Electron. Commun. Probab. **26** (2021), article no. 28, 1–13.
https://doi.org/10.1214/21-ECP393
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Limits of random walks with distributionally robust transition probabilities*

Daniel Bartl[†] Stephan Eckstein[‡] Michael Kupper[§]

Abstract

We consider a nonlinear random walk which, in each time step, is free to choose its own transition probability within a neighborhood (w.r.t. Wasserstein distance) of the transition probability of a fixed Lévy process. In analogy to the classical framework we show that, when passing from discrete to continuous time via a scaling limit, this nonlinear random walk gives rise to a nonlinear semigroup. We explicitly compute the generator of this semigroup and corresponding PDE as a perturbation of the generator of the initial Lévy process.

Keywords: nonlinear Lévy processes; Wasserstein distance; scaling limit.
MSC2020 subject classifications: 620G51; 60G50; 47H20.
Submitted to ECP on July 20, 2020, final version accepted on April 18, 2021.

1 Introduction and main results

Lévy processes are mathematically tractable and therefore often used to model certain real-world phenomena. This bears the task of correctly specifying / estimating the corresponding parameters, e.g., drift and variance in case of a Brownian motion. In many situations this can only be achieved up to a certain degree of *uncertainty*. For this reason, Peng [20] introduced his nonlinear Brownian motion and started a systematic investigation of this object. The nonlinear Brownian motion is defined via a nonlinear PDE and, heuristically speaking, within each infinitesimal time increment it is allowed to select its parameters (drift and variance) within a given fixed set. Accordingly, a nonlinear Feynman-Kac formula makes it possible to compute the worst case expectations of certain functions of the random process. Several works followed this *parametric* nonlinearization approach to Lévy processes, see, e.g., Hu and Peng [14], Neufeld and Nutz [18], Denk et al. [9] and Kühn [16].

On the other hand, in discrete time where no mathematical limitations force one to restrict to parametric uncertainty, a more natural and general nonlinearization of a given (baseline) random walk is of *nonparametric* nature. We start with a random walk which is the discrete-time restriction of an $\mathbb{R}^d$-valued Lévy process starting in zero, whose marginal laws we denote by $(\mu_t)_{t\geq 0}$. For instance, μ_t can be the normal distribution with mean 0 and variance t in which case we end up with a Gaussian random walk.

*Daniel Bartl is grateful for financial support through the Vienna Science and Technology Fund (WWTF) project MA16-021 and the Austrian Science Fund (FWF) project P28661.

[†]University of Vienna, Austria. E-mail: daniel.bartl@univie.ac.at
[‡]University of Hamburg, Germany. E-mail: stephan.eckstein@uni-hamburg.de
[§]University of Konstanz, Germany. E-mail: kupper@uni-konstanz.de

For a fixed parameter $\delta \geq 0$ representing the level of freedom (or uncertainty) and $n \in \mathbb{N}$, the *nonlinear random walk* with time index $\mathbb{T} = \{0 = t_0 < t_1 < t_2 < \cdots\} \subset \mathbb{R}_+$ is defined as follows: for each time step $t_n \rightsquigarrow t_{n+1}$, the nonlinear random walk is allowed to select its transition probability within the neighborhood of size $\delta \Delta t_{n+1}$ of the transition probability $\mu_{\Delta t_{n+1}}$ of our baseline random walk, where $\Delta t_{n+1} := t_{n+1} - t_n$ and the neighborhood is taken w.r.t. the p-th *Wasserstein distance* $\mathcal{W}_p$.[1] This means that, conditioned on the event that the nonlinear random walk takes the value $x \in \mathbb{R}^d$ at time t_n, the worst possible expected value of an arbitrary function $f \in C_0(\mathbb{R}^d)$ at time t_{n+1} is given by

$$S(\Delta t_{n+1})f(x) := \sup \Big\{ \int_{\mathbb{R}^d} f(x+y)\, \nu(dy) : \nu \text{ such that } \mathcal{W}_p(\mu_{\Delta t_{n+1}}, \nu) \leq \delta \Delta t_{n+1} \Big\}.$$

Recall here that $C_0(\mathbb{R}^d)$ is the set of continuous function vanishing at infinity. Iterating this scheme, conditioned on the event that the nonlinear random walk starts in x at time 0, the worst possible expectation at time $t_n \in \mathbb{T}$ is given by

$$\mathscr{S}^{\mathbb{T}}(t_n)f(x) := S(t_1 - t_0) \circ \cdots \circ S(t_n - t_{n-1})f(x). \tag{1.1}$$

In conclusion, the corresponding processes follow the same heuristics as the nonlinear Brownian motion and can be seen as a discrete time nonparametric reincarnation thereof.

Regarding the computation of $\mathscr{S}^{\mathbb{T}}$ we stumble on a recurring scheme in discrete time: while definitions are mathematically simple, explicit computations are often very challenging. Here this is evident as S and therefore $\mathscr{S}^{\mathbb{T}}$ are results of (iterated, nonparametric, and infinite dimensional) control problems. In the following, we shall show that when passing from small to infinitesimal time steps, the $\mathscr{S}^{\mathbb{T}}$'s give rise to a nonlinear semigroup and that a computation of the limit is possible via a nonlinear PDE.

For the rest of this article we shall fix $p \in (1, \infty)$ and assume that our initial Lévy process has finite p-th moment, i.e., $\int_{\mathbb{R}^d} |x|^p \, \mu_1(dx) < \infty$. For convenience, for every $n \in \mathbb{N}$ consider dyadic numbers $\mathbb{T}_n := 2^{-n}\mathbb{N}_0$ and set $\mathscr{S}^n(t) := \mathscr{S}^{\mathbb{T}_n}(t_n) \circ S(t - t_n)$ for $t \geq 0$, where $t_n \in \mathbb{T}_n$ is the closest dyadic number prior to t.

Proposition 1.1 (Semigroup). *Both $\mathscr{S}^n$ and $\mathscr{S} := \lim_{n\to\infty} \mathscr{S}^n$ are well-defined and the family $(\mathscr{S}(t))_{t\geq 0}$ defines a sublinear semigroup on $C_0(\mathbb{R}^d)$. More precisely, for every $s, t \geq 0$ and $x \in \mathbb{R}^d$,*

1. *$\mathscr{S}(t)$ maps $C_0(\mathbb{R}^d)$ to itself and $\mathscr{S}(t) \circ \mathscr{S}(s) = \mathscr{S}(t+s)$,*
2. *$\mathscr{S}(t)(\cdot)(x) \colon C_0(\mathbb{R}^d) \to \mathbb{R}$ is sublinear, increasing, and maps zero to zero, and $\mathscr{S}(t)$ is contractive, i.e., $\|\mathscr{S}(t)f - \mathscr{S}(t)g\|_\infty \leq \|f - g\|_\infty$ for all $f, g \in C_0(\mathbb{R}^d)$.*

Now that the semigroup property is established, we can state our main result.

Theorem 1.2 (Feynman-Kac). *Let $f \in C_0(\mathbb{R}^d)$ and define $u\colon [0,\infty)\times\mathbb{R}^d \to \mathbb{R}$ via $u(t,x) := \mathscr{S}(t)f(x)$. Then u is a viscosity solution of*

$$\begin{cases} \partial_t u(t,x) = A^\mu u(t,\cdot)(x) + \delta |\nabla u(t,x)| & \text{for } (t,x) \in (0,\infty) \times \mathbb{R}^d, \\ u(0,x) = f(x) & \text{for } x \in \mathbb{R}^d, \end{cases}$$

where A^μ is the generator of the initial Lévy process.

