\| > R$ and $\|y\| \leq M$, $a \in A$, and $t > 0$ sufficiently small.

Therefore, it suffices to estimate the first summand in the right-hand side of (A.1) for $\|x\| \leq R$, $t > 0$ sufficiently small, and $a \in A$.

Let (t, x, a) be a tuple of values satisfying the previous constraints and write

$$(A.3) \quad \begin{aligned} \int_{\|y\| \leq M} \frac{f(\psi_t^a(x) + y) - f(x)}{t} \mu_t^a(dy) &= \int_{\|y\| \leq M} \frac{f(\psi_t^a(x) + y) - f(x + y)}{t} \mu_t^a(dy) \\ &\quad + \int_{\|y\| \leq M} \frac{f(x + y) - f(x)}{t} \mu_t^a(dy). \end{aligned}$$

We tackle the two terms separately. For the first term, applying a Taylor expansion, we get

$$(A.4) \quad \begin{aligned} \int_{\|y\| \leq M} \frac{f(\psi_t^a(x) + y) - f(x + y)}{t} \mu_t^a(dy) &= \int_{\|y\| \leq M} \frac{1}{t} \langle \nabla f(x + y), \psi_t^a(x) - x \rangle \mu_t^a(dy) \\ &\quad + \int_{\|y\| \leq M} \int_0^1 \frac{1-s}{t} \langle \psi_t^a(x) - x, \nabla^2 f(x + y + s(\psi_t^a(x) - x))(\psi_t^a(x) - x) \rangle ds \mu_t^a(dy) \\ &\leq \int_{\|y\| \leq M} \|\nabla f(x + y)\| \frac{\|\psi_t^a(x) - x\|}{t} \mu_t^a(dy) + \|\nabla^2 f\|_\infty \frac{\|\psi_t^a(x) - x\|^2}{t} \\ &\leq C(1 + \|x\|) \int_{\|y\| \leq M} \|\nabla f(x + y)\| \mu_t^a(dy) + tC^2(1 + R)^2 \|\nabla^2 f\|_\infty, \end{aligned}$$

where $C > 0$ is the constant appearing in Assumption 2.2 (i). Since μ_t^a converges to δ_0 in the Wasserstein- p topology, uniformly over $a \in A$, as $t \rightarrow 0$ and f has bounded second derivatives, we obtain

$$C(1 + \|x\|) \int_{\|y\| \leq M} \|\nabla f(x + y)\| \mu_t^a(dy) \rightarrow C(1 + \|x\|) \|\nabla f(x)\| \quad \text{as } t \rightarrow 0$$

uniformly over $a \in A$ and $x \in \mathbb{R}^d$ with $\|x\| \leq R$. Since $f \neq 0$, i.e., $\|f\|_\infty > 0$, we can bound (A.4), for $t > 0$ small enough, by

$$(A.5) \quad \begin{aligned} \int_{\|y\| \leq M} \frac{f(\psi_t^a(x) + y) - f(x + y)}{t} \mu_t^a(dy) &\leq C(1 + \|x\|) \|\nabla f(x)\| + tC^2(1 + R)^2 \|\nabla^2 f\|_\infty + \frac{\varepsilon}{2} \|f\|_\infty \\ &\leq C((1 + \|x\|) \|\nabla f(x)\| + \|\nabla^2 f\|_\infty) + \frac{\varepsilon}{2} \|f\|_\infty. \end{aligned}$$

Note that, in the previous inequality, we have also chosen t small enough so that $tC(1 + R)^2 \leq 1$.

