

**The exponential resolvent of a markov process and large deviations for markov processes via hamilton-jacobi equations**

Kraaij, Richard C.

**DOI**

[10.1214/20-EJP539](https://doi.org/10.1214/20-EJP539)

**Publication date**

2020

**Document Version**

Final published version

**Published in**

Electronic Journal of Probability

**Citation (APA)**

Kraaij, R. C. (2020). The exponential resolvent of a markov process and large deviations for markov processes via hamilton-jacobi equations. *Electronic Journal of Probability*, 25, 1-39. Article 134. <https://doi.org/10.1214/20-EJP539>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

# The exponential resolvent of a Markov process and large deviations for Markov processes via Hamilton-Jacobi equations

Richard C. Kraaij\*

## Abstract

We study the Hamilton-Jacobi equation  $f - \lambda Hf = h$ , where  $Hf = e^{-f} Ae^f$  and where  $A$  is an operator that corresponds to a well-posed martingale problem.

We identify an operator that gives viscosity solutions to the Hamilton-Jacobi equation, and which can therefore be interpreted as the resolvent of  $H$ . The operator is given in terms of an optimization problem where the running cost is a path-space relative entropy.

Finally, we use the resolvents to give a new proof of the abstract large deviation result of Feng and Kurtz (2006).

**Keywords:** non-linear resolvent; Hamilton-Jacobi equations; large deviations; Markov processes.

**MSC2020 subject classifications:** Primary 60F10; 47H20, Secondary 60J25; 60J35; 49L25.

Submitted to EJP on January 8, 2020, final version accepted on October 22, 2020.

Supersedes arXiv:1910.03946.

## 1 Introduction

Let  $E$  be Polish and let  $A \subseteq C_b(E) \times C_b(E)$  be an operator such that the martingale problem for  $A$  is well posed. In this paper, we study the non-linear operator  $H \subseteq C_b(E) \times C_b(E)$  given by all pairs  $(f, g)$  such that

$$t \mapsto \exp \left\{ f(X(t)) - f(X(0)) - \int_0^t g(X(s)) ds \right\} \quad (1.1)$$

is a martingale with respect to  $\mathcal{F}_t := \sigma(X(s) \mid s \leq t)$  and where  $X$  is a solution of a well-posed martingale problem for  $A$  (If  $e^f \in \mathcal{D}(A)$ , then  $(f, e^{-f} Ae^f) \in H$ ).

The operator  $H$ , the martingales of (1.1) corresponding to  $H$ , and the semigroup

$$V(t)f(x) = \log \mathbb{E} \left[ e^{f(X(t))} \mid X(0) = x \right]. \quad (1.2)$$

---

\*Delft University of Technology, The Netherlands. E-mail: [r.c.kraaij@tudelft.nl](mailto:r.c.kraaij@tudelft.nl)

that formally correspond to  $H$  play (possibly after rescaling) a key role in the theory of stochastic control and large deviations of Markov processes, see e.g. [10, 22, 11, 7, 18, 4, 19, 5, 9].

Consider a sequence of Markov processes  $X_n$ . [9] showed in their extensive monograph on the large deviations for Markov processes that the convergence of the non-linear semigroups  $V_n(t)$  defined by  $V_n(t)f(x) = \frac{1}{n} \log \mathbb{E} [e^{nf(X_n(t))} | X_n(0) = x]$  to some appropriate limiting semigroup  $V(t)$  is a major step in establishing path-space large deviations for the sequence  $X_n$ .

It is well-known in the theory of linear semigroups that the convergence of semigroups  $V_n(t)$  to  $V(t)$  is essentially implied by the convergence of their infinitesimal generators ' $H_n f = \partial_t V_n(t)f|_{t=0}$ ' to ' $Hf = \partial_t V(t)f|_{t=0}$ ', see e.g. [23, 12, 16]. These results also hold for the non-linear context, but the relation between semigroup and generator is less clear. To be precise,  $V(t)$  is generated by  $H$  if we have a resolvent

$$R(\lambda) := (\mathbb{1} - \lambda H)^{-1}, \quad \lambda > 0, \tag{1.3}$$

which approximates the semigroup in the following way

$$V(t)h = \lim_m R\left(\frac{t}{m}\right)^m h, \quad \forall h \in C_b(E), t \geq 0. \tag{1.4}$$

To be able to effectively use the Trotter-Kato-Kurtz approximation results in the theory of large deviations or stochastic control, it is therefore important to have a grip on the resolvent that connects the semigroup  $V(t)$  to the operator  $H$  via (1.3) and (1.4).

An important first step in this direction was made in [9] by replacing the Markov process  $X$  by an approximating jump process with bounded generator. Indeed, in the case of bounded  $A$  one can establish the existence of (1.3) by using fix-point arguments. [9] then proceed to establish path-space large deviations for sequences of Markov processes using probabilistic approximation arguments, semigroup convergence (Trotter-Kato-Kurtz) and the theory of viscosity solutions to characterize the limiting semigroup.

A second observation is that in the context of diffusion processes, or for operators  $H$  that are first-order, it is not clear that one can actually invert  $(\mathbb{1} - \lambda H)$  due to issues with the domain: solutions of the Hamilton-Jacobi equation  $f - \lambda Hf = h$  can have non-differentiable points. However, one can often give a family of operators  $R(\lambda)$  in terms of a deterministic control problem that yield viscosity solutions to the equation  $f - \lambda Hf = h$ . An extension  $\widehat{H}$  of  $H$  can then be defined in terms of  $R$  such that the operator  $\widehat{H}$  and the semigroup  $V(t)$  are connected as in (1.3) and (1.4).

This paper therefore has a two-fold aim.

- (1) Identify an operator  $R(\lambda)$  in terms of a control problem, which yields viscosity solutions to  $f - \lambda Hf = h$  where  $H$  is in terms of the martingales of (1.1). This we aim to do in the context of *general* (Feller) Markov processes.
- (2) Give a new proof of the main large deviation result of [9] by using the operators  $R(\lambda)$ .

Regarding (1), we will show that the operators  $R(\lambda)$  defined as

$$R(\lambda)h(x) = \sup_{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+))} \left\{ \int_0^\infty \left( \int h(X(t))\mathbb{Q}(dX) - S_t(\mathbb{Q} | \mathbb{P}_x) \right) \tau_\lambda(dt) \right\} \tag{1.5}$$

give viscosity solutions to the Hamilton-Jacobi equation for  $H$ . That is:  $R(\lambda)h$  is a viscosity solution to

$$f - \lambda Hf = h, \quad h \in C_b(E), \lambda > 0. \tag{1.6}$$

Here  $S_t(\mathbb{Q} | \mathbb{P}_x)$  is the relative entropy of  $\mathbb{Q}$  with respect to the solution of the martingale problem started at  $x$  evaluated up to time  $t$ , and  $\tau_\lambda$  is the law of an exponential random variable with mean  $\lambda$ .

Our proof that  $R(\lambda)$  is a viscosity solutions to the Hamilton-Jacobi equation will be carried out using a variant of a result by [9] extended to an abstract context in [14]. The family  $\{R(\lambda)\}_{\lambda>0}$  of (1.5) gives viscosity solutions to (1.6) if

1. for all  $(f, g) \in H$  we have  $R(\lambda)(f - \lambda g) = f$ ,
2.  $R(\lambda)$  is contractive and a pseudo-resolvent. That is:  $\|R(\lambda)\| \leq 1$  and for all  $h \in C_b(E)$  and  $0 < \alpha < \beta$  we have

$$R(\beta)h = R(\alpha) \left( R(\beta)h - \alpha \frac{R(\beta)h - h}{\beta} \right).$$

In other words: if  $R(\lambda)$  serves as a classical left-inverse to  $\mathbb{1} - \lambda H$  and is also a pseudo-resolvent, then it is a viscosity right-inverse of  $(\mathbb{1} - \lambda H)$ .

To finish the analysis towards goal (1), we need to establish that our resolvent approximates the semigroup:

3. For the resolvent in (1.5) it holds that  $V(t)h = \lim_m R\left(\frac{t}{m}\right)^m h$ , where the semigroup is given by (1.2).

This result follows from the intuition that the sum of  $m$  independent exponential random variables of mean  $t/m$  converges to  $t$ . The difficulty lies in analysing the concatenation of suprema as in (1.5), which will be carried out using suitable upper and lower bounds.

The second goal, (2), of this paper is to reprove the main large deviation result of [9]. The general procedure is as follows:

- Given exponential tightness, one can restrict the analysis to the finite-dimensional distributions.
- One establishes the large deviation principle for finite-dimensional distributions by assuming this is true at time 0 and by proving that rescaled versions of the semigroups (1.2) of conditional log-moment generating functions converge.
- One proves convergence of the infinitesimal generators  $H_n \rightarrow H$  and establishes well-posedness of the Hamilton-Jacobi equation  $f - \lambda H f = h$  to obtain convergence of the semigroups.

This paper follows the same general strategy, but establishes the third step in a new way. Instead of working with the resolvent of approximating Markov jump processes, the proof in this paper is based on a semigroup approximation argument of [14] combined with the explicit identification of the resolvents corresponding to the non-linear operators  $H_n$ .

We give a short comparison of the result in this paper to the main result in [9]. Our condition on the convergence of Hamiltonians  $H_n \rightarrow H$  is slightly simpler than the one in [9]. This is due to being able to work with the Markov process itself instead of approximating jump process. The result in this paper is a bit weaker in the sense that we assume the solutions to the martingale problems are continuous in the starting point, as opposed to only assuming measurability in [9]. This is to keep the technicalities as simple as possible, and it is expected this can be generalized. In addition, [9] establishes a result for discrete time processes, which we do not carry out here. This extension should be possible too.

The paper is organized as follows. We start in Section 2 with preliminary definitions. In Section 3 we state the main results on the resolvent. In addition to the announced results 1, 2 and 3 we also obtain that  $R(\lambda)$  is a continuous map on  $C_b(E)$ . Proofs of continuity of  $R(\lambda)$  in addition to various other regularity properties are given in Section 5, the proofs of 1, 2 and 3 are given in Section 6.

In Section 4 we state a simple version of the large deviation result. A more general version and its proof are given in Section 7.

## 2 Preliminaries

Let  $E$  be a Polish space.  $C_b(E)$  denotes the space of continuous and bounded functions. Denote by  $\mathcal{B}(E)$  the Borel  $\sigma$ -algebra of  $E$ . Denote by  $M(E)$  and  $M_b(E)$  the spaces of measurable and bounded measurable functions  $f : E \rightarrow [-\infty, \infty]$  and denote by  $\mathcal{P}(E)$  the space of Borel probability measures on  $E$ . We denote by  $\|\cdot\|$  the supremum norm on  $C_b(E)$ . In addition to considering uniform convergence we consider the compact-open and strict topologies:

- The *compact open* topology  $\kappa$  on  $C_b(E)$  is generated by the semi-norms  $p_K(f) = \sup_{x \in K} |f(x)|$ , where  $K$  ranges over all compact subsets of  $E$ .
- The *strict* topology  $\beta$  on  $C_b(E)$  is generated by all semi-norms

$$p_{K_n, a_n}(f) := \sup_n a_n \sup_{x \in K_n} |f(x)|$$

varying over non-negative sequences  $a_n$  converging to 0 and sequences of compact sets  $K_n \subseteq E$ . See e.g. [21, 24, 15].

As we will often work with the convergence of sequences for the strict topology, we characterize this convergence and give a useful notion of taking closures.

A sequence  $f_n$  converges to  $f$  for the strict topology if and only if  $f_n$  converges to  $f$  bounded and uniformly on compacts (buc):

$$\sup_n \|f_n\| < \infty, \quad \forall \text{ compact } K \subseteq E : \quad \lim_{n \rightarrow \infty} \sup_{x \in K} |f_n(x) - f(x)| = 0.$$

Let  $B_r \subseteq C_b(E)$  be the collection of functions  $f$  such that  $\|f\| \leq r$ . We say that  $\widehat{D}$  is the *quasi-closure* of  $D \subseteq C_b(E)$  if  $\widehat{D} = \bigcup_{r>0} \widehat{D}_r$ , where  $\widehat{D}_r$  is the strict closure of  $D \cap B_r$ .

We denote by  $D_E(\mathbb{R}^+)$  the *Skorokhod space* of trajectories  $X : \mathbb{R}^+ \rightarrow E$  that have left limits and are right-continuous. We equip this space with its usual topology, see [8, Chapter 3]. As  $D_E(\mathbb{R}^+)$  is our main space of interest, we write  $\mathcal{P} := \mathcal{P}(D_E(\mathbb{R}^+))$ .

Let  $\mathcal{X}$  be a general Polish space (e.g.  $E$  or  $D_E(\mathbb{R}^+)$ ). For two measures  $\mu, \nu \in \mathcal{P}(\mathcal{X})$  we denote by

$$S(\nu | \mu) = \begin{cases} \int \log \frac{d\nu}{d\mu} d\nu & \text{if } \nu \ll \mu, \\ \infty & \text{otherwise,} \end{cases}$$

the *relative entropy* of  $\nu$  with respect to  $\mu$ . For any sub-sigma algebra  $\mathcal{F}$  of  $\mathcal{B}(\mathcal{X})$ , we denote by  $S_{\mathcal{F}}$  the relative entropy when the measures are restricted to the  $\sigma$ -algebra  $\mathcal{F}$ . In the text below, we will often work with the space  $D_E(\mathbb{R}^+)$ . We will then write  $S_t$  for the relative entropy when we restrict to  $\mathcal{F}_t := \sigma(X(s) | s \leq t)$ .

