NONLINEAR SEMIGROUPS AND LIMIT THEOREMS FOR
CONVEX EXPECTATIONS

JONAS BLESSING\textsuperscript{*1} AND MICHAEL KUPPER\textsuperscript{**2}

Abstract. Based on the Chernoff approximation, we provide a general approximation result for convex monotone semigroups which are continuous w.r.t. the mixed topology on suitable spaces of continuous functions. Starting with a family \((I(t))_{t \geq 0}\) of operators, the semigroup is constructed as the limit \(S(t)f := \lim_{n \to \infty} I(t/n)^n f\) and is uniquely determined by the time derivative \(I'(0)f\) for smooth functions. We identify explicit conditions for the generating family \((I(t))_{t \geq 0}\) that are transferred to the semigroup \((S(t))_{t \geq 0}\) and can easily be verified in applications. Furthermore, there is a structural link between Chernoff type approximations for nonlinear semigroups and law of large numbers and central limit theorem type results for convex expectations. The framework also includes large deviation results.

Key words: Nonlinear semigroup, Chernoff approximation, comparison principle, convex expectation, robust limit theorem, G-distribution, large deviations.

MSC 2020: Primary 47H20; 60F05; Secondary 47J25; 60F10; 60G50.

1. Introduction

In this article, we establish a general approximation result for convex monotone semigroups which is based on the Chernoff approximation, see [9,11,12], that generalizes the Trotter–Kato product formula for linear semigroups, see [31,43]. The idea is to start with a generating family \((I(t))_{t \geq 0}\) of operators \(I(t) \colon C_b \to C_b\) which do not form a semigroup but have, for smooth functions, the derivative \(I'(0)f := \lim_{h \downarrow 0} \frac{I(h)f - f}{h} \in C_b\).

Here, the space \(C_b\) consists of all bounded continuous functions \(f : \mathbb{R}^d \to \mathbb{R}\). In order to obtain a corresponding semigroup \((S(t))_{t \geq 0}\) on \(C_b\) with prescribed generator \(Af = I'(0)f\), we iterate the operators \((I(t))_{t \geq 0}\) over a sequence of equidistant partitions with mesh size tending to zero and define the semigroup as the limit

\[
S(t)f := \lim_{n \to \infty} \left( I\left( \frac{t}{n} \right) \right)^n f \in C_b.
\]

Recall that in the case of the Trotter–Kato formula the choice \(I(t) := e^{tA_1}e^{tA_2}\) leads to a semigroup with generator \(A = A_1 + A_2\). There are two major questions to address.

First, we need a suitable topology for the convergence in equation (1.1). For Nisio semigroups, see [16,36,37], it follows from \(I(s + t)f \leq I(s)I(t)f\) for all \(s, t \geq 0\) and \(f \in C_b\) that the sequence \((I(2^{-n}t)2^n f)_{n \in \mathbb{N}}\) is non-decreasing and the semigroup can be
defined pointwise as the monotone limit $S(t)f := \sup_{n \in \mathbb{N}} I(2^{-n}t)2^n f$. Typical examples have the form

$$I(t)f := \sup_{\lambda \in \Lambda} (S_{\lambda}(t)f - \varphi(\lambda)t),$$

for a family $(S_{\lambda})_{\lambda \in \Lambda}$ of linear semigroups $(S_{\lambda}(t))_{t \geq 0}$ and a function $\varphi: \Lambda \to [0, \infty]$. It is also possible to define the semigroup as the limit of a decreasing sequence, see [3,25], where the authors study semigroups corresponding to Markov processes with uncertain transition probabilities. While Chernoff’s original work requires convergence w.r.t. a metric, convergence w.r.t. the supremum norm cannot be expected in the present setting. On the other hand, it has been recently shown in [26] that the non-metrizable mixed topology between the supremum norm and the topology of uniform convergence of compacts is suitable to study (linear) semigroups and characterize Markov processes via their infinitesimal generators. Note that a sequence $(f_n)_{n \in \mathbb{N}}$ converges in the mixed topology if and only if it is bounded w.r.t. the (weighted) supremum norm and converges uniformly on compacts. In particular, by Dini’s theorem, the previously mentioned monotone convergence implies convergence in the mixed topology. Moreover, compactness in the mixed topology can be characterized by Arzéla-Ascoli’s theorem, i.e., equicontinuity of the sequence $(I(t/n)^nf)_{n \in \mathbb{N}}$ rather than monotonicity. For the intended application of the approximation result to limit theorems for convex expectations, it can not be guaranteed that the sequence $(I(t/n)^nf)_{n \in \mathbb{N}}$ is non-decreasing.

Second, since the convergence in equation (1.1) is based on relative compactness, we obtain a priori only convergence for a subsequence, i.e.,

$$S(t)f := \lim_{l \to \infty} I\left(\frac{l}{m}\right)^nf. \tag{1.2}$$

Following the arguments in [7,9], one can only show that the convergent subsequence can be chosen independent of $f \in C_b$ and $t \in \mathcal{T}$, where $\mathcal{T} \subset [0, \infty)$ is countable and dense. However, if the semigroup $(S(t))_{t \geq 0}$ was uniquely determined by its infinitesimal generator $Af = I'(0)f$, one could argue that the limit in equation (1.2) does not depend on the choice of the convergent subsequence and thus obtain the desired convergence in equation (1.1). In view of [26, Theorem 6.2], a possible approach to show that the semigroup is uniquely determined by its generator could rely on comparison principles for viscosity solutions of fully nonlinear equations, see, e.g., [13,20]. In this article, we use instead the comparison principle from [7] for strongly continuous convex monotone semigroups, where the generator is defined as a limit w.r.t. $\Gamma$-convergence for functions in the upper Lipschitz set. It is important to note that the upper Lipschitz set is larger than the domain of the generator defined w.r.t. the mixed topology. In general, the latter is not invariant under the semigroup and the classical theory of m-accretive operators cannot be applied, see [7] for a detailed discussion.

By the comparison principle in [7], two strongly continuous convex monotone semigroups which are uniformly continuous w.r.t. the mixed topology coincide if they have the same upper Lipschitz set and their generators evaluated at smooth functions coincide. While the latter property can often be verified by straightforward computations, it remained an open problem how to show equality of the upper Lipschitz sets even in the case where the two semigroups are obtained by choosing different convergent subsequences in equation (1.2). Hence, the first main contribution of this article is to provide explicit conditions under which two semigroups whose generators evaluated at smooth functions coincide also have the same upper Lipschitz set, see Theorem 2.5. By applying this result on semigroups which are obtained from choosing different convergent subsequences in equation (1.2), we obtain that the limit does not depend on the choice.
of the convergent subsequence. Hence, the desired convergence in equation (1.1) follows and the semigroup \((S(t))_{t \geq 0}\) is uniquely determined by the derivative \(I'(0)f\) for smooth functions \(f\), see Theorem 2.9 and Corollary 2.10. The conditions required in Assumption 2.7 are explicit and can be verified for a variety of examples by straightforward computations.

In Section 3 and Section 4, we show that there is a structural link between Chernoff type approximations for convex monotone semigroups of the form (1.1) and law of large numbers (LLN) and central limit theorem (CLT) type results for convex expectations. The following diagram depicts this for the classical CLT: for every iid sequence \((\xi_n)_{n \in \mathbb{N}}\) with finite second moments,

\[
\mathbb{E} \left[ f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \right) \right] = (I(\frac{1}{n})^n)(0) \quad \text{CLT}
\]

\[
\int_{\mathbb{R}^d} f(y) \mathcal{N}(0, 1)(dy) = (S(1)f)(0), \quad \text{Chernoff}
\]

where \((I(t)f)(x) := \mathbb{E}[f(x + \sqrt{t}\xi_1)]\) and \((S(t)f)(x) := \int_{\mathbb{R}^d} f(x + y) \mathcal{N}(0, t)(dy)\) for all \(t \geq 0\), \(f \in C_b\) and \(x \in \mathbb{R}^d\). While in the linear case it is natural to use the convergence on the left-hand side to obtain the convergence on the right-hand side, in the nonlinear case we will use Theorem 2.9 to show the reverse implication instead. To be precise: we will replace the linear expectation \(\mathbb{E}[\cdot]\) by a convex expectation \(\mathcal{E}[\cdot]\) and show that

\[
(S(1)f)(0) = \lim_{n \to \infty} \frac{1}{n} \mathcal{E} \left[ nf \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \right) \right] \quad \text{for all } f \in C_b.
\]

Furthermore, the generator of \((S(t))_{t \geq 0}\) given by

\[
(Af)(x) = \mathcal{E} \left[ \frac{1}{2} \xi_1^T D^2 f(x) \xi_1 \right] \quad \text{for all } f \in C^2_b \text{ and } x \in \mathbb{R}^d
\]

uniquely characterizes the semigroup, i.e., the limit is \(G\)-distributed with

\[
G: \mathbb{S}^d \to \mathbb{R}, \ a \mapsto \mathcal{E} \left[ \frac{1}{2} \xi_1^T a \xi_1 \right].
\]

This result is stated in Theorem 4.1 and is illustrated by an application to uncertain samples in Wasserstein spaces, see Theorem 4.3 and Theorem 4.4. For a brief overview on convex expectations, we refer to Appendix B. To the best of our knowledge, this is also the first result of this kind for convex rather than sublinear expectations. In the sublinear case, Peng [38] introduced the \(G\)-distribution by

\[
F_G: \text{Lip}_b \to \mathbb{R}, \ f \mapsto u^f(1, 0),
\]

where \(u^f\) denotes the unique viscosity solution of the fully nonlinear PDE

\[
\begin{cases}
\partial_t u(t, x) = G(D^2 u(x)), \quad (t, x) \in (0, \infty) \times \mathbb{R}^d, \\
u(0, x) = f(x), \quad x \in \mathbb{R}^d
\end{cases}
\]

with \(G(a) := \frac{1}{2} \sup_{\lambda \in \Lambda} \text{tr}(\lambda \lambda^T a)\) for a bounded closed non-empty set \(\Lambda \subset \mathbb{R}^{d \times d}\). The \(G\)-distribution is interpreted as a normal distribution with uncertain covariance and the corresponding CLT was proved in [39–41]. We also want to mention extensions to Lévy processes, see [4,28], and results regarding convergence rates, see [27,29,32,42]. Note that the proofs given in [39,41] rely heavily on a deep interior estimate of a fully nonlinear PDE while the more probabilistic approach in [40] based on tightness and
weak compactness requires an additional moment condition. Since [26, Theorem 6.2] guarantees $F_c(f) = (S(1)f)(0)$ for all $f \in \text{Lip}_b$, Theorem 4.1 is consistent with previous results for the sublinear expectations. Moreover, the proof of Theorem 4.1 resembles the approach in [40], but covers the sublinear case under the more natural moment condition $\lim_{c \to \infty} \mathcal{E}[|\xi_1^2 - c|^+] = 0$ from [41].

Similarly, we can obtain LLN type results of the form

$$(S(1)f)(0) = \lim_{n \to \infty} \frac{1}{n} \mathcal{E} \left[ n f \left( \frac{1}{n} \sum_{i=1}^{n} \xi_i \right) \right]$$

for all $f \in \text{C}_b$, where the generator is given by $(Af)(x) = \mathcal{E}[\nabla f(x)\xi_1]$ for all $f \in \text{C}_b^1$ and $x \in \mathbb{R}^d$. This result is stated in Theorem 3.2 and Theorem 3.4 shows that the limit is maximally distributed, i.e.,

$$(S(1)f)(0) = \sup_{x \in \mathbb{R}^d} (f(x) - \varphi(y)) \quad \text{for all } f \in \text{C}_b,$$

where $\varphi(y) := \sup_{z \in \mathbb{R}^d} (yz - \mathcal{E}[z\xi])$ and $\xi := \text{id}_{\mathbb{R}^d}$. In the sublinear case, this result goes back to Peng [39, 41] who introduced the maximal distribution by

$$F_\Lambda : \text{Lip}_b \to \mathbb{R}, \ f \mapsto \sup_{\lambda \in \Lambda} f(\lambda)$$

for a bounded closed non-empty set $\Lambda \subset \mathbb{R}^d$. There are two major differences between the framework of this article and previous works. First, we consider a sequence $(\psi_n)_{n \in \mathbb{N}}$ of recursively defined functions and are interested in the limit behaviour of

$$X_n := \psi_n \left( \frac{1}{n}, 0, \xi_1, \ldots, \xi_n \right).$$

While the choice $\psi_n \left( \frac{1}{n}, 0, \xi_1, \ldots, \xi_n \right) := \frac{1}{n} \sum_{i=1}^{n} \xi_i$ is clearly admissible, the sample $\xi_{n+1}$ can also be randomly shifted by a nonlinear function depending on the average of the previous samples $\xi_1, \ldots, \xi_n$, see Corollary 3.3. Second, our results are not restricted to sublinear expectations and therefore create a uniform framework to cover both LLN type results for samples with uncertain distribution and large deviation results. For instance, Cramér’s theorem can be seen as LLN for the entropic risk measure, see Corollary 3.5. In view of this connection, the previously possibly unexpected scaling $\frac{1}{n} \mathcal{E}[n f(X_n)]$ instead of $\mathcal{E}[f(X_n)]$ suddenly appears to be very natural. We further extend Cramér’s theorem to samples that are perturbed by a nonlinear function as in Corollary 3.3. In this case, the asymptotic convergence rate is lowered according to the size of perturbation, see Theorem 3.7. Finally, we pick up the setting from Lacker [34], based on the weak convergence approach in [17], who has previously worked in a similar setting with convex expectations and established a non-exponential extension of Sanov’s theorem. This leads to polynomial convergence rates that only require the existence of finite $p$-moments, see Theorem 3.8. Large deviation theory based on suitable classes of risk measures has recently also been explored by several other authors, see, e.g., [2, 18, 21, 33]. The abstract results are again illustrated by uncertain samples in Wasserstein spaces, see Theorem 3.9. Finally, we want to mention related LLN type results for capacities, see [35], and for coherent lower previsions, see [14].

The large deviation results obtained in Section 3 show that the theory of nonlinear semigroups developed in Section 2 and the previous works [7, 9] can be used to obtain asymptotic convergence rates. Furthermore, it is even possible to derive non-asymptotic rates for the limit in equation (1.1) which yields error bounds for the LLN and CLT type results in Section 3 and Section 4. These results require additional moment conditions
and are consistent with the ones in [27,29,32,42] for sublinear expectations. For details, we refer to [8].

2. Convex monotone semigroups

The results in this section rely strongly on the previous works [7,9] which have, despite their generality, two major drawbacks. First, we only obtain a convergent subsequence that exists due to a relative compactness argument. Furthermore, the comparison principle in [7] uniquely characterizes semigroups by their upper $\Gamma$-generator defined on their upper Lipschitz set. This theoretical result is an analogue to the classical statement that strongly continuous linear semigroups are uniquely determined by their generators evaluated at smooth functions. This resembles the concept of a core in the theory of linear semigroups. While [7, Section 3] provides results that allow to approximate the $\Gamma$-generator with smooth functions, the question whether two semigroups have the same upper Lipschitz set remained open. Theorem 2.5 now shows that, under conditions which are consistent with the framework in [7], two semigroups whose generators evaluated at smooth functions coincide have the same upper Lipschitz set. In view of the results in [7] these semigroups are then uniquely determined by their generators evaluated at smooth functions, see Theorem 2.6. Moreover, we provide sufficient conditions on the family $(I(t))_{t \geq 0}$ that transfer to the corresponding semigroup given by equation (1.2) so that Theorem 2.6 applies and equation (1.1) is valid. Throughout this article, we consider semigroups that are defined on spaces of continuous functions which do not exceed a certain growth rate at infinity. For that purpose, let $\kappa: \mathbb{R}^d \to (0, \infty)$ be a bounded continuous function with

$$
c_\kappa := \sup_{x \in \mathbb{R}^d} \sup_{|y| \leq 1} \frac{\kappa(x)}{\kappa(x - y)} < \infty
$$

and denote by $C_\kappa$ the space of all continuous functions $f: \mathbb{R}^d \to \mathbb{R}$ with

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

In Section 3 and Section 4, we will choose $\kappa(x) := (1 + |x|^p)^{-1}$ for $p = 1$ and $p = 2$, respectively. Moreover, we endow $C_\kappa$ with the mixed topology between $\| \cdot \|_\kappa$ and the topology of uniform convergence on compacts sets. It is well-known, see [26, Proposition A.4], that a sequence $(f_n)_{n \in \mathbb{N}} \subset C_\kappa$ converges to $f \in C_\kappa$ w.r.t. the mixed topology if and only if

$$
\sup_{n \in \mathbb{N}} \|f_n\|_\kappa < \infty \quad \text{and} \quad \lim_{n \to \infty} \|f - f_n\|_{\infty,K} = 0
$$

for all compact subsets $K \subset \mathbb{R}^d$, where $\|f\|_{\infty,K} := \sup_{x \in K} |f(x)|$. Similarly, for a family $(f_h)_{h > 0} \subset C_\kappa$ and $f \in C_\kappa$, it holds that $f = \lim_{h \downarrow 0} f_h$ w.r.t. the mixed topology if and only if $\sup_{h \in (0,h_0]} \|f_h\|_\infty < \infty$ for some $h_0 > 0$ and, for every $\varepsilon > 0$ and compact $K \subset \mathbb{R}^d$, there exists $h_\varepsilon > 0$ with $\|f - f_h\|_{\infty,K} < \varepsilon$ for all $h \in (0,h_\varepsilon]$. Subsequently, if not stated otherwise, all limits in $C_\kappa$ are understood w.r.t. the mixed topology and compact subsets are denoted by $K \Subset \mathbb{R}^d$. For more details, we refer to [26] and the references therein. Moreover, the set of all real-valued functions is endowed with the pointwise order, i.e., for functions $f, g: \mathbb{R}^d \to \mathbb{R}$ we write $f \leq g$ if and only if $f(x) \leq g(x)$ for all $x \in \mathbb{R}^d$. All order-related notions (sup, inf, max, min, lim sup, etc.) for such
functions are understood with respect to this order and the positive part of a function $f : \mathbb{R}^d \to \mathbb{R}$ is denoted by $f^+ := \max\{f, 0\}$. Let

$$B_{C_\kappa}(r) := \{f \in C_\kappa : \|f\|_\infty \leq r\} \quad \text{and} \quad B_{\mathbb{R}^d}(r) := \{x \in \mathbb{R}^d : |x| \leq r\}$$

be the closed balls with radius $r \geq 0$ around zero, where $| \cdot |$ denotes the Euclidean distance. The space $C_b$ consists of all bounded continuous functions $f : \mathbb{R}^d \to \mathbb{R}$ and $C_b^\infty$ is the space of all infinitely differentiable functions $f \in C_b$ such that all partial derivatives are in $C_b$. Let $\text{Lip}_b$ be the space of all bounded Lipschitz continuous functions $f : \mathbb{R}^d \to \mathbb{R}$ and, for every $r \geq 0$, the set $\text{Lip}_r(\kappa)$ consists of all bounded $r$-Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}$ with $\|f\|_\infty \leq r$. Here, $\|f\|_\infty := \sup_{x \in \mathbb{R}^d} |f(x)|$ denotes the usual supremum norm. Let

$$(\tau_{x}f)(y) := f(x + y) \quad \text{for all } x, y \in \mathbb{R}^d \text{ and } f : \mathbb{R}^d \to \mathbb{R}$$

be the shift operators and $\mathbb{R}_+: = \{x \in \mathbb{R} : x \geq 0\}$ the positive real numbers including zero.

2.1. **The upper Lipschitz set.** We start with a formal definition of the (upper) Lipschitz set and some basic terminology concerning nonlinear semigroups. Then, after an auxiliary lemma, we give explicit conditions which ensure that two semigroups whose generators coincide for smooth functions also have the same upper Lipschitz set. Recall that limits (and thus continuity) in $C_\kappa$ are understood w.r.t. the mixed topology rather than the norm $\| \cdot \|_\kappa$.

**Definition 2.1.** Let $(I(t))_{t \geq 0}$ be a family of operators $I(t) : C_\kappa \to C_\kappa$ with $I(0)f = f$ for all $f \in C_\kappa$. The Lipschitz set $\mathcal{L}^I$ consists of all $f \in C_\kappa$ such that there exist $c \geq 0$ and $t_0 > 0$ with

$$\|I(t)f - f\|_\kappa \leq ct \quad \text{for all } t \in [0, t_0].$$

The upper Lipschitz set $\mathcal{L}^I_+$ contains all $f \in C_\kappa$ such that there exist $c \geq 0$ and $t_0 > 0$ with

$$\|(I(t)f - f)^+\|_\kappa \leq ct \quad \text{for all } t \in [0, t_0].$$

For $f \in C_\kappa$ such that the following limit exists and lies in $C_\kappa$, we define the derivative of the mapping $t \mapsto I(t)f$ at zero by

$$I'(0)f := \lim_{h \downarrow 0} \frac{I(h)f - f}{h}.$$

While the constant $c \geq 0$ in the previous definition typically depends on $f$, the parameter $t_0 > 0$ can often be chosen arbitrarily. Furthermore, since convergence w.r.t. the mixed topology incorporates norm boundedness, it holds that $f \in \mathcal{L}^I$ for all $f \in C_\kappa$ such that the limit $I'(0)f$ exists.