For the second term in (A.3), we use a Taylor expansion for f around x to get

$$\begin{aligned} \int_{\|y\| \leq M} \frac{f(x + y) - f(x)}{t} \mu_t^a(dy) &= \frac{1}{t} \int_{\|y\| \leq M} \langle \nabla f(x), y \rangle \mu_t^a(dy) \\ &\quad + \frac{1}{t} \int_{\|y\| \leq M} \int_0^1 (1-s) \langle \nabla^2 f(x + sy) y, y \rangle ds \mu_t^a(dy). \end{aligned}$$

Again we tackle the two terms separately. First, let $\chi \in C_c^\infty$ such that $0 \leq \chi \leq 1$, $\chi(0) = 1$, and $\chi(y) = 0$ for $\|y\| > M$. Note that the function $x \mapsto x\chi(x)$ is also in C_c^∞ . We then have some $C_2 > 0$ such that

$$\frac{1}{t} \int_{\|y\| \leq M} \langle \nabla f(x), y \rangle \mu_t^a(dy) \leq \|\nabla f\|_\infty \left| \int_{\|y\| \leq M} \frac{y\chi(y) + y(1-\chi(y))}{t} \mu_t^a(dy) \right|$$

$$\begin{aligned}
&\leq \|\nabla f\|_\infty \left(\left| \int_{\mathbb{R}^d} \frac{y\chi(y)}{t} \mu_t^a(dy) \right| + M \left| \int_{\mathbb{R}^d} \frac{\chi(y) - \chi(0)}{t} \mu_t^a(dy) \right| \right) \\
(A.6) \quad &\leq C_2 \|\nabla f\|_\infty,
\end{aligned}$$

where the last inequality follows from Assumption 2.2 (iii).

Moreover, since the function $x \mapsto \|x\|^2 \chi(x)$ is in C_c^∞ , we can use the same arguments presented for (A.6) to have some $C_3 > 0$ satisfying

$$\begin{aligned}
&\frac{1}{t} \int_{\|y\| \leq M} \int_0^1 (1-s) \langle \nabla^2 f(x+sy) y, y \rangle ds \mu_t^a(dy) \\
&\leq \|\nabla^2 f\|_\infty \int_{\|y\| \leq M} \frac{\|y\|^2 \chi(y) + \|y\|^2 (1 - \chi(y))}{t} \mu_t^a(dy) \\
&\leq \|\nabla^2 f\|_\infty \left(\int_{\mathbb{R}^d} \frac{\|y\|^2 \chi(y)}{t} \mu_t^a(dy) + M^2 \left| \int_{\mathbb{R}^d} \frac{\chi(y) - \chi(0)}{t} \mu_t^a(dy) \right| \right) \\
(A.7) \quad &\leq C_3 \|\nabla^2 f\|_\infty.
\end{aligned}$$

Combining (A.5), (A.6), and (A.7), there exists a constant $C' > 0$ such that

$$\int_{\|y\| \leq M} \frac{f(\psi_t^a(x) + y) - f(x)}{t} \mu_t^a(dy) \leq C' \left(\sup_{y \in \mathbb{R}^d} (1 + \|y\|) \|\nabla f(y)\| + \|\nabla^2 f\|_\infty \right) + \frac{\varepsilon}{2} \|f\|_\infty$$

for all $t > 0$ small enough. The proof is complete. \square

A.3. Continuity from above. The following lemma is a particular version of [16, Corollary C.4] tailored to families of subadditive, positive homogeneous, and monotone operators. It provides a criterion to verify the continuity from above of such families of operators, uniformly on compacts, by means of suitable cut-off functions. Note that we work with the sup-norm, which corresponds to the setting $\kappa \equiv 1$ in [16]. For a nonempty index set \mathcal{K} , we consider a family $\{\mathfrak{I}_k \mid k \in \mathcal{K}\}$ of subadditive, positive homogeneous, and monotone operators $\mathfrak{I}_k: C_b \rightarrow F$ with $\mathfrak{I}_k 0 = 0$ and

$$(A.8) \quad \sup_{k \in \mathcal{K}} \sup_{\|f\|_\infty \leq r} \|\mathfrak{I}_k f\|_\infty < \infty \quad \text{for all } r \geq 0,$$

where F denotes the space of all bounded functions $\mathbb{R}^d \rightarrow \mathbb{R}$.