Finally, for  $\lambda > 0$ , denote by  $\tau_\lambda \in \mathcal{P}(\mathbb{R}^+)$  the law of an exponential random variable with mean  $\lambda$ :

$$\tau_\lambda(dt) = \mathbb{1}_{\{t \geq 0\}} \lambda^{-1} e^{-\lambda^{-1}t} dt.$$

### 2.1 The martingale problem

**Definition 2.1** (The martingale problem). Let  $A: \mathcal{D}(A) \subseteq C_b(E) \rightarrow C_b(E)$  be a linear operator. For  $(A, \mathcal{D}(A))$  and a measure  $\nu \in \mathcal{P}(E)$ , we say that  $\mathbb{P} \in \mathcal{P}(D_E(\mathbb{R}^+))$  solves the martingale problem for  $(A, \nu)$  if  $\mathbb{P} \circ X(0)^{-1} = \nu$  and if for all  $f \in \mathcal{D}(A)$

$$M_f(t) := f(X(t)) - f(X(0)) - \int_0^t Af(X(s))ds$$

is a martingale with respect to its natural filtration  $\mathcal{F}_t := \sigma(X(s) \mid s \leq t)$  under  $\mathbb{P}$ .

We say that uniqueness holds for the martingale problem if for every  $\nu \in \mathcal{P}(E)$  the set of solutions of the martingale problem that start at  $\nu$  has at most one element. Furthermore, we say that the martingale problem is well-posed if this set contains exactly one element for every  $\nu$ .

### 2.2 Viscosity solutions of Hamilton-Jacobi equations

Consider an operator  $B \subseteq C_b(E) \times C_b(E)$ . If  $B$  is single valued and  $(f, g) \in B$ , we write  $Bf := g$ . We denote  $\mathcal{D}(B)$  for the domain of  $B$  and  $\mathcal{R}(B)$  for the range of  $B$ .

**Definition 2.2.** Let  $B \subseteq C_b(E) \times C_b(E)$ . Fix  $h \in C_b(E)$  and  $\lambda > 0$ . Consider the equation

$$f - \lambda Bf = h \tag{2.1}$$

- We say that a bounded upper semi-continuous function  $u : E \rightarrow \mathbb{R}$  is a subsolution of equation (2.1) if for all  $(f, g) \in B$  there is a sequence  $x_n \in E$  such that

$$\lim_{n \rightarrow \infty} u(x_n) - f(x_n) = \sup_x u(x) - f(x), \tag{2.2}$$

and

$$\limsup_{n \rightarrow \infty} u(x_n) - \lambda g(x_n) - h(x_n) \leq 0. \tag{2.3}$$

- We say that a bounded lower semi-continuous function  $v : E \rightarrow \mathbb{R}$  is a supersolution of equation (2.1) if for all  $(f, g) \in B$  there is a sequence  $x_n \in E$  such that

$$\lim_{n \rightarrow \infty} v(x_n) - f(x_n) = \inf_x v(x) - f(x), \tag{2.4}$$

and

$$\liminf_{n \rightarrow \infty} v(x_n) - \lambda g(x_n) - h(x_n) \geq 0. \tag{2.5}$$

- We say that  $u$  is a solution of (2.1) if it is both a subsolution and a supersolution.
- We say that (2.1) satisfies the comparison principle if for all  $h_1, h_2 \in C_b(E)$  and every subsolution  $u$  to (2.1) with  $h$  replaced by  $h_1$ , and supersolution  $v$  to (2.1) with  $h$  replaced by  $h_2$ , we have

$$\sup_x u(x) - v(x) \leq \sup_x h_1(x) - h_2(x). \tag{2.6}$$

Note that the comparison principle implies uniqueness of viscosity solutions.

### 2.3 Convergence of operators

**Definition 2.3.** For a sequence of operators  $B_n \subseteq C_b(E) \times C_b(E)$  and  $B \subseteq C_b(E) \times C_b(E)$  we say that  $B$  is subset of the extended limit of  $B_n$ , denoted by  $B \subseteq ex - \text{LIM } B_n$  if for all  $(f, g) \in B$  there are  $(f_n, g_n) \in B_n$  such that  $\beta - \lim_n f_n = f$  and  $\beta - \lim_n g_n = g$ .

### 2.4 Large deviations

**Definition 2.4.** Let  $\{X_n\}_{n \geq 1}$  be a sequence of random variables on a Polish space  $\mathcal{X}$ . Furthermore, consider a function  $I : \mathcal{X} \rightarrow [0, \infty]$  and a sequence  $\{r_n\}_{n \geq 1}$  of positive real numbers such that  $r_n \rightarrow \infty$ . We say that

- the function  $I$  is a rate-function if the set  $\{x \mid I(x) \leq c\}$  is closed for every  $c \geq 0$ . We say  $I$  is good if the sub-level sets are compact.
- the sequence  $\{X_n\}_{n \geq 1}$  is exponentially tight at speed  $r_n$  if, for every  $a \geq 0$ , there exists a compact set  $K_a \subseteq \mathcal{X}$  such that  $\limsup_n r_n^{-1} \log \mathbb{P}[X_n \notin K_a] \leq -a$ .
- the sequence  $\{X_n\}_{n \geq 1}$  satisfies the large deviation principle with speed  $r_n$  and good rate-function  $I$  if for every closed set  $A \subseteq \mathcal{X}$ , we have

$$\limsup_{n \rightarrow \infty} r_n^{-1} \log \mathbb{P}[X_n \in A] \leq - \inf_{x \in A} I(x),$$

and if for every open set  $U \subseteq \mathcal{X}$ ,

$$\liminf_{n \rightarrow \infty} r_n^{-1} \log \mathbb{P}[X_n \in U] \geq - \inf_{x \in U} I(x).$$

## 3 The non-linear resolvent of a Markov process

Our main result is based on the assumption that the martingale problem is well-posed and that the solution map in terms of the starting point is continuous.

**Condition 3.1.**  $A \subseteq C_b(E) \times C_b(E)$  is an operator such that the martingale problem for  $A \subseteq C_b(E) \times C_b(E)$  is well-posed. Denote by  $\mathbb{P}_x \in \mathcal{P}(D_E(\mathbb{R}^+))$  the solution that satisfies  $X(0) = x$ ,  $\mathbb{P}_x$  almost surely. The map  $x \mapsto \mathbb{P}_x$  is assumed to be continuous for the weak topology on  $\mathcal{P} = \mathcal{P}(D_E(\mathbb{R}^+))$ .

We introduce the triplet of key objects in semi-group theory: generator, resolvent, and semigroup.

**Definition 3.2.** 1. Let  $H$  be a collection of pairs  $(f, g) \in C_b(E) \times C_b(E)$  such that for all  $x \in E$

$$t \mapsto \exp \left\{ f(X(t)) - f(X(0)) - \int_0^t g(X(s)) ds \right\}$$

are martingales with respect to the filtration  $\mathcal{F}_t := \sigma(X(s) \mid s \leq t)$  and law  $\mathbb{P}_x$ .

2. For  $\lambda > 0$  and  $h \in C_b(E)$ , define

$$R(\lambda)h(x) = \sup_{\mathbb{Q} \in \mathcal{P}} \left\{ \int_0^\infty \left( \int h(X(t)) \mathbb{Q}(dX) - S_t(\mathbb{Q} \mid \mathbb{P}_x) \right) \tau_\lambda(dt) \right\}.$$

3. For  $t \geq 0$  and  $h \in C_b(E)$ , define

$$V(t)h(x) = \log \mathbb{E}_x \left[ e^{h(X(t))} \right] = \sup_{\mathbb{Q} \in \mathcal{P}} \left\{ \int h(X(t)) \mathbb{Q}(dX) - S(\mathbb{Q} \mid \mathbb{P}_x) \right\}. \quad (3.1)$$

Note that the final equality follows by Lemma A.1.

Before giving our main results, we make two remarks on the definitions of  $R(\lambda)$  and  $V(t)$  respectively.

**Remark 3.3.** As for linear semigroups, the corresponding resolvent is in terms of the exponential distribution. That the resolvent inverts the Hamiltonian is crucially dependent on the properties of these distributions, as will be discussed briefly Section 3.1 and features prominently in the proofs. The key question in the definition is where to

place the supremum over  $\mathbb{Q} \in \mathcal{P}$ . Here we follow typical results in control theory, see e.g. [2, Section III.2], where the optimization<sup>1</sup> is placed outside of the evaluation of the pay-off.

**Remark 3.4.** The relative entropy  $S(\mathbb{Q} | \mathbb{P}_x)$  that features on the right-hand side of (3.1) can be replaced by the relative entropy  $S_t(\mathbb{Q} | \mathbb{P}_x)$  up to time  $t$ .

This is a result of Proposition A.3 on the decomposition of relative entropy. Conditioning on the path up to time  $t$ , one obtains that tilting the measure away from the Markovian dynamics corresponding to the operator  $A$  beyond time  $t$  increases relative entropy. As this tilt does not influence the pay-off  $h$ , it follows that an optimizing measure  $\mathbb{Q}^*$  only tilts the dynamics up to time  $t$ . Thus, for optimizing measures, we have  $S(\mathbb{Q}^* | \mathbb{P}_x) = S_t(\mathbb{Q}^* | \mathbb{P}_x)$ .

The following is an immediate consequence of [8, Lemma 4.3.2].

**Lemma 3.5.** *Let Condition 3.1 be satisfied. We have*

$$\{(f, e^{-f}g) \mid (e^f, g) \in A\} \subseteq H.$$

The first main result of this paper is the following.

**Theorem 3.6.** *Let Condition 3.1 be satisfied. For each  $h \in C_b(E)$  and  $\lambda > 0$  the function  $R(\lambda)h$  is a viscosity solution to  $f - \lambda Hf = h$ .*

The proof of this result follows in Section 6. To facilitate further use of the non-linear resolvent, we establish also that

1. The map  $R(\lambda)$  maps  $C_b(E)$  into  $C_b(E)$ .
2. The operators  $R(\lambda)$  act as the resolvent of the semigroup  $\{V(t)\}_{t \geq 0}$ .

These properties will allow us to use our main result to establish large deviations in a later part of the paper, see Section 7. We state (a) and (b) as Propositions.

**Proposition 3.7.** *For every  $\lambda > 0$  and  $h \in C_b(E)$  we have  $R(\lambda)h \in C_b(E)$ .*

**Proposition 3.8.** *For each  $h \in C_b(E)$ ,  $t \geq 0$  and  $x \in E$  we have*

$$\lim_{m \rightarrow \infty} R\left(\frac{t}{m}\right)^m h = V(t)h.$$

for the strict topology.

Proposition 3.7 will be verified in Section 5, in which we will also verify other regularity properties of  $R(\lambda)$ . Proposition 3.8 is a part of our main results connecting the resolvent and semigroup and will be established in Section 6.

### 3.1 Strategy of the proof of Theorem 3.6 and discussion on extensions

Theorem 3.6 will follow as a consequence of Proposition 3.4 of [14]. We therefore have to check three properties of  $R(\lambda)$ :

1. For all  $(f, g) \in H$ , we have  $f = R(\lambda)(f - \lambda g)$ ;
2. The pseudo-resolvent property: for all  $h \in C_b(E)$  and  $0 < \alpha < \beta$  we have

$$R(\beta)h = R(\alpha) \left( R(\beta)h - \alpha \frac{R(\beta)h - h}{\beta} \right).$$

3.  $R(\lambda)$  is contractive.

<sup>1</sup>In [2] the authors consider equations of the type  $f + \lambda Hf = h$ . The change in sign in front of the Hamiltonian translates to the replacement of a sup by an inf.

We verify (c) in Section 5 as it relates to the regularity of the resolvent. We verify (a) and (b) in Sections 6.1 and 6.2 respectively.

As is known from the theory of weak convergence, the resolvent is related to exponential integrals.

- (a) is related to integration by parts: for bounded measurable functions  $z$  on  $\mathbb{R}^+$ , we have

$$\lambda \int_0^\infty z(t) \tau_\lambda(dt) = \int_0^\infty \int_0^t z(s) ds \tau_\lambda(dt).$$

- (b) is related to a more elaborate property of exponential random variables. Let  $0 < \alpha < \beta$  then

$$\int_0^\infty z(s) \tau_\beta(ds) = \frac{\alpha}{\beta} \int_0^\infty z(s) \tau_\alpha(ds) + \left(1 - \frac{\alpha}{\beta}\right) \int_0^\infty \int_0^\infty z(s+u) \tau_\beta(du) \tau_\alpha(ds).$$

- Finally, the approximation property of Proposition 3.8 is essentially a law of large numbers. The sum of  $m$  independent random variables of mean  $t/m$  converges to  $t$ .

In the non-linear setting, our resolvent is given in terms of an optimization problem over an exponential integral. Thus, our method is aimed towards treating the optimisation procedures by careful choices of measures and decomposition and concatenation or relative entropies by using Proposition A.3 and then using the properties of exponential integrals.

Any of the results mentioned in the above section can also be carried out if one introduces an extra scaling parameter into the operators.