Here ∇ denotes the spatial derivative. Moreover, the notion of viscosity solution we consider here is that of [9], and we refer to the discussion before and after Theorem 2.12 below for the definition and comments on uniqueness.

[1] For $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ (the set of Borel probabilities on $\mathbb{R}^d$ with finite p-th moment), define $\mathcal{W}_p(\mu,\nu) := \inf\big\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^p \, \gamma(dx,dy) : \gamma \in \mathrm{Cpl}(\mu,\nu)\big\}^{1/p}$ where $\mathrm{Cpl}(\mu,\nu)$ is the set of all Borel probabilities on $\mathbb{R}^d \times \mathbb{R}^d$ with first and second marginal μ and ν, respectively. Throughout, $|\cdot|$ is the Euclidean norm.

Remark 1.3. Starting with an arbitrary Lévy process, Theorem 1.2 ensures that the limiting semigroup $\mathscr{S}$ corresponds to a nonlinear Lévy process with parametric drift uncertainty. The interesting feature of Theorem 1.2 is that, even though we consider *nonparametric uncertainty* in discrete time leading to robustifications that are structurally unconstrained, in the limit we end up with a process bearing only *parametric uncertainty*, which is a drift-perturbed version of the initial Lévy process. For instance, if we start with a Brownian motion with generator $A^\mu = \frac{1}{2}\Delta$, we end up with a g-Brownian motion with generator $\frac{1}{2}\Delta + \delta|\nabla|$. More generally, if the initial Lévy process is a Brownian motion with drift $\gamma \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$, then a quick computation shows that the PDE of Theorem 1.2 takes the form

$$\partial_t u(t,x) = \frac{1}{2}\sum_{i,j=1}^d \Sigma_{ij}\partial_{ij}u(t,x) + \max_{\eta\in\Gamma}\sum_{i=1}^d \eta_i\partial_i u(t,x),$$

where $\Gamma := \{\eta \in \mathbb{R}^d : |\eta-\gamma| \le \delta\}$. In case of a g-Brownian motion, the solution of the PDE can be represented in terms of a g-expectation, see [19]. Likewise, based on the theory of backward stochastic differential equations with jumps, g-expectations also exist for certain jump filtrations (e.g., the one generated by a Brownian motion and a Poisson random measure), see [15, 17, 23]. This leads to corresponding nonlinear processes with parametric drift uncertainty. Finally, we note that similar drift perturbations also arise for related scaling limits, see, e.g., [21, Proposition 11].

In the following chapter we consider a convex generalization of the above setting: in the definition of $S(\Delta t_{n+1})$, instead of considering all ν in the $\delta\Delta t_{n+1}$ neighborhood of $\mu_{\Delta t_{n+1}}$, we take into account all ν but penalized by their distance to $\mu_{\Delta t_{n+1}}$. In the limit this gives a convex semigroup for which the generator includes a convex perturbation in ∇u (instead of the absolute value), see Theorem 2.12.

Finally, let us point out that numerical computation of nonlinear PDEs like the ones resulting from Theorem 1.2 and Remark 1.3 has received a lot of attention in recent years and by now efficient methods are available, see, e.g., [5, 22] and references therein.

Possible extensions and related literature. There are several natural variations of the results in this paper. For instance, one can ask which effect additional constraints on the measures ν appearing in the definition of $S(t)$ might have. Concretely, what would happen if one allows only for those ν which (additional to being in a Wasserstein neighborhood of μ_t) have the same mean as μ_t, or if one replaces the Wasserstein distance by its martingale version [6]. In the latter case, when changing the scaling of the radius from δt to δt^2, one could guess the PDE to be

$$\begin{cases}\partial_t u(t,x) = \Delta u(t,\cdot)(x) + \delta|\nabla^2 u(t,x)| & \text{for } (t,x)\in(0,\infty)\times\mathbb{R}^d,\\ u(0,x) = f(x) & \text{for } x\in\mathbb{R}^d,\end{cases}$$

in case that the underlying Lévy process is the Brownian motion. However, with the exact methods of this paper, this can be made rigorous only with a (unnatural) technical twist in definition of $S(t)$ and understanding the full picture would require considerations beyond the scope of this paper.

In a similar spirit, it would be interesting to start with transition probabilities $(\mu_t^n)_t$ which approximate $(\mu_t)_t$ (e.g. a Binomial random walk which converges to a Brownian motion). A (parametric) variant of this was done by Dolinsky, Nutz, and Soner [10] for Binomial random walks with freedom in the Bernoulli-parameter. Related, one could ask whether Donsker-type results hold, i.e., whether the family of laws of the nonlinear random walks (on the path space) has a limiting family.

Finally, let us highlight the connection to distributionally robust optimization (DRO) using the Wasserstein distance. In DRO, the basic task consists of computing $\inf_\lambda S(t)f^\lambda$, where $(f^\lambda)_\lambda$ is a parametrized family of function; we refer to [4, 7, 12, 13] for recent results and applications. Here duality arguments often help to compute the (infinite dimensional) optimization problem appearing in the definition of $S(t)$. In multi-step versions of DRO (e.g., time-consistent utility maximization with Markovian endowment [2]), the computation of $\mathscr{S}^n(t)f$ is the key element. Related multi-step versions also occur in the literature on robust Markov chains with interval probabilities (see [11, 25] and references therein), and in particular on robust Markov decision processes [26, 27]. As $\mathscr{S}(t)f$ can be seen as a proxy for $\mathscr{S}^n(t)f$ for large n, a natural question is whether the results in the current paper can be used as an approximation tool for these multi-step versions of DRO. This also motivates studying the speed of convergence $\mathscr{S}^n(t)f \to \mathscr{S}(t)f$.

2 Convex version and proof of main results

Let $\varphi\colon [0,+\infty) \to [0,+\infty]$ be a lower semicontinuous, convex, and increasing function which is not constant and such that $\varphi(0) = 0$. Assume that $x \mapsto \varphi(x^{1/p})$ is convex, denote by $\varphi^*(y) := \sup_{x\geq 0}(xy - \varphi(x))$ for $y \geq 0$ its convex conjugate, and set $\varphi(+\infty) := \varphi^*(+\infty) = +\infty$. For every $f \in C_0(\mathbb{R}^d)$ and $t \geq 0$, we define

$$S(t)f(x) = \sup_{\nu\in\mathcal{P}_p(\mathbb{R}^d)} \Big(\int_{\mathbb{R}^d} f(x+y)\,\nu(dy) - \varphi_t(\mathcal{W}_p(\mu_t,\nu))\Big), \tag{2.1}$$

where $\varphi_t(\cdot) := t\varphi(\cdot/t)$ for $t > 0$, $\varphi_0(a) = +\infty$ for $a > 0$, and $\varphi_0(0) = 0$. The results stated in the introduction will follow from the choice $\varphi := +\infty 1_{(\delta,+\infty)}$, in which case $\varphi^*(y) = \delta y$ for all $y \geq 0$. Notice that the supremum in (2.1) can also be taken over the set

$$\Delta_{f,t} := \{\nu \in \mathcal{P}_p(\mathbb{R}^d) : \varphi_t(\mathcal{W}_p(\mu_t,\nu)) \leq \|f\|_\infty + 1\}. \tag{2.2}$$

As φ_t is increasing and unbounded, this implies that there is a uniform upper bound on $\mathcal{W}_p(\mu_t,\nu)$ over $\nu \in \Delta_{f,t}$. Hence, by the following simple observation, the set $\Delta_{f,t}$ is tight, and therefore the supremum in (2.1) is attained[2].