Lemma A.4. *Assume that for every $\varepsilon > 0$ and $K \Subset \mathbb{R}^d$, there exist a function $\zeta \in C_b$ and $\tilde{K} \Subset \mathbb{R}^d$ with*

- (i) $0 \leq \zeta(x) \leq 1$ for all $x \in \mathbb{R}^d$,
- (ii) $\zeta(x) \leq \varepsilon$ for all $x \in \tilde{K}^c$,
- (iii) $(\mathfrak{I}_k(1 - \zeta))(x) \leq \varepsilon$ for all $k \in \mathcal{K}$ and $x \in K$.

Then for every $\varepsilon > 0$, $r \geq 0$, and $K \Subset \mathbb{R}^d$, there exist some $c' \geq 0$ and $K' \Subset \mathbb{R}^d$ such that, for all $k \in \mathcal{K}$ and $f, g \in C_b$ with $\|f\|_\infty \leq r$ and $\|g\|_\infty \leq r$,

$$\|\mathfrak{I}_k f - \mathfrak{I}_k g\|_{\infty, K} \leq c' \|f - g\|_{\infty, K'} + \varepsilon.$$

Proof. Suppose the result does not hold, in particular, suppose we find $\varepsilon > 0$, $r \geq 0$, and $K \Subset \mathbb{R}^d$, such that, for all $n \in \mathbb{N}$, there exist an index $k_n \in \mathcal{K}$ and functions $f_n, g_n \in C_b$ with $\|f_n\|_\infty \leq r$ and $\|g_n\|_\infty \leq r$, satisfying

$$\|\mathfrak{I}_{k_n} f_n - \mathfrak{I}_{k_n} g_n\|_{\infty, K} > n \|f_n - g_n\|_{\infty, K_n} + \varepsilon,$$

where $K_n := \{x \in \mathbb{R}^d : \|x\| \leq n\}$. Then, by [16, Corollary C.4] it is not possible to find a function $\zeta \in C_b$ satisfying the assumptions (i)-(iii) in the statement above. As a matter of fact, if such a function existed, by setting $\zeta_x := \zeta$ for all $x \in K$, we would also have a family of functions

$(\zeta_x)_{x \in K}$ satisfying the assumptions of [16, Corollary C.4] for the countable set $\{k_n : n \in \mathbb{N}\} \subset \mathcal{K}$, thus implying that it is possible to find $c' \geq 0$ and $K' \subseteq \mathbb{R}^d$ such that

$$\|\mathfrak{I}_k f - \mathfrak{I}_k g\|_{\infty, K} \leq c' \|f - g\|_{\infty, K'} + \varepsilon$$

for all $k \in \{k_n : n \in \mathbb{N}\}$ and $f, g \in C_b$ with $\|f\|_\infty \leq r$ and $\|g\|_\infty \leq r$. By choosing $n \in \mathbb{N}$ such that $n \geq c'$ and $K' \subseteq K_n$, we obtain that the index k_n and the functions f_n, g_n , violate the previous condition, leading to a contradiction. \square

We end this section with a technical result borrowed from [60], which provides the existence of suitable cut-off functions.

Lemma A.5. *For all $\varepsilon > 0$ and $n \in \mathbb{N}$, there exists a function $\zeta \in C_c^\infty$ with $0 \leq \zeta \leq 1$, $\zeta(x) = 1$ for all $x \in \mathbb{R}^d$ with $\|x\| \leq n$, and*

$$(A.9) \quad \sup_{x \in \mathbb{R}^d} ((1 + \|x\|) \|\nabla \zeta(x)\| + \|\nabla^2 \zeta(x)\|) \leq \varepsilon.$$

Proof. It is enough to note that the function $x \mapsto 1 + \|x\|$ satisfies the hypothesis of [60, Lemma A.1]. The result is then a particular version of the aforementioned lemma. \square