**Remark 3.9.** Fix  $r > 0$ . Let  $H[r]$  be a collection of pairs  $(f, g)$  such that

$$t \mapsto \exp \left\{ r \left( f(X(t)) - f(X(0)) - \int_0^t g(X(s)) ds \right) \right\}$$

are martingales. As above, we have  $\{(f, r^{-1}e^{-rf}g) \mid (e^{rf}, g) \in A\} \subseteq H[r]$ . Relatively straightforwardly, chasing the constant  $r$ , one can show that  $R[r](\lambda)h$  gives viscosity solutions to  $f - \lambda H[r]f = h$ , where

$$R[r](\lambda)h(x) = \sup_{\mathbb{Q} \in \mathcal{P}} \left\{ \int_0^\infty \lambda^{-1} e^{-\lambda^{-1}t} \left( \int h(X(t)) \mathbb{Q}(dX) - \frac{1}{r} S_t(\mathbb{Q} \mid \mathbb{P}_x) \right) dt \right\}.$$

We also have

$$\lim_{m \rightarrow \infty} R[r] \left( \frac{t}{m} \right)^m h(x) = r^{-1} V(t)(rh)(x).$$

**Question 3.10.** To some extent one could wonder whether Theorem 3.6 has an extension where  $H_\dagger$  is a collection of pairs  $(f, g)$  such that

$$t \mapsto \exp \left\{ f(X(t)) - f(X(0)) - \int_0^t g(X(s)) ds \right\}$$

are supermartingales, and where  $H_\ddagger$  is a collection of pairs  $(f, g)$  such that

$$t \mapsto \exp \left\{ f(X(t)) - f(X(0)) - \int_0^t g(X(s)) ds \right\}$$

are submartingales.

The statement would become that for each  $h \in C_b(E)$  and  $\lambda > 0$  the function  $R(\lambda)h$  is a viscosity subsolution to  $f - \lambda H_\dagger f = h$  and a viscosity supersolution to  $f - \lambda H_\ddagger f = h$ .

Indeed, some of the arguments in Section 6 can be carried out for sub- and supermartingales respectively. Certain arguments, however, use that we work with martingales. For example, Lemma A.1 holds for probability measures only.

#### 4 Large deviations for Markov processes

In this section, we consider the large deviations on  $D_E(\mathbb{R}^+)$  of a sequence of Markov processes  $X_n$ . In Section 7 below, we will instead consider the more general framework where the  $X_n$  take their values in a sequence of spaces  $E_n$  that are embedded in  $E$  by a map  $\eta_n$  and where the images  $\eta_n(E_n)$  converge in some appropriate way to  $E$ . As this introduces a whole range of technical complications, we restrict ourselves in this section to the most simple case.

**Condition 4.1.** Let  $A_n \subseteq C_b(E) \times C_b(E)$  be linear operators and let  $r_n$  be positive real numbers such that  $r_n \rightarrow \infty$ . Suppose that

- The martingale problems for  $A_n$  are well-posed. Denote by  $x \mapsto \mathbb{P}_x^n$  the solution to the martingale problem for  $A_n$ .
- For each  $n$  that  $x \mapsto \mathbb{P}_x^n$  is continuous for the weak topology on  $\mathcal{P}(D_E(\mathbb{R}^+))$ .
- for all compact sets  $K \subseteq E$  and  $a \geq 0$  there is a compact set  $K_a \subseteq D_E(\mathbb{R}^+)$  such that

$$\limsup_n \sup_{x \in K} \frac{1}{r_n} \log \mathbb{P}_x^n [K_a^c] \leq -a.$$

The first two conditions correspond to Condition 3.1. The final one states that we have exponential tightness of the processes  $X_n$  uniformly in the starting position in a compact set.

Corresponding to the previous section, define the operators  $H_n$  consisting of pairs  $(f, g) \in C_b(E) \times C_b(E)$  such that

$$t \mapsto \exp \left\{ r_n \left( f(X_n(t)) - f(X_n(0)) - \int_0^t g(X_n(s)) ds \right) \right\}$$

are martingales. Also define the rescaled log moment-generating functions

$$V_n(t)f(x) := \frac{1}{r_n} \log \mathbb{E} \left[ e^{r_n f(X_n(t))} \mid X_n(0) = x \right].$$

**Theorem 4.2.** Let Condition 4.1 be satisfied. Let  $r_n > 0$  be some sequence such that  $r_n \rightarrow \infty$ . Suppose that

1. The large deviation principle holds for  $X_n(0)$  on  $E$  with speed  $r_n$  and good rate function  $I_0$ .
2. There is an operator  $H \subseteq ex - \text{LIM } H_n$ .
3. The comparison principle holds for  $f - \lambda H f = h$ .

Then there exists a semigroup  $V(t)$  on  $C_b(E)$  such that if  $\beta - \lim f_n = f$  and  $t_n \rightarrow t$  if holds that  $\beta - \lim V_n(t_n)f_n = V(t)f$ .

In addition, the processes  $X_n$  satisfy a large deviation principle on  $D_E(\mathbb{R}^+)$  with speed  $r_n$  and rate function

$$I(\gamma) = I_0(\gamma(0)) + \sup_{k \geq 1} \sup_{\substack{0=t_0 < t_1 < \dots < t_k \\ t_i \in \Delta_\gamma^c}} \sum_{i=1}^k I_{t_i - t_{i-1}}(\gamma(t_i) \mid \gamma(t_{i-1})). \tag{4.1}$$

Here  $\Delta_\gamma^c$  is the set of continuity points of  $\gamma$ . The conditional rate functions  $I_t$  are given by

$$I_t(y \mid x) = \sup_{f \in C_b(E)} \{ f(y) - V(t)f(x) \}.$$

**Remark 4.3.** A representation for  $I$  in a Lagrangian form can be obtained by the analysis in Chapter 8 of [9]. To some extent the analysis is similar to the one of this paper. First, one identifies the resolvent as the optimum of a deterministic control problem by showing that it solves the Hamilton-Jacobi equation in the viscosity sense. Second, one shows that it approximates a control-semigroup. Third, one uses the control-semigroup to show that (4.1) is also given in terms of the control problem.

## 5 Regularity of the semigroup and resolvent

The main object of study of this paper is the resolvent introduced in Definition 3.2. Before we start with the main results, we first establish that the resolvent itself is ‘regular’:

- We establish that  $R(\lambda)h \in C_b(E)$  for  $\lambda > 0$  and  $h \in C_b(E)$ .
- We establish that  $h \mapsto R(\lambda)h$  is sequentially continuous for the strict topology.
- We establish that  $\lim_{\lambda \downarrow 0} R(\lambda)h = h$  for the strict topology.

Before starting with analysing the resolvent, we establish regularity properties for the cost function that appears in the definition of  $R(\lambda)$ .

### 5.1 Properties of relative entropy

A key property of Legendre transformation is that convergence of convex functionals implies (and is often equivalent to) Gamma convergence of their convex duals. This can be derived from a paper of Zabell [25]. In the context of weak convergence of measures this has recently been established with a direct proof by Mariani in Proposition 3.2 of [17].

We state the result for completeness.

**Proposition 5.1.** *Let  $\mathcal{X}$  be some Polish space. Then (a) and (b) are equivalent:*

1.  $\mu_n \rightarrow \mu$  weakly,
2. The functionals  $S(\cdot | \mu_n)$  Gamma converge to  $S(\cdot | \mu)$ : that is:
  - (1) The Gamma lower bound: for any sequence  $\nu_n \rightarrow \nu$  we have  $\liminf_n S(\nu_n | \mu_n) \geq S(\nu | \mu)$ .
  - (2) The Gamma upper bound: for any  $\nu$  there are  $\nu_n$  such that  $\nu_n \rightarrow \nu$  and such that  $\limsup_n S(\nu_n | \mu_n) \leq S(\nu | \mu)$ .

Our resolvent is given in terms of the cost functional

$$\mathcal{S}_\lambda(\mathbb{Q} | \mathbb{P}_x) := \int_0^\infty S_t(\mathbb{Q} | \mathbb{P}_x) \tau_\lambda(dt). \tag{5.1}$$

Below, we establish Gamma convergence for  $\mathcal{S}_\lambda$ .

- The Gamma  $\liminf_n$  inequality, in addition to the compactness of the level sets (coercivity) of  $\mathcal{S}_\lambda$  is, established in Lemma 5.2.
- In Proposition 5.3 we strengthen the coercivity to allow for compactness of the level sets of  $\mathcal{S}_\lambda$  uniformly for small  $\lambda$  (equi-coercivity). This property will allow us to study  $R(\lambda)$  uniformly for small  $\lambda$ .
- The Gamma  $\limsup_n$  inequality is established in Proposition 5.4.

**5.1.1 The  $\Gamma$  – lim inf inequality and coercivity**

**Lemma 5.2.** For any  $\lambda > 0$  the map

$$(\mathbb{P}, \mathbb{Q}) \mapsto \mathcal{S}_\lambda(\mathbb{Q} | \mathbb{P}) = \int_0^\infty S_t(\mathbb{Q} | \mathbb{P}) \tau_\lambda(dt)$$

is lower semi-continuous. In addition, the map has compact sublevel sets in the following sense: fix a compact set  $K \subseteq \mathcal{P}(D_E(\mathbb{R}^+))$  and  $c \geq 0$ . Then the set

$$A(c) := \{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)) \mid \exists \mathbb{P} \in K : \mathcal{S}_\lambda(\mathbb{Q} | \mathbb{P}) \leq c\}$$

is compact.

*Proof.* The first claim follows by lower semi-continuity of  $(\mathbb{P}, \mathbb{Q}) \mapsto S_t(\mathbb{Q} | \mathbb{P})$  and Fatou’s lemma. For the second claim note that a set  $A \subseteq \mathcal{P}(D_E(\mathbb{R}^+))$  is compact if the set of measures in  $A$  restricted to  $\mathcal{P}(D_E([0, t]))$  is compact for all  $t$ , see Theorem 3.7.2 in [8].

Thus, fix  $t$  and suppose  $\mathbb{Q} \in A(c)$ . Then there is some  $\mathbb{P} \in K$  such that

$$\begin{aligned} S_t(\mathbb{Q} | \mathbb{P}) &\leq \int_0^\infty S_{t+s}(\mathbb{Q} | \mathbb{P}) \tau_\lambda(ds) \\ &= e^{\lambda^{-1}t} \int_t^\infty S_u(\mathbb{Q} | \mathbb{P}) \tau_\lambda(du) \\ &\leq e^{\lambda^{-1}t} c. \end{aligned}$$

The result now follows by Proposition A.4. □

The final estimate in the above proof is not uniform for small  $\lambda$ . this is due to the fact that the exponential random variables  $\tau_\lambda$  concentrate near 0. Thus, we can only control the relative entropies for small intervals of time after which the measure  $\mathbb{Q}$  is essentially free to do what it wants. Equi-coercivity of the level sets can be recovered to some extent by restricting the interval on which one is allowed to tilt the measure.

**Proposition 5.3.** Fix a compact set  $K \subseteq \mathcal{P}$ ,  $\lambda_0 > 0$  and constants  $c \geq 0$  and  $\varepsilon \in (0, 1)$ . Let  $T(\lambda) := -\lambda \log \varepsilon$ . Then the set

$$\bigcup_{0 < \lambda \leq \lambda_0} \bigcup_{\mathbb{P} \in K} \{\mathbb{Q} \in \mathcal{P} \mid \mathcal{S}_\lambda(\mathbb{Q} | \mathbb{P}) \leq c, \quad S_{T(\lambda)}(\mathbb{Q} | \mathbb{P}) = S(\mathbb{Q} | \mathbb{P})\}$$

is relatively compact in  $\mathcal{P}$ .

*Proof.* First recall that a set of measures in  $\mathcal{P}$  is compact if the set of their restrictions to a finite time interval is relatively compact.

Pick  $\mathbb{P} \in K$  and  $0 < \lambda \leq \lambda_0$  and let  $\mathbb{Q}^* \in \mathcal{P}$  be such that  $\mathcal{S}_\lambda(\mathbb{Q}^* | \mathbb{P}) \leq c$ . We obtain

$$\begin{aligned} S_{T(\lambda)}(\mathbb{Q}^* | \mathbb{P}) &= \int_0^\infty S_{T(\lambda)+s}(\mathbb{Q}^* | \mathbb{P}) \tau_\lambda(ds) \\ &= e^{\lambda^{-1}T(\lambda)} \int_{T(\lambda)}^\infty S_u(\mathbb{Q}^* | \mathbb{P}) \tau_\lambda(du) \\ &\leq e^{\lambda^{-1}T(\lambda)} c = \frac{c}{\varepsilon} \end{aligned}$$

which is uniformly bounded in  $\lambda$ . Note that as  $S_{T(\lambda)}(\mathbb{Q}^* | \mathbb{P}) = S(\mathbb{Q}^* | \mathbb{P})$ , we have for all  $t \geq 0$  that  $S_{T(\lambda)}(\mathbb{Q}^* | \mathbb{P}) = S_{T(\lambda)+t}(\mathbb{Q}^* | \mathbb{P})$ . The map  $\lambda \mapsto T(\lambda)$  is increasing, so if  $t \geq T(\lambda_0)$  then  $t \geq T(\lambda)$  and  $S_t(\mathbb{Q}^* | \mathbb{P}) = S_{T(\lambda)}(\mathbb{Q}^* | \mathbb{P})$ . This implies that the measure  $\mathbb{Q}^*$  is contained in the set

$$\bigcup_{\mathbb{P} \in K} \left\{ \mathbb{Q} \in \mathcal{P} \mid \forall t \geq T(\lambda_0) : S_t(\mathbb{Q} | \mathbb{P}) = S_{T(\lambda)}(\mathbb{Q}^* | \mathbb{P}) \leq \frac{c}{\varepsilon} \right\}.$$

By the remark at the start of the proof, this set is compact by Proposition A.4. □

**5.1.2 The  $\Gamma$  – lim sup inequality: construction of a recovery sequence**

For the proof of the  $\Gamma$  – lim inf inequality, we could use Proposition 5.1 and Fatou. In the context of the  $\Gamma$  – lim sup inequality, we run into the following issue.