**Definition 2.2.** A family $(S(t))_{t \geq 0}$ of operators $S(t) : C_\kappa \to C_\kappa$ is called a semigroup if $S(0) = \text{id}_{C_\kappa}$ and $S(s + t)f = S(s)S(t)f$ for all $s, t \geq 0$ and $f \in C_\kappa$. The semigroup is called convex (monotone) if the mapping $C_\kappa \to \mathbb{R}$, $f \mapsto (S(t)f)(x)$ is convex (monotone) for all $t \geq 0$ and $x \in \mathbb{R}^d$. Moreover, the semigroup is called strongly continuous if the mapping $\mathbb{R}_+ \to C_\kappa$, $t \mapsto S(t)f$ is continuous for all $f \in C_\kappa$. The generator is defined by

$$A : D(A) \to C_\kappa, \quad f \mapsto \lim_{h \downarrow 0} \frac{S(h)f - f}{h},$$

where the domain $D(A)$ consists of all $f \in C_\kappa$ such that the previous limit exists.
Since convergence w.r.t. the mixed topology incorporates norm boundedness, we obtain $D(A) \subseteq \mathcal{L}^S$. For every $t \geq 0$, $f \in C_\kappa$ and $x \in \mathbb{R}^d$ such that the mapping $s \mapsto (S(s)f)(x)$ is measurable and bounded from above, we define the pointwise integral

$$
\left( \int_0^t S(s)f \, ds \right)(x) := \int_0^t (S(s)f)(x) \, ds.
$$

**Lemma 2.3.** Let $(S(t))_{t \geq 0}$ be a convex semigroup on $C_\kappa$ such that $S(t) : C_\kappa \to C_\kappa$ is continuous for all $t \geq 0$. For every $r, t \geq 0$, we assume that there exists $c \geq 0$ with

$$
\|S(s)f - S(s)g\|_\kappa \leq c\|f - g\|_\kappa
$$

for all $s \in [0, t]$ and $f, g \in B_{C_\kappa}(r)$. Then, for every $t \geq 0$ and $f \in D(A)$,

$$
S(t)f - f \leq \int_0^t S(s)(f + Af) - S(s)f \, ds.
$$

**Proof.** For every $x \in \mathbb{R}^d$, it follows from $f \in D(A) \subseteq \mathcal{L}^S$ and inequality (2.2) that the mapping $\mathbb{R}_+ \to \mathbb{R}$, $t \mapsto (S(t)f)(x)$ is locally Lipschitz continuous and thus differentiable almost everywhere. Hence, by Rademacher’s theorem,

$$
(S(t)f - f)(x) = \int_0^t \frac{d}{ds}(S(s)f)(x) \, ds \quad \text{for all } t \geq 0.
$$

For every $t \geq 0$ and $h \in (0, 1]$, Lemma A.1 implies

$$
\frac{S(t)S(h)f - S(t)f}{h} - S(t)(f + Af) + S(t)f
\leq S(t) \left( \frac{S(h)f - f}{h} + f \right) - S(t)(f + Af)
\leq \frac{1}{2} S(t) \left( 2 \left( \frac{S(h)f - f}{h} - Af \right) + f + Af \right) - \frac{1}{2} S(t)(f + Af).
$$

Since $S(t)$ is continuous, the right-hand side converges to zero as $h \downarrow 0$ which yields the claim. \qed

**Assumption 2.4.** Let $(S(t))_{t \geq 0}$ be a convex monotone semigroup on $C_\kappa$ which satisfies the following conditions:

(i) For every $r, t \geq 0$, there exists $c \geq 0$ with

$$
\|S(s)f - S(s)g\|_\kappa \leq c\|f - g\|_\kappa
$$

for all $s \in [0, t]$ and $f, g \in B_{C_\kappa}(r)$. Moreover, it holds that $S(t)0 = 0$ for all $t \geq 0$.

(ii) The operator $S(t) : C_\kappa \to C_\kappa$ is continuous for all $t \geq 0$.

(iii) For every $f \in \mathcal{L}^S_+ \cap \text{Lip}_0$, there exist $L \geq 0$, $t_0 > 0$ and $\delta > 0$ with

$$
\|S(t)(\tau_x f) - \tau_x S(t)f\|_\kappa \leq Lt
$$

for all $t \in [0, t_0]$ and $x \in B_{\mathbb{R}^d}(\delta)$.

Condition (i) means that the semigroup is locally uniformly continuous w.r.t. the weighted supremum norm while condition (ii) states that the semigroup is continuous at a fixed time w.r.t. the mixed topology. Condition (iii) is, in particular, satisfied if the semigroup is translation-invariant, i.e., $S(t)(\tau_x f) = \tau_x S(t)f$ for all $t \geq 0$, $f \in C_\kappa$ and $x \in \mathbb{R}^d$. 
Theorem 2.5. Let \((S(t))_{t \geq 0}\) and \((T(t))_{t \geq 0}\) be two semigroups satisfying Assumption 2.4. Moreover, we assume that \(C_0^\infty \subset D(A) \cap D(B)\) and \((Af)^+ \leq (Bf)^+\) for all \(f \in C_0^\infty\), where \(A\) and \(B\) denote the generators of \((S(t))_{t \geq 0}\) and \((T(t))_{t \geq 0}\), respectively. Then,

\[
\mathcal{L}_+^T \cap \text{Lip}_b \subset \mathcal{L}_+^S \cap \text{Lip}_b.
\]

Proof. Fix \(f \in \mathcal{L}_+^T \cap \text{Lip}_b\) and choose \(L \geq 0\), \(t_0 > 0\) and \(\delta \in (0,1]\) satisfying Assumption 2.4(iii). Let \(\eta : \mathbb{R}^d \to \mathbb{R}_+\) be an infinitely differentiable function with \(\text{supp}(\eta) \subset B_{\mathbb{R}^d}(\delta)\) and \(\int_{\mathbb{R}^d} \eta(x) \, dx = 1\). For every \(n \in \mathbb{N}\) and \(x \in \mathbb{R}^d\), we define \(\eta_n(x) := n^d \eta(nx)\) and

\[
f_n(x) := (f \ast \eta_n)(x) = \int_{\mathbb{R}^d} f(x-y) \eta_n(y) \, dy.
\]

For every \(t \in [0,t_0]\), \(n \in \mathbb{N}\) and \(x \in \mathbb{R}^d\), we use Jensen’s inequality, Assumption 2.4(iii) and inequality (2.1) to estimate

\[
(T(t)f_n - f_n)(x) \leq \int_{B(\delta)} (T(t)(\tau_y f) - \tau_y f)(x) \eta_n(y) \, dy
\]

\[
\leq \int_{B(\delta)} (T(t)f - f)(x-y) \eta_n(y) \, dy + Lt
\]

\[
\leq c_\kappa \int_{B(\delta)} (T(t)f - f)(x-y) \eta_n(y) \, dy + Lt
\]

\[
\leq c_\kappa ((T(t)f)^+ \kappa) \ast \eta_n + Lt,
\]

where \(B(\delta) := B_{\mathbb{R}^d}(\delta)\). Taking the supremum over \(x \in \mathbb{R}^d\) yields

\[
\|(T(t)f_n - f_n)^+\|_\kappa \leq c_\kappa \|(T(t)f - f)^+\|_\kappa + Lt.
\]

Hence, it follows from \(f \in \mathcal{L}_+^T\) that there exist \(c \geq 0\) and \(t_1 \in (0,t_0]\) with

\[
\|(T(t)f_n - f_n)^+\|_\kappa \leq (c_\kappa + L)t \quad \text{for all } n \in \mathbb{N} \text{ and } t \in [0,t_1].
\]

Since \(f_n \in C_0^\infty\), we obtain \(\|(Af_n)^+\|_\kappa \leq \|(Bf_n)^+\|_\kappa \leq c_\kappa + L\) for all \(n \in \mathbb{N}\). Moreover, inequality (2.1) guarantees that \(\|f_n\|_\kappa \leq c_\kappa \|f\|_\kappa\) for all \(n \in \mathbb{N}\). Lemma 2.3, the monotonicity of \(S(s)\), Assumption 2.4(i) and inequality (2.1) imply that there exists \(c' \geq 0\) with

\[
(S(t)f_n - f_n)\kappa \leq \int_0^t (S(s)(f_n + (Af_n)^+) - S(s)f_n)\kappa \, ds
\]

\[
\leq \int_0^t c'\|(Af_n)^+\|_\kappa \, ds \leq c'(c_\kappa + L)t
\]

for all \(n \in \mathbb{N}\) and \(t \in [0,t_1]\). Taking the limit \(n \to \infty\) yields

\[
\|(S(t)f - f)^+\|_\kappa \leq c'(c_\kappa + L)t \quad \text{for all } t \in [0,t_1]
\]

which shows that \(\mathcal{L}_+^T \cap \text{Lip}_b \subset \mathcal{L}_+^S \cap \text{Lip}_b\). \(\Box\)

While the following comparison principle does not have the same generality as \([7,\ \text{Theorem 2.10}]\), it has the advantage that the conditions (i)–(v) can easily be verified in many applications and that the semigroups are uniquely determined by their generators evaluated at smooth functions. Moreover, Assumption 2.7 provides sufficient conditions on the generating family \((I(t))_{t \geq 0}\) such that Theorem 2.6 applies to the corresponding semigroup \((S(t))_{t \geq 0}\). In Subsection 2.2, we also discuss how the conditions (i)–(v) can be verified.
\section*{Limit Theorems}

\begin{theorem}
Let \((S_1(t))_{t \geq 0}\) and \((S_2(t))_{t \geq 0}\) be two strongly continuous convex monotone semigroups on \(C_{\kappa}\) with \(S_1(t)0 = 0\) and generators \(A_1\) and \(A_2\), respectively, such that the following conditions are satisfied for \(i = 1, 2\):
\begin{enumerate}[(i)]
  \item It holds \(C_0^\infty \subset D(A_i)\) and \(A_1f = A_2f\) for all \(f \in C_0^\infty\).
  \item For every \(r, T \geq 0\), there exists \(c \geq 0\) with
    \[\|S_i(t)f - S_i(t)g\|_\kappa \leq c\|f - g\|_\kappa \quad \text{for all } t \in [0, T] \text{ and } f, g \in B_{C_\kappa}(r)\]
  \item For every \(\varepsilon > 0\), \(r, T \geq 0\) and \(K \in \mathbb{R}^d\), there exist \(c \geq 0\) and \(K' \in \mathbb{R}^d\) with
    \[\|S_i(t)f - S_i(t)g\|_{\infty, K} \leq c\|f - g\|_{\infty, K'} + \varepsilon\]
    for all \(t \in [0, T]\) and \(f, g \in B_{C_\kappa}(r)\).
  \item For every \(f \in \text{Lip}_b\) and \(\varepsilon > 0\), there exist \(\delta, t_0 > 0\) with
    \[\|S_i(t)(\tau_x f) - \tau_x S_i(t)f\|_\kappa \leq \varepsilon t\]
    for all \(t \in [0, t_0]\) and \(x \in B_{\mathbb{R}^d}(\delta)\).
  \item It holds \(S_i(t) : \text{Lip}_b \to \text{Lip}_b\) for all \(t \geq 0\).
\end{enumerate}
Then, it holds \(S_1(t)f = S_2(t)f\) for all \(t \geq 0\) and \(f \in C_\kappa\).
\end{theorem}

\begin{proof}
It follows from Theorem 2.5 that
\[\mathcal{L}^{C_1}_+ \cap \text{Lip}_b = \mathcal{L}^{C_2}_+ \cap \text{Lip}_b\]
and therefore [7, Theorem 2.10] implies
\[S_1(t)f = S_2(t)f \quad \text{for all } (f, t) \in C_0^\infty \times \mathbb{R}_+.\]
Indeed, the semigroups \((S_1(t))_{t \geq 0}\) and \((S_2(t))_{t \geq 0}\) clearly satisfy [7, Assumption 2.4] while equation (2.3), condition (ii) and condition (v) imply \(\mathcal{L}^{C_1}_+ \cap \text{Lip}_b \subset \mathcal{L}^{C_2}_+\) and
\[S_1(t) : \mathcal{L}^{C_1}_+ \cap \text{Lip}_b \to \mathcal{L}^{C_1}_+ \cap \text{Lip}_b \quad \text{for all } t \geq 0.\]
Moreover, due to condition (i) and (iv), we can apply [7, Lemma 3.6] to conclude that \((S_1(t))_{t \geq 0}\) and \((S_2(t))_{t \geq 0}\) have the same upper \(\Gamma\)-generator on \(\mathcal{L}^{C_1}_+ \cap \text{Lip}_b\). Hence,
\[S_1(t)f \leq S_2(t)f \quad \text{for all } t \geq 0 \text{ and } f \in C_0^\infty\]
and reversing the roles of \((S_1(t))_{t \geq 0}\) and \((S_2(t))_{t \geq 0}\) yields that equation (2.4) is valid. Since \(C_0^\infty \subset C_{\kappa}\) is dense, it follows from condition (iii) that
\[S_1(t)f = S_2(t)f \quad \text{for all } (f, t) \in C_{\kappa} \times \mathbb{R}_+.\]
\end{proof}

\subsection{The Chernoff approximation}
Due to Theorem 2.5 and the resulting Theorem 2.6, we are now able to improve the approximation result for convex monotone semigroups from [7]. Let \((I(t))_{t \geq 0}\) be a family of operators \(I(t) : C_{\kappa} \to C_{\kappa}\) and \((h_n)_{n \in \mathbb{N}} \subset (0, 1]\) be a sequence with \(h_n \to 0\). For every \(t \geq 0\), \(n \in \mathbb{N}\) and \(f \in C_{\kappa}\), we define
\[I(h_n^{k_n} f) := I(h_n)^{k_n} f = \underbrace{(I(h_n) \circ \ldots \circ I(h_n))}_{k_n \text{ times}} f,\]
where \(k_n := \max\{k \in \mathbb{N}_0 : kh_n \leq t\}\) and \(\pi_n^t := \{kh_n \land t : k \in \mathbb{N}_0\}\) denotes the equidistant partition of \([0, t]\) with mesh size \(h_n\).

\begin{assumption}
Let \((I(t))_{t \geq 0}\) be a family of operators \(I(t) : C_{\kappa} \to C_{\kappa}\) which satisfy the following conditions:
\begin{enumerate}[(i)]
  \item \(I(0) = \text{id}_{C_{\kappa}}\).
  \item \(I(t)\) is convex and monotone with \(I(t)0 = 0\) for all \(t \geq 0\).
\end{enumerate}
\end{assumption}
(iii) There exists $\omega \geq 0$ with
\[
\|I(t)f - I(t)g\|_\kappa \leq e^{\omega t}\|f - g\|_\kappa \quad \text{for all } t \in [0,1] \text{ and } f, g \in C_\kappa.
\]
(iv) There exist $t_0 > 0$, $\delta \in (0,1]$ and $L \geq 0$ with
\[
\|I(t)(\tau_x f) - \tau_x I(t)f\|_\kappa \leq Lrt|x|
\]
for all $t \in [0,t_0]$, $x \in B_{\mathbb{R}^d}(\delta)$, $r \geq 0$ and $f \in \text{Lip}_b(r)$.
(v) It holds that $C^\infty_0 \subset \mathcal{L}^1$ and $I'(0)f \in C_\kappa$ exists for all $f \in C^\infty_0$.
(vi) For every $\varepsilon > 0$, $r, T \geq 0$ and $K \subset \mathbb{R}^d$, there exist $c \geq 0$ and $K' \subset \mathbb{R}^d$ with
\[
\|I(\pi^t_n)f - I(\pi^t_n)g\|_{\infty,K} \leq \|f - g\|_{\infty,K'} + \varepsilon
\]
for all $t \in [0,T]$, $f, g \in B_{C_\kappa}(r)$ and $n \in \mathbb{N}$.
(vii) It holds that $I(t) : \text{Lip}_b(r) \to \text{Lip}_b(e^{\omega t}r)$ for all $r, t \geq 0$.

The key idea here is to identify conditions on the one-step operators $I(t)$ that are preserved during the iteration and thus transfer to the associated semigroup $(S(t))_{t \geq 0}$. Although condition (vi) is stated for the iterated operators $I(\pi^t_n)$, there are several sufficient conditions for the one-step operators $I(t)$ that can be verified in applications. One of them is given in Remark 2.8(b) and such conditions are more systematically studied in [10, Subsection 2.5]. In many applications such as the ones presented in Section 3 and Section 4, most of the previous conditions can easily be verified while the computation of the derivative $I'(0)f$ usually takes some computational effort. However, for smooth functions this can often be achieved by elementary arguments such as Taylor’s formula with a suitable remainder estimate. In Theorem 3.2 and Theorem 4.1, we apply Theorem 2.9 with $(I(t)f)(x) := \frac{1}{t} \mathbb{E}[tf(x + t\xi_1)]$ and $(I(t)f)(x) := \frac{1}{t} \mathbb{E}[tf(x + \sqrt{t}\xi_1)]$, respectively. The corresponding proofs consist in verifying Assumption 2.7, where the most of the conditions are immediately satisfied and the computation of the derivative $I'(0)f$ is the main part of the proof. Furthermore, the need to satisfy Assumption 2.7 requires the random variables $(\xi_n)_{n \in \mathbb{N}}$ to satisfy certain natural moment conditions that also become apparent in the proofs. Further examples that involve similar computations can be found in [7, Section 5] and [9, Section 6]. Since monotone convergence implies convergence w.r.t. the mixed topology, the Nisio semigroups studied in [16, 36] are also covered by the results of this section. Finally, we want to emphasize that, in contrast to the previous applications presented in [7, Section 5] and [9, Section 6], the uniqueness of the semigroup $(S(t))_{t \geq 0}$ follows immediately now. In Theorem 3.4, this allows us to explicitly represent the semigroup by the Hopf–Lax formula.

**Remark 2.8.** We want to discuss some of the previous conditions in more detail.

(a) Condition (iii) guarantees that the iterated operators $I(\pi^t_n)$ are uniformly Lipschitz continuous. In many examples, one can choose $\omega := 0$.

(b) Condition (vi) can be verified as follows: suppose that there exists another bounded continuous function $\tilde{\kappa} : \mathbb{R}^d \to (0, \infty)$ such that, for every $\varepsilon > 0$, there exists $K \subset \mathbb{R}^d$ with $\sup_{x \in K} \frac{\tilde{\kappa}(x)}{\kappa(x)} \leq \varepsilon$. Moreover, we assume that there exists $c \geq 0$ with
\[
\|I(t)f\|_{\tilde{\kappa}} \leq e^{ct}\|f\|_{\tilde{\kappa}}
\]
for all $t \in [0,1]$ and $f \in C_\kappa$ with $\|f\|_{\tilde{\kappa}} \leq 1$. By induction, one can show that
\[
\|I(\pi^t_n)f\|_{\tilde{\kappa}} \leq e^{ct}\|f\|_{\tilde{\kappa}}
\]
for all $t \geq 0$, $n \in \mathbb{N}$ and $f \in C_\kappa$ with $\|f\|_\kappa \leq e^{-ct}$. Let $(f_n)_{n \in \mathbb{N}} \subset C_\kappa$ be a sequence with $f_n \downarrow 0$. For every $\varepsilon > 0$, there exists $K \subset \mathbb{R}^d$ with
\[
|f_n(x)| \kappa(x) = |f_n(x)| \kappa(x) \frac{\kappa(x)}{\kappa(x)} \leq \|f_1\|_\kappa \frac{\kappa(x)}{\kappa(x)} < \varepsilon
\]
for all $n \in \mathbb{N}$ and $x \in K^c$. Moreover, by Dini’s theorem, there exists $n_0 \in \mathbb{N}$ with
\[
|f_n(x)| \kappa(x) \leq \|f_n\|_{\infty,K} \|\kappa\|_{\infty} < \varepsilon
\]
for all $n \geq n_0$ and $x \in K$. This implies
\[
\sup_{s \in [0,t]} \|I(\pi^t_s) f_n\|_\kappa \leq e^{ct} \|f_n\|_\kappa \to 0 \quad \text{as } n \to \infty.
\]
We obtain $\sup_{(t,x) \in [0,T] \times K} \sup_{k \in \mathbb{N}} (I(\pi^t_k) f_n)(x) \to 0$ as $n \to \infty$ for all $T \geq 0$ and $K \subset \mathbb{R}^d$ and [7, Lemma C.2] yields that condition (vi) is satisfied. In the subsequent sections, we choose $\kappa(x) := (1 + |x|^p)^{-1}$ for some $p \geq 1$ and define $I(t)$ via a nonlinear expectation. In this case, inequality (2.5) is a moment condition.

(c) If $\kappa \equiv 1$, the conditions (iii) and (iv) imply
\[
I(t) : \text{Lip}_b(r) \to \text{Lip}_b(e^{(\omega + L)t}r) \quad \text{for all } r,t \geq 0.
\]

**Theorem 2.9.** Let $(I(t))_{t \geq 0}$ be a family of operators satisfying Assumption 2.7. Then, there exists a strongly continuous convex monotone semigroup $(S(t))_{t \geq 0}$ on $C_\kappa$ with
\[
S(t)f = \lim_{{n \to \infty}} I(\pi^t_n) f \quad \text{for all } t \geq 0 \text{ and } f \in C_\kappa. \tag{2.6}
\]
Furthermore, the semigroup has the following properties:

(i) It holds that $f \in D(A)$ and $Af = I'(0)f$ for all $f \in C_\kappa$ such that $I'(0)f \in C_\kappa$ exists. In particular, this is valid for all $f \in C_\kappa^c$.

(ii) It holds that $\|S(t)f - S(t)g\|_{K} \leq e^{ct}\|f - g\|_{K}$ for all $t \geq 0$ and $f,g \in C_\kappa$.

(iii) For every $\varepsilon > 0$, $r,T \geq 0$ and $K \subset \mathbb{R}^d$, there exist $K' \subset \mathbb{R}^d$ and $c \geq 0$ with
\[
\|S(t)f - S(t)g\|_{\infty,K} \leq c\|f - g\|_{\infty,K'} + \varepsilon
\]
for all $t \in [0,T]$ and $f,g \in B_{C_\kappa}(r)$.

(iv) It holds that $\mathcal{L}^1 \subset \mathcal{L}^S$ and $\mathcal{L}^1_+ \subset \mathcal{L}^S_+$. Moreover, for every $t \geq 0$,
\[
S(t) : \mathcal{L}^S \to \mathcal{L}^S \quad \text{and} \quad S(t) : \mathcal{L}^S_+ \to \mathcal{L}^S_+.
\]

(v) For every $r,t \geq 0$, $f \in \text{Lip}_b(r)$ and $x \in B_{\mathbb{R}^d}(\delta)$,
\[
\|S(t)(\tau_x f) - \tau_x S(t)f\|_\kappa \leq Lrte^{2\omega t}|x|.
\]
Furthermore, it holds that $S(t) : \text{Lip}_b(r) \to \text{Lip}_b(e^{\omega t}r)$ for all $r,t \geq 0$.