Lemma 2.1. *For every $\nu \in \mathcal{P}_p(\mathbb{R}^d)$, $t \geq 0$, and $c > 0$, we have that*

$$\nu(\{y \in \mathbb{R}^d : |y| \geq c\}) \leq \frac{1}{c}\Big(\mathcal{W}_p(\mu_t,\nu) + \Big(\int_{\mathbb{R}^d} |x|^p\,\mu_t(dx)\Big)^{1/p}\Big).$$

Proof. An application of Markov's and Hölder's inequality implies that $\nu(\{y \in \mathbb{R}^d : |y| \geq c\}) \leq c^{-1}(\int_{\mathbb{R}^d} |y|^p\,\nu(dy))^{1/p}$. The latter equals $c^{-1}\mathcal{W}_p(\delta_0,\nu)$, so that the proof is completed by the triangle inequality for $\mathcal{W}_p$. □

For further reference, we provide the proof of the following simple observation.

Lemma 2.2. *For every $c \geq 0$, we have that $\lim_{t\downarrow 0}\sup\{r \geq 0 : \varphi_t(r) \leq c\} = 0$.*

Proof. By assumption $x \mapsto \varphi(\max\{x,0\}^{1/p})$ is a convex lower semicontinuous function, which is not constant equal to zero. Therefore, by the Fenchel-Moreau theorem there exist $a > 0$ and $b \in \mathbb{R}$, such that $\varphi(x^{1/p}) \geq ax + b$ for all $x \geq 0$. Thus, for every given $r > 0$, we conclude that

$$\varphi_t(r) = t\varphi\big(\tfrac{r}{t}\big) \geq ta\big(\tfrac{r}{t}\big)^p + tb.$$

As $p > 1$, this term converges to infinity when t converges to zero. □

[2] Indeed, the set $\Delta_{f,t}$ is weakly compact by Prokhorov's theorem and lower semicontinuity of $\nu \mapsto \varphi_t(\mathcal{W}_p(\mu_t,\nu))$.

Directly from the definition, the operator $S(t)$ has the following properties.

Lemma 2.3. *Let $t \geq 0$ and $f, g \in C_0(\mathbb{R}^d)$ such that $f \leq g$. Then, $S(t)$ is a convex contraction on $C_0(\mathbb{R}^d)$, which satisfies $S(t)0 = 0$, $S(t)f \leq S(t)g$. Further, $S(t)f$ has the same modulus of continuity as f.*

Proof. It is clear by definition that $S(t)$ is convex and monotone. Moreover, as $\inf \varphi_t = 0$, it follows that $S(t)0 = 0$. To show that $S(t)$ is a contraction, note that

$$\int_{\mathbb{R}^d} f(x+y)\,\nu(dy) \leq \int_{\mathbb{R}^d} g(x+y)\,\nu(dy) + \|f-g\|_\infty$$

for all $\nu \in \mathcal{P}_p(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$. Hence, $S(t)f(x) \leq S(t)g(x) + \|f-g\|_\infty$ for all $x \in \mathbb{R}^d$, and changing the role of f and g yields contractivity.

It remains to prove that $S(t)f \in C_0(\mathbb{R}^d)$. First, since f is in particular uniformly continuous it follows that $S(t)f$ is also uniformly continuous. To that end, let $\varepsilon > 0$ be arbitrary and fix $\delta > 0$ such that $|f(x) - f(y)| \leq \varepsilon$ for $x, y \in \mathbb{R}^d$ with $|x-y| \leq \delta$. Then, for every such pair x, y, contractivity of $S(t)$ implies that

$$\begin{aligned} S(t)f(x) &= S(t)f(x+\cdot)(0) \\ &\leq S(t)f(y+\cdot)(0) + \|f(x+\cdot) - f(y+\cdot)\|_\infty \\ &\leq S(t)f(y) + \varepsilon. \end{aligned}$$

Replacing the role of x and y shows that $S(t)f$ is uniformly continuous with the same modulus of continuity as f.

Second, we prove that $S(t)f$ is vanishing at infinity. Let $\varepsilon > 0$ be arbitrary and fix $a \geq 0$ such that $|f(x)| \leq \varepsilon$ for all $x \in \mathbb{R}^d$ with $|x| \geq a$. Since $\mathcal{W}_p(\mu_t, \nu)$ is uniformly bounded over $\nu \in \Delta_{f,t}$, it follows form Lemma 2.1 that there is $b > 0$ such that $\nu(\{y \in \mathbb{R}^d : |y| > b\}) \leq \varepsilon$ uniformly over $\nu \in \Delta_{f,t}$. Hence,

$$S(t)f(x) \leq \sup_{\nu \in \Delta_t} \int_{\mathbb{R}^d} f(x+y)1_{\{|y| \leq b\}} + f(x+y)1_{\{|y|>b\}}\,\nu(dy) \leq \varepsilon + \varepsilon\|f\|_\infty$$

for all $x \in \mathbb{R}^d$ such that $|x| \geq a+b$. For the reverse inequality, use that $S(t)f(x) \geq \int_{\mathbb{R}^d} f(x+y)\,\mu_t(dy)$ for all $x \in \mathbb{R}^d$, which follows from $\varphi(0) = 0$. Therefore the same arguments as above show that $S(t)f(x) \geq -\varepsilon - \varepsilon\|f\|_\infty$ for all $x \in \mathbb{R}^d$ such that $|x| \geq a+b$. As ε was arbitrary, the claim follows. ☐

At this point we know that $S(t)$ maps $C_0(\mathbb{R}^d)$ to itself, which allows us to define $S(t) \circ S(s)$, or more generally $\mathscr{S}^n$ as in (1.1). The following is the key result for our analysis, and allows in particular to define the limit $\lim_{n\to\infty} \mathscr{S}^n$.

Lemma 2.4. *For every $0 < s < t$ and $f \in C_0(\mathbb{R}^d)$, we have that*

$$S(s)S(t-s)f \leq S(t)f.$$

Proof. Fix $f \in C_0(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$. Let $\nu_s(db) \in \mathcal{P}_p(\mathbb{R}^d)$ such that

$$S(s)S(t-s)f(x) = \int_{\mathbb{R}^d} S(t-s)f(x+b)\,\nu_s(db) - \varphi_s(\mathcal{W}_p(\mu_s, \nu_s))$$

and let $\gamma_s(da, db) \in \mathcal{P}_p(\mathbb{R}^d \times \mathbb{R}^d)$ be an optimal coupling between $\mu_s(da)$ and $\nu_s(db)$. Similarly, for each $b \in \mathbb{R}^d$, let $\nu_{t-s}^b(de) \in \mathcal{P}_p(\mathbb{R}^d)$ be such that

$$S(t-s)f(b) = \int_{\mathbb{R}^d} f(b+e)\,\nu_{t-s}^b(de) - \varphi_{t-s}(\mathcal{W}_p(\mu_{t-s}, \nu_{t-s}^b))$$

and let $\gamma^b_{t-s}(dc, de) \in \mathcal{P}_p(\mathbb{R}^d \times \mathbb{R}^d)$ be an optimal coupling between $\mu_{t-s}(dc)$ and $\nu^b_{t-s}(de)$. Now define the measure $\gamma_t(dy, dz) \in \mathcal{P}_p(\mathbb{R}^d \times \mathbb{R}^d)$ by