Given a sequence  $x_n \rightarrow x$  and fixed time  $t$ , the result of Proposition 5.1 will allow to construct a sequence  $Q_n$  converging to  $Q$  such that  $\limsup_n S_t(Q_n | P_{x_n}) \leq S_t(Q | P_x)$ . This statement can, however, not immediately be lifted to the functional  $S_\lambda$  as the construction gives no information on times  $s \neq t$ .

But, using the Markovian structure of the family  $\{P_y\}_{y \in E}$  and continuity of these measures in  $y$  will allow us to construct measures  $Q_n$  converging to  $Q$  such that also  $\limsup_n S_\lambda(Q_n | P_{x_n}) \leq S_\lambda(Q | P_x)$ . This construction will be carried out via a projective limit argument.

**Proposition 5.4.** *Let  $Q$  be such that  $S_\lambda(Q | P_x) = \int_0^\infty S_t(Q | P_x) \tau_\lambda(dt) < \infty$ .*

*Then, there are measures  $Q_n \in \mathcal{P}(D_E(\mathbb{R}^+))$  that converge to  $Q$ . In addition*

$$\begin{aligned} S_t(Q_n | P_{x_n}) &\leq S_t(Q | P_x) + 1, && \forall n, \forall t \\ \limsup_n S_t(Q_n | P_{x_n}) &\leq S_t(Q | P_x), && \forall t. \end{aligned}$$

We infer from Fatou’s lemma that also

$$\limsup_{n \rightarrow \infty} S_\lambda(Q_n | P_{x_n}) = \limsup_{n \rightarrow \infty} \int_0^\infty S_t(Q_n | P_{x_n}) \tau_\lambda(dt) \leq \int_0^\infty S_t(Q | P_x) \tau_\lambda(dt) = S_\lambda(Q | P_x).$$

We will construct the measures  $Q_n$  by arguing via appropriately chosen finite-dimensional projections of  $Q$ . Thus, we need to establish a conditional version of the  $\limsup_n$  inequality for Gamma convergence of relative entropy functionals. We state and prove this conditional result first, after which we prove Proposition 5.4.

**Lemma 5.5.** *Let  $\mathcal{X}, \mathcal{Y}$  be Polish spaces. Let  $\nu, \mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$  and suppose that  $\mu_n$  are measures on  $\mathcal{X} \times \mathcal{Y}$  converging to  $\mu$ . Denote by  $\mu_{n,0}, \mu_0, \nu_0 \in \mathcal{P}(\mathcal{X})$  the restrictions of  $\mu_n, \mu, \nu$  to  $\mathcal{X}$ .*

*Suppose that*

1. *There are measures  $\nu_{n,0}$  on  $\mathcal{X}$  such that  $\nu_{n,0}$  converges weakly to  $\nu_0$  and such that  $\limsup_{n \rightarrow \infty} S_{\mathcal{X}}(\nu_{n,0} | \mu_{n,0}) \leq S_{\mathcal{X}}(\nu_0 | \mu_0)$ .*
2. *Suppose there is a family of measures  $\{\hat{\mu}(\cdot | x)\}_{x \in \mathcal{X}}$  on  $\mathcal{Y}$  that is weakly continuous in  $x$ . Suppose that this family of measures is a version of the regular conditional measures  $\mu_n(\cdot | x)$  and also of  $\{\mu(\cdot | x)\}_{x \in \mathcal{X}}$ .*

*Then there are measures  $\nu_n \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$  converging to  $\nu$  such that the restriction of  $\nu_n$  to  $\mathcal{X}$  equals  $\nu_{n,0}$  and  $\limsup_{n \rightarrow \infty} S(\nu_n | \mu_n) \leq S(\nu | \mu)$ .*

*Proof.* First of all, note that if  $S(\nu | \mu) = \infty$ , the proof is trivial. Thus, assume  $S(\nu | \mu) < \infty$ .

Denote by  $\nu(\cdot | x)$  a version of the regular conditional probability of  $\nu$  conditional on  $x \in \mathcal{X}$ . By the Skorokhod representation theorem, [3, Theorem 8.5.4], we can find a probability space  $(\Omega, \mathcal{A})$  and a measure  $\kappa$  on  $(\Omega, \mathcal{A})$ , and random variables  $X_n, X : \Omega \rightarrow \mathcal{X}$  such that the random variables  $X_n$  and  $X$  under the law  $\kappa$  have distributions  $\nu_{n,0}$  and  $\nu_0$  and such that  $X_n$  converges to  $X$   $\kappa$  almost surely.

Thus, by assumption, there is a set  $B \in \mathcal{A}$  of  $\kappa$  measure 1 on which  $X_n \rightarrow X$  and on which  $\mu_n(\cdot | X_n) = \hat{\mu}(\cdot | X_n)$  converges to  $\mu(\cdot | X) = \hat{\mu}(\cdot | X)$ . It follows by Proposition 5.1 that on this set there are measures  $\pi_n(\cdot | X_n)$  such that:

$$\begin{aligned} \lim_{n \rightarrow \infty} \pi_n(\cdot | X_n) &= \nu(\cdot | X) \quad \text{weakly,} \\ \limsup_n S(\pi_n(\cdot | X_n) | \mu_n(\cdot | X_n)) &\leq S(\nu(\cdot | X) | \mu(\cdot | X)). \end{aligned}$$

We could construct a sequence of measures  $\nu_n$  out of  $\nu_0$  and the conditional kernels  $\pi_n$ . To establish the  $\limsup_n$  inequality for the relative entropies, however, we will need to interchange a  $\limsup_n$  and an integral by using Fatou's lemma. At this point, we are not able to give a dominating function that will allow the application of Fatou. To solve this issue, we will use  $\pi_n$  only when its relative entropy is not too large.

Set  $A_n := \{\omega \in \Omega \mid S(\pi_n(\cdot \mid X_n(\omega)) \mid \mu_n(\cdot \mid X_n(\omega))) \leq S(\nu(\cdot \mid X(\omega)) \mid \mu(\cdot \mid X(\omega))) + 1\}$ . Note that  $\liminf_n A_n$  has  $\kappa$  measure 1. Now set

$$\nu_n(\cdot \mid X_n) := \begin{cases} \pi_n(\cdot \mid X_n) & \text{if } X_n \in A_n, \\ \mu_n(\cdot \mid X_n) & \text{if } X_n \notin A_n, \end{cases}$$

and define  $\nu_n(dx, dy) = \int \nu_n(dy \mid x)\nu_0(dx)$ . We will establish that

- (1)  $\limsup_n S(\nu_n \mid \mu_n) \leq S(\nu \mid \mu)$ ,
- (2)  $\nu_n \rightarrow \nu$ .

We start with the proof of (1). By construction and Proposition A.3, we have

$$\begin{aligned} \limsup_n S(\nu_n \mid \mu_n) &\leq \limsup_n S_{\mathcal{X}}(\nu_{0,n} \mid \mu_{0,n}) + \limsup_n \int S(\nu_n(\cdot \mid x) \mid \mu_n(\cdot \mid x))\nu_{n,0}(dx) \\ &\leq S_{\mathcal{X}}(\nu_0 \mid \mu_0) + \limsup_n \mathbb{E}_{\kappa} [S(\nu_n(\cdot \mid X_n) \mid \mu_n(\cdot \mid X_n))] \\ &\leq S_{\mathcal{X}}(\nu_0 \mid \mu_0) + \mathbb{E}_{\kappa} \left[ \limsup_n S(\nu_n(\cdot \mid X_n) \mid \mu_n(\cdot \mid X_n)) \right] \\ &\leq S_{\mathcal{X}}(\nu_0 \mid \mu_0) + \mathbb{E}_{\kappa} [S(\nu(\cdot \mid X) \mid \mu(\cdot \mid X))] \\ &= S(\nu \mid \mu). \end{aligned}$$

In line 3, we used Fatou's lemma, using as an upper bound the function  $S(\nu(\cdot \mid X) \mid \mu(\cdot \mid X)) + 1$ . This function has finite  $\kappa$  integral as

$$\mathbb{E}_{\kappa} [S(\nu(\cdot \mid X) \mid \mu(\cdot \mid X))] = S(\nu \mid \mu) - S(\nu_0 \mid \mu_0) < \infty.$$

Next, we establish (2):  $\nu_n \rightarrow \nu$ . By (1) and Proposition A.4 the collection of measures  $\nu_n$  is tight. As a consequence, it suffices to establish that  $\int h d\nu_n \rightarrow \int h d\nu$  for a strictly dense set of functions  $h$  that is also an algebra by the Stone-Weierstrass theorem for the strict topology. Clearly, the set of linear combinations of functions of the form  $h(x, y) = f(x)g(y)$  is an algebra that separates points. Thus, it suffices to establish convergence for  $h(x, y) = f(x)g(y)$  only. For  $h$  of this form, we have

$$\begin{aligned} \int f(x)g(y)\nu_n(dx, dy) &= \int f(x) \left( \int g(y)\nu_n(dy \mid x) \right) \nu_{n,0}(dx) \\ &= \mathbb{E}_{\kappa} \left[ f(X_n) \left( \int g(y)\nu_n(dy \mid X_n) \right) \right] \end{aligned}$$

By the weak convergence of  $\nu_n(\cdot \mid X_n)$  to  $\nu(\cdot \mid X)$  on a set of  $\kappa$  measure 1, we find by the dominated convergence theorem that

$$\mathbb{E}_{\kappa} \left[ f(X_n) \left( \int g(y)\nu_n(dy \mid X_n) \right) \right] \rightarrow \mathbb{E}_{\kappa} \left[ f(X) \left( \int g(y)\nu(dy \mid X) \right) \right].$$

This establishes that  $\int h d\nu_n \rightarrow \int h d\nu$  for  $h(x, y) = f(x)g(y)$  and thus that  $\nu_n \rightarrow \nu$ . □

*Proof of Proposition 5.4.* First of all: we can choose finite collections of times  $\mathcal{T}_k := \{0 = t_0^k < t_1^k < \dots, < t_{i_{\max(k)}}^k\}$ ,  $k \in \{1, 2, \dots\}$  such that:

- $\mathcal{T}_k \subseteq \mathcal{T}_{k+1}$ ,
- $t_{i_{\max}(k)} \geq k$ ,
- For all  $k$ , and  $i \leq i_{\max}(k)$ :  $t_{i+1}^k \leq t_i^k + k^{-1}$ ,
- For all  $k$ , and  $i \leq i_{\max}(k)$ :  $S_{t_{i+1}^k}(\mathbb{Q} | \mathbb{P}_x) \leq S_{t_i^k}(\mathbb{Q} | \mathbb{P}_x) + k^{-1}$ .

For any  $k$ , we find by Lemma 5.5 and induction over the finite collection of times in  $\mathcal{T}_k$  that there are measures  $\mathbb{Q}_n^k \in \mathcal{P}(D_E(\mathbb{R}^+))$  such that

- (1) for all  $t \geq t_{i_{\max}(k)}$ :

$$\begin{aligned} \limsup_n S_t(\mathbb{Q}_n^k | \mathbb{P}_{x_n}) &= \limsup_n S_{\mathcal{T}_k}(\mathbb{Q}_n^k | \mathbb{P}_{x_n}) \\ &\leq S_{\mathcal{T}_k}(\mathbb{Q} | \mathbb{P}_x) \leq S_{t_{i_{\max}(k)}^k}(\mathbb{Q} | \mathbb{P}_x) \leq S_t(\mathbb{Q} | \mathbb{P}_x). \end{aligned}$$

- (2) If  $t < t_{i_{\max}(k)}$ , let  $t_{i^*}^k$  be the smallest time in  $\mathcal{T}_k$  such that  $t_{i^*}^k \geq t$ . Then:

$$\begin{aligned} \limsup_n S_t(\mathbb{Q}_n^k | \mathbb{P}_{x_n}) &\leq \limsup_n S_{t_{i^*}^k}(\mathbb{Q}_n^k | \mathbb{P}_{x_n}) = \limsup_n S_{\mathcal{T}_k \cap [0, t_{i^*}^k]}(\mathbb{Q}_n^k | \mathbb{P}_{x_n}) \\ &\leq S_{\mathcal{T}_k \cap [0, t_{i^*}^k]}(\mathbb{Q} | \mathbb{P}_x) \leq S_{t_{i^*}^k}(\mathbb{Q} | \mathbb{P}_x) \leq S_t(\mathbb{Q} | \mathbb{P}_x) + k^{-1}. \end{aligned}$$

Thus, we obtain for all  $t \geq 0$  that

$$\sup_n \sup_k S_t(\mathbb{Q}_n^k | \mathbb{P}_{x_n}) < \infty$$

which implies by Proposition A.4 that the family  $\mathbb{Q}_n^k$  is tight. By construction, i.e. Lemma 5.5, the restrictions of the measures  $\mathbb{Q}_n^k$  to the set of times  $\mathcal{T}_k$  converge to the restriction of  $\mathbb{Q}$  to the times in  $\mathcal{T}_k$ . A straightforward diagonal argument can be used to find  $k(n)$  such that restriction of the measures  $\mathbb{Q}_n := \mathbb{Q}_n^{k(n)}$  to the union  $\bigcup_k \mathcal{T}_k$  to  $\mathbb{Q}$  restricted to the union  $\bigcup_k \mathcal{T}_k$ . This however, establishes that  $\mathbb{Q}_n$  converges to  $\mathbb{Q}$  by Theorem 3.7.8 of [8].  $\square$

## 5.2 Regularity of the resolvent in $x$

We proceed with the proof of Proposition 3.7: establishing  $R(\lambda)h \in C_b(E)$ . For the proof of upper semi-continuity of  $x \mapsto R(\lambda)h(x)$  we use the following technical result that we state for completeness.