**Proof.** First, we verify the assumptions of [7, Section 4]. Assumption 2.7 guarantees that the conditions (i)-(iv) and (vi) of [7, Assumption 4.1] and [7, Assumption 4.4] are satisfied. Moreover, by induction, we obtain
\[
I(\pi^t_n) : \text{Lip}_b(r) \to \text{Lip}_b(e^{\omega t}r) \quad \text{for all } r,t \geq 0 \text{ and } n \in \mathbb{N}. \tag{2.7}
\]
Since $C_\kappa^c \subset \mathcal{L}^1 \subset C_\kappa$ is separable and dense, this shows that [7, Assumption 4.1(v)] is satisfied. Now, let $\mathcal{T} \subset \mathbb{R}_+$ be a countable dense set including zero. By [7, Theorem 4.3], there exist strongly continuous convex monotone semigroup $(S(t))_{t \geq 0}$ on $C_\kappa$ and a subsequence $(n_l)_{l \in \mathbb{N}} \subset \mathbb{N}$ with
\[
S(t)f = \lim_{{l \to \infty}} I(\pi^t_{n_l}) f \quad \text{for all } (f,t) \in C_\kappa \times \mathcal{T}. \tag{2.8}
\]
Moreover, the semigroup has the properties (i)-(iii) and it holds that \( \mathcal{L}^I \subset \mathcal{L}^S \) while property (v) follows from [7, Theorem 4.5]. For the invariance of the (upper) Lipschitz set, we refer to [7, Lemma 2.8]. In order to show that the convergence in the previous equation holds for arbitrary points where \((t, h)\) depend on the choice of the convergent subsequence. For every subsequence \(\tilde{\pi}_{k_l}\) that holds that inequality (2.9) transfers to \(S(t)\), i.e.,

\[
\|(S(t)f - f)^+\|_\kappa \leq cte^{\omega t} \quad \text{for all } t \geq 0.
\]

Second, we verify equation (2.6) by showing that the limit in equation (2.8) does not depend on the choice of the convergent subsequence. For every subsequence \((\tilde{n}_k)_{k \in \mathbb{N}} \subset \mathbb{N}\), there exist a further subsequence \((\tilde{n}_{k_l})_{l \in \mathbb{N}}\) and a strongly continuous convex monotone semigroup \((\tilde{S}(t))_{t \geq 0}\) on \(C_\kappa\) with

\[
\tilde{S}(t)f = \lim_{l \to \infty} I(\pi_{\tilde{n}_{k_l}}^t) f \quad \text{for all } (f, t) \in C_\kappa \times \mathcal{T}
\]

and the properties (i)-(v). Hence, Theorem 2.6 implies

\[
S(t)f = \tilde{S}(t)f \quad \text{for all } t \geq 0 \text{ and } f \in C_\kappa.
\]

Since every subsequence has a further subsequence which converges to a limit that is independent of the choice of the subsequence, we conclude that

\[
S(t)f = \lim_{n \to \infty} I(\pi_n^t) f \quad \text{for all } (f, t) \in C_\kappa \times \mathcal{T}.
\]

In order to show that the convergence in the previous equation holds for arbitrary points in time, let \(t \geq 0\) and define \(\mathcal{T} := \mathcal{T} \cup \{t\}\). Analogously to the previous arguments, we obtain a semigroup \((\tilde{S}(t))_{t \geq 0}\) with

\[
S(t)f = \tilde{S}(t)f = \lim_{n \to \infty} I(\pi_n^t) f \quad \text{for all } (f, t) \in C_\kappa \times \mathcal{T}.
\]

Since \(t \geq 0\) was arbitrary, we obtain equation (2.6). \(\square\)

**Corollary 2.10.** Let \((I(t))_{t \geq 0}\) and \((J(t))_{t \geq 0}\) be two families of operators satisfying Assumption 2.7 with \(I'(0)f \leq J'(0)f\) for all \(f \in C_\kappa^\infty\). Then,

\[
S(t)f \leq T(t)f \quad \text{for all } t \geq 0 \text{ and } f \in C_\kappa,
\]

where \((S(t))_{t \geq 0}\) and \((T(t))_{t \geq 0}\) are the semigroups associated to \((I(t))_{t \geq 0}\) and \((J(t))_{t \geq 0}\), respectively.
Proof. This follows immediately from Theorem 2.6. □

3. First order scaling limits

Throughout this section, we choose \( \kappa \equiv 1 \) and \( \tilde{\kappa} : \mathbb{R}^d \to (0, \infty), \ x \mapsto (1 + |x|)^{-1} \).

Recall that the function \( \tilde{\kappa} \) has been introduced in Remark 2.8(b) in order to verify Assumption 2.7(vi) by means of a suitable moment condition. In Appendix B, we gather some basic definitions and properties of convex expectations such as independence and weak convergence which resemble their classical analogues from probability theory and that are frequently used in the sequel. Let \((\xi_n)_{n \in \mathbb{N}} \subset \mathcal{H}^d\) be an iid sequence of random vectors on a convex expectation space \((\Omega, \mathcal{H}, \bar{\mathcal{E}})\) with finite first moments and \((\psi_n)_{n \in \mathbb{N}}\) be a sequence of recursively defined functions \(\psi_n : \mathbb{R}_+ \times \mathbb{R}^d \times (\mathbb{R}^d)^n \to \mathbb{R}^d\). We define a sequence \((X_n)_{n \in \mathbb{N}}\) of random vectors by

\[
X_n := \psi_n \left( \frac{1}{n}, 0, \xi_1, \ldots, \xi_n \right)
\]

whose behaviour in the limit we are interested in. For \(\psi_n(t, x, y_1, \ldots, y_n) := x + t \sum_{i=1}^{n} y_i\) we end up with an averaged sum of iid samples, i.e.,

\[
X_n = \frac{1}{n} \sum_{i=1}^{n} \xi_i \quad \text{for all } n \in \mathbb{N}.
\]

Since we are only interested in weak convergence, we mainly consider the distributions of the vectors \(X_n\). In this framework, requiring finite first moments means that the distribution of \(\xi_1\) is well-defined even for continuous functions with at most linear growth at infinity, i.e.,

\[
F_{\xi_1} : C_{\tilde{\kappa}} \to \mathbb{R}, \ f \mapsto \bar{\mathcal{E}}[f(\xi_1)].
\]

This functional is subsequently simply denoted by \(\mathcal{E}[\cdot]\) and is supposed to be continuous from above in order to guarantee Assumption 2.7(vi). We remark that, in the linear case, continuity from above is always given due the fact that \(\lim_{c \to \infty} \mathbb{E}[(|\xi_1|1_{\{|\xi_1| \geq c\}})] = 0\). Moreover, continuity from above on \(C_{\tilde{\kappa}}\) is equivalent to the uniform integrability condition \(\lim_{c \to \infty} \bar{\mathcal{E}}[(|\xi_1| - c)^+] = 0\) from [41]. We define \(I(0) := \text{id}_{C_b}\) and

\[
(I(t)f)(x) := tf\left(\psi_1(t, x, \xi_1)\right)
\]

for all \(t > 0, \ f \in C_b\) and \(x \in \mathbb{R}^d\). Then, the independence of \((\xi_n)_{n \in \mathbb{N}}\) implies

\[
(I(t))^n f(0) = \frac{1}{n} \mathcal{E}[nf(X_n)].
\]

Hence, our operator theoretic result can be formulated equivalently by using classic probabilistic notation. In Subsection 3.1, we prove the previously discussed main result for first order scaling limits as well as some immediate consequences. Motivated by Cramér’s theorem, we then show that similar convergence rates can be extended beyond the case of averaged sums of iid samples, see Subsection 3.2. Moreover, we use a clever estimate from [34] to obtain polynomial convergence rates without requiring exponential moments. Finally, as an illustration of the previous rather abstract results, we consider convex expectations that are defined as a (weighted) supremum over a set of probability measures, see Subsection 3.3.
3.1. LLN for convex expectations and large deviations. Let \( \mathcal{E} : C_\tilde{\kappa} \to \mathbb{R} \) be a convex expectation which is continuous from above, i.e.,

- \( \mathcal{E}[c] = c \) for all \( c \in \mathbb{R} \),
- \( \mathcal{E}[f] \leq \mathcal{E}[g] \) for all \( f, g \in C_\tilde{\kappa} \) with \( f \leq g \),
- \( \mathcal{E}[\lambda f + (1 - \lambda)g] \leq \lambda \mathcal{E}[f] + (1 - \lambda)\mathcal{E}[g] \) for all \( f, g \in C_\tilde{\kappa} \) and \( \lambda \in [0, 1] \),
- \( \mathcal{E}[f_n] \downarrow 0 \) for all \( (f_n)_{n \in \mathbb{N}} \subset C_\tilde{\kappa} \) with \( f_n \downarrow 0 \).

The convergence \( f_n \downarrow 0 \) is understood pointwise. However, it follows from Dini’s theorem and \( \|f_n\|_\tilde{\kappa} \leq \|f_1\|_\tilde{\kappa} \) for all \( n \in \mathbb{N} \) that \( f_n \to 0 \) in the mixed topology. By [7, Theorem C.1], there exists a convex monotone extension \( \mathcal{E} : B_\tilde{\kappa} \to \mathbb{R} \) such that, for every \( \varepsilon > 0 \) and \( c \geq 0 \), there exists \( K \in \mathbb{R}^d \) with

\[
\mathcal{E} \left[ |x|_{K^c} \right] < \varepsilon. \tag{3.1}
\]

Here, \( B_\tilde{\kappa} \) denotes the space of all Borel measurable functions \( f : \mathbb{R}^d \to \mathbb{R} \) with \( \|f\|_\tilde{\kappa} < \infty \). Furthermore, it holds that \( \mathcal{E}[f_n] \downarrow \mathcal{E}[f] \) for all \( (f_n)_{n \in \mathbb{N}} \subset C_\tilde{\kappa} \) and \( f \in C_\tilde{\kappa} \) with \( f_n \downarrow f \).

**Assumption 3.1.** Let \( \psi : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \) be a continuous function which satisfies the following conditions:

1. There exists \( L \geq 0 \) such that, for all \( t \in [0, 1] \) and \( x, y, z \in \mathbb{R}^d \),
   \[ |x + \psi(t, y, z) - \psi(t, x + y, z)| \leq Lt|x|. \]
   Furthermore, it holds that \( |\psi(t, x, y) - x| \leq L(1 + |y|)t \) for all \( t \in [0, 1] \) and \( x, y \in \mathbb{R}^d \).
2. There exists a continuous function \( \psi_0 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \) with
   \[
   \lim_{h \downarrow 0} \sup_{x,y \in K} \frac{|\psi(h, x, y) - x|}{h} - \psi_0(x, y) = 0 \quad \text{for all } K \subset \mathbb{R}^d.
   \]

The previous conditions are clearly satisfied for \( \psi(t, x, y) := x + ty \) and \( \psi_0(x, y) := y \) in which case it holds \( \psi_n(\frac{1}{n}, x, y_1, \ldots, y_n) = x + \frac{1}{n} \sum_{i=1}^n y_i \). Another valid choice would be \( \psi(t, x, y) := x + t\varphi(x) + ty \) and \( \psi_0(x, y) := \varphi(x) + y \) for some Lipschitz continuous function \( \varphi : \mathbb{R}^d \to \mathbb{R}^d \) describing a perturbation of the sample \( \xi_{n+1} \) by a nonlinear function depending on the average of previous samples. In Section 4, the choice \( \psi(t, x, y) := x + \sqrt{t}y \) leads to a CLT type result but here Assumption 3.1(i) is not satisfied since \( |\psi(t, x, y) - x| = \sqrt{t}|y| \). We define \( I(0) := \text{id}_{C_b} \) and, for every \( t > 0 \), \( f \in C_b \) and \( x \in \mathbb{R}^d \),

\[
(I(t)f)(x) := t\mathcal{E} \left[ \frac{1}{t} f(\psi(t, x, \cdot)) \right].
\]

Let \( C_0^1 \) consist of all differentiable functions \( f \in C_b \) such that all partial derivatives are in \( C_b \) and denote by \( x y := \langle x, y \rangle \) the Euclidean inner product on \( \mathbb{R}^d \). We want to point out that \( C_\tilde{\kappa} \) appears in this section only to ensure that \( \mathcal{E}[\cdot] \) is defined and continuous from above on functions with at most linear growth at infinity. However, we consider \((I(t))_{t \geq 0} \) and \((S(t))_{t \geq 0} \) as operator families on \( C_b \). In particular, the definition of the (upper) Lipschitz sets and the properties from Theorem 2.9 are understood w.r.t. the supremum norm \( \|\cdot\|_\infty \). Moreover, the convergence in equation (3.2) and the definition of the generator \( Af \) are understood as convergence in \( C_b \) w.r.t. the corresponding mixed topology. The sequence \((\psi_n)_{n \in \mathbb{N}}\) of functions \( \psi_n : \mathbb{R}_+ \times \mathbb{R}^d \times (\mathbb{R}^d)^n \to \mathbb{R}^d \) is recursively defined by \( \psi_1 := \psi \) and

\[
\psi_{n+1}(t, x, y_1, \ldots, y_n, y_{n+1}) := \psi(t, \psi_n(t, x, y_1, \ldots, y_n), y_{n+1}).
\]

The following theorem is the main result of this section.
Suppose that Assumption 3.1 holds. Then, there exists a strongly continuous convex monotone semigroup \((S(t))_{t \geq 0}\) on \(C_b\) with

\[
S(t)f = \lim_{n \to \infty} I(\frac{t}{n})^n f \in C_b \quad \text{for all } t \geq 0 \text{ and } f \in C_b
\]

(3.2)

which has the properties (i)-(v) from Theorem 2.9 with \(\omega = 0\) and the constant \(L \geq 0\) from Assumption 3.1. For every \(f \in C_b\), it holds that \(f \in D(A)\) and

\[
(Af)(x) = \mathcal{E}[\nabla f(x)\psi_0(x, \cdot)] \quad \text{for all } x \in \mathbb{R}^d.
\]

In addition, for every convex expectation space \((\Omega, \mathcal{H}, \bar{\mathcal{E}})\) and iid sequence \((\xi_n)_{n \in \mathbb{N}} \subset \mathcal{H}^d\) with \(\bar{\mathcal{E}}[f(\xi_n)] = \mathcal{E}[f]\) for all \(f \in C_b\),

\[
(S(1)f)(0) = \lim_{n \to \infty} \frac{1}{n} \bar{\mathcal{E}}[nf(X_n)] \quad \text{for all } f \in C_b,
\]

(3.3)

where \(X_n := \psi_n(\frac{1}{n}, 0, \xi_1, \ldots, \xi_n)\).

**Proof.** First, we verify Assumption 2.7(i)-(iv), (vi) and (vii). Lemma B.2(ii) implies that Assumption 2.7(i)-(iii) are satisfied with \(\omega = 0\). For every \(t \in (0, 1], r \geq 0, f \in \text{Lip}_b(r)\) and \(x,y \in \mathbb{R}^d\), it follows from Lemma B.2(ii) and Assumption 3.1(i) that

\[
|I(t)(\tau_x f) - \tau_x I(t)f(x)| = t \mathcal{E} \left[ \frac{1}{t} f(x + \psi(t, y, \cdot)) \right] - \mathcal{E} \left[ \frac{1}{t} f(\psi(t, x + y, \cdot)) \right] \\
\leq \|f(x + \psi(t, y, \cdot)) - f(\psi(t, x + y, \cdot))\|_\infty \\
\leq r \|x + \psi(t, y, \cdot) - \psi(t, x + y, \cdot)\|_\infty \leq Lr|t|x|.
\]

(3.4)

Thus, Assumption 2.7(iv) is satisfied. For every \(f \in C_b\) with \(\|f\|_{\bar{\mathcal{H}}} \leq 1, t \in (0, 1]\) and \(x \in \mathbb{R}^d\), we use Lemma B.2(iii) and (v) and Assumption 3.1(i) to estimate

\[
|(I(t)f)(x)| \leq t \mathcal{E} \left[ \frac{1}{t} f(\psi(t, x, \cdot)) \right] \leq \|f\|_{\bar{\mathcal{H}}} \mathcal{E} \left[ \frac{1}{t} (1 + |\psi(t, x, \cdot)|) \right] \\
\leq \|f\|_{\bar{\mathcal{H}}} (1 + |x| + t \mathcal{E} [\frac{1}{t}]) \leq e^{ct}\|f\|_{\bar{\mathcal{H}}} (1 + |x|),
\]

where \(c := \mathcal{E}[t/\bar{\mathcal{H}}]\). This shows that \(\|I(t)f\|_{\bar{\mathcal{H}}} \leq e^{ct}\|f\|_{\bar{\mathcal{H}}}\) and Remark 2.8(b) yields that Assumption 2.7(vi) is satisfied. Concerning Assumption 2.7(vii), we remark that inequality (3.4) implies

\[
|(I(t)f)(x + y) - (I(t)f)(x)| \leq \|\tau_y I(t)f - I(t)(\tau_y f)\|_\infty + \|I(t)(\tau_y f) - I(t)f\|_\infty \\
\leq Lr|t|y| + \|\tau_y f - f\|_\infty \leq e^{Lt}r|y|
\]

for all \(r,t \geq 0, f \in \text{Lip}_b(r)\) and \(x,y \in \mathbb{R}^d\) and therefore \(I(t): \text{Lip}_b(r) \to \text{Lip}_b(e^{Lt}r)\). Moreover, for every \(r \geq 0\) and \(f \in \text{Lip}_b(r)\), it follows from Assumption 3.1(i) that

\[
|f(\psi(t, x, y)) - f(x)| \leq r|\psi(t, x, y) - x| \leq Lr(1 + |y|)t
\]

for all \(r,t \geq 0, f \in \text{Lip}_b(r)\) and \(x,y \in \mathbb{R}^d\). We use Lemma B.2(v) to obtain

\[
|(I(t)f - f)(x)| \leq t \mathcal{E} \left[ \frac{|f(\psi(t, x, \cdot)) - f(x)|}{t} \right] \leq \mathcal{E}[\frac{Lt}{\bar{\mathcal{H}}}]t
\]

and therefore \(\text{Lip}_b \subset \mathcal{L}^I\).
Second, we show that \((I'(0)f)(x) = \mathcal{E}[\nabla f(x)\psi_0(x, \cdot)]\) for all \(f \in \mathcal{C}^1_b\) and \(x \in \mathbb{R}^d\). For every \(h > 0\), \(f \in \mathcal{C}^1_b\), \(x \in \mathbb{R}^d\) and \(\lambda \in (0, 1]\), we use Lemma A.1 to estimate

\[
\left( \frac{I(h)f - f}{h} \right)(x) - \mathcal{E}[\nabla f(x)\psi_0(x, \cdot)] = \mathcal{E}\left[ \left( \frac{\psi(h, x, \cdot) - x}{h} \right) \int_0^1 \nabla f(x_s) \, ds \right] - \mathcal{E}[\nabla f(x)\psi_0(x, \cdot)] \\
\leq \lambda \mathcal{E}\left[ \frac{1}{\lambda} \left( \frac{\psi(h, x, \cdot) - x}{h} \right) \int_0^1 \nabla f(x_s) \, ds - \nabla f(x)\psi_0(x, \cdot) \right] + \nabla f(x)\psi_0(x, \cdot) \\
- \lambda \mathcal{E}[\nabla f(x)\psi_0(x, \cdot)],
\]

where \(x_s := x + s(\psi(t, x, \cdot) - x)\). Furthermore, the convexity of \(\mathcal{E}\) implies

\[
\lambda \mathcal{E}\left[ \frac{1}{\lambda} \left( \frac{\psi(h, x, \cdot) - x}{h} \right) \int_0^1 \nabla f(x_s) \, ds - \nabla f(x)\psi_0(x, \cdot) \right] + \nabla f(x)\psi_0(x, \cdot) \\
\leq \frac{\lambda}{4} \mathcal{E}\left[ \frac{4}{\lambda} \left( \frac{\psi(h, x, \cdot) - x}{h} \right) \int_0^1 \nabla f(x_s) - \nabla f(x) \, ds \right] 1_{B(n)}(\cdot) \\
+ \frac{\lambda}{4} \mathcal{E}\left[ \frac{4}{\lambda} \left( \frac{\psi(h, x, \cdot) - x}{h} \right) \int_0^1 \nabla f(x_s) - \nabla f(x) \, ds \right] 1_{B(n)^c}(\cdot) \\
+ \frac{\lambda}{4} \mathcal{E}\left[ \frac{4}{\lambda} \left( \frac{\psi(h, x, \cdot) - x}{h} - \psi_0(x, \cdot) \right) \nabla f(x) \right] + \frac{\lambda}{4} \mathcal{E}[4\nabla f(x)\psi_0(x, \cdot)]
\]

for all \(n \in \mathbb{N}\) and \(B(n) := B_{\mathbb{R}^d}(n)\). Let \(\varepsilon > 0\) and \(K \subseteq \mathbb{R}^d\). By Assumption 3.1, there exists \(\lambda \in (0, 1]\) with

\[
\sup_{x \in \mathbb{R}^d} \lambda \mathcal{E}[4\|\nabla f\|_\infty|\psi_0(x, \cdot)|] \leq \sup_{x \in \mathbb{R}^d} \lambda \mathcal{E}\left[ \frac{4L\|\nabla f\|_\infty}{\kappa} \right] \leq \frac{\varepsilon}{2}.
\]