REFERENCES

- [1] L. Ambrosio, E. Brué, and D. Semola. *Lectures on optimal transport*, volume 130 of *Unitext*. Springer, Cham, [2021] ©2021. La Matematica per il 3+2.
- [2] G. Barles, R. Buckdahn, and E. Pardoux. Backward stochastic differential equations and integral-partial differential equations. *Stochastics*, 60(1-2):57–83, 1997.
- [3] D. Bartl, S. Drapeau, J. Obłój, and J. Wiesel. Sensitivity analysis of Wasserstein distributionally robust optimization problems. *Proc. R. Soc. A*, 477(2256):20210176, 2021.
- [4] D. Bartl, S. Eckstein, and M. Kupper. Limits of random walks with distributionally robust transition probabilities. 2021.
- [5] D. Bartl, M. Kupper, and A. Neufeld. Duality theory for robust utility maximisation. *Finance Stoch.*, 25(3):469–503, 2021.
- [6] D. Bartl, A. Neufeld, and K. Park. Sensitivity of robust optimization problems under drift and volatility uncertainty. *Preprint arXiv:2311.11248*, 2023.
- [7] D. Bartl, A. Neufeld, and K. Park. Numerical method for nonlinear Kolmogorov PDEs via sensitivity analysis. *Preprint arXiv:2403.11910*, 2024.
- [8] D. Bartl and J. Wiesel. Sensitivity of multiperiod optimization problems with respect to the adapted Wasserstein distance. *SIAM J. Financial Math.*, 14(2):704–720, 2023.
- [9] E. Bayraktar and T. Chen. Nonparametric adaptive robust control under model uncertainty. *SIAM J. Control Optim.*, 61(5):2737–2760, 2023.
- [10] R. Blanchard and L. Carassus. Multiple-priors optimal investment in discrete time for unbounded utility function. *Ann. Appl. Probab.*, 28(3):1856–1892, 2018.
- [11] J. Blanchet, L. Chen, and X. Y. Zhou. Distributionally robust mean-variance portfolio selection with Wasserstein distances. *Manag. Sci.*, 68(9):6382–6410, 2022.
- [12] J. Blanchet and K. Murthy. Quantifying distributional model risk via optimal transport. *Math. Oper. Res.*, 44(2):565–600, 2019.
- [13] J. Blessing, R. Denk, M. Kupper, and M. Nendel. Convex monotone semigroups and their generators with respect to Γ -convergence. *J. Funct. Anal.*, 288(8):Paper No. 110841, 73, 2025.
- [14] J. Blessing and M. Kupper. Nonlinear Semigroups Built on Generating Families and their Lipschitz Sets. *Potential Anal.*, 59(3):857–895, 2023.
- [15] J. Blessing and M. Kupper. Nonlinear semigroups and limit theorems for convex expectations. *Ann. Appl. Probab.*, 35(2):779–821, 2025.
- [16] J. Blessing, M. Kupper, and M. Nendel. Convergence of infinitesimal generators and stability of convex monotone semigroups. *Preprint arXiv:2305.18981*, 2023.
- [17] J. Blessing, M. Kupper, and A. Sgarabottolo. Discrete approximation of risk-based prices under volatility uncertainty. *Preprint arXiv:2411.00713*, 2024.
- [18] R. C. Buck. Bounded continuous functions on a locally compact space. *Michigan Math. J.*, 5:95–104, 1958.