**Lemma 5.6** (Lemma 17.30 in [1]). *Let  $\mathcal{X}$  and  $\mathcal{Y}$  be two Polish spaces. Let  $\phi : \mathcal{X} \rightarrow \mathcal{K}(\mathcal{Y})$ , where  $\mathcal{K}(\mathcal{Y})$  is the space of non-empty compact subsets of  $\mathcal{Y}$ , be upper hemi-continuous. That is: if  $x_n \rightarrow x$  and  $y_n \rightarrow y$  and  $y_n \in \phi(x_n)$ , then  $y \in \phi(x)$ .*

*Let  $f : \text{Graph}(\phi) \rightarrow \mathbb{R}$  be upper semi-continuous. Then the map  $m(x) = \sup_{y \in \phi(x)} f(x, y)$  is upper semi-continuous.*

*Proof of Proposition 3.7.* Fix  $\lambda > 0$  and  $h \in C_b(E)$ . Denote as before

$$\mathcal{S}_\lambda(\mathbb{Q} | \mathbb{P}_x) := \int_0^\infty S_t(\mathbb{Q} | \mathbb{P}_x) \tau_\lambda(dt),$$

to shorten notation. By Lemma 5.2 the map  $\mathbb{Q} \mapsto \mathcal{S}_\lambda(\mathbb{Q} | \mathbb{P}_x)$  has compact sub-levelsets and is lower semi-continuous. As  $h$  is bounded we have

$$R(\lambda)h(x) = \sup_{\mathbb{Q} \in \Gamma_x} \left\{ \int h(X(t)) \mathbb{Q}(dX) \tau_\lambda(dt) - \mathcal{S}_\lambda(\mathbb{Q} | \mathbb{P}_x) \right\}$$

where  $\Gamma_x := \{\mathbb{Q} \in \mathcal{P} \mid \mathcal{S}_\lambda(\mathbb{Q} \mid \mathbb{P}_x) \leq 2\|h\|\}$ . Note that  $\Gamma_x$  is non-empty and compact.

Due to the lower semi-continuity of  $\mathcal{S}_\lambda$  and the continuity of the integral over  $h$ , it follows that  $x \mapsto R(\lambda)h(x)$  is upper semi-continuous by Lemma 17.30 of [1] if the collection of sets  $\Gamma_x$  is *upper hemi-continuous*; or in other words: if  $\mathbb{Q}_n \in \Gamma_{x_n}$  and  $(x_n, \mathbb{Q}_n) \rightarrow (x, \mathbb{Q})$  then  $\mathbb{Q} \in \Gamma_x$ . This, however, follows directly from the lower semi-continuity of  $\mathcal{S}_\lambda$ .

Next, we establish lower semi-continuity of  $x \mapsto R(\lambda)h(x)$ . Let  $x_n$  be a sequence converging to  $x$ . Pick  $\mathbb{Q}$  so that

$$R(\lambda)h(x) = \int_0^\infty \int h(X(t))\mathbb{Q}(dX)\tau_\lambda(dt) - \mathcal{S}_\lambda(\mathbb{Q} \mid \mathbb{P}_x)$$

It follows by Proposition 5.4 that there are  $\mathbb{Q}_n \in \mathcal{P}(D_E(\mathbb{R}^+))$  such that  $\mathbb{Q}_n \rightarrow \mathbb{Q}$  and  $\limsup_n \mathcal{S}_\lambda(\mathbb{Q}_n \mid \mathbb{P}_{x_n}) \leq \mathcal{S}_\lambda(\mathbb{Q} \mid \mathbb{P}_x)$ . We obtain that

$$\begin{aligned} \liminf_n R(\lambda)h(x_n) &\geq \liminf_n \int_0^\infty \int h(X(t))\mathbb{Q}_n(dX)\tau_\lambda(dt) - \mathcal{S}_\lambda(\mathbb{Q}_n \mid \mathbb{P}_{x_n}) \\ &\geq \int_0^\infty \int h(X(t))\mathbb{Q}(dX)\tau_\lambda(dt) - \mathcal{S}_\lambda(\mathbb{Q} \mid \mathbb{P}_x) = R(\lambda)h(x) \end{aligned}$$

establishing lower semi-continuity. □

### 5.3 Regularity of the resolvent in $h$

We proceed with establishing that the resolvent is sequentially strictly continuous in  $h$ , uniformly for small  $\lambda$ .

**Lemma 5.7.** *For every  $\lambda_0 > 0$  the family of maps  $\{R(\lambda)\}_{0 < \lambda \leq \lambda_0}$  is sequentially strictly equi-continuous. That is: for every  $h_1, h_2 \in C_b(E)$ , every compact set  $K \subseteq E$  and  $\delta > 0$  there is a compact set  $\hat{K} \subseteq E$  such that*

$$\sup_{x \in \hat{K}} \{R(\lambda)h_1(x) - R(\lambda)h_2(x)\} \leq \delta \|h_1 - h_2\| + \sup_{y \in \hat{K}} \{h_1(y) - h_2(y)\}$$

for all  $0 < \lambda \leq \lambda_0$ .

As above denote by  $\mathcal{S}_\lambda(\mathbb{Q} \mid \mathbb{P}) := \int_0^\infty S_t(\mathbb{Q} \mid \mathbb{P})\tau_\lambda(dt)$ .

*Proof.* Fix  $h_1, h_2 \in C_b(E)$ ,  $\lambda_0 > 0$ ,  $\delta > 0$  and a compact set  $K \subseteq E$ .

Pick an arbitrary  $\lambda$  such that  $0 < \lambda \leq \lambda_0$ . For  $x \in K$ , let  $\mathbb{Q}_{x,\lambda} \in \mathcal{P}$  be the measure such that

$$R(\lambda)h_1(x) = \left\{ \int_0^\infty \int h_1(X(t))\mathbb{Q}_{x,\lambda}(dX)\tau_\lambda(dt) - \mathcal{S}_\lambda(\mathbb{Q}_{x,\lambda} \mid \mathbb{P}_x) \right\}$$

and note that  $\mathcal{S}_\lambda(\mathbb{Q}_{x,\lambda} \mid \mathbb{P}_x) \leq 2\|h_1\|$ . It follows that

$$\begin{aligned} &\sup_{x \in K} R(\lambda)h_1(x) - R(\lambda)h_2(x) \\ &\leq \sup_{x \in K} \left\{ \sup_{\mathbb{Q}_1 \in \mathcal{P}} \left\{ \int_0^\infty \int h_1(X(t))\mathbb{Q}_1(dX)\tau_\lambda(dt) - \mathcal{S}_\lambda(\mathbb{Q}_1 \mid \mathbb{P}_x) \right\} \right. \\ &\quad \left. - \sup_{\mathbb{Q}_2 \in \mathcal{P}} \left\{ \int_0^\infty \int h_2(X(t))\mathbb{Q}_2(dX)\tau_\lambda(dt) - \mathcal{S}_\lambda(\mathbb{Q}_2 \mid \mathbb{P}_x) \right\} \right\} \\ &\leq \sup_{x \in K} \int_0^\infty \int h_1(X(t)) - h_2(X(t))\mathbb{Q}_{x,\lambda}(dX)\tau_\lambda(dt). \end{aligned}$$

Denote by  $T(\lambda) := -\lambda \log \frac{\delta}{2}$ . Then it follows that

$$\begin{aligned} \sup_{x \in K} \{R(\lambda)h_1(x) - R(\lambda)h_2(x)\} &\leq \frac{\delta}{2} \|h_1 - h_2\| \\ &+ \sup_{x \in K} \int_0^{T(\lambda)} \int h_1(X(t)) - h_2(X(t)) \mathbb{Q}_{x,\lambda}(dX) \tau_\lambda(dt). \end{aligned}$$

Now denote by  $\widehat{\mathbb{Q}}_{x,\lambda}$  the measure that equals  $\mathbb{Q}_{x,\lambda}$  on the time interval  $[0, T(\lambda)]$  and satisfies  $S_{T(\lambda)}(\widehat{\mathbb{Q}}_{x,\lambda} | \mathbb{P}_x) = S(\widehat{\mathbb{Q}}_{x,\lambda} | \mathbb{P}_x)$ . By Proposition 5.3 the set of the measures  $\widehat{\mathbb{Q}}_{x,\lambda}$ ,  $x \in K, 0 < \lambda \leq \lambda_0$ , is relatively compact, which implies we can find  $\widehat{K} \subseteq E$  such that, with probability at least  $(1 - \frac{\delta}{2} \|h_1 - h_2\|)$ , the trajectories stay in  $\widehat{K}$ . We conclude that

$$\sup_{x \in K} \{R(\lambda)h_1(x) - R(\lambda)h_2(x)\} \leq \delta \|h_1 - h_2\| + \sup_{y \in \widehat{K}} \{h_1(y) - h_2(y)\}$$

for all  $\lambda$  such that  $0 < \lambda \leq \lambda_0$ . □

#### 5.4 Strong continuity of the resolvent and semigroup

We establish that as  $\lambda \downarrow 0$  the resolvents converge to the identity operator. We also establish strict continuity of the semigroup.

**Lemma 5.8.** *For  $h \in C_b(E)$  we have  $\lim_{\lambda \rightarrow 0} R(\lambda)h = h$  for the strict topology.*

*Proof.* As  $\|R(\lambda)h\| \leq \|h\|$  strict convergence  $\lim_{\lambda \rightarrow 0} R(\lambda)h = h$  follows by proving uniform convergence on compact sets  $K \subseteq E$ .

If we choose for  $\mathbb{Q}$  the measure  $\mathbb{P}_x$  in the defining supremum of  $R(\lambda)h(x)$ , we obtain the upper bound

$$R(\lambda)h(x) - h(x) \geq \int_0^\infty \int h(X(t)) - h(x) \mathbb{P}_x(dX) \tau_\lambda(dt).$$

As the measures  $\{\mathbb{P}_x\}_{x \in K}$  are tight, we have control on the modulus of continuity of the trajectories  $t \mapsto X(t)$ . This implies that the right-hand side converges to 0 as  $\lambda \downarrow 0$  uniformly for  $x \in K$ .

We prove the second inequality. Fix  $\varepsilon \in (0, 4\|h\|)$ , we prove that for  $\lambda$  sufficiently small, we have  $\sup_{x \in K} R(\lambda)h(x) - h(x) \leq \varepsilon$ . First of all, let  $T(\lambda) := -\lambda \log \left( \frac{\varepsilon}{4\|h\|} \right)$  and let  $\mathbb{Q}_{x,\lambda}$  optimize  $R(\lambda)h(x)$ . We then have

$$R(\lambda)h(x) - h(x) \leq \frac{1}{2}\varepsilon + \int_0^{T(\lambda)} \int h(X(t)) - h(x) \mathbb{Q}_{x,\lambda}(dX) \tau_\lambda(dt). \tag{5.2}$$

Also note that as in Lemma 5.7 we have  $S_\lambda(\mathbb{Q}_{x,\lambda} | \mathbb{P}_x) \leq 2\|h\|$ . This implies, using that  $t \mapsto S_t$  is increasing in  $t$ , that

$$S_{T(\lambda)}(\mathbb{Q}_{x,\lambda} | \mathbb{P}_x) \leq 8\|h\|^2 \varepsilon^{-1}. \tag{5.3}$$

Denote by  $\widehat{\mathbb{Q}}_{x,\lambda}$  the measures that equal  $\mathbb{Q}_{x,\lambda}$  up to time  $T(\lambda)$  and satisfy  $S_{T(\lambda)}(\widehat{\mathbb{Q}}_{x,\lambda} | \mathbb{P}_x) = S(\widehat{\mathbb{Q}}_{x,\lambda} | \mathbb{P}_x)$ .

Now let  $\lambda \leq \lambda^* := (\log 4\|h\| \varepsilon^{-1})^{-1}$ . Then  $T(\lambda) \leq 1$  and we obtain for all  $s \geq 1$  that

$$S_s(\widehat{\mathbb{Q}}_{x,\lambda} | \mathbb{P}_x) \leq 8\|h\|^2 \varepsilon^{-1}. \tag{5.4}$$

By Proposition A.4, the measures  $\{\widehat{\mathbb{Q}}_{x,\lambda}\}_{0 < \lambda \leq \lambda^*, x \in K}$  form a tight family. Replacing  $\mathbb{Q}_{x,\lambda}$  by  $\widehat{\mathbb{Q}}_{x,\lambda}$  in (5.2), using the tightness of the family of measures, the upper bound follows as for the lower bound. □

**Lemma 5.9.** *For each  $h \in C_b(E)$  we have that  $t \mapsto V(t)h$  is continuous for the strict topology.*

*Proof.* Let  $S(t) : C_b(E) \rightarrow C_b(E)$  be the semigroup of conditional expectations  $S(t)f(x) = \mathbb{E}_x[f(X(t))]$  corresponding to the process with generator  $A$ . It follows that  $V(t)f = \log S(t)e^f$ . The map  $t \mapsto S(t)e^f$  is strictly continuous by Theorem 3.2 of [15] and bounded away from 0. Thus a straightforward verification shows that also  $V(t)f = \log S(t)e^f$  is strictly continuous.  $\square$

### 5.5 Measurability of the optimal measure

In Section 6 below, we will apply the resolvent to the resolvent. This means we have to perform an optimization procedure twice. In particular, this implies we have to integrate over the outcome of the first supremum. To treat this procedure effectively, we need measurability of the optimizing measure.