Furthermore, due to Assumption 3.1 and inequality (3.1), there exist \(K_1 \subseteq \mathbb{R}^d\) and \(h_0 \in (0, 1]\) with

\[
\frac{\lambda}{4} \mathcal{E}\left[ \frac{4}{\lambda} \left( \frac{\psi(h, x, \cdot) - x}{h} - \psi_0(x, \cdot) \right) \nabla f(x) \right] \\
\leq \frac{\lambda}{4} \mathcal{E}\left[ \frac{4}{\lambda} \left| \frac{\psi(h, x, \cdot) - x}{h} - \psi_0(x, \cdot) \right| \cdot \|\nabla f\|_\infty \right] \\
\leq \frac{\lambda}{8} \mathcal{E}\left[ \frac{8\|\nabla f\|_\infty}{\lambda} \left| \frac{\psi(h, x, \cdot) - x}{h} - \psi_0(x, \cdot) \right| \right] 1_{K_1}(\cdot) + \frac{\lambda}{16L\|\nabla f\|_\infty} \mathcal{E}\left[ \frac{16L\|\nabla f\|_\infty}{\lambda\kappa} 1_{K_1}(\cdot) \right] \leq \frac{\varepsilon}{8}
\]

for all \(x \in K\) and \(h \in (0, h_0]\). By Assumption 3.1(i) and inequality (3.1), there exist \(n \in \mathbb{N}\) with

\[
\frac{\lambda}{4} \mathcal{E}\left[ \frac{4}{\lambda} \left( \frac{\psi(h, x, \cdot) - x}{h} \right) \int_0^1 (\nabla f(x_s) - \nabla f(x)) \, ds \right] 1_{B(n)^c}(\cdot) \\
\leq \frac{\lambda}{4} \mathcal{E}\left[ \frac{8L\|\nabla f\|_\infty\kappa}{\lambda\kappa} 1_{B(n)^c}(\cdot) \right] \leq \frac{\varepsilon}{8},
\]

for all \(x \in \mathbb{R}^d\) and \(h \in (0, h_0]\). Since \(K \subseteq \mathbb{R}^d\) is compact, there exists \(\delta > 0\) with

\[
|\nabla f(x + y) - \nabla f(x)| < \frac{\varepsilon}{8Ln} \quad \text{for all } x \in K \text{ and } y \in B(\delta).
\]
Hence, by Assumption 3.1(i), there exist $h_1 \in (0, h_0]$ with
\[
\frac{\lambda}{4} \mathcal{E} \left[ \frac{4}{\lambda} \left( \frac{\psi(h, x, \cdot) - x}{h} \right) \int_0^1 \nabla f(x_s) - \nabla f(x) \, ds \right] \mathbf{1}_{B(n)}(\cdot) \leq \frac{\lambda}{4} \mathcal{E} \left[ \frac{4L_n}{\lambda} \frac{\varepsilon}{8L_n} \right] = \varepsilon
\]
for all $x \in K$ and $h \in (0, h_1]$. It follows from the previous estimates that
\[
\left( \frac{I(h)f - f}{h} \right)(x) - \mathcal{E}[^{\nabla f(x)}_{\psi(x, \cdot)}] \leq \varepsilon
\]
for all $x \in K$ and $h \in (0, h_1]$. Concerning the lower bound, Lemma A.1 yields
\[
\mathcal{E} \left[ \left( \frac{\psi(h, x, \cdot) - x}{h} \right) \int_0^1 \nabla f(x_s) \, ds \right] - \mathcal{E}[^{\nabla f(x)}_{\psi(x, \cdot)}] \\
\geq -\lambda \mathcal{E} \left[ \frac{\nabla f(x)\psi(x, \cdot) - y_s + y_s}{\lambda} \right] + \lambda \mathcal{E}[y_s]
\]
for all $h > 0$, $x \in \mathbb{R}^d$ and $\lambda \in (0, 1]$, where
\[
y_s := \left( \frac{\psi(h, x, \cdot) - x}{h} \right) \int_0^1 \nabla f(x_s) \, ds.
\]
Furthermore, we use the convexity of $\mathcal{E}$ to estimate
\[
\lambda \mathcal{E} \left[ \frac{\nabla f(x)\psi(x, \cdot) - y_s + y_s}{\lambda} \right] \\
\leq \frac{\lambda}{4} \mathcal{E} \left[ \frac{4}{\lambda} \left( \psi_0(x, \cdot) - \frac{\psi(h, x, \cdot) - x}{h} \right) \int_0^1 \nabla f(x_s) \, ds \right] \\
+ \frac{\lambda}{4} \mathcal{E} \left[ \frac{4}{\lambda} \left( \psi_0(x, \cdot) \int_0^1 \left( \nabla f(x) - \nabla f(x_s) \right) \, ds \right) \mathbf{1}_{B(n)}(\cdot) \right] \\
+ \frac{\lambda}{4} \mathcal{E} \left[ \frac{4}{\lambda} \left( \psi_0(x, \cdot) \int_0^1 \left( \nabla f(x) - \nabla f(x_s) \right) \, ds \right) \mathbf{1}_{B(n)\cap(\cdot)} \right] \\
+ \frac{\lambda}{4} \mathcal{E} \left[ \frac{4}{\lambda} \left( \frac{\psi(h, x, \cdot) - x}{h} \right) \int_0^1 \nabla f(x_s) \, ds \right].
\]
for all $n \in \mathbb{N}$. Analogously to the upper bound, one can estimate the terms on the right-hand side of the previous inequality to obtain
\[
\left( \frac{I(h)f - f}{h} \right)(x) - \mathcal{E}[^{\nabla f(x)}_{\psi(x, \cdot)}] \geq -\varepsilon
\]
for all $x \in K$ and sufficiently small $h > 0$. This shows
\[
\limsup_{h \downarrow 0} \left| \left( \frac{I(h)f - f}{h} \right)(x) - \mathcal{E}[^{\nabla f(x)}_{\psi(x, \cdot)}] \right| = 0 \quad \text{for all } K \subseteq \mathbb{R}^d.
\]
We obtain that $I'(0)f \in C_b$ exists and is given by
\[
(I'(0)f)(x) = \mathcal{E}[^{\nabla f(x)}_{\psi(x, \cdot)}] \quad \text{for all } x \in \mathbb{R}^d.
\]
Now, the first part of the claim follows from Theorem 2.9.

Third, we verify equation (3.3). Choose $t := 1$ and $h_n := 1/n$ for all $n \in \mathbb{N}$. Due to Corollary 2.10, the limit in equation (3.2) does not depend on the choice of the partition. In particular, for every $f \in C_b$ and $x \in \mathbb{R}^d$,
\[
(S(1)f)(x) = \lim_{n \to \infty} (I(\frac{1}{n})^n f)(x).
\]
For fixed $n \in \mathbb{N}$ and $h := 1/n$, we show by induction that

$$(I(h)^k f)(x) = h\mathcal{E} \left[ \frac{1}{h} f(\psi_k(h, x, \xi_1, \ldots, \xi_k)) \right]$$

for all $f \in C_b$, $x \in \mathbb{R}^d$ and $k \in \mathbb{N}$. For $k = 1$, we have

$$(I(h)f)(x) = h\mathcal{E} \left[ \frac{1}{h} f(\psi(h, x, \cdot)) \right] = h\mathcal{E} \left[ \frac{1}{h} f(\psi_1(h, x, \xi_1)) \right].$$

For the induction step, we use that $\xi_k$ is independent of $(\xi_1, \ldots, \xi_k)$ and has the same distribution as $\xi_1$ to obtain

$$(I(h)^{k+1} f)(x) = (I(h)^k I(h) f)(x) = h\mathcal{E} \left[ \frac{1}{h} (I(h)f)(\psi_k(h, x, \xi_1, \ldots, \xi_k)) \right]$$

$$= h\mathcal{E} \left[ \frac{1}{h} f(\psi(h, \psi_k(h, x, y_1, \ldots, y_k), \xi_{k+1})) \right] \bigg|_{(y_1, \ldots, y_k) = (\xi_1, \ldots, \xi_k)}$$

$$= h\mathcal{E} \left[ \frac{1}{h} f(\psi(h, \psi_k(h, x, \xi_1, \ldots, \xi_k), \xi_{k+1})) \right]$$

$$= h\mathcal{E} \left[ \frac{1}{h} f(\psi_{k+1}(h, x, \xi_1, \ldots, \xi_k, \xi_{k+1})) \right].$$

□

The existence of $(\Omega, \mathcal{H}, \mathcal{E})$ and $(\xi_n)_{n \in \mathbb{N}} \subset \mathcal{H}^d$ as in the previous theorem is always guaranteed by a nonlinear version of Kolmogorov’s extension theorem, see Theorem B.6.

For $\psi(t, x, y) := x + ty$ and a sublinear expectation $\mathcal{E}[\cdot]$, one can estimate

$$\left( \frac{I(h) f - f}{h} \right)(x) - \mathcal{E}[\nabla f(x) \xi] \leq \mathcal{E} \left[ \int_0^1 \left| \nabla f(x_s) - \nabla f(x) \right| |\xi| ds \right]$$

with $\xi := \text{id}_{\mathbb{R}^d}$ and the computation of the generator simplifies accordingly. However, the main ideas of the proof and the result remain the same. To give an explicit example for the sequence $(\psi_n)_{n \in \mathbb{N}}$, we consider the case that the sample $\xi_{n+1}$ is randomly shifted by a nonlinear function depending on the average of the previous samples $\xi_1, \ldots, \xi_n$.

**Corollary 3.3.** Let $\psi(t, x, y) := x + \varphi(t, x) + ty$ for a function $\varphi : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ such that there exist $L \geq 0$ with $\varphi(t, \cdot) \in \text{Lip}_p(Lt)$ for all $t \geq 0$. Furthermore, we assume that the limit $\varphi_0 := \lim_{h \to 0} \varphi(h\cdot)/h \in C_b$ exists. Then, there exists a strongly continuous convex monotone semigroup $(S(t))_{t \geq 0}$ on $C_b$ with

$$S(t) f = \lim_{n \to \infty} I \left( \frac{t}{n} \right)^n f \in C_b \quad \text{for all } t \geq 0 \text{ and } f \in C_b$$

which has the properties (i)-(v) from Theorem 2.9 with $\omega = 0$ and the constant $L \geq 0$ from above. For every $f \in C^1_b$ and $x \in \mathbb{R}^d$, it holds that $f \in D(A)$ and

$$(Af)(x) = \mathcal{E}[\nabla f(x) \xi] + \varphi_0(x) \nabla f(x), \quad \text{where } \xi := \text{id}_{\mathbb{R}^d}.$$  

In addition, for every convex expectation space $(\Omega, \mathcal{H}, \mathcal{E})$ and iid sequence $(\xi_n)_{n \in \mathbb{N}} \subset \mathcal{H}^d$ with $\mathcal{E}[\xi(\xi_n)] = \mathcal{E}[f]$ for all $f \in C_b$,

$$(S(1)f)(0) = \lim_{n \to \infty} \frac{1}{n} \mathcal{E}[nf(X_n)] \quad \text{for all } f \in C_b, \quad (3.5)$$

where $X_n := \psi_n \left( \frac{1}{n}, 0, \xi_1, \ldots, \xi_n \right)$.

**Proof.** It is straightforward to show that Assumption 3.1 is satisfied with

$$\psi_0(x, y) := \varphi_0(x) + y \quad \text{for all } x, y \in \mathbb{R}^d.$$  

Hence, the claim follows from Theorem 3.2. □
For non-perturbed averaged sums of iid samples, the semigroup can be represented explicitly by the Hopf–Lax formula.

**Theorem 3.4.** Let \( \psi(t, x, y) := x + ty \) for all \( t \geq 0 \) and \( x, y \in \mathbb{R}^d \). Then, the semigroup \((S(t))_{t \geq 0}\) from Theorem 3.2 has the representation

\[
(S(t)f)(x) = \sup_{y \in \mathbb{R}^d} \{ f(x + ty) - \varphi(y)t \} \quad \text{for all } t \geq 0, f \in C_b \text{ and } x \in \mathbb{R}^d,
\]

where \( \varphi(y) := \sup_{z \in \mathbb{R}^d} (yz - \mathcal{E}[z\xi]) \) and \( \xi := \text{id}_{\mathbb{R}^d} \). For every \( f \in C_b^1 \), it holds that

\[
(Af)(x) = \mathcal{E} [\nabla f(x) \xi] \quad \text{for all } x \in \mathbb{R}^d.
\]

In addition, for every convex expectation space \((\Omega, \mathcal{H}, \mathcal{E})\) and iid sequence \((\xi_n)_{n \in \mathbb{N}} \subset \mathcal{H}^d\) with \( \mathcal{E}[f(\xi_n)] = \mathcal{E}[f] \) for all \( f \in C_b \),

\[
\lim_{n \to \infty} \frac{1}{n} \mathcal{E} \left[ n f \left( \frac{1}{n} \sum_{i=1}^{n} \xi_i \right) \right] = \sup_{x \in \mathbb{R}^d} (f(x) - \varphi(x)) \quad \text{for all } f \in C_b.
\]

**Proof.** For every \( t \geq 0, f \in C_b \) and \( x \in \mathbb{R}^d \), we define

\[
(T(t)f)(x) := \sup_{y \in \mathbb{R}^d} \{ f(x + ty) - \varphi(y)t \}.
\]

We show that \((T(t))_{t \geq 0}\) is a semigroup satisfying Assumption 2.7. The convexity of the mapping \( \mathbb{R}^d \to \mathbb{R}, x \mapsto \mathcal{E}[x\xi] \) and Fenchel–Moreau’s theorem yield

\[
\mathcal{E}[x\xi] = \sup_{y \in \mathbb{R}^d} (xy - \varphi(y)) \quad \text{for all } x \in \mathbb{R}^d.
\]

In particular, we obtain \( \inf_{x \in \mathbb{R}^d} \varphi(x) = -\mathcal{E}[0] = 0 \). For every \( t \geq 0 \) and \( f \in C_b \), it follows from \( \lim_{|x| \to \infty} \varphi(x) = \infty \) that there exists \( K \in \mathbb{R}^d \) with

\[
(T(t)f)(x) = \sup_{y \in K} (f(x + ty) - \varphi(y)t) \quad \text{for all } x \in \mathbb{R}^d.
\]

Since \( f \) is uniformly continuous on compacts, we obtain \( T(t)f \in C_b \). Clearly, Assumption 2.7(i)-(iv) and (vii) are satisfied with \( \omega = L = 0 \), where \( T(t)0 = 0 \) follows from \( \inf_{x \in \mathbb{R}^d} \varphi(x) = 0 \). For every \( r \geq 0 \) and \( f \in \text{Lip}_b(r) \), we use

\[
f(x + ty) - \varphi(y)t \leq f(x) + (r|y| - \varphi(y)t)
\]

and \( \lim_{|x| \to \infty} \varphi(x)/|x| = \infty \) to choose \( K \in \mathbb{R}^d \) with

\[
(T(t)f)(x) = \sup_{y \in K} (f(x + ty) - \varphi(y)t) \quad \text{for all } t \geq 0 \text{ and } x \in \mathbb{R}^d. \tag{3.6}
\]

We obtain \( 0 \leq T(t)f - f \leq \sup_{y \in K} rt|y| \) for all \( t \geq 0 \) which shows \( \text{Lip}_b \subset C^T \). Let \( f \in C_b^1 \) and choose \( K \in \mathbb{R}^d \) such that equation (3.6) holds. For every \( x \in \mathbb{R}^d \), we use Fenchel–Moreau’s theorem to estimate

\[
\left| \frac{(T(t)f - f)(x)}{t} - \mathcal{E} [\nabla f(x) \xi] \right| \leq \sup_{y \in K} \left| \frac{f(x + ty) - f(x)}{t} - \nabla f(x)y \right|
\]

\[
\leq \sup_{y \in K} \frac{1}{t} \int_0^t |\nabla f(x + sy) - \nabla f(x)| \cdot |y| \, ds.
\]
Since $\nabla f$ is uniformly continuous on compacts, we obtain $f \in D(B)$ and $Af = Bf$, where $B$ denotes the generator of $(T(t))_{t \geq 0}$. In particular, Assumption 2.7(v) is satisfied. Next, we show that $(T(t))_{t \geq 0}$ forms a semigroup. For every $s, t \geq 0$, $f \in C_b$ and $x \in \mathbb{R}^d$,\[
abla (T(s + t)f)(x) = \sup_{y \in \mathbb{R}^d} \left( f(x + sy + ty) - \varphi(y)s - \varphi(y)t \right) \leq \sup_{y \in \mathbb{R}^d} \sup_{z \in \mathbb{R}^d} \left( f(x + sy + tz) - \varphi(y)s - \varphi(z)t \right) = (T(s)T(t)f)(x).\]
Furthermore, due to $\lim_{|x| \to \infty} \varphi(x) = \infty$, there exists $y_s, y_t \in \mathbb{R}^d$ with $(T(s)T(t)f)(x) = (T(t)f)(x + sy_s) - \varphi(y_s)s = f(x + sy_s + ty_t) - \varphi(y_s)s - \varphi(y_t)t$.
For $y_{s+t} := \frac{s}{s+t}y_s + \frac{t}{s+t}y_t$, it follows from the convexity of $\varphi$ that\[
\varphi(y_{s+t})(s+t) = \varphi \left( \frac{s}{s+t}y_s + \frac{t}{s+t}y_t \right) \leq \varphi(y_s)s + t\varphi(y_t)t.
\]
We obtain\[
(T(s)T(t)f)(x) = f(x + (s + t)y_{s+t}) - \varphi(y_s)s - \varphi(y_t)t \leq f(x + (s + t)y_{s+t}) - \varphi(y_{s+t})(s + t) \leq (T(s + t)f)(x).
\]
It remains to verify Assumption 2.7(vi). For every $t \geq 0$, $x \in \mathbb{R}^d$ and $(f_n)_{n \in \mathbb{N}} \subset C_b$ with $f_n \downarrow 0$, there exists $K \subset \mathbb{R}^d$ with $(T(t)f_n)(x) = \sup_{y \in K} \left( f_n(x + ty) - \varphi(y)t \right)$ for all $n \in \mathbb{N}$ such that Dini’s theorem implies $(T(t)f_n)(x) \downarrow 0$ as $n \to \infty$. Since the mapping $\mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$, $(t, x) \mapsto (T(t)f_n)(x)$ is continuous for all $n \in \mathbb{N}$, we can use Dini’s theorem again to obtain\[
\sup_{(s, x) \in [0, t] \times K} (T(s)f)(x) \downarrow 0 \text{ for all } t \geq 0 \text{ and } K \subset \mathbb{R}^d.
\]
The previous statement remains valid for sequences $(f_n)_{n \in \mathbb{N}} \subset C_b$ because every function in $C_b$ can be approximated from above by a decreasing sequence in $\text{Lip}_b$. Finally, we can apply Corollary 2.10 to obtain $S(t)f = T(t)f$ for all $t \geq 0$ and $f \in C_b$. Furthermore,\[
\psi_n(t, x, y_1, \ldots, y_n) = x + t \sum_{i=1}^n y_i
\]
for all $n \in \mathbb{N}$, $t \geq 0$ and $x, y_1, \ldots, y_n \in \mathbb{R}^d$. Thus, the second part of the claim follows from Theorem 3.2. \hfill \Box

The previous theorem extends Peng’s results from the sublinear to the convex case. Indeed, if $\mathcal{E}[\cdot]$ is a sublinear expectation, it holds that\[
\lim_{n \to \infty} \mathcal{E} \left[ f \left( \frac{1}{n} \sum_{i=1}^n \xi_i \right) \right] = \sup_{\{x \in \mathbb{R}^d : \varphi(x) = 0\}} f(x) \text{ for all } f \in C_b.
\]
Hence, we obtain convergence to a maximal distribution as in [39, 41] and the limit in Theorem 3.4 can be seen as a convex version thereof. Furthermore, as an immediate consequence of the previous result, we obtain Cramér’s theorem as LLN for the entropic
risk measure. Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space and \((\xi_n)_{n \in \mathbb{N}}\) be an iid sequence of random vectors \(\xi_n : \Omega \to \mathbb{R}^d\) with \(\mathbb{E}[e^{c\xi_n}] < \infty\) for all \(c \geq 0\). Let
\[
\Lambda : \mathbb{R}^d \to \mathbb{R}, \ x \mapsto \log \left( \mathbb{E}[e^{x\xi}] \right)
\]
be the logarithmic moment generating function and denote by
\[
\Lambda^* : \mathbb{R}^d \to \mathbb{R}, \ x \mapsto \sup_{y \in \mathbb{R}^d} (xy - \Lambda(y))
\]
its convex conjugate. Moreover, we define \(X_n := \frac{1}{n} \sum_{i=1}^n \xi_i\) for all \(n \in \mathbb{N}\).

**Corollary 3.5** (Cramér). For every \(f \in C_b\),
\[
\lim_{n \to \infty} \frac{1}{n} \log \left( \mathbb{E}[e^{nf(X_n)}] \right) = \sup_{x \in \mathbb{R}^d} (f(x) - \Lambda^*(x)).
\]

**Proof.** Clearly, the functional \(\mathcal{E} : C_{\bar{R}} \to \mathbb{R}, \ f \mapsto \log(\mathbb{E}[e^{f(\xi)}])\) is convex and monotone with \(\mathcal{E}[c] = c\) for all \(c \in \mathbb{R}\). In addition, the dominated convergence theorem implies \(\lim_{n \to \infty} \mathcal{E}[\cdot | 1_{B(n,c)}] = 0\) which shows that \(\mathcal{E}\) is continuous from above. Hence, we can apply Theorem 3.4 to obtain the claim. \(\square\)

### 3.2. Upper bounds and convergence rates.

Let \(\mathcal{E} : C_{\bar{R}} \to \mathbb{R}\) be a convex expectation which is continuous from above and \(\psi : \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d\) be a function satisfying Assumption 3.1. For \(\psi(t, x, y) := x + ty\), the semigroup that we obtain in the limit can explicitly be represented by the Hopf–Lax formula; see Theorem 3.4. In general, such a formula is not available but we can still provide explicit (upper) bounds and resulting convergence rates. Let \(H_\pm : \mathbb{R}^d \to \mathbb{R}\) be convex functions with
\[
H_-(y) \leq \mathcal{E}[\psi_0(y, \cdot)] \leq H_+(y) \quad \text{for all} \quad x, y \in \mathbb{R}^d
\]
and \(H_+(0) = 0\). Note that at least an upper bound always exists, since we can choose
\[
H_+(y) := \sup_{x \in \mathbb{R}^d} \mathcal{E}[\psi_0(x, \cdot)] \leq \mathcal{E}[L(1 + |\cdot|)]
\]
Moreover, under the assumptions of Corollary 3.3 we can choose
\[
H_+(y) := \mathcal{E}[y\xi] + \sup_{x \in \mathbb{R}^d} \varphi_0(x)y \leq \mathcal{E}[y\xi] + L|y|, \quad \text{where} \quad \xi := \text{id}_{\mathbb{R}^d}.
\]

For the following lemma, let \((S(t))_{t \geq 0}\) be the semigroup from Theorem 3.2 given by
\[
S(t)f = \lim_{n \to \infty} I_n(t) f \quad \text{for all} \quad t \geq 0 \quad \text{and} \quad f \in C_b,
\]
where \(I(0) := \text{id}_{C_b}\) and, for every \(t > 0\), \(f \in C_b\) and \(x \in \mathbb{R}^d\),
\[
(I(t)f)(x) := t\mathcal{E} \left[ \frac{1}{t} f(\psi(t, x, \cdot)) \right].
\]

**Lemma 3.6.** It holds that \(S_-(t)f \leq S(t)f \leq S_+(t)f\) for all \(t \geq 0\) and \(f \in C_b\), where
\[
(S_\pm(t)f)(x) := \sup_{y \in \mathbb{R}^d} \left( f(x + ty) - H^*_\pm(y)t \right)
\]
and \(H^*_\pm(y) := \sup_{z \in \mathbb{R}^d} (yz - H_\pm(z))\).