- [19] R. Buckdahn and J. Li. Stochastic differential games and viscosity solutions of Hamilton–Jacobi–Bellman–Isaacs equations. *SIAM J. Control Optim.*, 47(1):444–475, 2008.
- [20] S. Cerreia-Vioglio, L. P. Hansen, F. Maccheroni, and M. Marinacci. Making decisions under model misspecification. *The Review of Economic Studies*, page rdaf046, 06 2025.
- [21] Z. Chen and L. Epstein. Ambiguity, risk, and asset returns in continuous time. *Econometrica*, 70(4):1403–1443, 2002.
- [22] 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.
- [23] A. Compain, N. Saulduois, and N. Touzi. Sensitivity Analysis of Distributionally Robust BSDEs and RBS-DEs. *Preprint arXiv:2511.01828*, 2025.
- [24] F. Coquet, Y. Hu, J. Mémin, and S. Peng. Filtration-consistent nonlinear expectations and related g -expectations. *Probab. Theory Relat. Fields*, 123(1):1–27, 2002.
- [25] M. G. Crandall, H. Ishii, and P.-L. Lions. User’s guide to viscosity solutions of second order partial differential equations. *Bull. Am. Math. Soc., New Ser.*, 27(1):1–67, 1992.
- [26] M. G. Crandall and P.-L. Lions. Viscosity solutions of Hamilton–Jacobi equations. *Trans. Amer. Math. Soc.*, 277(1):1–42, 1983.
- [27] D. Criens and M. Kupper. Representation Theorems for Convex Expectations and Semigroups on Path Space. *Preprint arXiv:2503.10572*, 2025.
- [28] D. Criens and L. Niemann. A stochastic representation theorem for sublinear semigroups with non-local generators. *Electron. J. of Probab.*, 30:1–36, 2025.
- [29] L. Denis, M. Hu, and S. Peng. Function spaces and capacity related to a sublinear expectation: application to G -brownian motion paths. *Potential Anal.*, 34:139–161, 2011.
- [30] L. Denis and M. Kervarec. Optimal investment under model uncertainty in nondominated models. *SIAM J. Control Optim.*, 51(3):1803–1822, 2013.
- [31] R. Denk, M. Kupper, and M. Nendel. Kolmogorov-type and general extension results for nonlinear expectations. *Banach J. Math. Anal.*, 12(3):515–540, 2018.
- [32] R. Denk, M. Kupper, and M. Nendel. A semigroup approach to nonlinear Lévy processes. *Stochastic Process. Appl.*, 130(3):1616–1642, 2020.
- [33] J. Dow and S. R. da Costa Werlang. Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica*, pages 197–204, 1992.
- [34] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Math. Finance*, 7(1):1–71, 1997.
- [35] W. H. Fleming and H. M. Soner. *Controlled Markov processes and viscosity solutions*, volume 25. Springer Science & Business Media, 2006.
- [36] W. H. Fleming and P. E. Souganidis. On the existence of value functions of two-player, zero-sum stochastic differential games. *Indiana Univ. Math. J.*, 38(2):293–314, 1989.
- [37] D. H. Fremlin, D. J. H. Garling, and R. G. Haydon. Bounded measures on topological spaces. *Proc. London Math. Soc. (3)*, 25:115–136, 1972.
- [38] A. Friedman. Stochastic differential games. *J. Differential Equations*, 11(1):79–108, 1972.
- [39] F. Fuchs and M. Nendel. Existence of Viscosity Solutions to Abstract Cauchy Problems via Nonlinear Semigroups. *Preprint arXiv:2502.18270*, 2025.
- [40] S. Fuhrmann, M. Kupper, and M. Nendel. Wasserstein perturbations of Markovian transition semigroups. *Ann. Inst. Henri Poincaré Probab. Stat.*, 59(2):904–932, 2023.
- [41] R. Gao, X. Chen, and A. J. Kleywegt. Wasserstein distributionally robust optimization and variation regularization. *Oper. Res.*, 72(3):1177–1191, 2024.
- [42] R. Gao and A. Kleywegt. Distributionally robust stochastic optimization with Wasserstein distance. *Math. Oper. Res.*, 48(2):603–655, 2023.
- [43] L. Garlappi, R. Uppal, and T. Wang. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Rev. Financial Stud.*, 20(1):41–81, 2007.
- [44] I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *J. Math. Econom.*, 18(2):141–153, 1989.
- [45] B. Goldys and M. Kocan. Diffusion semigroups in spaces of continuous functions with mixed topology. *J. Differential Equations*, 173(1):17–39, 2001.
- [46] B. Goldys, M. Nendel, and M. Röckner. Operator semigroups in the mixed topology and the infinitesimal description of Markov processes. *J. Differential Equations*, 412:23–86, 2024.
- [47] J.-y. Gotoh, M. J. Kim, and A. E. Lim. Distributionally Robust Optimization is a Multi-Objective Problem. *Preprint arXiv:2507.11350*, 2025.