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} h(y,z)\,\gamma_t(dy,dz) := \int_{\mathbb{R}^d\times\mathbb{R}^d}\int_{\mathbb{R}^d\times\mathbb{R}^d} h(a+c, b+e)\,\gamma^b_{t-s}(dc,de)\,\gamma_s(da,db)$$

for all $h\colon \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$ bounded and Borel (we have ignored the fact that $b \mapsto \gamma^b_{t-s}$ needs to be γ_s-measurable for this expression to make sense, but this can be shown by usual measurable selection arguments). Denoting by $\nu_t(dz) := \gamma_t(dz)$ the second marginal of γ_t, it holds

$$S(t)f(x) \geq \int_{\mathbb{R}^d} f(x+z)\,\nu_t(dz) - \varphi_t(\mathcal{W}_p(\mu_t,\nu_t)).$$

Further, $\gamma_t(dy, dz)$ is a coupling between $\mu_t(dy)$ and $\nu_t(dz)$. Indeed, by definition $\gamma_t(dz) = \nu_t(dz)$, and $\gamma_t(dy) = \mu_t(dy)$ as

$$\begin{aligned}\gamma_t(A\times\mathbb{R}^d) &= \int_{\mathbb{R}^d\times\mathbb{R}^d}\int_{\mathbb{R}^d\times\mathbb{R}^d} 1_A(a+c)\,\gamma^b_{t-s}(dc,de)\,\gamma_s(da,db)\\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1_A(a+c)\,\mu_{t-s}(dc)\,\mu_s(da) = (\mu_s * \mu_{t-s})(A) = \mu_t(A)\end{aligned}$$

for every Borel set $A \subset \mathbb{R}^d$. Similarly, we obtain

$$\nu_t(B) = \gamma_t(\mathbb{R}^d\times B) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1_B(b+e)\,\nu^b_{t-s}(de)\,\nu_s(db)$$

for every Borel set $B \subset \mathbb{R}^d$. Moreover, by definition of the p-th Wasserstein distance it holds

$$\begin{aligned}&\mathcal{W}_p(\mu_t,\nu_t) \leq \Big(\int_{\mathbb{R}^d\times\mathbb{R}^d} |y-z|^p\,\gamma_t(dy,dz)\Big)^{1/p}\\ &= \Big(\int_{\mathbb{R}^d\times\mathbb{R}^d}\int_{\mathbb{R}^d\times\mathbb{R}^d} |(b-a)+(e-c)|^p\,\gamma^b_{t-s}(dc,de)\gamma_s(da,db)\Big)^{1/p}\\ &\leq \Big(\int_{\mathbb{R}^d\times\mathbb{R}^d} |b-a|^p\,\gamma_s(da,db)\Big)^{1/p} + \Big(\int_{\mathbb{R}^d\times\mathbb{R}^d}\int_{\mathbb{R}^d\times\mathbb{R}^d} |e-c|^p\,\gamma^b_{t-s}(dc,de)\,\gamma_s(da,db)\Big)^{1/p}.\end{aligned}$$

Denote by I the first term in the above equation and by J the second one. By definition of $\varphi_t = t\varphi(\cdot/t)$, together with the fact that φ is convex and increasing, it holds

$$\begin{aligned}\varphi_t(\mathcal{W}_p(\mu_t,\nu_t)) &\leq t\varphi\Big(\frac{s}{t}\frac{1}{s}I + \frac{t-s}{t}\frac{1}{t-s}J\Big)\\ &\leq s\varphi\Big(\frac{1}{s}I\Big) + (t-s)\varphi\Big(\frac{1}{t-s}J\Big) = \varphi_s(I) + \varphi_{t-s}(J).\end{aligned}$$

Moreover, convexity of $x \mapsto \varphi(x^{1/p})$ implies convexity of $x \mapsto \varphi_{t-s}(x^{1/p})$. Therefore, by Jensen's inequality, we obtain

$$\varphi_{t-s}(J) \leq \int_{\mathbb{R}^d\times\mathbb{R}^d} \varphi_{t-s}\bigg(\Big(\int_{\mathbb{R}^d\times\mathbb{R}^d} |e-c|^p\,\gamma^b_{t-s}(dc,de)\Big)^{1/p}\bigg)\,\gamma_s(da,db).$$

Recalling the definitions of I and J and that γ_s and γ^b_{t-s} are optimal couplings, we conclude

$$\varphi_t(\mathcal{W}_p(\mu_t,\nu_t)) \leq \varphi_s(\mathcal{W}_p(\mu_s,\nu_s)) + \int_{\mathbb{R}^d} \varphi_{t-s}(\mathcal{W}_p(\mu_{t-s},\nu^b_{t-s}))\,\nu_s(db).$$

Putting everything together, we obtain

$$\begin{aligned}S(t)f(x) &\geq \int_{\mathbb{R}^d} f(x+z)\,\nu_t(dz) - \varphi_s(\mathcal{W}_p(\mu_s,\nu_s)) - \int_{\mathbb{R}^d} \varphi_{t-s}(\mathcal{W}_p(\mu_{t-s},\nu_{t-s}^{b+x}))\,\nu_s(db)\\ &= \int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d} f(x+b+e)\,\nu_{t-s}^{b+x}(de) - \varphi_{t-s}(\mathcal{W}_p(\mu_{t-s},\nu_{t-s}^{b+x}))\Big)\,\nu_s(db) - \varphi_s(\mathcal{W}_p(\mu_s,\nu_s))\\ &= S(s)S(t-s)f(x)\end{aligned}$$

This completes the proof. □

Lemma 2.5. *For all $f \in C_0(\mathbb{R}^d)$, $t \geq 0$ and $n \in \mathbb{N}$, we have that $\mathscr{S}^{n+1}(t)f \leq \mathscr{S}^n(t)f$. Further, $\mathscr{S}^n(t)$ is a contraction on $C_0(\mathbb{R}^d)$, and $\mathscr{S}^n(t)f$ has the same modulus of continuity as f.*

Proof. Both statements follow from Lemma 2.4 (respectively Lemma 2.3) together with an induction. □

Corollary 2.6. *Let $t \geq 0$ and $f, g \in C_0(\mathbb{R}^d)$ such that $f \leq g$. Then, the pointwise limit $\mathscr{S}(t)f := \lim_{n\to\infty} \mathscr{S}^n(t)f$ exists and is in fact uniform. Moreover, $\mathscr{S}(t)$ is a convex contraction on $C_0(\mathbb{R}^d)$ such that $\mathscr{S}(t)0 = 0$ and $\mathscr{S}(t)f \leq \mathscr{S}(t)g$.*

Proof. We will use the semigroup $(S^\mu(t))_{t\geq 0}$ corresponding to our initial Lévy process, given by $S^\mu(t)f(x) := \int f(x+y)\,\mu_t(dy)$ and often use that it is a Feller semigroup (see, e.g., [1, Theorem 3.1.9]).