**Proof.** As seen during the proof of Theorem 3.4, the families \((S_\pm(t))_{t \geq 0}\) are semigroups satisfying Assumption 2.7 with generators
\[
(A_\pm f)(x) = H_\pm(\nabla f(x)) \quad \text{for all} \quad f \in C_b^1 \quad \text{and} \quad x \in \mathbb{R}^d.
\]

In particular, it holds that \(A_-f \leq Af \leq A_+f\) for all \(f \in C_b^1\) and Corollary 2.10 implies
\[
S_-(t)f \leq S(t)f \leq S_+(t)f \quad \text{for all} \quad t \geq 0 \quad \text{and} \quad f \in C_b. \quad \square
\]
We illustrate the previous estimate for the entropic risk measure which yields exponential convergence rates in the situation of Corollary 3.3, where we considered averaged sums of perturbed iid samples. Recall that the perturbation of the $\xi_{n+1}$ consisted of a random shift by a nonlinear depending on the average of the previous samples $\xi_1, \ldots, \xi_n$. It turns out that the well-known convergence rate from the case of unperturbed iid samples is reduced according to the size of the shift, see Theorem 3.7. Furthermore, if we only require that $(\xi_n)_{n \in \mathbb{N}}$ has finite $p$-th moments instead of finite exponential moments, we still obtain polynomial convergence rates as in [34], see Theorem 3.8. For the following two theorems, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\xi_n)_{n \in \mathbb{N}}$ be an iid sequence of random vectors $\xi_n : \Omega \to \mathbb{R}^d$. Furthermore, let $\varphi : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ be a function such that there exist $L \geq 0$ with $\varphi(t, \cdot) \in \text{Lip}_b(Lt)$ for all $t \geq 0$ and suppose that the limit $\varphi_0 := \lim_{h \downarrow 0} \varphi(h, \cdot)/h \in C_b$ exists. We define $\psi(t, x, y) := x + \varphi(t, x) + ty$ and recursively a sequence $(\psi_n)_{n \in \mathbb{N}}$ of functions $\psi_n : \mathbb{R}_+ \times \mathbb{R}^d \times (\mathbb{R}^d)^n \to \mathbb{R}^d$ by $\psi_1 := \psi$ and

\[ \psi_{n+1}(t, x, y_1, \ldots, y_{n+1}) := \psi(t, \psi_n(t, x, y_1, \ldots, y_n), y_{n+1}). \]

Finally, let $X_n := \psi_n(\frac{1}{n}, 0, \xi_1, \ldots, \xi_n)$ for all $n \in \mathbb{N}$.

**Theorem 3.7.** Assume that $\mathbb{E}[e^{\xi_1}] < \infty$ for all $c \geq 0$. Denote by

\[ \Lambda : \mathbb{R}^d \to \mathbb{R}, \ x \mapsto \log (\mathbb{E}[e^{\xi_1}]) \]

the logarithmic moment generating function and by

\[ \Lambda^* : \mathbb{R}^d \to \mathbb{R}, \ x \mapsto \sup_{y \in \mathbb{R}^d} (xy - \Lambda(y)) \]

its convex conjugate. Then,

\[ \lim_{n \to \infty} \frac{1}{n} \log (\mathbb{E}[e^{nf(X_n)}]) \leq \sup_{x \in \mathbb{R}^d} (f(x) - H_+(x)) \]

for all $f \in C_b$, where $H_+(x) := \inf_{y \in B(L)} \Lambda^*(x + y)$. Furthermore,

\[ \limsup_{n \to \infty} \frac{1}{n} \log (\mathbb{P}(X_n \in A)) \leq - \inf_{x \in A_L} \Lambda^*(x) \]

for all closed sets $A \subset \mathbb{R}^d$, where $A_L := \{x + y : x \in A, y \in B(L)\}$.

**Proof.** Applying Corollary 3.3 with $\mathcal{E}[f] := \log (\mathbb{E}[e^{f(\xi_1)}])$ yields a semigroup $(S(t))_{t \geq 0}$ with generator $(Af)(x) = \Lambda(\nabla f(x)) + \varphi_0(x) \nabla f(x)$ for all $f \in C^1_b$ and $x \in \mathbb{R}^d$. Moreover, for every $t \geq 0$, $f \in C_b$ and $x \in \mathbb{R}^d$, we define $H_+(x) := \Lambda(x) + L|x|$ and

\[ (S_+(t)f)(x) := \sup_{y \in \mathbb{R}^d} (f(x + ty) - H_+(y)t) \]

where $H_+(y) := \sup_{z \in \mathbb{R}^d} (yz - H_+(z))$. Lemma 3.6 implies $S(t)f \leq S_+(t)f$ for all $t \geq 0$ and $f \in C_b$. Moreover, by Fenchel–Moreau’s theorem, the function

\[ f : \mathbb{R}^d \to \mathbb{R}, \ x \mapsto \inf_{x = y + z} (\Lambda^*(y) + \infty \mathbb{1}_{B(L)^c}(z)) \]

satisfies $f^*(x) = \Lambda^*(x) + L|x| = H_+(x)$ and thus

\[ H_+(x) = f(x) = \inf_{y \in B(L)} \Lambda^*(x + y). \]

Let $f := -\infty \mathbb{1}_{A^c} \in U_b$ for a closed subset $A \subset \mathbb{R}^d$ and $(f_n)_{n \in \mathbb{N}} \subset C_b$ be a sequence with $f_n \downarrow f$, where $U_b$ consists of all upper semicontinuous functions $g : \mathbb{R}^d \to [-\infty, \infty)$.
with \( \|g^+\|_\infty < \infty \). Due to Dini’s theorem and \( \lim_{|x| \to \infty} H_+^*(x) = \infty \), the functionals
\[
\Phi: U_b \to [-\infty, \infty), \ g \mapsto \sup_{x \in \mathbb{R}^d} \left( f(x) - H_+^*(x) \right),
\]
\[
\Phi_n: U_b \to [-\infty, \infty), \ g \mapsto \frac{1}{n} \log \left( \mathbb{E}[e^{nf(X_n)}] \right) \quad \text{for all } n \in \mathbb{N}
\]
are continuous from above. Hence, by changing a supremum with an infimum at the cost of an inequality, we obtain
\[
\limsup_{n \to \infty} \frac{1}{n} \log(\mathbb{P}(X_n \in A)) = \limsup_{n \to \infty} \Phi_n(f) = \limsup_{n \to \infty} \inf_{k \in \mathbb{N}} \Phi_n(f_k)
\]
\[
\leq \inf_{k \in \mathbb{N}} \sup_{n \in \mathbb{N}} \Phi_n(f_k) = \inf_{k \in \mathbb{N}} \limsup_{n \to \infty} \Phi_n(f_k)
\]
\[
\leq \inf_{k \in \mathbb{N}} \Phi(f_k) = \Phi(f) = - \inf_{x \in A_L} \Lambda^*(x). \quad \square
\]

Replacing the entropic risk measure by the short fall risk measure leads to polynomial rather than exponential convergence rates.

**Theorem 3.8.** Assume that \( \mathbb{E}[|\xi|^p] < \infty \) for some \( p \in (1, \infty) \). Define
\[
\Lambda: \mathbb{R}^d \to \mathbb{R}, \ x \mapsto \inf\{m \in \mathbb{R}: \mathbb{E}[(1 + x\xi - m)^+] \leq 1\}
\]
and the convex conjugate
\[
\Lambda^*: \mathbb{R}^d \to \mathbb{R}, \ x \mapsto \sup_{y \in \mathbb{R}^d} (xy - \Lambda(y)).
\]

Then, for every closed subset \( A \subset \mathbb{R}^d \),
\[
\limsup_{n \to \infty} n^{p-1} \mathbb{P}(X_n \in A) \leq \left( \inf_{x \in A_L} \Lambda^*(x) \right)^{-p},
\]
where \( A_L := \{x + y: x \in A, y \in B(L)\} \).

**Proof.** We consider the short fall risk measure
\[
\mathcal{E}: C_{\mathbb{R}} \to \mathbb{R}, \ f \mapsto \inf\{m \in \mathbb{R}: \mathbb{E}[(1 + f(\xi) - m)^+] \leq 1\}
\]
which is indeed a convex expectation and continuous from above, see [23, Chapter 4.9]. Hence, Corollary 3.3 yields a corresponding semigroup \( (S(t))_{t \geq 0} \) with generator
\[
(Af)(x) = \Lambda(\nabla f(x)) + \varphi_0(x)\nabla f(x) \quad \text{for all } f \in C_b^1 \text{ and } x \in \mathbb{R}^d.
\]

Similar to the proof of Theorem 3.7, by using that the functionals \( \Phi(f) := (S(1)f)(0) \) and \( \Phi_n(f) := (I(\frac{1}{n})^n f)(0) \) are continuous from above on \( U_b \), one can show that
\[
\limsup_{n \to \infty} (I(\frac{1}{n})^n f)(0) \leq - \inf_{x \in A_L} \Lambda^*(x),
\]
where \( f := -\infty 1_{A^c} \) and \( (I(t)f)(x) := t\mathcal{E}[\frac{1}{t} f(x + t\xi_1)] \). It remains to show that
\[
-n^{-\frac{p-1}{p}} \mathbb{P}(X_n \in A)^{-\frac{1}{p}} \leq (I(\frac{1}{n})^n f)(0) \quad \text{for all } n \in \mathbb{N}.
\]

To do so, we want to apply [34, Lemma 4.2]. Let \( \mathcal{P} \) be the set of all probability measures on the Borel-\( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}^d) \). By Fenchel–Moreau’s theorem, we have
\[
\mathcal{E}[f] = \sup_{\nu \in \mathcal{P}} \left( \int_{\mathbb{R}^d} f \, d\nu - \alpha(\nu) \right) \quad \text{for all } f \in C_b,
\]
where \( \alpha(\nu) := \sup_{f \in C_b} \left( \int_{\mathbb{R}^d} f \, d\nu - \mathcal{E}[f] \right) \). By induction, one can show that
\[
(I \left( \frac{1}{n} n \right)n)(0) = \sup_{\nu \in \mathcal{P}^n} \left( \int_{\mathbb{R}^d} \psi \left( \frac{1}{n}, 0, x_1, \ldots, x_n \right) \nu(dx_1, \ldots, dx_n) - \frac{1}{n} \alpha_n(\nu) \right)
\]
for all \( n \in \mathbb{N} \) and \( f \in C_b \), where \( \mathcal{P}^n \) consists of all probability measures on the Borel-\( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}^d)^n \). Moreover, the penalization functions \( \alpha_n : \mathcal{P}^n \to [0, \infty] \) are defined by
\[
\alpha_n(\nu) := \int_{\mathbb{R}^d} \sum_{i=1}^n \alpha(\nu_{i-1,i}(x_1, \ldots, x_{i-1})) \nu(dx_1, \ldots, dx_n),
\]
where the kernels \( \nu_{i-1,i} \) are determined by the disintegration
\[
\nu(dx_1, \ldots, dx_n) = \nu_{0,1}(dx_1) \prod_{i=2}^n \nu_{i-1,i}(x_1, \ldots, x_{i-1})(dx_i).
\]
It follows from [34, Lemma 4.2] that
\[
\alpha_n(\nu) \leq n^{\frac{1}{p}} \int \left\| \frac{d\nu}{d\mu^n} \right\|_{L^p(\mu^n)} \quad \text{for all } \nu \ll \mu,
\]
where \( \mu := \mathbb{P} \circ \xi_1^{-1} \) and \( \mu^n := \mu \otimes \cdots \otimes \mu \) denotes the \( n \)-fold product measure. Now, the claim follows similarly to the proof of [34, Theorem 1.2]. \( \square \)

3.3. **Uncertain samples in Wasserstein space.** As an illustration of the previous abstract results, we consider convex expectations that are defined as a supremum over a set of probability measures which are weighted according to their Wasserstein distance to a fixed reference model. This type of uncertainty has previously been studied in a framework with nonlinear semigroups corresponding to Markov processes with uncertain transition probabilities, see [7,25]. Let \( p \in (1, \infty) \) and denote by \( \mathcal{P}_p \) the \( p \)-Wasserstein space consisting of all probability measures on the Borel-\( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}^d) \) with finite \( p \)-th moment. We endow \( \mathcal{P}_p \) with the \( p \)-Wasserstein distance
\[
\mathcal{W}_p(\mu, \nu) := \left( \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^\frac{1}{p} \pi(dx, dy) \right)^{\frac{p}{n}} \quad \text{for all } \mu, \nu \in \mathcal{P}_p,
\]
where \( \Pi(\mu, \nu) \) consists of all probability measures on \( \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d) \) with first marginal \( \mu \) and second marginal \( \nu \). Furthermore, let \( \varphi : \mathbb{R}_+ \to [0, \infty] \) be a function with \( \varphi(0) = 0 \) and \( \lim_{c \to \infty} \varphi(c)/c = \infty \). In the sequel, we fix \( \mu \in \mathcal{P}_p \) and define
\[
\mathcal{E} : C_\mathcal{K} \to \mathbb{R}, \ f \mapsto \sup_{\nu \in \mathcal{P}_p} \left( \int_{\mathbb{R}^d} f(x) \nu(dx) - \varphi(\mathcal{W}_p(\mu, \nu)) \right). \tag{3.7}
\]
In the case \( \varphi := \infty \chi_{[0,\infty]} \), the functional \( \mathcal{E} \) is defined as supremum over all measures in a Wasserstein ball with radius \( r \geq 0 \) around the reference model \( \mu \). Moreover, for every \( t > 0, f \in C_b \) and \( x \in \mathbb{R}^d \),
\[
t\mathcal{E} \left[ \frac{1}{t} f(x + t\xi) \right] = \sup_{\nu \in \mathcal{P}_p} \left( \int_{\mathbb{R}^d} f(x + ty) \nu(dy) - t \varphi(\mathcal{W}_p(\mu, \nu)) \right) = \sup_{\nu \in \mathcal{P}_p} \left( \int_{\mathbb{R}^d} f(x + y) \nu_t(dy) - t \varphi \left( \frac{\mathcal{W}_p(\mu_t, \nu_t)}{t} \right) \right) = \sup_{\nu \in \mathcal{P}_p} \left( \int_{\mathbb{R}^d} f(x + y) \nu(dy) - t \varphi \left( \frac{\mathcal{W}_p(\mu_t, \nu)}{t} \right) \right),
\]
where $\xi := \text{id}_{\mathbb{R}^d}$, $\nu_t := \nu \circ (t\xi)^{-1}$ and $\mu_t := \mu \circ (t\xi)^{-1}$. Hence, due to the expression in the last line, the definition

$$(I(t)f)(x) := t\mathcal{E} \left[ \frac{1}{t} f(x + t\xi) \right]$$

is consistent with the scaling of the penalization function $\varphi$ in [7, 25]. In the following theorem, choosing $p, q \in (1, \infty)$ with $1/p + 1/q = 1$, $\varphi(c) := c^q/p$ and $m = 0$ yields the generator $(Af)(x) = |\nabla f(x)|^q/q$ for all $f \in C^1_b$ and $x \in \mathbb{R}^d$.

**Theorem 3.9.** The functional $\mathcal{E}$ defined by equation (3.7) is a convex expectation which is continuous from above. Hence, there exists a strongly continuous convex monotone semigroup $(S(t))_{t \geq 0}$ on $C_b$ with

$$S(t)f = \lim_{n \to \infty} I \left( \frac{t}{n} \right)^n f \in C_b \quad \text{for all } t \geq 0 \text{ and } f \in C_b$$

which has the properties (i)-(v) from Theorem 2.9 with $\omega = L = 0$. For every $f \in C^1_b$, it holds that $f \in D(A)$ and

$$(Af)(x) = \sup_{c \geq 0} (c|\nabla f(x)| - \varphi(c)) + m \nabla f(x) \quad \text{for all } x \in \mathbb{R}^d,$$

where $m := \int_{\mathbb{R}^d} y \mu(dy)$. Furthermore, let $(\Omega, \mathcal{H}, \mathcal{E})$ be a convex expectation space and $(\xi_n)_{n \in \mathbb{N}}$ be an iid sequence of random vectors $\xi_n : \Omega \to \mathbb{R}^d$ with $\mathcal{E}[f(\xi_n)] = \mathcal{E}[f]$ for all $f \in C_b$. Then, for every $f \in C_b$, $t > 0$ and $x \in \mathbb{R}^d$,\n
$$\lim_{n \to \infty} \frac{t}{n} \mathcal{E} \left[ \frac{n}{t} f \left( x + \frac{t}{n} \sum_{i=1}^n \xi_i \right) \right] = \sup_{y \in \mathbb{R}^d} (f(x + t(m + y)) - \varphi(|y|)t).$$

**Proof.** First, we show that $\mathcal{E}$ is a convex expectation which is continuous from above. Let $f \in C^\infty$ and choose $c \geq 0$ with $|f(x)| \leq c(1 + |x|)$ for all $x \in \mathbb{R}^d$. We use

$$\left| \int_{\mathbb{R}^d} |x| \mu(dx) - \int_{\mathbb{R}^d} |y| \nu(dy) \right| \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |x| - |y| |\pi(dx, dy)$$

for all $\pi \in \Pi(\mu, \nu)$, Hölder’s inequality and $\lim_{c \to \infty} \varphi(c)/c = \infty$ to estimate

$$\mathcal{E}[|f|] \leq \sup_{\nu \in \mathcal{P}_p} \left( \int_{\mathbb{R}^d} c(1 + |x|) \nu(dx) - \varphi(W_p(\mu, \nu)) \right)$$

$$\leq \sup_{\nu \in \mathcal{P}_p} \left( \int_{\mathbb{R}^d} c(1 + |x|) \mu(dx) + \int_{\mathbb{R}^d} c|x| \nu(dx) - \int_{\mathbb{R}^d} c|x| \mu(dx) - \varphi(W_p(\mu, \nu)) \right)$$

$$\leq \int_{\mathbb{R}^d} c(1 + |x|) \mu(dx) + \sup_{\nu \in \mathcal{P}_p} (cW_1(\mu, \nu) - \varphi(W_p(\mu, \nu)))$$

$$\leq \int_{\mathbb{R}^d} c(1 + |x|) \mu(dx) + \sup_{\nu \in \mathcal{P}_p} (cW_p(\mu, \nu) - \varphi(W_p(\mu, \nu))) < \infty.$$
where \( M := \{ \nu \in \mathcal{P}_p : \mathcal{W}_p(\mu, \nu) \leq R \} \). For every \( \varepsilon > 0 \), Lemma C.1 implies that there exists \( r \geq 0 \) with

\[
\sup_{\nu \in M} \int_{B(r)^c} c(1 + |x|) \nu(dx) \leq \frac{\varepsilon}{2}.
\]

Moreover, we can use Dini’s theorem to choose \( n_0 \in \mathbb{N} \) with

\[
\int_{\mathbb{R}^d} f_n(x) \nu(dx) \leq \int_{B(r)} f_n(x) \nu(dx) + \int_{B(r)^c} c(1 + |x|) \nu(dx) \leq \varepsilon
\]

for all \( n \geq n_0 \) and \( \nu \in M \). We obtain \( \mathcal{E}[f_n] \downarrow 0 \) as \( n \to \infty \). Now, Theorem 3.2 yields the existence of the semigroup \( (S(t))_{t \geq 0} \).

Second, for every \( x \in \mathbb{R}^d \), we show that

\[
\mathcal{E}[x\xi] = \sup_{c \geq 0} (c|x| - \varphi(c)) + \int_{\mathbb{R}^d} xy \mu(dy).
\]

W.l.o.g., let \( x \neq 0 \). For every \( c \geq 0 \), we choose \( \nu := \mu * \delta_{\frac{x}{|x|}} \). Then,

\[
\int_{\mathbb{R}^d} xy \nu(dy) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} x(y+z) \mu(dy) \delta_{\frac{z}{|z|}}(dz) = c|x| + \int_{\mathbb{R}^d} xy \mu(dy).
\]

We take the supremum over \( c \geq 0 \) and use \( \mathcal{W}_p(\mu, \nu) = c \) to conclude

\[
\mathcal{E}[x\xi] \geq \sup_{c \geq 0} (c|x| - \varphi(c)) + \int_{\mathbb{R}^d} xy \mu(dy).
\]

For every \( \nu \in \mathcal{P}_p \), it follows from

\[
\left| \int_{\mathbb{R}^d} xy \mu(dy) - \int_{\mathbb{R}^d} xz \nu(dz) \right| \leq |x| \int_{\mathbb{R}^d \times \mathbb{R}^d} |y-z| \pi(dy, dz)
\]

for all \( \pi \in \Pi(\mu, \nu) \) and Hölder’s inequality that

\[
\int_{\mathbb{R}^d} xy \nu(dy) - \varphi(\mathcal{W}_p(\mu, \nu)) = \int_{\mathbb{R}^d} xy \mu(dy) + \int_{\mathbb{R}^d} xy \nu(dy) - \int_{\mathbb{R}^d} xy \mu(dy) - \varphi(c)
\]

\[
\leq \int_{\mathbb{R}^d} xy \mu(dy) + \mathcal{W}_1(\mu, \nu)|x| - \varphi(c)
\]

\[
\leq \int_{\mathbb{R}^d} xy \mu(dy) + c|x| - \varphi(c),
\]

where \( c := \mathcal{W}_p(\mu, \nu) \). Taking the supremum over \( \nu \in \mathcal{P}_p \) yields

\[
\mathcal{E}[x\xi] \leq \sup_{c \geq 0} (c|x| - \varphi(c)) + \int_{\mathbb{R}^d} xy \mu(dy).
\]

In particular, for every \( f \in C^1_b \) and \( x \in \mathbb{R}^d \), we obtain from Theorem 3.4 that

\[
(Af)(x) = \mathcal{E}[\nabla f(x)\xi] = \sup_{c \geq 0} (c|\nabla f(x)| - \varphi(c)) + m \nabla f(x),
\]

where \( m := \int_{\mathbb{R}^d} y \mu(dy) \). Moreover, for every \( x \in \mathbb{R}^d \),

\[
\mathcal{E}[x\xi] = \sup_{y \in \mathbb{R}^d} (xy - \psi(y)), \text{ where } \psi(y) := \varphi(|y - m|).
\]

Hence, Fenchel–Moreau’s theorem and Theorem 3.4 imply

\[
(S(t)f)(x) = \sup_{y \in \mathbb{R}^d} \left( f(x + t(m + y)) - \varphi(|y|) \right)
\]

for all \( t \geq 0 \), \( f \in C_b \) and \( x \in \mathbb{R}^d \). We apply again Theorem 3.4 to obtain the last part of the statement. \( \square \)
We conclude this subsection by showing that the semigroup (and thus the distribution) which we obtain in the limit is the same whether we define the convex expectation as the supremum over an uncertainty set of measures or a set of parameters in \( \mathbb{R}^d \). The latter corresponds to shifting the fixed measures \( \mu \) in all possible deterministic directions. We define
\[
\tilde{\mathcal{E}} : C_{\tilde{\kappa}} \to \mathbb{R}, \ f \mapsto \sup_{\lambda \in \mathbb{R}^d} \left( \int_{\mathbb{R}^d} f(x + \lambda) \mu(dx) - \varphi(|\lambda|) \right),
\]
\( J(0) := \text{id}_{C_b} \) and, for every \( t > 0, f \in C_b \) and \( x \in \mathbb{R}^d \),
\[
(J(t)f)(x) := t\tilde{\mathcal{E}} \left[ \frac{1}{t} f(x + t\xi) \right].
\]

**Corollary 3.10.** Denoting by \((S(t))_{t \geq 0}\) the semigroup from Theorem 3.9, we have
\[
S(t)f = \lim_{n \to \infty} J(\frac{t}{n})^n f \quad \text{for all } t \geq 0 \text{ and } f \in C_b.
\]

**Proof.** It follows from \( \mathcal{W}_p(\mu, \nu) = |\lambda| \) for \( \nu := \mu * \delta_\lambda \) with \( \lambda \in \mathbb{R}^d \) that \( \tilde{\mathcal{E}}[[f]] \leq \mathcal{E}[[f]] \) for all \( f \in C_{\tilde{\kappa}} \). Hence, \( \tilde{\mathcal{E}} : C_{\tilde{\kappa}} \to \mathbb{R} \) is a well-defined convex expectation which is continuous from above. By Theorem 3.4, there exists a semigroup \((T(t))_{t \geq 0}\) on \( C_{\tilde{\kappa}} \) with
\[
T(t)f = \lim_{n \to \infty} J(\frac{t}{n})^n f \quad \text{for all } t \geq 0 \text{ and } f \in C_b
\]
and generator \((Bf)(x) = \tilde{\mathcal{E}}[\nabla f(x)\xi]\) for all \( f \in C^1_b \) and \( x \in \mathbb{R}^d \). By Theorem 3.9 and a straightforward computation, it holds that
\[
(Af)(x) = \sup_{c \geq 0} \left( c\nabla f(x) - \varphi(c) \right) + m\nabla f(x) = (Bf)(x)
\]
for all \( f \in C^1_b \) and \( x \in \mathbb{R}^d \). Hence, Corollary 2.10 implies \( S(t)f = T(t)f \) for all \( t \geq 0 \) and \( f \in C_b \).