- [48] S. Hamadene, J.-P. Lepeltier, S. Peng, et al. BSDEs with continuous coefficients and stochastic differential games. *Pitman Research Notes in Mathematics Series*, pages 115–128, 1997.
- [49] L. P. Hansen and T. J. Sargent. Structured ambiguity and model misspecification. *J. Econ. Theory*, 199:32, 2022. Id/No 105165.
- [50] D. Hernández-Hernández and A. Schied. A control approach to robust utility maximization with logarithmic utility and time-consistent penalties. *Stochastic Process. Appl.*, 117(8):980–1000, 2007.
- [51] S. Herrmann and J. Muhle-Karbe. Model uncertainty, recalibration, and the emergence of delta–vega hedging. *Finance Stoch.*, 21:873–930, 2017.
- [52] S. Herrmann, J. Muhle-Karbe, and F. T. Seifried. Hedging with small uncertainty aversion. *Finance Stoch.*, 21(1):1–64, 2017.
- [53] N. Jacob. *Pseudo differential operators and Markov processes. In 3 vol. Vol. 1: Fourier analysis and semigroups*. London: Imperial College Press, 2001.
- [54] Y. Jiang and J. Obloj. Sensitivity of causal distributionally robust optimization. *arXiv preprint arXiv:2408.17109*, 2024.
- [55] M. Kunze. Continuity and equicontinuity of semigroups on norming dual pairs. *Semigroup Forum*, 79(3):540–560, 2009.
- [56] C. Liu and A. Neufeld. Compactness criterion for semimartingale laws and semimartingale optimal transport. *Trans. Amer. Math. Soc.*, 372(1):187–231, 2019.
- [57] F. Maccheroni, M. Marinacci, and A. Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006.
- [58] A. Matoussi, D. Possamaï, and C. Zhou. Robust utility maximization in nondominated models with 2BSDE: the uncertain volatility model. *Math. Finance*, 25(2):258–287, 2015.
- [59] P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Programming*, 171(1):115–166, 2018.
- [60] M. Nendel. Chernoff-Mehler Approximation for Lévy Processes with Drift. *Preprint arXiv:2511.19414*, 2025.
- [61] M. Nendel. Lower semicontinuity of monotone functionals in the mixed topology on C_b . *Finance Stoch.*, 29(1):261–287, 2025.
- [62] M. Nendel and M. Röckner. Upper envelopes of families of Feller semigroups and viscosity solutions to a class of nonlinear Cauchy problems. *SIAM J. Control Optim.*, 59(6):4400–4428, 2021.
- [63] M. Nendel and A. Sgarabottolo. A parametric approach to the estimation of convex risk functionals based on Wasserstein distance. *Preprint arXiv:2210.14340*, 2022.
- [64] A. Neufeld and M. Nutz. Measurability of semimartingale characteristics with respect to the probability law. *Stochastic Process. Appl.*, 124(11):3819–3845, 2014.
- [65] A. Neufeld and M. Nutz. Nonlinear Lévy processes and their characteristics. *Trans. Amer. Math. Soc.*, 369(1):69–95, 2017.
- [66] A. Neufeld and M. Nutz. Robust utility maximization with Lévy processes. *Math. Finance*, 28(1):82–105, 2018.
- [67] A. Neufeld, J. Sester, and M. Šikić. Markov decision processes under model uncertainty. *Math. Finance*, 33(3):618–665, 2023.
- [68] M. Nutz. Random G -expectations. *Ann. Appl. Probab.*, 23(5):1755–1777, 2013.
- [69] J. Obłój and J. Wiesel. Distributionally robust portfolio maximization and marginal utility pricing in one period financial markets. *Math. Finance*, 31(4):1454–1493, 2021.
- [70] E. Pardoux and S. Peng. Adapted solution of a backward stochastic differential equation. *Syst. Control Lett.*, 14(1):55–61, 1990.
- [71] K. Park, K. Chen, and H. Y. Wong. Irreversible consumption habit under ambiguity: Singular control and optimal G -stopping time. *Ann. Appl. Probab.*, 35(4):2471–2525, 2025.
- [72] K. Park and H. Y. Wong. Robust consumption-investment with return ambiguity: A dual approach with volatility ambiguity. *SIAM J. Financial Math.*, 13(3):802–843, 2022.
- [73] K. Park and H. Y. Wong. Robust retirement with return ambiguity: Optimal G -stopping time in dual space. *SIAM J. Control Optim.*, 61(3):1009–1037, 2023.
- [74] S. Peng. Backward SDE and related g -expectation. *Pitman research notes in mathematics series*, pages 141–160, 1997.
- [75] S. Peng. G -expectation, G -Brownian motion and related stochastic calculus of Itô type. In *Stochastic Analysis and Applications: The Abel Symposium 2005*, pages 541–567. Springer, 2007.
- [76] S. Peng. Multi-dimensional G -brownian motion and related stochastic calculus under G -expectation. *Stochastic Process. Appl.*, 118(12):2223–2253, 2008.
- [77] H. Pham. *Continuous-time stochastic control and optimization with financial applications*, volume 61. Springer Science & Business Media, 2009.