By Lemma 2.5, the sequence $(\mathscr{S}^n(t)f)_{n\in\mathbb{N}}$ is decreasing, hence the limit $\mathscr{S}(t)f = \lim_{n\to\infty} \mathscr{S}^n(t)f$ exists pointwise. Also, the limit $\mathscr{S}(t)f$ is vanishing at infinity. Indeed, from the semigroup property of $(S^\mu(t))_{t\geq 0}$, it follows that $S^\mu(t)f \leq \mathscr{S}^n(t)f \leq S(t)f$ for all $n \in \mathbb{N}$, and therefore $S^\mu(t)f \leq \mathscr{S}(t)f \leq S(t)f$. Since by Lemma 2.3, $S(t)f$ is vanishing at infinity, and $(S^\mu(t))_{t\geq 0}$ is a Feller semigroup, we conclude that $\mathscr{S}(t)f$ is vanishing at infinity. Further, by Lemma 2.5, the sequence $\mathscr{S}^n(t)f$ is uniformly equicontinuous on every compact subset of $\mathbb{R}^d$, which by the Arzelà-Ascoli theorem and the fact that $\mathscr{S}(t)f$ is vanishing at infinity, implies that $\lim_{n\to\infty} \|\mathscr{S}^n(t)f - \mathscr{S}(t)f\|_\infty = 0$.

Finally, by induction over $n \in \mathbb{N}$, it follows from Lemma 2.3 that $\mathscr{S}^n(t)$ is a convex contraction on $C_0(\mathbb{R}^d)$, which satisfies $\mathscr{S}^n(t)0 = 0$ and $\mathscr{S}^n(t)f \leq \mathscr{S}^n(t)g$. These properties remain true for the limit $\mathscr{S}(t)$. The proof is complete. □

Denote by $X_t \sim \mu_t$ the initial Lévy process. In the following, we shall often use that $t \mapsto \mu_t$ is continuous w.r.t. $\mathcal{W}_p$ (at $t = 0$). To see that this is true, use the assumption $E[|X_1|^p] < \infty$ and [24, Theorem 25.18] to obtain $E[\sup_{t\in[0,1]} |X_t|^p] < \infty$. As X has càdlàg paths, dominated convergence implies that $\mathcal{W}_p(\mu_t, \delta_0)^p = \int_{\mathbb{R}^d} |x|^p\,\mu_t(dx) = E[|X_t|^p] \to 0$ as $t \downarrow 0$.

The next result states the strong continuity of the family $(\mathscr{S}(t))_{t\geq 0}$ at zero.

Lemma 2.7. *For every $f \in C_0(\mathbb{R}^d)$, we have that*

$$\lim_{t\downarrow 0} \|\mathscr{S}(t)f - f\|_\infty = 0.$$

Proof. Let $f \in C_0(\mathbb{R}^d)$ and $\varepsilon > 0$.

We first show an upper bound, namely that there is $t_0 > 0$ such that $\mathscr{S}(t)f \leq f + 2\varepsilon$ for all $t < t_0$. As functions in $C_0(\mathbb{R}^d)$ are uniformly continuous, there is $\delta > 0$ such that

$|f(x+y)-f(x)|\le\varepsilon$ for all $x,y\in\mathbb{R}^d$ with $|y|\le\delta$. Then,

$$\begin{aligned}\mathscr{S}(t)f(x)\le S(t)f(x)&=\sup_{\nu\in\Delta_{f,t}}\Big(\int_{\mathbb{R}^d}f(x+y)\nu(dy)-\varphi_t(\mathcal{W}_p(\mu_t,\nu))\Big)\\&\le\sup_{\nu\in\Delta_{f,t}}\Big(\int_{\mathbb{R}^d}f(x+y)1_{\{|y|\le\delta\}}+f(x+y)1_{\{|y|>\delta\}}\,\nu(dy)\Big)\\&\le f(x)+\varepsilon+\|f\|_\infty\sup_{\nu\in\Delta_{f,t}}\nu(\{y\in\mathbb{R}^d:|y|>\delta\}).\end{aligned}\tag{2.3}$$

By Lemma 2.2, it holds that

$$\lim_{t\downarrow0}\sup_{\nu\in\Delta_{f,t}}\mathcal{W}_p(\mu_t,\nu)\le\lim_{t\downarrow0}\sup\big\{r\ge0:\varphi_t(r)\le\|f\|_\infty+1\big\}=0.$$

As argued before this lemma, we have $\lim_{t\downarrow0}\int_{\mathbb{R}^d}|y|^p\,\mu_t(dy)=0$ and thus it follows from Lemma 2.1 that

$$\lim_{t\downarrow0}\sup_{\nu\in\Delta_{f,t}}\nu(\{y\in\mathbb{R}^d:|y|>\delta\})=0,$$

which yields the upper bound.

As for the lower bound, similarly as in the proof of Corollary 2.6, we make use of the fact that $S^\mu\le\mathscr{S}$. Since $(S^\mu(t))_{t\ge0}$ is a Feller semigroup, it holds $f-\varepsilon\le\mathscr{S}(t)f$ for all $0\le t<t_0$ for a suitable $t_0>0$. This completes the proof. □

For later reference, let us point out that the same proof as given for Lemma 2.7 yields the following result.

Corollary 2.8. *We have that* $\lim_{n\to\infty}\|\mathscr{S}^n(t_n)f-f\|_\infty=0$ *for all* $f\in C_0(\mathbb{R}^d)$ *and all sequences* $(t_n)_{n\in\mathbb{N}}$ *in* $[0,\infty)$ *with* $\lim_{n\to\infty}t_n=0$.

Lemma 2.9. *Let* $t,t_n\ge0$ *with* $t_n\in\mathbb{T}_n$*,* $t_n\le t$*, and* $g,g_n\in C_0(\mathbb{R}^d)$ *for all* $n\in\mathbb{N}$*. If* $\lim_{n\to\infty}t_n=t$ *and* $\lim_{n\to\infty}\|g_n-g\|_\infty=0$*, then*

$$\lim_{n\to\infty}\|\mathscr{S}^n(t_n)g_n-\mathscr{S}(t)g\|_\infty\to0.$$

Proof. As $\mathscr{S}^n(t)=\mathscr{S}^n(t_n)\mathscr{S}^n(t-t_n)$ by definition, the triangle inequality implies that

$$\begin{aligned}&\|\mathscr{S}(t)g-\mathscr{S}^n(t_n)g_n\|_\infty\le\|\mathscr{S}(t)g-\mathscr{S}^n(t)g\|_\infty\\&\quad+\|\mathscr{S}^n(t_n)\mathscr{S}^n(t-t_n)g-\mathscr{S}^n(t_n)g\|_\infty+\|\mathscr{S}^n(t_n)g-\mathscr{S}^n(t_n)g_n\|_\infty.\end{aligned}$$

By Corollary 2.6 the first term converges to zero as $n\to\infty$. As for the middle term, by Lemma 2.5 we have that $\mathscr{S}^n(t_n)$ is a contraction, so that

$$\|\mathscr{S}^n(t_n)\mathscr{S}^n(t-t_n)g-\mathscr{S}^n(t_n)g\|_\infty\le\|\mathscr{S}^n(t-t_n)g-g\|_\infty.$$

The latter converges to zero by Corollary 2.8. Again by Lemma 2.5, the last term converges to zero as $n\to\infty$. This completes the proof. □

Now, we are ready to state our first main result (the convex generalization of Proposition 1.1).