4. **Second order scaling limits**

Throughout this section, we choose the weight function
\[
\kappa : \mathbb{R}^d \to (0, \infty), \ x \mapsto (1 + |x|)^{-2}.
\]
In analogy to the previous section, let \((\xi_n)_{n \in \mathbb{N}} \subset \mathcal{H}^d\) be a sequence of iid random vectors on a convex expectation space \((\Omega, \mathcal{H}, \mathcal{E})\) with finite second moments and \( \mathcal{E}[a\xi_1] = 0 \) for all \( a \in \mathbb{R}^d \). For the sake of readability, we refrain from introducing a sequence \((\psi_n)_{n \in \mathbb{N}}\) of recursively defined functions as in Section 3 and study only the limit behaviour of the sequence
\[
X_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i
\]
which corresponds to choosing \( \psi(t, x, y) := x + \sqrt{t}y \). Again, we are interested in the behaviour of the distributions of \( X_n \) and requiring finite second moments means that the distribution of \( \xi_1 \) is well-defined even for continuous functions with at most quadratic growth at infinity, i.e.,
\[
F_{\xi_1} : C_{\tilde{\kappa}} \to \mathbb{R}, \ f \mapsto \tilde{\mathcal{E}}[f(\xi_1)].
\]
Denoting this functional by \( \mathcal{E}[-] \), we remark that continuity from above on \( C_{\tilde{\kappa}} \) is equivalent to the uniform integrability condition \( \lim_{c \to \infty} \tilde{\mathcal{E}}[|\xi_1|^2 - c] = 0 \) from [11]. Defining
\[
(I(t)f)(x) := t\tilde{\mathcal{E}} \left[ \frac{1}{t} f(x + \sqrt{t}\xi_1) \right]
\]
for all \( t > 0, f \in C_b \) and \( x \in \mathbb{R}^d \) leads to
\[
(S(1)f)(0) = \lim_{n \to \infty} \left( I \left( \frac{1}{n} \right) \right)^n f(0) = \lim_{n \to \infty} \frac{1}{n} \bar{\mathcal{E}} \left[ n f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \right) \right].
\]

This result is proven in Subsection 4.1 analogously to Theorem 3.2. Hence, it seems very likely that a more general result including a sequence \((\psi_n)_{n \in \mathbb{N}}\) of recursively defined functions can be obtained mainly at the cost of additional terms appearing in the estimates which would make the proof quite lengthy and might disguise the main differences between the first and second order case. In Subsection 4.2, we illustrate again the abstract results by considering convex expectations that are defined as a (weighted) supremum over a set of probability measures. The necessary modifications of the setting will be explained in full detail. We also want to mention that this approach allows for a generalization of the previous results from [3,25] about Markov processes with uncertain transition probabilities, which were restricted to first order perturbations, to the second order case. The possibility of this extension was already conjectured in [3] but whether the original construction of the semigroup based on monotone convergence can be transferred remains an open question.

4.1. CLT for convex expectations. Let \( \mathcal{E} : C_{\bar{\kappa}} \to \mathbb{R} \) be a convex expectation which is continuous from above. By [7, Theorem C.1], there exists a convex monotone extension \( \mathcal{E} : B_{\bar{\kappa}} \to \mathbb{R} \) such that, for every \( \varepsilon > 0 \) and \( c \geq 0 \), there exists \( K \in \mathbb{R}^d \) with
\[
\mathcal{E} \left[ \kappa K \right] < \varepsilon.
\]

Here, \( B_{\bar{\kappa}} \) denotes the space of all Borel measurable functions \( f : \mathbb{R}^d \to \mathbb{R} \) with \( \| f \|_{\bar{\kappa}} < \infty \).

We define \( I(0) := \text{id}_{C_b} \) and, for every \( t > 0, f \in C_b \) and \( x \in \mathbb{R}^d \),
\[
(I(t)f)(x) := tf \left[ \frac{1}{t} f(x + \sqrt{t} \xi) \right], \quad \text{where} \quad \xi := \text{id}_{\mathbb{R}^d}.
\]

Moreover, we denote by \( C_b^2 \) the space of all twice differentiable functions \( f \in C_b \) such that all partial derivatives are in \( C_b \) and by \( D^2 f(x) = (\partial_i \partial_j f)_{i,j=1,\ldots,d} \in \mathbb{R}^{d \times d} \) the second derivative. Let \( x^T \in \mathbb{R}^{1 \times d} \) be the transposed vector for all \( x \in \mathbb{R}^d \cong \mathbb{R}^{d \times 1} \). The space \( \mathbb{R}^{d \times d} \) is subsequently endowed with the Frobenius norm.

**Theorem 4.1.** Assume that \( \mathcal{E}[a \xi] = 0 \) for all \( a \in \mathbb{R}^d \). Then, there exists a strongly continuous convex monotone semigroup \((S(t))_{t \geq 0}\) on \( C_b \) with
\[
S(t)f = \lim_{n \to \infty} I \left( \frac{1}{n} \right)^n f \in C_b \quad \text{for all} \ t \geq 0 \text{ and} \ f \in C_b
\]
which has the properties (i)-(v) from Theorem 2.9 with \( \omega = L = 0 \). For every \( f \in C_b^2 \) and \( x \in \mathbb{R}^d \), it holds that \( f \in D(A) \) and
\[
(Af)(x) = \mathcal{E} \left[ \frac{1}{2} \xi^T D^2 f(x) \xi \right].
\]

In addition, for every convex expectation space \((\Omega,H,\bar{\mathcal{E}})\) and iid sequence \((\xi_n)_{n \in \mathbb{N}} \subset H^d \) with \( \mathcal{E}[f(\xi_n)] = \mathcal{E}[f] \) for all \( f \in C_b \),
\[
(S(1)f)(0) = \lim_{n \to \infty} \frac{1}{n} \bar{\mathcal{E}} \left[ n f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \right) \right] \quad \text{for all} \ f \in C_b.
\]
Proof.  It follows from Lemma B.2(ii) that Assumption 2.7(i)-(iv) and (vii) are satisfied with \( \omega = L = 0 \). Moreover, similar to the proof of Theorem 3.2, one can show that \( \|I(t)f)\|_{\tilde{\mathcal{H}}} \leq e^{ct}\|f\|_{\tilde{\mathcal{H}}} \) for all \( f \in C_0 \) with \( \|f\|_{\tilde{\mathcal{H}}} \leq 1 \) and \( t \in (0,1] \), where \( c := \frac{1}{2}\mathcal{E}[\xi^2]. \)
Hence, Remark 2.8(b) yields that Assumption 2.7(vi) is satisfied. Next, we show that
\[
(I'(0)f)(x) = \mathcal{E}\left[ \frac{1}{2} \xi^T D^2 f(x) \xi \right] \quad \text{for all} \quad f \in C_0^2 \quad \text{and} \quad x \in \mathbb{R}^d.
\]
For every \( h > 0 \) and \( x \in \mathbb{R}^d \),
\[
f(x + \sqrt{h} \xi) = f(x) + \nabla f(x) \sqrt{h} \xi + \int_0^1 \int_0^1 sh \xi^T D^2 f(x + rs\sqrt{h} \xi) \xi \, dr \, ds.
\]
Hence, it follows from Lemma B.2(vi) and Lemma A.1 that
\[
\left( \frac{I(h)f - f}{h} \right)(x) - \mathcal{E}\left[ \frac{1}{2} \xi^T D^2 f(x) \xi \right] = \mathcal{E}\left[ \frac{1}{\sqrt{h}} \nabla f(x) \xi + \int_0^1 \int_0^1 s \xi^T D^2 f(x + rs\sqrt{h} \xi) \xi \, dr \, ds \right] - \mathcal{E}\left[ \frac{1}{2} \xi^T D^2 f(x) \xi \right] \\
\leq \lambda \mathcal{E}\left[ \frac{1}{\lambda} \left( \int_0^1 \int_0^1 s \xi^T D^2 f(x + rs\sqrt{h} \xi) \xi \, dr \, ds - \frac{1}{2} \xi^T D^2 f(x) \xi \right) + \frac{1}{2} \xi^T D^2 f(x) \xi \right] \\
- \lambda \mathcal{E}\left[ \frac{1}{2} \xi^T D^2 f(x) \xi \right] \\
\leq \frac{\lambda}{2} \mathcal{E}\left( \int_0^1 \int_0^1 s \xi^T D^2 f(x + rs\sqrt{h} \xi) \xi \, dr \, ds - \frac{1}{2} \xi^T D^2 f(x) \xi \right) + \frac{\lambda}{2} \mathcal{E}\left[ \xi^T D^2 f(x) \xi \right] - \lambda \mathcal{E}\left[ \frac{1}{2} \xi^T D^2 f(x) \xi \right].
\]
Let \( \varepsilon > 0 \) and choose \( \lambda \in (0,1) \) with \( \mathcal{E}[\|D^2 f\|_{\infty} |\xi|^2] < \varepsilon/2 \). Lemma B.2(iii) and (v) imply
\[
\frac{\lambda}{2} \mathcal{E}\left[ \xi^T D^2 f(x) \xi \right] - \lambda \mathcal{E}\left[ \frac{1}{2} \xi^T D^2 f(x) \xi \right] \leq \frac{\varepsilon}{2}.
\]
Furthermore, by inequality (4.1), there exists \( n \in \mathbb{N} \) with
\[
\frac{\lambda}{2} \mathcal{E}\left[ \frac{2}{\lambda} \left( \int_0^1 \int_0^1 s \xi^T D^2 f(x + rs\sqrt{h} \xi) \xi \, dr \, ds - \frac{1}{2} \xi^T D^2 f(x) \xi \right) \mathbb{1}_{B(n)^c}(\xi) \right] \\
\leq \frac{\lambda}{2} \mathcal{E}\left[ \frac{2}{\lambda} \left( \int_0^1 s \, ds + \frac{1}{2} \right) \|D^2 f\|_{\infty} |\xi|^2 \mathbb{1}_{B(n)^c}(\xi) \right] \leq \frac{\varepsilon}{4}
\]
for all \( x \in \mathbb{R}^d \). Now, let \( K \subseteq \mathbb{R}^d \). Since \( D^2 f \) is uniformly continuous on compacts, there exists \( \delta > 0 \) with
\[
|D^2 f(x + y) - D^2 f(x)| < \frac{\varepsilon}{2n^2} \quad \text{for all} \quad x \in K \quad \text{and} \quad y \in B(\delta).
\]
Then, for every \( h \in (0,\delta^2/n^2] \) and \( x \in K \),
\[
\frac{\lambda}{2} \mathcal{E}\left[ \frac{2}{\lambda} \left( \int_0^1 \int_0^1 s \xi^T D^2 f(x + rs\sqrt{h} \xi) \xi \, dr \, ds - \frac{1}{2} \xi^T D^2 f(x) \xi \right) \mathbb{1}_{B(n)^c}(\xi) \right] \\
= \frac{\lambda}{2} \mathcal{E}\left[ \frac{2}{\lambda} \left( \int_0^1 \int_0^1 s \xi^T \left( D^2 f(x + rs\sqrt{h} \xi) - D^2 f(x) \right) \xi \, dr \, ds \right) \mathbb{1}_{B(n)^c}(\xi) \right] \\
\leq \frac{\lambda}{2} \mathcal{E}\left[ \frac{2}{\lambda} \left( \int_0^1 s |\xi|^2 \frac{\varepsilon}{2n^2} \, ds \right) \mathbb{1}_{B(n)^c}(\xi) \right] \leq \frac{\varepsilon}{4}.
\]
Hence, for every $\varepsilon > 0$ and $K \subseteq \mathbb{R}^d$, there exists $h_0 > 0$ with
\[
\left( \frac{I(h)f - f}{h} \right) (x) - \mathcal{E} \left[ \frac{1}{2} \xi^T D^2 f(x) \xi \right] \leq \varepsilon
\]
for all $x \in K$ and $h \in (0, h_0]$. The lower bound follows by similar arguments. Furthermore,
\[
\|I(t) - f\| \leq \mathcal{E} \left[ \frac{1}{2} \|D^2 f\| \|\xi\|^2 \right] t \quad \text{for all } t \geq 0.
\]
This shows that Assumption 2.7(v) is satisfied. Now, the first part of the claim follows from Theorem 2.9 and the second part follows similarly to the proof of Theorem 3.2. □

Similar to Theorem 3.2, for a sublinear expectation $\mathcal{E}[\cdot]$, the computation of the generator simplifies but this does not affect the result. Furthermore, in the sublinear case, the semigroup $(S(t))_{t \geq 0}$ is the unique viscosity solution of the PDE $u_t = G(D^2 u)$ with $G(a) := \mathcal{E}[\frac{1}{2} \xi^T a \xi]$ and therefore Theorem 4.1 is consistent with previous results in that direction. If we weaken the condition $\mathcal{E}[a \xi] = 0$ for all $a \in \mathbb{R}^d$ by merely requiring
\[
\mathcal{E}[a \xi] \geq 0 \quad \text{for all } a \in \mathbb{R}^d, \tag{4.4}
\]
one can still apply the previous result on the transformed expectation
\[
\tilde{\mathcal{E}} : \mathcal{H} \to \mathbb{R}, \quad X \mapsto \inf_{a \in \mathbb{R}^d} \mathcal{E}[X + a \xi]. \tag{4.5}
\]
In the particular case $\mathcal{E}[f] = \sup_{\mu \in M} \int_{\mathbb{R}^d} f(x) \mu(dx)$ for a set $M$ of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, condition (4.4) is satisfied if $\int_{\mathbb{R}^d} x \mu(dx) = 0$ for some $\mu \in M$.

**Corollary 4.2.** Assume that $\mathcal{E}[a \xi] \geq 0$ for all $a \in \mathbb{R}^d$. Define $I(0) := \text{id}_{C_b}$ and
\[
(I(t)f)(x) := \inf_{a \in \mathbb{R}^d} t \mathcal{E} \left[ \frac{1}{2} f(x + \sqrt{t} \xi) + a \xi \right]
\]
for all $t > 0$, $f \in C_b$ and $x \in \mathbb{R}^d$. Then, there exists a strongly continuous convex monotone semigroup $(S(t))_{t \geq 0}$ on $C_b$ with
\[
S(t)f = \lim_{n \to \infty} I \left( \frac{1}{n} \right)^n f \in C_b \quad \text{for all } t \geq 0 \text{ and } f \in C_b
\]
which has the properties (i)-(v) from Theorem 2.9 with $\omega = L = 0$. For every $f \in C_b^2$ and $x \in \mathbb{R}^d$, it holds that $f \in D(A)$ and
\[
(Af)(x) = \inf_{a \in \mathbb{R}^d} \mathcal{E} \left[ \frac{1}{2} (\xi^T D^2 f(x) + a \xi) \right].
\]
**Proof.** By Lemma B.7, the convex expectation $\tilde{\mathcal{E}}$ defined by equation (4.5) satisfies $\tilde{\mathcal{E}}[a \xi] = 0$ for all $a \in \mathbb{R}^d$. Hence, we can apply Theorem 4.1 to obtain the claim. □

### 4.2. Uncertain samples in Wasserstein space.

Similar to Subsection 3.3, we consider convex expectations that are defined as a supremum over a set of probability measures but some natural modifications of the setting are necessary. To ensure that the convex expectation is continuous from above on functions with quadratic growth at infinity, let $p > 2$ and $\varphi : [0, \infty) \to [0, \infty]$ be a non-decreasing function with $\varphi(0) = 0$, $\varphi(\infty) = \infty$ and $\lim_{c \to \infty} \frac{\varphi(c)}{c^2} = \infty$. Moreover, we fix a reference measure $\mu \in \mathcal{P}_p$ with mean zero, i.e., $\int_{\mathbb{R}^d} x \mu(dx) = 0$. In view of Lemma B.7 it now seems natural to define
\[
\tilde{\mathcal{E}} : C_{\tilde{K}} \to \mathbb{R}, \quad f \mapsto \inf_{a \in \mathbb{R}^d} \sup_{\nu \in \mathcal{P}_p} \left( \int_{\mathbb{R}^d} f(x) + ax \nu(dx) - \varphi(W_p(\mu, \nu)) \right).
\]
Using [19, Theorem 2] to interchange the supremum with the infimum, one can show that

\[ \mathcal{E}[f] = \sup_{\nu \in \mathcal{P}_p^0} \left( \int_{\mathbb{R}^d} f(x) \nu(dx) - \varphi(\mathcal{W}_p(\mu, \nu)) \right) \quad \text{for all } f \in C_{\tilde{\mathcal{K}}}, \]

where

\[ \mathcal{P}_p^0 := \left\{ \nu \in \mathcal{P}_p : \int_{\mathbb{R}^d} x \nu(dx) = 0 \right\}. \]

Furthermore, Theorem 4.1 yields a corresponding semigroup \((S(t))_{t \geq 0}\) with generator

\[ (Af)(x) = \mathcal{E} \left[ \frac{1}{2} \xi^T D^2 f(x) \xi \right] \quad \text{for all } f \in C_b^2 \text{ and } x \in \mathbb{R}^d. \]

However, if we want to give an explicit formula for the generator, i.e., the generator should be a given as a supremum over a set of parameters in \(\mathbb{R}^d\) rather than a set of measures, it seems necessary to replace the Wasserstein distance \(\mathcal{W}_p(\mu, \nu)\) by a transport cost which is given as the infimum over a smaller set of couplings. One possible natural choice are martingale couplings \([5,6]\). We call \(\pi \in \Pi(\mu, \nu)\) a martingale coupling between \(\mu\) and \(\nu\) if there exist random variables \(X, Y\) on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) with \(\pi = \mathbb{P} \circ (X, Y)^{-1}\) and \(\mathbb{E}[Y|X] = X\). Equivalently, we could require that

\[ \int_{\mathbb{R}^d} f(x)(y - x) \pi(dx, dy) = 0 \quad \text{for all } f \in C_b. \quad (4.6) \]

Denoting by \(\Pi_M(\mu, \nu)\) the set of all martingale couplings between \(\mu\) and \(\nu\), it follows from Strassen’s theorem that \(\Pi_M(\mu, \nu) \neq \emptyset\) if and only if \(\mu \leq \nu\) in convex order. Moreover, we define the corresponding transport cost by

\[ C_M(\mu, \nu) := \left( \inf_{\pi \in \Pi_M(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy) \right)^{\frac{1}{p}}, \]

where \(\inf \emptyset = \infty\). The convex expectation is then defined by

\[ \mathcal{E} : C_{\tilde{\mathcal{K}}} \to \mathbb{R}, \quad f \mapsto \sup_{\nu \in \mathcal{P}_p^0} \left( \int_{\mathbb{R}^d} f(x) \nu(dx) - \varphi(C_M(\mu, \nu)) \right). \quad (4.7) \]

As before, we define \(I(0) := \text{id}_{C_b}\) and, for every \(t > 0\), \(f \in C_b\) and \(x \in \mathbb{R}^d\),

\[ (I(t)f)(x) := t\mathcal{E} \left[ \frac{1}{t} f \left( x + \sqrt{t} \xi \right) \right], \quad \text{where } \xi := \text{id}_{\mathbb{R}^d}. \]

Similar to Subsection 3.3, one can show that

\[ (I(t)f)(x) = \sup_{\nu \in \mathcal{P}_p^0} \left( \int_{\mathbb{R}^d} f(x + y) \nu(dy) - t\varphi \left( \frac{C_M(\mu_t, \nu)}{\sqrt{t}} \right) \right) \]

for all \(t > 0\), \(f \in C_b\) and \(x \in \mathbb{R}^d\), where \(\mu_t := \mu \circ (\sqrt{t} \xi)^{-1}\). Let \(\text{tr}(a) := \sum_{i=1}^n a_{ii}\) be the matrix trace for all \(a \in \mathbb{R}^{d \times d}\). Note that \(\text{tr}(\lambda \lambda^T a) = \lambda^T a \lambda\) and \(\lambda \lambda^T \in S_d^+\) for all \(\lambda \in \mathbb{R}^d\) and \(a \in \mathbb{R}^{d \times d}\), where \(S_d^+\) consists of all positive semi-definite symmetric \(d \times d\)-matrices. The proof of the following theorem is omitted since it is almost identical with the one of Theorem 4.4 below. For details, we refer to discussion after Theorem 4.4. In Theorem 4.3 and Theorem 4.4, choosing \(d = 1, p, q \in (1, \infty)\) with \(1/p + 1/q = 1\), \(\varphi(c) := c^{2p}/p\) and \(\Sigma := 1\) yields the generator \((Af)(x) = \frac{1}{2} f''(x) + (f''(x))^q/q\) for all \(f \in C_b^2\) and \(x \in \mathbb{R}\).
Theorem 4.3. The functional $\mathcal{E}$ defined by equation (4.7) is a convex expectation which is continuous from above and satisfies $\mathcal{E}[a\xi] = 0$ for all $a \in \mathbb{R}^d$. Hence, there exists a strongly continuous convex monotone semigroup $(S(t))_{t \geq 0}$ on $C_b$ with

$$S(t) f = \lim_{n \to \infty} I(n) f \in C_b$$

for all $t \geq 0$ and $f \in C_b$ which has the properties (i)-(v) from Theorem 2.9 with $\omega = L = 0$. For every $f \in C_b^2$ and $x \in \mathbb{R}^d$, it holds that $f \in D(A)$ and

$$(Af)(x) = \sup_{\lambda \in \mathbb{R}^d} \left( \frac{1}{2} \text{tr}(\lambda \lambda^T D^2 f(x)) - \varphi(|\lambda|) \right) + \frac{1}{2} \text{tr}(\Sigma D^2 f(x)),$$

where $\Sigma := \int_{\mathbb{R}^d} yy^T \mu(dy) \in S^+_d$. Moreover, for every convex expectation space $(\Omega, \mathcal{H}, \mathcal{E})$ and iid sequence $(\xi_n)_{n \in \mathbb{N}} \subset \mathcal{H}^d$ with $\mathcal{E}[f(\xi_n)] = \mathcal{E}[f]$ for all $f \in C_b$,

$$(S(1)f)(0) = \lim_{n \to \infty} \frac{1}{n} \mathcal{E} \left[ n \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \right]$$

for all $f \in C_b$.