- [78] H. Pham, X. Wei, and C. Zhou. Portfolio diversification and model uncertainty: A robust dynamic mean-variance approach. *Math. Finance*, 32(1):349–404, 2022.
- [79] N. Sauldubois and N. Touzi. First order martingale model risk and semi-static hedging. *Preprint arXiv:2410.06906*, 2024.
- [80] A. Schied. Risk measures and robust optimization problems. *Stochastic Models*, 22(4):753–831, 2006.
- [81] A. Schied. Optimal investments for risk-and ambiguity-averse preferences: a duality approach. *Finance Stoch.*, 11(1):107–129, 2007.
- [82] F. D. Sentilles. Bounded continuous functions on a completely regular space. *Trans. Amer. Math. Soc.*, 168:311–336, 1972.
- [83] H. M. Soner, N. Touzi, and J. Zhang. Quasi-sure stochastic analysis through aggregation. *Electron. J. of Probab.*, 16:1844–1879, 2011.
- [84] H. M. Soner, N. Touzi, and J. Zhang. Wellposedness of second order backward SDEs. *Probab. Theory Relat. Fields*, 153(1):149–190, 2012.
- [85] H. M. Soner, N. Touzi, and J. Zhang. Dual formulation of second order target problems. *Ann. Appl. Probab.*, pages 308–347, 2013.
- [86] D. Talay and Z. Zheng. Worst case model risk management. *Finance Stoch.*, 6:517–537, 2002.
- [87] R. Tevzadze, T. Toronjadze, and T. Uzunashvili. Robust utility maximization for a diffusion market model with misspecified coefficients. *Finance Stoch.*, 17:535–563, 2013.
- [88] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
- [89] R. F. Wheeler. A survey of Baire measures and strict topologies. *Exposition. Math.*, 1(2):97–190, 1983.
- [90] A. Wiweger. Linear spaces with mixed topology. *Studia Math.*, 20:47–68, 1961.
- [91] J. Yong and X. Y. Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*, volume 43. Springer Science & Business Media, 1999.
- [92] M.-C. Yue, D. Kuhn, and W. Wiesemann. On linear optimization over Wasserstein balls. *Math. Programming*, 195(1):1107–1122, 2022.

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE, UNIVERSITY OF WATERLOO
Email address: mnendel@uwaterloo.ca

DIVISION OF MATHEMATICAL SCIENCES, NANYANG TECHNOLOGICAL UNIVERSITY
Email address: ariel.neufeld@ntu.edu.sg

DIVISION OF MATHEMATICAL SCIENCES, NANYANG TECHNOLOGICAL UNIVERSITY
Email address: kyunghyun.park@ntu.edu.sg

DEPARTMENT OF MATHEMATICS, LMU MÜNCHEN
Email address: sgarabottolo@math.lmu.de