**Lemma 5.10.** *Let  $h \in C_b(E)$  and  $\lambda > 0$ . There exists a measurable map  $x \mapsto \mathbb{Q}_x$  such that  $\mathbb{Q}_x \in \mathcal{P}$  and*

$$R(\lambda)h(x) = \int_0^\infty \left[ \int h(Y(t)) \mathbb{Q}_x(dY) - S_t(\mathbb{Q}_x | \mathbb{P}_x) \right] \tau_\lambda(dt).$$

We base the proof of this result on a measurable-selection theorem. We state it for completeness.

**Theorem 5.11** (Theorem 6.9.6 in [3]). *Let  $X, Y$  be Polish spaces and let  $\Gamma$  be a measurable subset of  $X \times Y$ . Suppose that the set  $\Gamma_x := \{y \mid (x, y) \in \Gamma\}$  is non-empty and  $\sigma$ -compact for all  $x \in X$ . Then  $\Gamma$  contains the graph of a Borel measurable mapping  $f : X \rightarrow Y$ .*

We will apply this result below by using the following argument. Let  $f, g$  be measurable maps  $f, g : X \times Y \rightarrow (-\infty, \infty]$ . The set  $\{(x, y) \mid f(x, y) = g(x, y)\}$  is measurable as it equals  $\{(x, y) \mid f(x, y) - g(x, y) = 0\}$  which is the inverse image of  $\{0\}$  and hence measurable.

*Proof of Lemma 5.10.* We aim to apply Theorem 5.11. Thus, we have to establish that the set  $\Gamma \subseteq E \times \mathcal{P}$  defined by

$$\Gamma := \left\{ (x, \mathbb{Q}) \mid R(\lambda)h(x) = \int_0^\infty \left[ \int h(Y(t)) \mathbb{Q}(dY) - S_t(\mathbb{Q} | \mathbb{P}_x) \right] \tau_\lambda(dt) \right\}$$

is measurable and that  $\Gamma_x := \{\mathbb{Q} \mid (x, \mathbb{Q}) \in \Gamma\}$  is non-empty and  $\sigma$ -compact.

Similarly as in the proof of Proposition 3.7, we find that  $\Gamma_x$  is compact and non-empty. We also saw in that proof that the map  $(x, \mathbb{Q}) \mapsto \int h(Y(t)) \mathbb{Q}(dY) - S(\mathbb{Q} | \mathbb{P}_x) \tau_\lambda(dt)$  is upper semi-continuous. As  $x \mapsto R(\lambda)h(x)$  is continuous by Proposition 3.7 we see that the set  $\Gamma$  is the set of points where two measurable functions agree implying that  $\Gamma$  is measurable. An application of Theorem 5.11 concludes the proof.  $\square$

## 6 Proofs of the main results

In this section, we prove the two main results: Theorem 3.6 and Proposition 3.8. We argued in Section 3.1 that the first result follows by establishing that  $R(\lambda)$  is a classical left-inverse of  $(\mathbb{1} - \lambda H)$  and that the family  $R(\lambda)$  is a pseudo-resolvent. We establish these two properties in Sections 6.1 and 6.2. The proof of Proposition 3.8 is carried out in Section 6.3.

**6.1  $R(\lambda)$  is a classical left-inverse of  $\mathbb{1} - \lambda H$**

The proof that  $R(\lambda)$  is a classical left-inverse of  $\mathbb{1} - \lambda H$  is based on a well known integration by parts formula for the exponential distribution. For bounded measurable functions  $z$  on  $\mathbb{R}^+$  we have

$$\lambda \int_0^\infty z(t) \tau_\lambda(dt) = \int_0^\infty \int_0^t z(s) ds \tau_\lambda(dt). \tag{6.1}$$

A generalization is given by the following lemma.

**Lemma 6.1.** Fix  $\lambda > 0$  and  $\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+))$ . Let  $z$  be a measurable function on  $E$ . Then we have

$$\lambda \int_0^\infty \int z(X(t)) \mathbb{Q}(dX) \tau_\lambda(dt) = \int_0^\infty \int \int_0^t z(X(s)) ds \mathbb{Q}(dX) \tau_\lambda(dt).$$

The lemma allows us to rewrite the application of  $R(\lambda)$  to  $f - \lambda g$  in integral form. The integral that comes out can be analyzed using the definition of  $H$  in terms of exponential martingales. This leads to the desired result.

**Proposition 6.2.** Let  $\mathbb{P}_x$  be a collection of Markov measures as in Condition 3.1. For all  $\lambda > 0$ ,  $x \in E$  and  $(f, g) \in H$ , we have  $R(\lambda)(f - \lambda g)(x) = f(x)$ .

*Proof.* Fix  $\lambda > 0$ ,  $x \in E$  and  $(f, g) \in H$ . We start by proving  $R(\lambda)(f - \lambda g)(x) \leq f(x)$ . Set  $h = f - \lambda g$ . By Lemma 6.1 we have

$$\begin{aligned} R(\lambda)h(x) &= \sup_{\mathbb{Q} \in \mathcal{P}} \left\{ \int_0^\infty \tau_\lambda(dt) \left[ \int (f(X(t)) - \lambda g(X(t))) \mathbb{Q}(dX) - S_t(\mathbb{Q} | \mathbb{P}_x) \right] \right\} \\ &= \sup_{\mathbb{Q} \in \mathcal{P}} \left\{ \int_0^\infty \tau_\lambda(dt) \left[ \int \left( f(X(t)) - \int_0^t g(X(s)) ds \right) \mathbb{Q}(dX) - S_t(\mathbb{Q} | \mathbb{P}_x) \right] \right\}. \end{aligned}$$

By optimizing the integrand, we find by Lemma A.1

$$\begin{aligned} &R(\lambda)(f - \lambda g)(x) \\ &\leq \int_0^\infty \tau_\lambda(dt) \left\{ \sup_{\mathbb{Q} \in \mathcal{P}} \left[ \int \mathbb{Q}(dX) \left( f(X(t)) - \int_0^t g(X(s)) ds \right) - S_t(\mathbb{Q} | \mathbb{P}_x) \right] \right\} \\ &= \int_0^\infty \tau_\lambda(dt) \left\{ \log \mathbb{E} \left[ e^{f(X(t)) - \int_0^t g(X(s)) ds} \mid X(0) = x \right] \right\}. \end{aligned}$$

As  $(f, g) \in H$  we can reduce the inner expectation to time 0 by using the martingale property. This yields

$$R(\lambda)(f - \lambda g)(x) \leq \int \tau_\lambda(dt) \left\{ \log \mathbb{E} \left[ e^{f(X(0))} \mid X(0) = x \right] \right\} = f(x),$$

establishing the first inequality.

We now prove the reverse inequality  $R(\lambda)(f - \lambda g)(x) \geq f(x)$ . To do so, we construct a measure  $\mathbb{Q}$  that achieves the supremum. For each time  $t \geq 0$ , define the measure  $\mathbb{Q}^t$  via the Radon-Nykodim derivative

$$\frac{d\mathbb{Q}^t}{d\mathbb{P}_x}(X) = \exp \left\{ f(X(t)) - f(X(0)) - \int_0^t g(X(s)) ds \right\}.$$

Note that as  $t \mapsto \exp \left\{ f(X(t)) - f(X(0)) - \int_0^t g(X(s)) ds \right\}$  is a  $\mathbb{P}_x$  martingale, we have for  $s \leq t$  that  $\mathbb{Q}^t|_{\mathcal{F}_s} = \mathbb{Q}^s|_{\mathcal{F}_s}$ . Thus, standard arguments show that there is a measure

$\mathbb{Q} \in \mathcal{P}$  such that  $\mathbb{Q}|_{\mathcal{F}_t} = \mathbb{Q}^t|_{\mathcal{F}_t}$ . Note that by construction, we have  $\mathbb{Q}(X(0) = x) = 1$ . Using this measure  $\mathbb{Q}$ , applying Lemma 6.1, we obtain

$$\begin{aligned} R(\lambda)(f - \lambda Hf)(x) &\geq \int_0^\infty \int \left[ \left( f(X(t)) - \int_0^t g(X(s)) ds \right) \right. \\ &\quad \left. - \left( f(X(t)) - f(X(0)) - \int_0^t g(X(s)) ds \right) \right] \mathbb{Q}(dX) \tau_\lambda(dt) \\ &= f(x) \end{aligned}$$

establishing the second inequality. □

### 6.2 $R$ is a pseudo-resolvent

The next step is the verification that the family of operators  $R(\lambda)$  is a pseudo-resolvent. As in the previous section, this property is essentially an extension of a key property of the exponential distribution. We state it as a lemma that can be verified using basic calculus.

**Lemma 6.3.** *Let  $z : \mathbb{R}^+ \rightarrow \mathbb{R}$  be a bounded and measurable function. Let  $0 < \alpha < \beta$ . Then*

$$\int_0^\infty z(s) \tau_\beta(ds) = \frac{\alpha}{\beta} \int_0^\infty z(s) \tau_\alpha(ds) + \left(1 - \frac{\alpha}{\beta}\right) \int_0^\infty \int_0^\infty z(s+u) \tau_\beta(du) \tau_\alpha(ds).$$

Lifting this property to the family  $R(\lambda)$  yields the pseudo-resolvent property.

**Proposition 6.4.** *For all  $h \in C_b(E)$ ,  $x \in E$ , and  $0 < \alpha < \beta$ , we have*

$$R(\beta)h(x) = R(\alpha) \left( R(\beta)h - \alpha \frac{R(\beta)h - h}{\beta} \right) (x). \tag{6.2}$$

Note that (6.2) is equivalent to

$$R(\beta)h(x) = R(\alpha) \left( \frac{\alpha}{\beta} h - \left(1 - \frac{\alpha}{\beta}\right) R(\beta)h \right) (x), \tag{6.3}$$

which is indeed the operator analogue of the outcome of Lemma 6.3. Before going into the proof of Proposition 6.4, we first rewrite (6.2) to identify the key problem that we need to solve in the proof.

The right-hand side of (6.3) can be written as

$$\begin{aligned} R(\alpha) \left( \frac{\alpha}{\beta} h - \left(1 - \frac{\alpha}{\beta}\right) R(\beta)h \right) (x) &= \sup_{\mathbb{Q} \in \mathcal{P}} \left\{ \int_0^\infty \left[ \int \left( \frac{\alpha}{\beta} h(X(t)) + \left(1 - \frac{\alpha}{\beta}\right) R(\beta)h(X(t)) \right) \mathbb{Q}(dX) - S_t(\mathbb{Q} | \mathbb{P}_x) \right] \tau_\alpha(dt) \right\} \\ &= \sup_{\mathbb{Q} \in \mathcal{P}} \left\{ \frac{\alpha}{\beta} \int_0^\infty \left( \int h(X(t)) \mathbb{Q}(dX) - S_t(\mathbb{Q} | \mathbb{P}_x) \right) \tau_\alpha(dt) \right. \\ &\quad \left. + \left(1 - \frac{\alpha}{\beta}\right) \int_0^\infty \left( \int R(\beta)h(X(t)) \mathbb{Q}(dX) - S_t(\mathbb{Q} | \mathbb{P}_x) \right) \tau_\alpha(dt) \right\}. \end{aligned} \tag{6.4}$$

The integrand of the final integral can be rewritten as

$$\begin{aligned} &\int R(\beta)h(X(t)) \mathbb{Q}(dX) - S_t(\mathbb{Q} | \mathbb{P}_x) \\ &\int \mathbb{Q}(dX) \sup_{\mathbb{Q}^t \in \mathcal{P}} \left\{ \int_0^\infty \int h(Y(s)) \mathbb{Q}^t(dY) - S_s(\mathbb{Q}^t | \mathbb{P}_{X(t)}) \tau_\beta(ds) \right\} - S_t(\mathbb{Q} | \mathbb{P}_x). \end{aligned} \tag{6.5}$$

We will prove (6.2) by proving two inequalities.

- For the proof of

$$R(\beta)h(x) \geq R(\alpha) \left( R(\beta)h - \alpha \frac{R(\beta)h - h}{\beta} \right) (x)$$

we concatenate the optimizing measure in (6.4) with the optimizers for (6.5). Working with a random choice over the time at which we concatenate, and then combining the relative entropies and integrals over  $h$  of the two separate parts, leads to a result that is dominated by  $R(\beta)h$ . We carry out this proof in Section 6.2.1 below.

- For the proof of

$$R(\beta)h(x) \leq R(\alpha) \left( R(\beta)h - \alpha \frac{R(\beta)h - h}{\beta} \right) (x)$$

we start with the optimizer for  $R(\beta)h(x)$ . The time integral that we end up with can be decomposed in the inverse method as for the proof above. We end up with a measure and a second set of measures; these are then dominated by the suprema in for the resolvents on the right of the inequality. We carry out the proof in Section 6.2.2 below.

Both proofs are inspired by the proof of Lemma 8.20 of [9] where the pseudo-resolvent property is established for the deterministic case.