Proposition 2.10. *The family* $(\mathscr{S}(t))_{t\ge0}$ *is a strongly continuous, convex, monotone and normalized contraction semigroup on* $C_0(\mathbb{R}^d)$*, i.e., for every* $s,t\ge0$ *and* $f\in C_0(\mathbb{R}^d)$*, we have that*

(i) $\mathscr{S}(t)\colon C_0(\mathbb{R}^d)\to C_0(\mathbb{R}^d)$ *is a convex and monotone contraction such that* $\mathscr{S}(t)0=0$*,*

(ii) $\mathscr{S}(0)f=f$*,*

(iii) $\mathscr{S}(t) \circ \mathscr{S}(s) = \mathscr{S}(t+s)$,

(iv) $\lim_{t\downarrow 0} \|\mathscr{S}(t)f - f\|_\infty = 0$.

Proof. In view of Corollary 2.6 and Lemma 2.7, it remains to prove the semigroup property $\mathscr{S}(t) \circ \mathscr{S}(s) = \mathscr{S}(t+s)$. To that end, fix some $f \in C_0(\mathbb{R}^d)$ and $s, t \geq 0$ and set denote by $s_n, t_n \in \mathbb{T}_n$ the closest dyadic elements prior to s and t, respectively. By Lemma 2.9 (applied with $g = g_n = f$) we have

$$\mathscr{S}(t+s)f = \lim_{n\to\infty} \mathscr{S}^n(t_n + s_n)f = \lim_{n\to\infty} \mathscr{S}^n(t_n) \circ \mathscr{S}^n(s_n)f,$$

where the last equality follows by definition of $\mathscr{S}^n$. Further, Lemma 2.9 also implies that $\mathscr{S}^n(s_n)f$ converges uniformly to $\mathscr{S}(s)f$. Therefore, we may apply Lemma 2.9 again (with $g = \mathscr{S}(s)f$, and $g_n = \mathscr{S}^n(s_n)f$) and obtain

$$\lim_{n\to\infty} \mathscr{S}^n(t_n) \circ \mathscr{S}^n(s_n)f = \mathscr{S}(t) \circ \mathscr{S}(s)f.$$

This completes the proof. □

Proposition 2.11. *For every $f \in D(A^\mu) \cap C_0^1(\mathbb{R}^d)$, we have that*

$$\mathscr{A}f := \lim_{t\downarrow 0} \frac{\mathscr{S}(t)f - f}{t} = A^\mu f + \varphi^*(|\nabla f|)$$

and the limit is uniform.

Proof. Fix $f \in D(A^\mu) \cap C_0^1(\mathbb{R}^d)$.

(a) We start by showing that

$$\mathscr{S}(t)f - f \geq tA^\mu f + t\varphi^*(|\nabla f|) + o(t) \quad \text{as } t \downarrow 0. \tag{2.4}$$

To that end, let $t > 0$. For notational simplicity we assume that t is a dyadic number, say $t = k_0 2^{-n_0}$ for some $k_0, n_0 \in \mathbb{N}$; the general case (is only notationally heavier but) works analogously. Then, $\mathscr{S}^{n_0}(t)$ is just the convolution of $S(2^{-n_0})$ with itself k_0 times. For every $x \in \mathbb{R}^d$, let $r = r_x \in \mathbb{R}^d$ with $|r| = 1$ and $a = a_x \geq 0$ be such that

$$r\nabla f(x) = |\nabla f(x)| \quad \text{and} \quad \varphi^*(|\nabla f(x)|) = a|\nabla f(x)| - \varphi(a), \tag{2.5}$$

where the product between elements in $\mathbb{R}^d$ is understood as the scalar product. Note that such r exists as $|\cdot|$ is its own dual norm, and such a exists as $\lim_{y\to\infty} \varphi(y)/y = \infty$ which follows from the assumption that $y \mapsto \varphi(y^{1/p})$ is convex. Moreover, since $|\nabla f(x)|$ is uniformly bounded over $x \in \mathbb{R}^d$, the same holds for $a = a_x$.

Now, for each $n \geq n_0$, set

$$\nu^n_{2^{-n}} := \mu_{2^{-n}} * \delta_{a2^{-n}r}.$$

Then, one can compute that $\mathcal{W}_p(\mu_{2^{-n}}, \nu_{2^{-n}}) = a2^{-n}$, and therefore

$$S(2^{-n})g(y) \geq \int_{\mathbb{R}^d} g(y + z + a2^{-n}r)\, \mu_{2^{-n}}(dz) - \varphi_{2^{-n}}(a2^{-n}) \tag{2.6}$$

for all $g \in C_0(\mathbb{R}^d)$ and $y \in \mathbb{R}^d$. Note that $t = k_n 2^{-n}$ for $k_n := k_0 2^{n-n_0}$, and that the measure which results in taking the convolution of $\nu_{2^{-n}}$ with itself k_n times, is equal to $\mu_t * \delta_{atr}$. As further,

$$k_n \varphi_{2^{-n}}(a2^{-n}) = k_0 2^{n-n_0} 2^{-n} \varphi(a) = t\varphi(a) = \varphi_t(at)$$

and each $\varphi_{2^{-n}}(a2^{-n})$ does not depend on the state variable, estimating every $S(2^{-n})$ which appears in the definition of $\mathscr{S}^n(t)$ (as the convolution of $S(2^{-n})$ with itself k_n times) by (2.6) gives

$$\mathscr{S}^n(t)f(x) \geq \int_{\mathbb{R}^d} f(x+y+atr)\,\mu_t(dy) - \varphi_t(at)$$

for all $n \geq n_0$. The right hand side does not depend on n, so that the definition of $\mathscr{S}(t)f$ as the limit of $\mathscr{S}^n(t)f$ therefore implies that

$$\begin{aligned}\mathscr{S}(t)f(x) - f(x) \geq &\int_{\mathbb{R}^d} f(x+y) - f(x)\,\mu_t(dy)\\ &+ \int_{\mathbb{R}^d} f(x+y+atr) - f(x+y)\,\mu_t(dy) - \varphi_t(at) =: I_1 + I_2.\end{aligned}$$

By definition of the infinitesimal generator A^μ of $(S^\mu(t))_{t\geq 0}$ and $f \in D(A^\mu)$, the first term I_1 equals $tA^\mu f + o(t)$ (uniformly over $x \in \mathbb{R}^d$). The second term I_2 is estimated by a Taylor's expansion: for some (measurable) $\xi = \xi(x,y)$ with $|\xi| \leq ta$, we may write

$$\begin{aligned}I_2 &= \int_{\mathbb{R}^d} atr\nabla f(x+y+\xi(x,y))\,\mu_t(dy) - t\varphi(a)\\ &\geq ar\nabla f(x) + o(1) - t\varphi(a),\end{aligned}$$

uniformly over $x \in \mathbb{R}^d$, where we need to justify the last step. Indeed, this follows as in the proof of Lemma 2.7 by splitting the $\mu_t(dy)$ integral into two parts (close to zero $\{|y| \leq b\}$ and its complement $\{|y| > b\}$), and using uniform continuity of $ar\nabla f(x+\cdot)$ together with the fact that $\mu_t(\{|y| > b\}) \to 0$ and $\lim_{t\downarrow 0}\sup_{x,y\in\mathbb{R}^d}|\xi(x,y)| = 0$ as $a = a_x$ is bounded uniformly over $x \in \mathbb{R}^d$. Recalling (2.5) we conclude that

$$\int_{\mathbb{R}^d} f(x+y+atr) - f(x+y)\,\mu_t(dy) - \varphi_t(at) \geq t\varphi^*(|\nabla f(x)|) + o(t)$$

uniformly over $x \in \mathbb{R}^d$, which shows (2.4).