It turns out that, instead of the martingale constraint, it is sufficient to require that the couplings used to define the transport cost satisfy the condition

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} x^T a(y - x) \pi(dx, dy) = 0 \text{ for all } a \in \mathbb{S}_d, \tag{4.8}$$

where $\mathbb{S}_d$ denotes the set of all symmetric $d \times d$-matrices. We denote the set of all couplings $\pi \in \Pi(\mu, \nu)$ satisfying equation (4.8) by $\Pi_0(\mu, \nu)$ and define the corresponding transport cost by

$$C_0(\mu, \nu) := \left( \inf_{\pi \in \Pi_0(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy) \right)^{\frac{1}{p}},$$

where $\inf \emptyset = \infty$. The convex expectation is then defined by

$$\mathcal{E}: C_{\mathbb{R}} \to \mathbb{R}, \ f \mapsto \sup_{\nu \in \Pi_0} \left( \int_{\mathbb{R}^d} f(x) \nu(dx) - \varphi(C_0(\mu, \nu)) \right). \tag{4.9}$$

Let $I(0) := \text{id}_{C_b}$ and, for every $t > 0$, $f \in C_b$ and $x \in \mathbb{R}^d$,

$$(I(t)f)(x) := t \mathcal{E} \left[ \frac{1}{t} f(x + \sqrt{t} \xi) \right], \text{ where } \xi := \text{id}_{\mathbb{R}^d}.$$ 

Again, one can show that

$$(I(t)f)(x) = \sup_{\nu \in \Pi_0} \left( \int_{\mathbb{R}^d} f(x + y) \nu(dy) - t \varphi \left( \frac{C_0(\mu_t, \nu)}{\sqrt{t}} \right) \right),$$

where $\mu_t := \mu \circ (\sqrt{t} \xi)^{-1}$. This setting leads to the following result.

Theorem 4.4. The functional $\mathcal{E}$ defined by equation (4.9) is a convex expectation which is continuous from above and satisfies $\mathcal{E}[a\xi] = 0$ for all $a \in \mathbb{R}^d$. Hence, there exists a strongly continuous convex monotone semigroup $(S(t))_{t \geq 0}$ on $C_b$ with

$$S(t) f = \lim_{n \to \infty} I(n) f \in C_b$$

for all $t \geq 0$ and $f \in C_b$ which has the properties (i)-(v) from Theorem 2.9 with $\omega = L = 0$. For every $f \in C_b^2$ and $x \in \mathbb{R}^d$, it holds that $f \in D(A)$ and

$$(Af)(x) = \sup_{\lambda \in \mathbb{R}^d} \left( \frac{1}{2} \text{tr}(\lambda \lambda^T D^2 f(x)) - \varphi(|\lambda|) \right) + \frac{1}{2} \text{tr}(\Sigma D^2 f(x)),$$
Moreover, we can use Dini’s theorem to choose \( \Sigma := \left| f \right| \nu \in R \). For every \( \nu \in P_0^d \), it follows from equation (4.8) that
\[
\int_{R^d} \left| y \right|^2 \nu(dy) - \int_{R^d} \left| x \right|^2 \mu(dx) = \int_{R^d \times R^d} \left| x - y \right|^2 \pi(dx, dy) + 2 \int_{R^d \times R^d} x(y - x) \pi(dx, dy)
\]
for all \( \pi \in \Pi_0(\mu, \nu) \) and therefore Hölder’s inequality yields
\[
\int_{R^d} \left| y \right|^2 \nu(dy) - \int_{R^d} \left| x \right|^2 \mu(dx) \leq C_0(\mu, \nu)^2.
\]
Hence, we can use \( \lim_{c \to \infty} \varphi(c)/c^2 = \infty \) to obtain
\[
\mathcal{E}[f] \leq \sup_{\nu \in P_0^d} \left( \int_{R^d} c(1 + \left| x \right|^2) \nu(dx) - \varphi(C_0(\mu, \nu)) \right)
\]
\[
= \sup_{\nu \in P_0^d} \left( \int_{R^d} c(1 + \left| x \right|^2) \mu(dx) + \int_{R^d} c\left| x \right|^2 \nu(dx) - \int_{R^d} c\left| x \right|^2 \mu(dx) - \varphi(C_0(\mu, \nu)) \right)
\]
\[
\leq \int_{R^d} c(1 + \left| x \right|^2) \mu(dx) + \sup_{\nu \in P_0^d} \left( C_0(\mu, \nu)^2 - \varphi(C_0(\mu, \nu)) \right)
\]
\[
\leq \int_{R^d} c(1 + \left| x \right|^2) \mu(dx) + \sup_{\nu \in P_0^d} \left( C_0(\mu, \nu)^2 - \varphi(C_0(\mu, \nu)) \right) < \infty.
\]
It follows from \( \varphi(0) = 0 \) that \( \mathcal{E}[c] = c \) for all \( c \in R \). Moreover, the functional \( \mathcal{E} \) is clearly convex and monotone. For every \( a \in R^d \), we use the fact that \( \int_{R^d} x \nu(dx) = 0 \) for all \( \nu \in P_0^d \) and the condition \( \varphi(0) = 0 \) to conclude \( \mathcal{E}[a\xi] = 0 \). Let \( (f_n)_{n \in N} \subset C_R \) be a sequence with \( f_n \downarrow 0 \) and choose \( c \geq 0 \) with \( f_1(x) \leq c(1 + \left| x \right|^2) \) for all \( x \in R^d \). Since
\[
\int_{R^d} f_n(x) \nu(dx) - \varphi(C_0(\mu, \nu)) \leq \int_{R^d} c(1 + \left| x \right|^2) \mu(dx) + C_0(\mu, \nu)^2 - \varphi(C_0(\mu, \nu))
\]
for all \( \nu \in P_0^d \) and \( \lim_{c \to \infty} \varphi(c)/c^2 = \infty \), there exists \( R \geq 0 \) with
\[
\mathcal{E}[f_n] = \sup_{\nu \in M} \left( \int_{R^d} f_n(x) \nu(dx) - \varphi(C_0(\mu, \nu)) \right) \leq \sup_{\nu \in M} \int_{R^d} f_n(x) \nu(dx),
\]
where \( M := \{ \nu \in P_0^d : C_0(\mu, \nu) \leq R \} \). For every \( \varepsilon > 0 \), we use \( W_0(\mu, \nu) \leq C_0(\mu, \nu) \) and Lemma C.1 to choose \( r \geq 0 \) with
\[
\sup_{\nu \in M} \int_{B(r)c} c(1 + \left| x \right|^2) \nu(dx) \leq \frac{\varepsilon}{2}.
\]
Moreover, we can use Dini’s theorem to choose \( n_0 \in N \) with
\[
\int_{R^d} f_n(x) \nu(dx) \leq \int_{B(r)} f_n(x) \nu(dx) + \int_{B(r)c} c(1 + \left| x \right|^2) \nu(dx) \leq \varepsilon.
\]
for all \( n \geq n_0 \) and \( \nu \in M \). We obtain \( E[f_n] \downarrow 0 \) as \( n \to \infty \). Now, Theorem 4.1 yields the existence of the semigroup \((S(t))_{t \geq 0}\). Furthermore, for every \( f \in C_b^2 \) and \( x \in \mathbb{R}^d \),
\[
(Af)(x) = E \left[ \frac{1}{2} \xi^T D^2 f(x) \xi \right].
\]

Second, for every \( f \in C_b \) and \( x \in \mathbb{R}^d \), we show that
\[
(Af)(x) = \sup_{\lambda \in \mathbb{R}^d} \left( \frac{1}{2} \text{tr}(\lambda \lambda^T D^2 f(x)) - \varphi(|\lambda|) \right) + \frac{1}{2} \text{tr}(\Sigma D^2 f(x)).
\]

For every \( \nu \in \mathcal{P}^0_\mu \) with \( c := C_0(\mu, \nu) < \infty \) and \( \pi \in \Pi_0(\mu, \nu) \), inequality (4.8) yields
\[
\frac{1}{2} \int_{\mathbb{R}^d} z^T D^2 f(x) \nu(dz) - \varphi(C_0(\mu, \nu)) = \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} z^T D^2 f(x) (z - y) y \pi(dy, dz) + \frac{1}{2} \int_{\mathbb{R}^d} y^T D^2 f(x) y \mu(dy) - \varphi(c)
\]
\[
= \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} (z - y)^T D^2 f(x) (z - y) \pi(dy, dz) + \int_{\mathbb{R}^d \times \mathbb{R}^d} y^T D^2 f(x) (z - y) \pi(dy, dz)
\]
\[
+ \frac{1}{2} \text{tr}(\Sigma D^2 f(x)) - \varphi(c)
\]
\[
\leq \frac{1}{2} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - z|^2 \pi(dy, dz) \right) \sup_{|\lambda| = 1} \lambda^T D^2 f(x) \lambda + \frac{1}{2} \text{tr}(\Sigma D^2 f(x)) - \varphi(c).
\]

We take the supremum over \( \pi \in \Pi_0(\mu, \nu) \) to obtain
\[
(Af)(x) \leq \sup_{c \geq 0} \left( \frac{1}{2} c^2 \sup_{|\lambda| = 1} \text{tr}(\lambda \lambda^T D^2 f(x)) - \varphi(c) \right) + \frac{1}{2} \text{tr}(\Sigma D^2 f(x))
\]
\[
= \sup_{\lambda \in \mathbb{R}^d} \left( \frac{1}{2} \text{tr}(\lambda \lambda^T D^2 f(x)) - \varphi(|\lambda|) \right) + \frac{1}{2} \text{tr}(\Sigma D^2 f(x)).
\]

In order to show the reverse inequality, let \( \lambda \in \mathbb{R}^d \) and \( \nu := \mu * \left( \frac{1}{2} \delta_{\lambda} + \frac{1}{2} \delta_{-\lambda} \right) \in \mathcal{P}^0_\mu \). Furthermore, let \( X, Z \) be independent random variables on a probability space \((\Omega, \mathcal{F}, P)\) with \( \mathbb{P} \circ X^{-1} = \mu \) and \( \mathbb{P} \circ Z^{-1} = \frac{1}{2} \delta_{\lambda} + \frac{1}{2} \delta_{-\lambda} \). Define \( Y := X + Z \) and \( \pi := \mathbb{P} \circ (X,Y)^{-1} \). Clearly, it holds that \( \pi \in \Pi(\mu, \nu) \) and
\[
C_0(\mu, \nu) \leq E[|X - Y|^2] = E[|Z|^2] = |\lambda|.
\]

Hence, we can use the independence of \( X \) and \( Z \) to estimate
\[
(Af)(x) \geq \frac{1}{2} \int_{\mathbb{R}^d} y^T D^2 f(x) y \nu(dy) - \varphi(C_0(\mu, \nu))
\]
\[
\geq \frac{1}{2} \mathbb{E}[|X + Z|^2 D^2 f(x)(X + Z)] - \varphi(|\lambda|)
\]
\[
= \frac{1}{2} \text{tr}(\lambda \lambda^T D^2 f(x)) - \varphi(|\lambda|) + \frac{1}{2} \text{tr}(\Sigma D^2 f(x)).
\]

Taking the supremum over \( \lambda \in \mathbb{R}^d \) yields
\[
(Af)(x) \geq \sup_{\lambda \in \mathbb{R}^d} \left( \frac{1}{2} \text{tr}(\lambda \lambda^T D^2 f(x)) - \varphi(|\lambda|) \right) + \frac{1}{2} \text{tr}(\Sigma D^2 f(x)). \quad \Box
\]
We want to remark that Theorem 4.3 can be obtained from the previous proof. Indeed, it follows from \( \Pi_M(\mu, \nu) \subseteq \Pi_0(\mu, \nu) \) that \( C_M(\mu, \nu) \geq C_0(\mu, \nu) \). Furthermore, the measure \( \nu \) and coupling \( \pi \) that we have chosen to show

\[
(Af)(x) \geq \sup_{\lambda \in \mathbb{R}^d} \left( \frac{1}{2} \text{tr}(\lambda \lambda^T D^2 f(x)) - \varphi(|\lambda|) \right) + \frac{1}{2} \text{tr}(\Sigma D^2 f(x))
\]

satisfy \( \pi \in \Pi_M(\mu, \nu) \) and \( C_M(\mu, \nu) \leq |\lambda| \). Corollary 2.10 implies that the semigroups from Theorem 4.3 and Theorem 4.4 coincide. We conclude this subsection by showing that the semigroup from Theorem 4.4 corresponding to a supremum over an infinite dimensional set of arbitrary distributions can be approximated by random walks. To do so, we define the convex expectation

\[
\tilde{\mathcal{E}}: \mathcal{C}_\kappa \to \mathbb{R}, \ f \mapsto \sup_{\lambda \in \mathbb{R}^d} \left( \frac{1}{2} \int_{\mathbb{R}^d} f(x + \lambda) + f(x - \lambda) \mu(dx) - \varphi(|\lambda|) \right).
\]

Moreover, we define

\[
(J(t)f)(x) := t\tilde{\mathcal{E}}\left[ \frac{1}{t} f(x + \sqrt{t} \xi) \right].
\]

**Corollary 4.5.** Denoting by \((S(t))_{t \geq 0}\) the semigroup from Theorem 4.4, we have

\[
S(t)f = \lim_{n \to \infty} J\left( \frac{t}{n} \right)^n f \quad \text{for all } t \geq 0 \text{ and } f \in \mathcal{C}_b.
\]

**Proof.** As seen during the proof of Theorem 4.4, it holds that

\[
\tilde{\mathcal{E}}[f] = \sup_{\lambda \in \mathbb{R}^d} \left( \int_{\mathbb{R}^d} f(x) \nu_\lambda(dx) - \varphi(|\lambda|) \right) \leq \sup_{\lambda \in \mathbb{R}^d} \left( \int_{\mathbb{R}^d} f(x) \nu_\lambda(dx) - \varphi(C_0(\mu, \nu_\lambda)) \right) \leq \mathcal{E}[f],
\]

where \( \nu_\lambda := \mu \ast \left( \frac{1}{2} \delta_{-\lambda} + \frac{1}{2} \delta_{\lambda} \right) \). Hence, the functional \( \tilde{\mathcal{E}}: \mathcal{C}_\kappa \to \mathbb{R} \) is a well-defined convex expectation which is continuous from above and satisfies \( \tilde{\mathcal{E}}[\alpha \xi] = 0 \) for all \( \alpha \in \mathbb{R}^d \). By Theorem 4.1, there exists a semigroup \((T(t))_{t \geq 0}\) on \( \mathcal{C}_\kappa \) with

\[
T(t)f = \lim_{n \to \infty} J(\pi_n^t)f \quad \text{for all } t \geq 0 \text{ and } f \in \mathcal{C}_b
\]

and generator \((Bf)(x) = \tilde{\mathcal{E}}[\frac{1}{2} \xi^T D^2 f(x) \xi]\) for all \( f \in \mathcal{C}_b^2 \) and \( x \in \mathbb{R}^d \). For every \( f \in \mathcal{C}_b^2 \) and \( x \in \mathbb{R}^d \), it follows from Theorem 4.4 and a straightforward computation that

\[
(Af)(x) = \sup_{\lambda \in \mathbb{R}^d} \left( \frac{1}{2} \text{tr}(\lambda \lambda^T D^2 f(x)) - \varphi(|\lambda|) \right) + \text{tr}(\Sigma D^2 f(x)) = (Bf)(x).
\]

Hence, Corollary 2.10 implies \( S(t)f = T(t)f \) for all \( t \geq 0 \) and \( f \in \mathcal{C}_b \). \( \square \)

5. Conclusion

Based on the abstract results in [7], we have derived explicit and verifiable conditions under which strongly continuous convex monotone semigroups are uniquely determined by their generators evaluated at smooth functions and can be constructed as the limit of iterative approximation schemes. Hence, we are able to extend previous works in this direction and can provide a semigroup approach to HJB equations which is independent from the theory of viscosity solutions. In the application to limit theorems we are, to the best of our knowledge, the first ones to work with convex rather than sublinear expectations. While this does not affect the proofs, apart from sometimes longer estimates, it opens the door to new applications in this context such as large deviations.
theory. A future potential application, where it is crucial to replace sublinearity by convexity, are transitions from discrete to continuous models in the context of mathematical finance. In the sequel, we briefly discuss some further questions. Extending the results in Section 3 and Section 4 to α-stable distributions, as it has been achieved in [4,28] for sublinear expectations, should be attainable with the tools developed in this paper. Furthermore, the corresponding convergence rates for LLN and CLT type results have been established in [8] and are consistent with the ones in [27,29,32,42]. While the results in this article are the same whether we consider sublinear or convex expectations and semigroups, the convergence rates in [8] strongly depend on how nonlinear the expectations and semigroups are. Finally, it would be desirable to have an in-depth comparison between the established viscosity approach to HJB equations and the novel semigroup approach which provides a comparison principle and stability results that are very similar to the well-known properties of viscosity solutions, see [10].

**Appendix A. A basic convexity estimate**

**Lemma A.1.** Let $X$ be a vector space and $\varphi : X \to \mathbb{R}$ be a convex functional. Then,

$$\varphi(x) - \varphi(y) \leq \lambda \left( \varphi \left( \frac{x - y}{\lambda} + y \right) - \varphi(y) \right) \quad \text{for all } x,y \in X \text{ and } \lambda \in (0,1].$$

**Proof.** We use the convexity to estimate

$$\varphi(x) - \varphi(y) = \varphi \left( \lambda \left( \frac{x - y}{\lambda} + y \right) + (1 - \lambda)y \right) - \varphi(y)$$

$$\leq \lambda \varphi \left( \frac{x - y}{\lambda} + y \right) + (1 - \lambda)\varphi(y) - \varphi(y)$$

$$= \lambda \left( \varphi \left( \frac{x - y}{\lambda} + y \right) - \varphi(y) \right)$$

for all $x, y \in X$ and $\lambda \in (0,1]$. \hfill \□

**Appendix B. Convex expectation spaces**

Nonlinear expectations were introduced by Peng, see [41] for a detailed discussion, and are closely related to several other concepts. For instance, sublinear expectations are called upper expectation in robust statistics [30], upper coherent prevision in the theory of imprecise probabilities [44] and coherent risk measure [1] in mathematical finance. Moreover, a convex expectation coincides (up to the sign) with the notion of a convex risk measure [22,24]. In the sequel, we mainly follow [41, Chapter 1] up to some direct transfers from the sublinear to the convex case. All inequalities and the monotone convergence in the following definition are understood pointwise.

**Definition B.1.** Let $\Omega$ be a set and $\mathcal{H}$ a linear space of functions $X : \Omega \to \mathbb{R}$ with $c \in \mathcal{H}$ and $|X| \in \mathcal{H}$ for all $c \in \mathbb{R}$ and $X \in \mathcal{H}$. A convex expectation is a functional $\mathcal{E} : \mathcal{H} \to \mathbb{R}$ with

- $\mathcal{E}[c] = c$ for all $c \in \mathbb{R}$,
- $\mathcal{E}[X] \leq \mathcal{E}[Y]$ for all $X,Y \in \mathcal{H}$ with $X \leq Y$,
- $\mathcal{E}[\lambda X + (1 - \lambda)Y] \leq \lambda \mathcal{E}[X] + (1 - \lambda)\mathcal{E}[Y]$ for all $X,Y \in \mathcal{H}$ and $\lambda \in [0,1]$.

\[1\]We identify $c$ with the constant function $c \mathbf{1}_\Omega$. 

The triplet \((\Omega, \mathcal{H}, \mathcal{E})\) is called a convex expectation space. Furthermore, we say that \(\mathcal{E}\) is continuous from above if \(\mathcal{E}[X_n] \downarrow 0\) for all \((X_n)_{n \in \mathbb{N}} \subset \mathcal{H}\) with \(X_n \downarrow 0\). If \(\mathcal{E}[\lambda X] = \lambda \mathcal{E}[X]\) for all \(\lambda \geq 0\) and \(X \in \mathcal{H}\) we say that \(\mathcal{E}\) is sublinear and call \((\Omega, \mathcal{H}, \mathcal{E})\) a sublinear expectation space.

If \(\mathcal{E}\) is continuous from above it follows from the convexity that \(\mathcal{E}[X_n] \downarrow \mathcal{E}[X]\) for all \((X_n)_{n \in \mathbb{N}} \subset \mathcal{H}\) and \(X \in \mathcal{H}\) with \(X_n \downarrow X\). We collect some elementary properties of convex expectations.

**Lemma B.2.** For a convex expectation space \((\Omega, \mathcal{H}, \mathcal{E})\) the following statements hold:

(i) \(\mathcal{E}[X + c] = \mathcal{E}[X] + c\) for all \(X \in \mathcal{H}\) and \(c \in \mathbb{R}\).

(ii) \(|\mathcal{E}[X] - \mathcal{E}[Y]| \leq \|X - Y\|_{\infty}\) for all bounded \(X, Y \in \mathcal{H}\).

(iii) \(\mathcal{E}[\lambda X] \leq \lambda \mathcal{E}[X]\) for all \(\lambda \in [0, 1]\) and \(X \in \mathcal{H}\).

(iv) \(-\mathcal{E}[X] \leq \mathcal{E}[X]\) for all \(X \in \mathcal{H}\).

(v) \(|\mathcal{E}[X]| \leq \mathcal{E}[|X|]\) for all \(X \in \mathcal{H}\).