### 6.2.1 Concatenating measures

In this section, we will prove that

$$R(\beta)h \geq R(\alpha) \left( R(\beta)h - \alpha \frac{R(\beta)h - h}{\beta} \right). \tag{6.6}$$

We start by introducing the procedure of concatenating measures. Even though the right-hand side of (6.6) involves the concatenation of measures at a random time, we will start by concatenating at a fixed time  $s$ . Using these particular concatenations, we obtain a lower bound for  $R(\lambda)h$  for each  $s$ . Afterwards, we take a convex combination over lower bounds, and recover the right-hand side.

For  $s \geq 0$  and  $X, Y \in D_E(\mathbb{R}^+)$  such that  $X(s) = Y(0)$ , define the concatenation  $\kappa_{X,Y}^s \in D_E(\mathbb{R}^+)$  of the trajectories  $X$  and  $Y$  at time  $s$  by

$$\kappa_{X,Y}^s(t) = \begin{cases} X(t) & \text{if } t \leq s, \\ Y(t-s) & \text{if } t \geq s. \end{cases}$$

Fix  $s \geq 0$  and let  $\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+))$  and let  $Q : E \rightarrow \mathcal{P}(D_E(\mathbb{R}^+))$  be a measurable map such that for all  $x \in E$  the measure  $Q^x := Q(x)$  satisfies  $Q^x(Y(0) = x) = 1$ .

For  $s, \mathbb{Q}$  and  $Q$  define the measure

$$\mathbb{Q} \odot_s Q(dZ) = \int \int \mathbb{Q}(dX) Q^{X(s)}(dY) \delta_{\kappa_{X,Y}^s}(dZ) \tag{6.7}$$

Before starting with the proof of (6.6), we start with the computation of integrals against  $\mathbb{Q} \odot_s Q$  and the computation of the relative entropy of  $\mathbb{Q} \odot_s Q$ .

**Lemma 6.5.** *Fix  $s \geq 0$ ,  $\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+))$  and a measurable map  $Q$  be a measurable map such that for all  $x \in E$  the measure  $Q^x := Q(x)$  satisfies  $Q^x(Y(0) = x) = 1$ . Define  $\mathbb{Q} \odot_s Q$  as in (6.7). Then the following hold:*

- We have

$$S_t(\mathbb{Q} \odot_s Q | \mathbb{P}_x) = \begin{cases} S_t(\mathbb{Q} | \mathbb{P}_x) & \text{if } t \leq s, \\ S_s(\mathbb{Q} | \mathbb{P}_x) + \int S_{t-s}(Q^{X(s)} | \mathbb{P}_{X(s)})\mathbb{Q}(dX) & \text{if } t > s. \end{cases}$$

- We have

$$\begin{aligned} & \int_0^\infty S_t(\mathbb{Q} \odot_s Q | \mathbb{P}_x)\tau_\beta(dt) \\ &= \int_0^s S_t(\mathbb{Q} | \mathbb{P}_x)\tau_\beta(dt) + e^{-\beta^{-1}s} \int_0^\infty \left( S_s(\mathbb{Q} | \mathbb{P}_x) + \int S_t(Q^{X(s)} | \mathbb{P}_{X(s)})\mathbb{Q}(dX) \right) \tau_\beta(dt). \end{aligned}$$

- For  $h \in C_b(E)$ , we have

$$\begin{aligned} & \int_0^\infty \int h(X(t)) (\mathbb{Q} \odot_s Q) (dX)\tau_\beta(dt) \\ &= \int_0^s \int h(X(t))\mathbb{Q}(dX)\tau_\beta(dt) + e^{-\beta^{-1}s} \int_0^\infty \int \int h(Y(t))Q^{X(s)}(dY)\mathbb{Q}(dX)\tau_\beta(dt). \end{aligned}$$

*Proof.* The case where  $t \leq s$  for the first item is immediate as  $\mathbb{Q} \odot_s Q$  equals  $\mathbb{Q}$  up until time  $s$ . Thus, fix  $s, t > s$ . For  $X \in D_E(\mathbb{R}^+)$ , define the measure  $\widehat{Q}^{s,X}(dZ) = \int Q^{X(s)}(dY)\delta_{\kappa_{X,Y}^s}(dZ)$ . It follows by definition that  $\mathbb{Q} \odot_s Q(dZ) = \int \mathbb{Q}(dX)\widehat{Q}^{s,X}(dZ)$  and that  $\widehat{Q}^{s,X}$  is the regular conditional measure of  $\mathbb{Q} \odot_s Q$  conditioned on  $\mathcal{F}_s$ . Denote by  $\mathbb{P}_{[0,s],X}$  the measure  $\mathbb{P}_x$  conditioned on  $\mathcal{F}_s$ .

Proposition A.3, applied for the conditioning on  $\mathcal{F}_s$  yields

$$S_t(\mathbb{Q} \odot_s Q | \mathbb{P}_x) = S_s(\mathbb{Q} | \mathbb{P}_x) + \int S_t(\widehat{Q}^{s,X} | \mathbb{P}_{[0,s],X})\mathbb{Q}(dX).$$

Both measures  $\widehat{Q}^{s,X}$  and  $\mathbb{P}_{[0,s],X}$  are supported by trajectories that equal  $X$  on the time interval  $[0, s]$ . Shifting both measures by  $s$ , we find  $Q^{X(s)}$  (as defined above) and by the Markov property  $\mathbb{P}_{X(s)}$ . As this shift is a isomorphism of measure spaces, we find

$$S_t(\mathbb{Q} \odot_s Q | \mathbb{P}_x) = S_s(\mathbb{Q} | \mathbb{P}_x) + \int S_{t-s}(Q^{X(s)} | \mathbb{P}_{X(s)})\mathbb{Q}(dX),$$

which establishes the first claim.

Integrating the first claim over  $\tau_\beta$ , splitting the integral into the integral over  $[0, s]$  and the part over  $(s, \infty)$ , we find

$$\begin{aligned} & \int_0^\infty S_t(\mathbb{Q} \odot_s Q | \mathbb{P}_x)\tau_\beta(dt) \\ &= \int_0^s S_t(\mathbb{Q} | \mathbb{P}_x)\tau_\beta(dt) + \int_s^\infty \left( S_s(\mathbb{Q} | \mathbb{P}_x) + \int S_{t-s}(Q^{X(s)} | \mathbb{P}_{X(s)})\mathbb{Q}(dX) \right) \tau_\beta(dt) \\ &= \int_0^s S_t(\mathbb{Q} | \mathbb{P}_x)\tau_\beta(dt) + e^{-\beta^{-1}s} \int_0^\infty \left( S_s(\mathbb{Q} | \mathbb{P}_x) + \int S_t(Q^{X(s)} | \mathbb{P}_{X(s)})\mathbb{Q}(dX) \right) \tau_\beta(dt) \end{aligned}$$

which yields the second claim. For the third claim, we have:

$$\begin{aligned} & \int_0^\infty \int h(X(t)) (\mathbb{Q} \odot_s Q) (dX)\tau_\beta(dt) \\ &= \int_0^s \int h(X(t)) (\mathbb{Q} \odot_s Q) (dX)\tau_\beta(dt) + \int_s^\infty \int h(X(t)) (\mathbb{Q} \odot_s Q) (dX)\tau_\beta(dt) \\ &= \int_0^s \int h(X(t))\mathbb{Q}(dX)\tau_\beta(dt) + \int_s^\infty \int \int h(Y(t))\widehat{Q}^{s,X}(dY)\mathbb{Q}(dX)\tau_\beta(dt) \\ &= \int_0^s \int h(X(t))\mathbb{Q}(dX)\tau_\beta(dt) + e^{-\beta^{-1}s} \int_0^\infty \int \int h(Y(t))Q^{X(s)}(dY)\mathbb{Q}(dX)\tau_\beta(dt). \quad \square \end{aligned}$$

We start with some first steps towards the proof of (6.6). As  $R(\beta)h(x)$  is given as the supremum over a collection of measures, it follows for any  $s \geq 0$ ,  $\mathbb{Q}$  and map  $Q$  that

$$R(\beta)h(x) \geq \int_0^\infty \int h(X(t)) (\mathbb{Q} \odot_s Q) (dX) - S_t(\mathbb{Q} \odot_s Q | \mathbb{P}_x) \tau_\beta(dt).$$

Applying Lemma 6.5, we find

$$\begin{aligned} R(\beta)h(x) & \geq \int_0^s \int h(X(t)) \mathbb{Q}(dX) \tau_\beta(dt) + e^{-\beta^{-1}s} \int_0^\infty \int \int h(Y(t)) Q^{X(s)}(dY) \mathbb{Q}(dX) \tau_\beta(dt) \\ & \quad - \int_0^s S_t(\mathbb{Q} | \mathbb{P}_x) \tau_\beta(dt) \\ & \quad - e^{-\beta^{-1}s} \int_0^\infty \left( S_s(\mathbb{Q} | \mathbb{P}_x) + \int S_t(\mathbb{Q}^X | \mathbb{P}_{X(s)}) \mathbb{Q}(dX) \right) \tau_\beta(dt) \end{aligned} \tag{6.8}$$

We thus obtain a lower bound for  $R(\lambda)h(x)$  for any  $s \geq 0$ . To perform the proof of (6.6), we make two additional steps:

- We pick  $\mathbb{Q}$  as the optimizer of the  $R(\alpha)$  operator in

$$R(\alpha) \left( R(\beta)h - \alpha \frac{R(\beta)h - h}{\beta} \right). \tag{6.9}$$

We construct  $Q$  from the optimizers for  $R(\beta)h$ .

- We take a convex combination of the lower bounds (6.8) obtained for these optimizers in a specifically chosen way as to recover (6.9) which will establish the inequality (6.6) that we aim to prove.

The particular convex combination is motivated by the following integral identity, of which the right-hand side of the final equation has a similar structure of the right-hand side in Lemma 6.3.

**Lemma 6.6.** *Let  $z_1, z_2 : \mathbb{R}^+ \rightarrow \mathbb{R}$  be two bounded and measurable functions. Let  $0 < \alpha < \beta$ . The measure*

$$\left(1 - \frac{\alpha}{\beta}\right) \alpha^{-1} e^{\beta^{-1}s - \alpha^{-1}s} \mathbb{1}_{\{s \geq 0\}} ds$$

is a probability measure and

$$\begin{aligned} & \int_0^\infty \left(1 - \frac{\alpha}{\beta}\right) \alpha^{-1} e^{\beta^{-1}s - \alpha^{-1}s} \left[ \int_0^s z_1(t) \tau_\beta(dt) + e^{-\beta^{-1}s} \int_0^\infty z_2(t) \tau_\beta(dt) \right] ds \\ & = \frac{\alpha}{\beta} \int_0^\infty z_1(s) \tau_\alpha(ds) + \left(1 - \frac{\alpha}{\beta}\right) \int_0^\infty \int_0^\infty z_2(t) \tau_\beta(dt) \tau_\alpha(ds). \end{aligned}$$

*Proof.* The part involving  $z_2$  is immediate. The part involving  $z_1$  follows from integration by parts over the  $s$  variable. □

*Proof of (6.6).* Fix  $h \in C_b(E)$ ,  $x \in E$  and  $0 < \alpha < \beta$ .

We aim to establish (6.6) by taking the optimizers for both optimization procedures on the right-hand side and concatenating them. This will yield a new measure that also turns up in the optimization procedure on the left-hand side, thus establishing the claim. For the concatenation, we use Lemma 6.5 to put together the relative entropies of both procedures and finish with Lemma 6.3 to obtain the correct integral form.

Thus, let  $\mathbb{Q} \in \mathcal{P}$  be the optimizer of

$$\sup_{\mathbb{Q} \in \mathcal{P}} \left\{ \int_0^\infty \int \left( 1 - \frac{\alpha}{\beta} \right) R(\beta)h(X(t)) - \frac{\alpha}{\beta} h(X(t)) \mathbb{Q}(dX) - S_t(\mathbb{Q} | \mathbb{P}_x) \tau_\alpha(dt) \right\}.$$

For any  $y \in E$  let  $Q(y) = Q^y \in \mathcal{P}$  be the optimizer of

$$\sup_{Q' \in \mathcal{P}} \left\{ \int h(Y(s)) Q'(dY) - S_s(Q' | \mathbb{P}_y) \tau_\beta(ds) \right\}.$$

Fix  $s \geq 0$ . It follows from Lemma 5.10 that the map  $Q$  is measurable. Thus, for  $s$ ,  $\mathbb{Q}$  and  $Q$ , we define  $\mathbb{Q} \odot_s Q$  as in (6.7). By definition of  $R(\beta)h(x)$ , we find

$$R(\beta)h(x) \geq \int_0^\infty \int h(X(t)) (\mathbb{Q} \odot_s Q) (dX) - S_t(\mathbb{Q} \odot_s Q | \mathbb{P}_x) \tau_\beta(dt). \tag{6.10}$$

and by Lemma 6.5:

$$\begin{aligned} R(\beta)h(x) &\geq \int_0^s \int h(X(t)) \mathbb{Q}(dX) \tau_\beta(dt) + e^{-\beta^{-1}s} \int_0^\infty \int \int h(Y(t)) Q^{X(s)}(dY) \mathbb{Q}(dX) \tau_\beta(dt) \\ &\quad - \int_0^s S_t(\mathbb{Q} | \mathbb{P}_x) \tau_\beta(dt) \\ &\quad - e^{-\beta^{-1}s} \int_0^\infty \left( S_s(\mathbb{Q} | \mathbb{P}_x) + \int S_t(Q^{X(s)} | \mathbb{P}_{X(s)}) \mathbb{Q}(dX) \right) \tau_\beta(dt) \end{aligned} \tag{6.11}$$

Thus, for each fixed  $s$ , we find a lower bound for  $R(\beta)h(x)$  that is in terms of the optimizing measures for (6.4) and (6.5). We proceed by showing that by taking a convex combination over  $s$  in the right way, we can obtain the desired lower bound.