(b) It remains to show that

$$\mathscr{S}(t)f - f \leq tA^\mu f + t\varphi^*(|\nabla f|) + o(t). \tag{2.7}$$

Since $\int_{\mathbb{R}^d} f(x+y)\,\mu_t(dy) - f(x) = tA^\mu f(x) + o(t)$ as $f \in D(A^\mu)$ and $\mathscr{S}(t)f \leq S(t)f$ by Corollary 2.6, it holds

$$\begin{aligned}\mathscr{S}(t)f(x) - f(x) &\leq S(t)f(x) - \int_{\mathbb{R}^d} f(x+y)\,\mu_t(dy) + tAf(x) + o(t)\\ &= \sup_{u,\nu}\Big(\int_{\mathbb{R}^d} f(x+z)\,\nu(dz) - \int_{\mathbb{R}^d} f(x+y)\,\mu_t(dy) - t\varphi\big(\tfrac{u}{t}\big)\Big) + tAf(x) + o(t)\end{aligned}$$

uniformly over $x \in \mathbb{R}^d$, where the supremum is taken over all $u \geq 0$ and $\nu \in \mathcal{P}_p(\mathbb{R}^d)$ with $\mathcal{W}_p(\mu_t,\nu) = u$. Actually, for every $t \geq 0$, one may restrict to those $u \geq 0$ for which $t\varphi(u/t) \leq \|f\|_\infty + 1$. As φ grows faster than linear, this implies that there is some u_0 (independent of t) for which the latter implies $u \leq u_0 t$.

Now, fix $0 \leq u \leq u_0 t$ and $\nu \in \mathcal{P}_p(\mathbb{R}^d)$ with $\mathcal{W}_p(\mu_t,\nu) = u$, and a coupling $\pi(dy,dz)$ between μ_t and ν which is optimal for $\mathcal{W}_p(\mu_t,\nu)$. By Taylor's theorem,

$$f(x+z) - f(x+y) = \nabla f(x+y+\xi)(z-y)$$

for all $x,y,z\in\mathbb{R}^d$, where $\xi=\xi(x,y,z)$ is a measurable function such that $|\xi|\leq|z-y|$. Hence, it follows from Hölder's inequality that

$$\begin{aligned}&\int_{\mathbb{R}^d} f(x+z)\,\nu(dz)-\int_{\mathbb{R}^d} f(x+y)\,\mu_t(dy)=\int_{\mathbb{R}^d}\nabla f(x+\xi)(z-y)\,\pi(dy,dz)\\&\leq\Big(\int_{\mathbb{R}^d\times\mathbb{R}^d}|\nabla f(x+\xi)|^{p^*}\,\pi(dy,dz)\Big)^{1/p^*}\Big(\int_{\mathbb{R}^d\times\mathbb{R}^d}|z-y|^p\,\pi(dy,dz)\Big)^{1/p},\end{aligned}$$

where $p^*=p/(p-1)$ is the conjugate Hölder exponent of p. For every $0\leq u\leq u_0t$ and ν as above, it follows from $\mathcal{W}_p(\delta_0,\nu)\leq\mathcal{W}_p(\delta_0,\mu_t)+u_0t=o(1)$, that

$$\Big(\int_{\mathbb{R}^d\times\mathbb{R}^d}|\nabla f(x+\xi)|^{p^*}\,\pi(dy,dz)\Big)^{1/p^*}\leq|\nabla f(x)|+o(1)$$

uniformly over $x\in\mathbb{R}^d$, again by the same arguments as in the proof of Lemma 2.7. Putting everything together yields

$$\begin{aligned}&\frac{1}{t}\Big(S(t)f(x)-\int_{\mathbb{R}^d} f(x+y)\,\mu_t(dy)\Big)\\&\leq\sup_{0\leq u\leq u_0t}\Big(\frac{u}{t}(|\nabla f(x)|+o(1))-\varphi\Big(\frac{u}{t}\Big)\Big)\leq\varphi^*(|\nabla f(x)|)\end{aligned}$$

uniformly over $x\in\mathbb{R}^d$, where the last inequality follows from the definition of the convex conjugate φ^*. This shows (2.7) and therefore completes the proof. □

This is a good place to mention the recent paper [3] in which related ideas as in Proposition 2.11 are applied in the context of stochastic optimization. With Proposition 2.11 at our disposal, we can finally prove our main result (Theorem 1.2), or rather its convex generalization (Theorem 2.12 below).

Before doing so, let us recall the notion of viscosity solution that we use: denote by $C_0^1(\mathbb{R}^d)$ the space of all continuously differentiable functions $f\in C_0(\mathbb{R}^d)$ whose gradient is vanishing at infinity, and call $v\colon(0,\infty)\to C_0(\mathbb{R}^d)$ *test function* if it is differentiable (w.r.t. the supremum norm) and satisfies $v(t)\in D(A^\mu)\cap C_0^1(\mathbb{R}^d)$ for every $t\in(0,\infty)$.

Then, following [9], we say that a continuous function $u:[0,\infty)\to C_0(\mathbb{R}^d)$ is a *viscosity subsolution* of

$$\begin{cases}\partial_t u(t,x)=\mathscr{A}u(t,x) & \text{for } (t,x)\in(0,\infty)\times\mathbb{R}^d,\\ u(0,x)=f(x) & \text{for } x\in\mathbb{R}^d,\end{cases}$$

if for every $(t,x)\in(0,\infty)\times\mathbb{R}^d$ and every test function v satisfying $u\leq v$ and $v(t,x)=u(t,x)$, it holds that $\partial_t v(t,x)\leq\mathscr{A}v(t,x)$. Similarly, u is called *viscosity supersolution* if the above holds with '$\leq$' replaced by '$\geq$' at both instances, and a *viscosity solution* if it is both a viscosity supersolution and subsolution.

As a consequence of the previous result we derive the following:

Theorem 2.12. *Let $f\in C_0(\mathbb{R}^d)$ and define $u\colon[0,\infty)\times\mathbb{R}^d\to\mathbb{R}$ via $u(t,x):=\mathscr{S}(t)f(x)$. Then u is a viscosity solution of*

$$\begin{cases}\partial_t u(t,x)=A^\mu u(t,x)+\varphi^*(|\nabla u(t,x)|) & \text{for } (t,x)\in(0,\infty)\times\mathbb{R}^d,\\ u(0,x)=f(x) & \text{for } x\in\mathbb{R}^d.\end{cases}\tag{2.8}$$

Proof. To show that u is a viscosity subsolution, let v be a test function such that $u\leq v$ and $v(t,x)=u(t,x)$ for some $(t,x)\in(0,\infty)\times\mathbb{R}^d$. Since v is differentiable at t there exists