(vi) Let \(X \in \mathcal{H}\) with \(\mathcal{E}[aX] = 0\) for all \(a \in \mathbb{R}\). Then, it holds that

\[
\mathcal{E}[X + Y] = \mathcal{E}[Y] \quad \text{for all } Y \in \mathcal{H}.
\]

**Proof.**

(i) For every \(X \in \mathcal{H}\), \(c \in \mathbb{R}\) and \(\lambda \in (0, 1)\), we use that \(\mathcal{E}\) is convex and preserves constants to estimate

\[
\mathcal{E}[X + c] \leq \lambda \mathcal{E} \left[\frac{X}{\lambda}\right] + (1 - \lambda) \mathcal{E} \left[\frac{c}{1 - \lambda}\right] = \lambda \mathcal{E} \left[\frac{X}{\lambda}\right] + c
\]

and

\[
\mathcal{E}[X] = \mathcal{E}[X + c - c] \leq \lambda \mathcal{E} \left[\frac{X + c}{\lambda}\right] + (1 - \lambda) \mathcal{E} \left[-\frac{c}{1 - \lambda}\right] = \lambda \mathcal{E} \left[\frac{X + c}{\lambda}\right] - c.
\]

Since the real-valued mapping \(\lambda \mapsto \mathcal{E}[\lambda X]\) is convex and therefore continuous, we obtain in the limit \(\lambda \to 1\) that \(\mathcal{E}[X + c] = \mathcal{E}[X] + c\).

(ii) It follows from the monotonicity and part (i) that

\[
\mathcal{E}[X] \leq \mathcal{E}[Y + \|X - Y\|_{\infty}] = \mathcal{E}[Y] + \|X - Y\|_{\infty}.
\]

Changing the roles of \(X\) and \(Y\) yields the claim.

(iii) For every \(\lambda \in [0, 1]\) and \(X \in \mathcal{H}\),

\[
\mathcal{E}[\lambda X] = \mathcal{E}[\lambda X + (1 - \lambda)0] \leq \lambda \mathcal{E}[X] + (1 - \lambda) \mathcal{E}[0] = \lambda \mathcal{E}[X].
\]

(iv) It holds that \(0 = \mathcal{E}[0] = \mathcal{E}[\frac{1}{2}X + \frac{1}{2}(-X)] \leq \frac{1}{2} \mathcal{E}[X] + \frac{1}{2} \mathcal{E}[-X]\).

(v) We use the monotonicity and part (iv) to estimate

\[
-\mathcal{E}[|X|] \leq -\mathcal{E}[-X] \leq \mathcal{E}[X] \leq \mathcal{E}[|X|].
\]

(vi) For every \(\lambda \in (0, 1]\), it follows from Lemma A.1 that

\[
\mathcal{E}[X + Y] - \mathcal{E}[Y] \leq \lambda \mathcal{E} \left[\frac{X}{\lambda} + Y\right] - \lambda \mathcal{E}[Y]
\leq \frac{\lambda}{2} \mathcal{E} \left[\frac{2X}{\lambda}\right] + \frac{\lambda}{2} \mathcal{E}[2Y] - \lambda \mathcal{E}[Y]
= \frac{\lambda}{2} \mathcal{E}[2Y] - \lambda \mathcal{E}[Y] \to 0 \quad \text{as } \lambda \downarrow 0.
\]
Similarly one can show that
\[ E[X + Y] - E[Y] \geq -\frac{\lambda}{2} E\left[-\frac{2X}{\lambda}\right] - \frac{\lambda}{2} E[2(X + Y)] + \lambda E[X + Y] \]
\[ = \frac{\lambda}{2} E[2(X + Y)] + \lambda E[X + Y] \to 0 \quad \text{as } \lambda \downarrow 0. \]

**Definition B.3.** Let \((\Omega, \mathcal{H}, E)\) be a convex expectation space with \(f(X_1, \ldots, X_n) \in \mathcal{H}\) for all \(n \in \mathbb{N}\), \(f \in C_b(\mathbb{R}^n)\) and \(X \in \mathcal{H}^n\). Let \(m, n \in \mathbb{N}\), \(X \in \mathcal{H}^m\) and \(Y \in \mathcal{H}^n\).

(i) The distribution of \(X\) is given by the functional
\[ F_X: C_b(\mathbb{R}^m) \to \mathbb{R}, \; f \mapsto E[f(X)]. \]

(ii) We say that \(X\) and \(Y\) are identically distributed if \(m = n\) and
\[ E[f(X)] = E[f(Y)] \quad \text{for all } f \in \text{Lip}_b(\mathbb{R}^m). \]

(iii) We say that \(Y\) is independent of \(X\) if
\[ E[f(X, Y)] = E[E[f(x, Y)]|x = X] \quad \text{for all } f \in \text{Lip}_b(\mathbb{R}^m \times \mathbb{R}^n). \]

In general, the statements “\(Y\) is independent of \(X\)” and “\(X\) is independent of \(Y\)” are not equivalent, see \([41, \text{Example 1.3.15}]\) for a counterexample.

**Definition B.4.** Let \((\Omega, \mathcal{H}, E)\) be a convex expectation space and \((X_n)_{n \in \mathbb{N}} \subset \mathcal{H}^d\) be a sequence of random vectors for some \(d \in \mathbb{N}\).

(i) We say that \((X_n)_{n \in \mathbb{N}}\) is independent and identically distributed (iiid) if \(X_m \) and \(X_n\) have the same distribution and \(X_{n+1}\) is independent of \((X_1, \ldots, X_n)\) for all \(m, n \in \mathbb{N}\).

(ii) We say that \((X_n)_{n \in \mathbb{N}}\) converges in distribution if the sequence \((E[f(X_n)])_{n \in \mathbb{N}}\) converges for all \(f \in \text{Lip}_b(\mathbb{R}^d)\).

Let \(\Omega_{\infty} := (\mathbb{R}^d)^\mathbb{N}\). For every \(n \in \mathbb{N}\), we define \(\pi_n: \Omega_{\infty} \to (\mathbb{R}^d)^n, \; x \mapsto (x_1, \ldots, x_n)\) and \(\mathcal{H}_n := \{f \circ \pi_n : f \in \text{Lip}_b((\mathbb{R}^d)^n)\}\). Furthermore, let \(\mathcal{H}_\infty := \bigcup_{n \in \mathbb{N}} \mathcal{H}_n\). For a convex expectation \(E\) on \((\mathbb{R}^d, \text{Lip}_b(\mathbb{R}^d))\), we define recursively a sequence of convex expectations \(E_n : \text{Lip}_b((\mathbb{R}^d)^n) \to \mathbb{R}\) by \(E_1 := E\) and
\[ E_{n+1}[f] := E_n[E[f(x, Y)]|x = X], \]
where \(X := \text{id}_{(\mathbb{R}^d)^n}\) and \(Y := \text{id}_{\mathbb{R}^d}\). Note that the function \(x \mapsto E[f(x, Y)]\) is Lipschitz continuous, because Lemma B.2(ii) implies
\[ |E[f(x, \cdot)] - E[f(y, \cdot)]| \leq \|f(x, \cdot) - f(y, \cdot)\|_\infty \leq r|x - y| \]
for all \(r \geq 0, f \in \text{Lip}_b(r)\) and \(x, y \in \mathbb{R}^d\). Furthermore, let
\[ E_\infty : \mathcal{H}_\infty \to \mathbb{R}, \; f \circ \pi_n \mapsto E_n[f]. \]

The next result is a direct transfer of \([41, \text{Proposition 1.3.17}]\) from the sublinear to the convex case.

**Lemma B.5.** Let \(E\) be a convex expectation on \((\mathbb{R}^d, \text{Lip}_b(\mathbb{R}^d))\) and define \((\Omega_{\infty}, \mathcal{H}_{\infty}, E_{\infty})\) as above. Furthermore, let
\[ \xi_n : \Omega_{\infty} \to \mathbb{R}, \; x \mapsto x_n \quad \text{for all } n \in \mathbb{N}. \]

Then, the sequence \((\xi_n)_{n \in \mathbb{N}}\) is iid and satisfies \(E_{\infty}[f(\xi_n)] = E[f] \) for all \(f \in \text{Lip}_b(\mathbb{R}^d)\).
Theorem B.6 (Kolmogorov). Let $\mathcal{E}$ be a convex expectation on $(\mathbb{R}^d, C_b(\mathbb{R}^d))$ which is continuous from above and define $(\Omega_\infty, \mathcal{H}_\infty, \mathcal{E}_\infty)$ as before. Then, there exists a unique convex expectation $\tilde{\mathcal{E}}_\infty : \mathcal{L}^\infty(\Omega_\infty) \to \mathbb{R}$ with $\tilde{\mathcal{E}}_\infty[f] = \mathcal{E}_\infty[f]$ for all $f \in \mathcal{H}_\infty$ which is continuous from below on $\mathcal{L}^\infty(\Omega_\infty)$ and continuous from above on 

$$(H_\infty)_b := \{ f \in \mathcal{L}_\infty(\Omega_\infty) : \text{there exists } (f_n)_{n \in \mathbb{N}} \subset \mathcal{H}_\infty \text{ with } f_n \downarrow f \}.$$ 

Moreover, the sequence $(\xi_n)_{n \in \mathbb{N}}$ defined by 

$$\xi_n : \Omega_\infty \to \mathbb{R}, \ x \mapsto x_n \ \text{ for all } n \in \mathbb{N}$$ 

is iid and satisfies $\mathcal{E}_\infty[f(\xi_n)] = \mathcal{E}[f]$ for all $f \in C_b(\mathbb{R}^d)$.

While the property that a random vector has mean zero w.r.t. a convex expectation might seem quite restrictive, this can always be achieved by a simple transformation of the convex expectation if the random vector has non-negative mean. If the convex expectation is defined as a supremum over a set of probability measures, the transformed expectation will be given by a supremum over a smaller set of measures which satisfy an additional constraint, see Subsection 4.2.

Lemma B.7. Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a convex expectation space and $\xi \in \mathcal{H}^d$ with $\mathcal{E}[a\xi] \geq 0$ for all $a \in \mathbb{R}^d$. Then, 

$$\tilde{\mathcal{E}} : \mathcal{H} \to \mathbb{R}, \ X \mapsto \inf_{a \in \mathbb{R}^d} \mathcal{E}[X + a\xi]$$ 

is a convex expectation with $\tilde{\mathcal{E}}[a\xi] = 0$ for all $a \in \mathbb{R}^d$. If $\mathcal{E}$ is continuous from above, the same holds for $\tilde{\mathcal{E}}$. Moreover, in the case $\mathcal{E}[a\xi] = 0$ for all $a \in \mathbb{R}^d$, it holds that $\mathcal{E} = \tilde{\mathcal{E}}$.

Proof. Clearly, the functional $\tilde{\mathcal{E}}$ is monotone and satisfies $\tilde{\mathcal{E}}[c] = c$ for all $c \in \mathbb{R}$. For every $X, Y \in \mathcal{H}$, $\lambda \in [0, 1]$ and $a, b \in \mathbb{R}^d$,

$$\tilde{\mathcal{E}}[\lambda X + (1 - \lambda)Y] \leq \mathcal{E}[\lambda X + (1 - \lambda)Y + (\lambda a + (1 - \lambda)b)\xi]$$

$$\leq \lambda \mathcal{E}[X + a\xi] + (1 - \lambda)\mathcal{E}[Y + b\xi].$$

Taking the infimum over $a, b \in \mathbb{R}^d$ yields

$$\tilde{\mathcal{E}}[\lambda X + (1 - \lambda)Y] \leq \lambda \tilde{\mathcal{E}}[X] + (1 - \lambda)\tilde{\mathcal{E}}[Y].$$

Since $\mathcal{E}[a\xi + b\xi] = \mathcal{E}[(a + b)\xi] \geq 0$ for all $a, b \in \mathbb{R}^d$ with equality for $b = -a$, we obtain $\tilde{\mathcal{E}}[a\xi] = 0$ for all $a \in \mathbb{R}^d$. Assume that $\mathcal{E}$ is continuous from above and let $(X_n)_{n \in \mathbb{N}} \subset \mathcal{H}$ with $X_n \downarrow 0$. Then,

$$\inf_{n \in \mathbb{N}} \tilde{\mathcal{E}}[X_n] = \inf_{n \in \mathbb{N}} \inf_{a \in \mathbb{R}^d} \mathcal{E}[X_n + a\xi] = \inf_{a \in \mathbb{R}^d} \inf_{n \in \mathbb{N}} \mathcal{E}[X_n + a\xi] = \inf_{a \in \mathbb{R}^d} \mathcal{E}[a\xi] = 0$$

which shows that $\tilde{\mathcal{E}}$ is continuous from above. If $\mathcal{E}[a\xi] = 0$ for all $a \in \mathbb{R}^d$, it follows from Lemma B.2(vi) that $\tilde{\mathcal{E}}[X] = \inf_{a \in \mathbb{R}^d} \mathcal{E}[X + a\xi] = \mathcal{E}[X]$ for all $X \in \mathcal{H}$. 

Appendix C. Tightness of Wasserstein balls

Let $p, q \in [1, \infty)$ with $p > q$ and denote by $\mathcal{P}_p$ the $p$-Wasserstein space endowed with the $p$-Wasserstein distance $W_p$.

Lemma C.1. Let $\mu \in \mathcal{P}_p$, $R \geq 0$ and $M := \{ \nu \in \mathcal{P}_p : W_p(\mu, \nu) \leq R \}$. Then, for every $\varepsilon > 0$, there exists $r \geq 0$ with

$$\sup_{\nu \in M} \int_{B(r)^c} |x|^q \nu(dx) \leq \varepsilon.$$
Proof. Let $\varepsilon > 0$ and choose $r_1 \geq 0$ with
\[
2^{q-1} \int_{B(r_1)^c} |x|^q \mu(dx) \leq \frac{\varepsilon}{4} \quad \text{and} \quad 2^{q-1} R^q \mu(B(r_1)^c) \frac{p-q}{p} \leq \frac{\varepsilon}{4}. \quad (C.1)
\]
Let $\nu \in M$ and choose an optimal coupling $\pi \in \Pi(\mu, \nu)$, i.e.,
\[
\mathcal{W}_p(\mu, \nu) = \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p \pi(dx, dy) \right)^{\frac{1}{p}}.
\]
It follows from Hölder’s inequality and inequality (C.1) that
\[
\int_{B(r_1)^c \times B(r_1)^c} |x-y|^q \pi(dx, dy)
\leq \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p \pi(dx, dy) \right)^{\frac{q}{p}} \pi(B(r_1)^c \times B(r_1)^c) \frac{p-q}{p}
\leq R^q \pi(B(r_1)^c \times \mathbb{R}^d) \frac{p-q}{p} = R^q \mu(B(r_1)^c) \frac{p-q}{p} \leq \frac{\varepsilon}{4}. \quad (C.2)
\]
Moreover, for every $x \in B(r_1)$,
\[
\frac{|y|^q}{|x-y|^p} \leq \frac{|y|^q}{(|y|-|x|)^p} \leq \frac{|y|^q}{(|y|-r_1)^p} \rightarrow 0 \quad \text{as} \quad |y| \rightarrow \infty.
\]
Hence, we can choose $r_2 \geq r_1$ with
\[
\frac{|y|^q}{(|x-y|)^p} \leq \frac{\varepsilon}{2R^p} \quad \text{for all} \quad x \in B(r_1) \quad \text{and} \quad y \in B(r_2)^c. \quad (C.3)
\]
It follows from inequality (C.1)-(C.3) that
\[
\begin{aligned}
\int_{B(r_2)^c} |y|^q \nu(dy) &= \int_{B(r_1)^c \times B(r_2)^c} |y|^q \pi(dx, dy) + \int_{B(r_1)^c \times B(r_2)^c} |y|^q \pi(dx, dy) \\
&\leq \int_{B(r_1)^c \times B(r_2)^c} \frac{|y|^q}{|x-y|^p} \cdot |x-y|^p \pi(dx, dy) \\
&\quad + 2^{q-1} \int_{B(r_1)^c \times B(r_2)^c} |x|^q + |x-y|^q \pi(dx, dy) \\
&\leq \frac{\varepsilon}{2R^p} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p \pi(dx, dy) + 2^{q-1} \int_{B(r_1)^c} |x|^q \mu(dx) \\
&\quad + 2^{q-1} \int_{B(r_1)^c \times B(r_1)^c} |x-y|^q \pi(dx, dy) \\
&\leq \frac{\varepsilon}{2R^p} \mathcal{W}_p(\mu, \nu)^p + \frac{\varepsilon}{2} \leq \varepsilon. \quad \Box
\end{aligned}
\]

References

[1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9(3):203–228, 1999.
[2] J. Backhoff-Veraguas, D. Lacker, and L. Tangpi. Nonexponential Sanov and Schilder theorems on Wiener space: BSDEs, Schrödinger problems and control. *Ann. Appl. Probab.*, 30(3):1321–1367, 2020.
[3] D. Bartl, S. Eckstein, and M. Kupper. Limits of random walks with distributionally robust transition probabilities. *Electron. Commun. Probab.*, 26:Paper No. 28, 13, 2021.
[4] E. Bayraktar and A. Munk. An $\alpha$-stable limit theorem under sublinear expectation. *Bernoulli*, 22(4):2548–2578, 2016.
[5] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices—a mass transport approach. Finance Stoch., 17(3):477–501, 2013.

[6] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. Ann. Probab., 44(1):42–106, 2016.

[7] J. Blessing, R. Denk, M. Kupper, and M. Nendel. Convex monotone semigroups and their generators with respect to Γ-convergence. Preprint arXiv:2202.08653, 2022.

[8] J. Blessing, L. Jiang, M. Kupper, and G. Liang. Convergence rates for chernoff-type approximations of convex monotone semigroups. Preprint arXiv:2310.09830, 2023.

[9] J. Blessing and M. Kupper. Nonlinear Semigroups Built on Generating Families and their Lipschitz Sets. Potential Anal., 59(3):857–895, 2023.

[10] J. Blessing, M. Kupper, and M. Nendel. Convergence of infinitesimal generators and stability of convex monotone semigroups. Preprint arXiv:2305.18981, 2023.

[11] P. R. Chernoff. Note on product formulas for operator semigroups. J. Functional Analysis, 2:238–242, 1968.

[12] P. R. Chernoff. Product formulas, nonlinear semigroups, and addition of unbounded operators, volume 140. American Mathematical Soc., 1974.

[13] M. G. Crandall, H. Ishii, and P.-L. Lions. User’s guide to viscosity solutions of second order partial differential equations. Bull. Amer. Math. Soc. (N.S.), 27(1):1–67, 1992.

[14] G. De Cooman and E. Miranda. Weak and strong laws of large numbers for coherent lower priors. J. Statist. Plann. Inference, 138(8):2409–2432, 2008.

[15] 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.

[16] R. Denk, M. Kupper, and M. Nendel. A semigroup approach to nonlinear Lévy processes. Stochastic Process. Appl., 130(3):1616–1642, 2020.

[17] P. Dupuis and R. S. Ellis. A weak convergence approach to the theory of large deviations. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, 1997. A Wiley-Interscience Publication.

[18] S. Eckstein. Extended Laplace principle for empirical measures of a Markov chain. Adv. in Appl. Probab., 51(1):136–167, 2019.

[19] K. Fan. Minimax theorems. Proc. Nat. Acad. Sci. U.S.A., 39:42–47, 1953.

[20] W. H. Fleming and H. M. Soner. Controlled Markov processes and viscosity solutions, volume 25 of Stochastic Modelling and Applied Probability. Springer, New York, second edition, 2006.

[21] H. Föllmer and T. Knispel. Entropic risk measures: coherence vs. convexity, model ambiguity, and robust large deviations. Stoch. Dyn., 11(2-3):333–351, 2011.

[22] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. Finance Stoch., 6(4):429–447, 2002.

[23] H. Föllmer and A. Schied. Stochastic finance. De Gruyter Graduate. De Gruyter, Berlin, 2016. An introduction in discrete time, Fourth revised and extended edition.

[24] M. Frittelli and E. R. Gianin. Putting order in risk measures. Journal of Banking & Finance, 26(7):1473–1486, 2002.

[25] S. Fuhrmann, M. Kupper, and M. Nendel. Wasserstein perturbations of Markovian transition semigroups. Ann. Inst. Henri Poincaré Probab. Stat., 50(2):904–932, 2023.

[26] 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.

[27] M. Hu, L. Jiang, and G. Liang. On the rate of convergence for an α-stable central limit theorem under sublinear expectation. To appear in J. Appl. Probab., 2024+.

[28] M. Hu, L. Jiang, G. Liang, and S. Peng. A universal robust limit theorem for nonlinear Lévy processes under sublinear expectation. Probab. Uncertain. Quant. Risk, 8(1):1–32, 2023.

[29] S. Huang and G. Liang. A monotone scheme for G-equations with application to the explicit convergence rate of robust central limit theorem. J. Differential Equations, 398:1–37, 2024.

[30] P. J. Huber. Robust statistics. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1981.

[31] T. Kato. Trotter’s product formula for an arbitrary pair of self-adjoint contraction semigroups. In Topics in functional analysis (essays dedicated to M. G. Krein on the occasion of his 70th birthday), volume 3 of Adv. in Math. Suppl. Stud., pages 185–195. Academic Press, New York-London, 1978.

[32] N. V. Krylov. On Shige Peng’s central limit theorem. Stochastic Process. Appl., 130(3):1426–1434, 2020.
[33] M. Kupper and J. M. Zapata. Large deviations built on max-stability. *Bernoulli*, 27(2):1001–1027, 2021.

[34] D. Lacker. A non-exponential extension of Sanov’s theorem via convex duality. *Adv. in Appl. Probab.*, 52(1):61–101, 2020.

[35] F. Maccheroni and M. Marinacci. A strong law of large numbers for capacities. *Ann. Probab.*, 33(3):1171–1178, 2005.

[36] 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.

[37] M. Nisio. On a non-linear semi-group attached to stochastic optimal control. *Publ. Res. Inst. Math. Sci.*, 12(2):513–537, 1976/77.

[38] S. Peng. Multi-dimensional G-Brownian motion and related stochastic calculus under G-expectation. *Stochastic Process. Appl.*, 118(12):2223–2253, 2008.

[39] S. Peng. A new central limit theorem under sublinear expectations. *Preprint arXiv:0803.2656*, 2008.

[40] S. Peng. Tightness, weak compactness of nonlinear expectations and application to clt. *Preprint arXiv:1006.2541*, 2010.

[41] S. Peng. *Nonlinear expectations and stochastic calculus under uncertainty*, volume 95 of *Probability Theory and Stochastic Modelling*. Springer, Berlin, 2019. With robust CLT and G-Brownian motion.

[42] Y. Song. Normal approximation by Stein’s method under sublinear expectations. *Stochastic Process. Appl.*, 130(5):2838–2850, 2020.

[43] H. F. Trotter. On the product of semi-groups of operators. *Proc. Amer. Math. Soc.*, 10:545–551, 1959.

[44] P. Walley. *Statistical reasoning with imprecise probabilities*, volume 42 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London, 1991.