$\partial_t v(t) \in C_0(\mathbb{R}^d)$ such that $v(t+h) = v(t) + h\partial_t v(t) + o(h)$ for $|h| \to 0$. Similar to the proof of Lemma 2.7 it follows that

$$\mathscr{S}(h)\big(v(t) - h\partial_t v(t) + o(h)\big) - S(h)(v(t)) = -h\partial_t v(t) + o(h)$$

for $|h| \to 0$. Hence, for $h > 0$ small enough, using Proposition 2.10, we have that

$$\begin{aligned} 0 &= \frac{\mathscr{S}(h)\mathscr{S}(t-h)f - \mathscr{S}(t)f}{h} = \frac{\mathscr{S}(h)u(t-h) - u(t)}{h} \le \frac{\mathscr{S}(h)v(t-h) - u(t)}{h} \\ &= \frac{\mathscr{S}(h)\big(v(t) - h\partial_t v(t) + o(h)\big) - \mathscr{S}(h)v(t)}{h} + \frac{\mathscr{S}(h)v(t) - u(t)}{h} \\ &= -\partial_t v(t) + \frac{\mathscr{S}(h)v(t) - v(t)}{h} + \frac{v(t) - u(t)}{h} + o(h). \end{aligned}$$

In particular, since $h^{-1}(\mathscr{S}(h)v(t,x) - v(t,x)) \to \mathscr{A}v(t,x)$ uniformly over $x \in \mathbb{R}^d$ by Proposition 2.11, and $v(t,x) = u(t,x)$ by assumption, we conclude that $0 \le -\partial_t v(t,x) + \mathscr{A}v(t,x)$. This shows that u is a viscosity subsolution of (2.8). The arguments that u is a viscosity supersolution follows along the same lines successively replacing '$\le$' by '$\ge$'. $\square$

Finally, we sketch how uniqueness of the solution of (2.8) in Theorem 2.12 may be obtained. Under certain conditions on the initial Lévy process, one obtains from [14, Corollary 53] uniqueness of the viscosity solution (2.8) by using the space $C_b^{2,3}((0,\infty) \times \mathbb{R}^d)$ as test functions. This requires an extension of the semigroup $(\mathscr{S}(t))_{t\ge 0}$ to the space $\mathrm{BUC}(\mathbb{R}^d)$ of all bounded and uniformly continuous functions, which may be achieved via monotone approximation and continuity from above of the operators $\mathscr{S}(t)$, $t \ge 0$, see also [8, Remark 5.4] Then, by adapting Proposition 2.11, it follows that Theorem 2.12 also holds for the test functions $v\colon (0,\infty) \to \mathrm{BUC}(\mathbb{R}^d)$ which are differentiable and $v(t) \in \mathrm{BUC}^2(\mathbb{R}^d)$ for all $t \ge 0$, where $\mathrm{BUC}^2(\mathbb{R}^d)$ denotes the space of all functions which are twice differentiable with bounded uniformly continuous derivatives up to order 2. Once this is done, the results in [14] may be used, since $C_b^{2,3}((0,\infty) \times \mathbb{R}^d)$ is a subset of the considered test functions, see [9, Remark 2.7].

References

[1] D. Applebaum. Lévy processes and stochastic calculus. *Cambridge university press*, 2009. MR-2512800

[2] D. Bartl. Exponential utility maximization under model uncertainty for unbounded endowments. *The Annals of Applied Probability*, 29(1):577–612, 2019. MR-3910012

[3] D. Bartl, S. Drapeau, J. Obłój, and J. Wiesel. Robust uncertainty sensitivity analysis. *arXiv preprint arXiv:2006.12022*, 2020.

[4] D. Bartl, S. Drapeau, and L. Tangpi. Computational aspects of robust optimized certainty equivalents and option pricing. *Mathematical Finance*, 30(1):287–309, 2020. MR-4067077

[5] C. Beck, E. Weinan, and A. Jentzen. Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. *Journal of Nonlinear Science*, 29(4):1563–1619, 2019. MR-3993178

[6] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *The Annals of Probability*, 44(1):42–106, 2016. MR-3456332

[7] J. Blanchet and K. Murthy. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600, 2019. MR-3959085

[8] R. Denk, M. Kupper, and M. Nendel. Kolmogorov-type and general extension results for nonlinear expectations. *Banach Journal of Mathematical Analysis*, 12(3):515–540, 2018. MR-3824739

[9] R. Denk, M. Kupper, and M. Nendel. A semigroup approach to nonlinear Lévy processes. *Stochastic Processes and their Applications*, 130(3):1616–1642, 2020. MR-4058284

[10] Y. Dolinsky, M. Nutz, and H. M. Soner. Weak approximation of G-expectations. *Stochastic Processes and their Applications*, 122(2):664–675, 2012. MR-2868935

[11] S. Eckstein. Extended laplace principle for empirical measures of a Markov chain. *Advances in Applied Probability*, 51(1):136–167, 2019. MR-3984013

[12] P. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166, 2018. MR-3844536

[13] Y. Feng, R. Rudd, C. Baker, Q. Mashalaba, M. Mavuso, and E. Schlögl. Quantifying the model risk inherent in the calibration and recalibration of option pricing models. *Available at SSRN 3267775*, 2018.

[14] M. Hu and S. Peng. G-Lévy processes under sublinear expectations. *arXiv preprint arXiv:0911.3533*, 2009.

[15] N. Kazi-Tani, D. Possamaï, Dylan and C. Zhou. Quadratic BSDEs with jumps: related nonlinear expectations *Stochastics and Dynamics*, 16(4), 2016. MR-3494687

[16] F. Kühn. Viscosity solutions to Hamilton-Jacobi-Bellman equations associated with sublinear Lévy (-type) processes. *Latin American Journal of Probability and Mathematical Statistics*, 16(1):531–559, 2019. MR-3941868

[17] J. Liu and S. Yao. Jump-filtration consistent nonlinear expectations with $\mathbb{L}^p$ domains. *Applied Mathematics & Optimization*, 79(1):87–129, 2019. MR-3903779

[18] A. Neufeld and M. Nutz. Nonlinear Lévy processes and their characteristics. *Transactions of the American Mathematical Society*, 369(1):69–95, 2017. MR-3557768

[19] S. Peng Backward SDE and Related g-Expectation. Pitman Research Notes in Mathematics Series. Longman, Harlow, 1997. MR-1752680

[20] S. Peng. G-expectation, G-Brownian motion and related stochastic calculus of Itô type. *Stochastic analysis and applications*, pages 541–567, 2007. MR-2397805

[21] A. Pichler and R. Schlotter. Quantification of risk in classical models of finance. *arXiv preprint arXiv:2004.04397*, 2020.

[22] M. Raissi. Deep hidden physics models: Deep learning of nonlinear partial differential equations. *The Journal of Machine Learning Research*, 19(1):932–955, 2018. MR-3862432

[23] M. Royer. Backward stochastic differential equations with jumps and related non-linear expectations. *Stochastic processes and their applications*, 116(10):1358–1676, 2006. MR-2260739

[24] K.-I. Sato. *Lévy processes and infinitely divisible distributions*. Cambridge university press, 1999. MR-1739520

[25] D. Škulj. Discrete time Markov chains with interval probabilities. *International journal of approximate reasoning*, 50(8):1314–1329, 2009. MR-2556123

[26] W. Wiesemann, D. Kuhn, and B. Rustem. Robust Markov decision processes. *Mathematics of Operations Research*, 38(1):153–183, 2013. MR-3029483

[27] I. Yang. A convex optimization approach to distributionally robust Markov decision processes with wasserstein distance. *IEEE control systems letters*, 1(1):164–169, 2017. MR-4208527

Acknowledgments. We thank the editor and an anonymous referee for valuable comments and feedback on an earlier version of the paper.