If we integrate (6.11) against the density  $\left( 1 - \frac{\alpha}{\beta} \right) \alpha^{-1} e^{\beta^{-1}s - \alpha^{-1}s}$  on  $\mathbb{R}^+$  introduced in Lemma 6.6 we find

$$\begin{aligned} R(\beta)h(x) &\geq \frac{\alpha}{\beta} \int_0^\infty \int h(X(s)) \mathbb{Q}(dX) \tau_\alpha(ds) \\ &\quad + \left( 1 - \frac{\alpha}{\beta} \right) \int_0^\infty \int_0^\infty \int \int h(Y(t)) Q^{X(s)}(dY) \mathbb{Q}(dX) \tau_\beta(dt) \tau_\alpha(ds) \\ &\quad - \frac{\alpha}{\beta} \int_0^\infty S_s(\mathbb{Q} | \mathbb{P}_x) \tau_\alpha(ds) \\ &\quad - \left( 1 - \frac{\alpha}{\beta} \right) \int_0^\infty \int_0^\infty \left( S_s(\mathbb{Q} | \mathbb{P}_x) + \int S_t(Q^{X(s)} | \mathbb{P}_{X(s)}) \mathbb{Q}(dX) \right) \tau_\beta(dt) \tau_\alpha(ds) \end{aligned}$$

Recall that the measures  $Q^{X(s)}$  arose from the optimizers for  $R(\beta)h(x)$ . We thus find that

$$R(\beta)h(x) \geq \int_0^\infty \left\{ \int \left( \left( 1 - \frac{\alpha}{\beta} \right) R(\beta)h(X(s)) + \frac{\alpha}{\beta} h(X(s)) \right) \mathbb{Q}(dX) - S_s(\mathbb{Q} | \mathbb{P}_x) \right\} \tau_\alpha(ds).$$

As  $\mathbb{Q}$  was chosen as the optimizer for

$$R(\alpha) \left( R(\beta)h - \alpha \frac{R(\beta)h - h}{\beta} \right)$$

we obtain (6.6) finalizing the proof. □

### 6.2.2 Decomposing measures

In this section, we will prove that

$$R(\beta)h \leq R(\alpha) \left( R(\beta)h - \alpha \frac{R(\beta)h - h}{\beta} \right). \quad (6.12)$$

The main step in the proof is to decompose the measure that turns up as the optimizer in the variational problem defining  $R(\beta)h$ . Fix  $x \in E$  and let  $\mathbb{Q} \in \mathcal{P}$  such that

$$R(\beta)h(x) = \int_0^\infty \int h(X(t))\mathbb{Q}(dX) - S_t(\mathbb{Q} | \mathbb{P}_x)\tau_\beta(dt).$$

By general measure theoretic arguments, we can find for every fixed  $t$  a  $\mathcal{F}_t$  measurable family of measures  $X \mapsto \mathbb{Q}^{t,X}$  such that

$$\mathbb{Q}(dY) = \int \mathbb{Q}^{t,X}(dY) \mathbb{Q}(dX) \quad (6.13)$$

and such that if  $\mathbb{Q}^{t,X}$  is restricted to trajectories up to time  $t$  we find  $\delta_X$ . Denote by  $\widehat{\mathbb{Q}}^{t,X}$  the measure that is obtained from  $\mathbb{Q}^{t,X}$  under the push-forward map

$$\theta^t(X)(s) = X(t+s).$$

Thus,  $\widehat{\mathbb{Q}}^{t,X}$  is supported by trajectories such that  $Y(0) = X(t)$  (for  $\mathbb{Q}$  almost all  $X$ ).

*Proof of (6.12).* Combining Lemma 6.3 and the decomposition of  $\mathbb{Q}$  following (6.13) yields that

$$\begin{aligned} \int_0^\infty \int h(X(t))\mathbb{Q}(dX)\tau_\beta(dt) &= \frac{\alpha}{\beta} \int_0^\infty \int h(X(t))\mathbb{Q}(dX)\tau_\alpha(dt) \\ &+ \left(1 - \frac{\alpha}{\beta}\right) \int_0^\infty \int_0^\infty \int \int h(Y(s))\widehat{\mathbb{Q}}^{t,X}(dY) \mathbb{Q}(dX) \tau_\beta(ds) \tau_\alpha(dt). \end{aligned}$$

Thus, if we can prove that

$$\begin{aligned} \int_0^\infty S_t(\mathbb{Q} | \mathbb{P}_x)\tau_\beta(dt) &= \frac{\alpha}{\beta} \int_0^\infty S_t(\mathbb{Q} | \mathbb{P}_x)\tau_\alpha(dt) \\ &+ \left(1 - \frac{\alpha}{\beta}\right) \int_0^\infty \left( S_t(\mathbb{Q} | \mathbb{P}_x) + \int \int_0^\infty S_s(\widehat{\mathbb{Q}}^{t,X} | \mathbb{P}_{X(t)})\tau_\beta(ds) \mathbb{Q}(dX) \right) \tau_\alpha(dt) \quad (6.14) \end{aligned}$$

then we obtain (6.12) by replacing  $\widehat{\mathbb{Q}}^{t,X}$  by its optimum to obtain  $R(\beta)h(X(t))$  in the integrand and afterwards optimizing to obtain  $R(\alpha)$ . This, however, follows similarly as in the proof of the first inequality in Section 6.2.1.  $\square$

### 6.3 A variational semigroup generated by the resolvent

We conclude this section by proving Proposition 3.8, that is, we establish that the resolvent approximates the semigroup.

Again, the key idea is to reduce to a property of exponential distributions. This time, we will use that the sum of  $n$  independent exponential random variables with mean  $t/n$  converges to  $t$ . As the resolvent is defined in terms of an optimization procedure, we cannot directly apply this intuition. However, we will use natural upper and lower bounds for concatenations of  $R(\lambda)$  that we can control.

**Proposition 6.7.** For each  $h \in C_b(E)$ ,  $t \geq 0$  and  $x \in E$  we have

$$\lim_{n \rightarrow \infty} R\left(\frac{t}{n}\right)^n h = V(t)h$$

for the strict topology.

The result will follow immediately from Lemma's 6.8 and 6.9 below. We start with the definition of some additional operators. For each distribution  $\tau \in \mathcal{P}(\mathbb{R}^+)$  and  $h \in C_b(E)$ , define

$$T^+(\tau)h(x) := \int_0^\infty \sup_{\mathbb{Q} \in \mathcal{P}} \left\{ \int h(X(t))\mathbb{Q}(dX) - S_t(\mathbb{Q} | \mathbb{P}_x) \right\} \tau(dt),$$

$$T^-(\tau)h(x) := \sup_{\mathbb{Q} \in \mathcal{P}} \int_0^\infty \left\{ \int h(X(t))\mathbb{Q}(dX) - S_t(\mathbb{Q} | \mathbb{P}_x) \right\} \tau(dt).$$

For all  $\tau$  and  $h$ , we have  $T^+(\tau)h \geq T^-(\tau)h$ . For exponential random variables  $\tau_\lambda$  or fixed times  $t$ , we find

$$T^+(\tau_\lambda)h \geq R(\lambda)h = T^-(\tau_\lambda)h, \quad T^+(\delta_t)h = V(t)h = T^-(\delta_t)h.$$

**Lemma 6.8.** For  $\tau_1, \tau_2$ , we have

$$T^+(\tau_1 * \tau_2)h \geq T^+(\tau_1)T^+(\tau_2)h,$$

$$T^-(\tau_1 * \tau_2)h \leq T^-(\tau_1)T^-(\tau_2)h.$$

*Proof.* The first claim follows by similar, but easier, arguments as in the proof of (6.6) in Section 6.2.1. Similarly, for the second claim, we refer to the arguments in Section 6.2.2. □

**Lemma 6.9.** Let  $h \in C_b(E)$  and  $t \in \mathbb{R}^+$  and let  $\tau_n \in \mathcal{P}(\mathbb{R}^+)$  be such that  $\tau_n \rightarrow \delta_t$ .

Then we have

$$\lim_n T^+(\tau_n)h = T^+(\delta_t)h = V(t)h$$

for the strict topology. In addition, we have for each sequence  $x_n \rightarrow x$  that

$$\liminf_{n \rightarrow \infty} T^-(\tau_n)h(x_n) \geq T^-(\delta_t)h(x) = V(t)h(x)$$

as well as  $\sup_n \|T^-(\tau_n)h\| \leq \|h\|$ .

*Proof.* Fix  $h \in C_b(E)$  and a sequence  $\tau_n$  and  $t$  such that  $\tau_n \rightarrow \delta_t$ . Note that it is immediate that  $\sup_n \|T^+(\tau_n)h\| \leq \|h\|$  and  $\sup_n \|T^-(\tau_n)h\| \leq \|h\|$ .

We proceed by establishing strict convergence for  $T^+(\tau_n)h$ . By Lemma A.1, we have

$$T^+(\tau_n)h(x) = \int_0^\infty V(t)h(x)\tau_n(dt).$$

By Lemma 5.9 the map  $t \mapsto V(t)f$  is continuous for the strict topology. Thus strict continuity of  $\tau \mapsto T^+(\tau)h$  follows.

For the second statement, fix  $x_n$  converging to  $x$  in  $E$ . Let  $\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+))$  such that

$$T^-(\delta_t)h(x) = V(t)h(x) = \int h(X(t))\mathbb{Q}(dX) - S_t(\mathbb{Q} | \mathbb{P}_x)$$

and such that  $S_t(\mathbb{Q} | \mathbb{P}_{x_n}) = S(\mathbb{Q} | \mathbb{P}_x)$ .

By Proposition 5.4, we can find  $\mathbb{Q}_n \in \mathcal{P}_n$  such that  $\mathbb{Q}_n \rightarrow \mathbb{Q}$  and such that for each  $s$  we have  $\limsup_s S_s(\mathbb{Q}_n | \mathbb{P}_{x_n}) \leq S_s(\mathbb{Q} | \mathbb{P}_x)$  and  $S_s(\mathbb{Q}_n | \mathbb{P}_{x_n}) \leq S_s(\mathbb{Q} | \mathbb{P}_x) + 1$  for all  $n$  and  $s$ . These properties imply that

$$S_s(\mathbb{Q}_n | \mathbb{P}_{x_n}) \leq \begin{cases} S_t(\mathbb{Q}_n | \mathbb{P}_{x_n}) & \text{if } s < t + 1, \\ S_t(\mathbb{Q}_n | \mathbb{P}_{x_n}) + 1 & \text{if } s \geq t + 1. \end{cases} \quad (6.15)$$

Thus, applying the  $\liminf_n$  to  $T^-(\tau_n)h(x_n)$ , we find

$$\begin{aligned} \liminf_{n \rightarrow \infty} T^-(\tau_n)h(x_n) &\geq \liminf_n \int_0^\infty \left\{ \int h(X(s))\mathbb{Q}_n(dX) - S_s(\mathbb{Q}_n | \mathbb{P}_{x_n}) \right\} \tau_n(ds) \\ &\geq \liminf_{n \rightarrow \infty} \int_0^\infty \int h(X(s))\mathbb{Q}_n(dX) \tau_n(ds) - \limsup_{n \rightarrow \infty} \int_0^\infty S_s(\mathbb{Q}_n | \mathbb{P}_{x_n}) \tau_n(ds). \end{aligned}$$

As  $\mathbb{Q}_n \rightarrow \mathbb{Q}$  and  $\tau_n \rightarrow \tau$  and the map  $s \mapsto X(s)$  is continuous at  $t$  for  $\mathbb{Q}$  almost every  $X$  as  $\mathbb{Q} \ll \mathbb{P}_x$ , the first term converges to  $\int h(X(t))\mathbb{Q}(dX)$ . For the second term, we obtain by (6.15) and the property that  $S_t(\mathbb{Q} | \mathbb{P}_x) = S(\mathbb{Q} | \mathbb{P}_x)$

$$\begin{aligned} \limsup_{n \rightarrow \infty} \int_0^\infty S_s(\mathbb{Q}_n | \mathbb{P}_{x_n}) \tau_n(ds) &\leq \limsup_{n \rightarrow \infty} S_{t+1}(\mathbb{Q}_n | \mathbb{P}_{x_n}) + \tau_n([t + 1, \infty)) \\ &\leq S_{t+1}(\mathbb{Q} | \mathbb{P}_x) \\ &= S_t(\mathbb{Q} | \mathbb{P}_x). \end{aligned}$$

We conclude that  $\liminf_n T^-(\tau_n)h(x_n) \geq V(t)f(x_n)$ . □

## 7 A large deviation principle for Markov processes

In Section 4, we considered a sequence of Markov processes on a Polish space  $E$  and stated a large deviation principle on  $D_E(\mathbb{R}^+)$ . In this section, we prove a more general version of this result that takes into account variations that one runs into in practice. As a first generalization, we consider Markov processes  $t \mapsto X_n(t)$  on a sequence of spaces  $E_n$  that are embedded into some space  $E$  using maps  $\eta_n : E_n \rightarrow E$ .

As an example  $X_n$  could be a process on  $E_n := \{-1, 1\}^n$ , whereas we are interested in the large deviation behaviour of the average of the  $n$  values which takes values in  $E = [-1, 1]$ .

In Theorem 4.2, we assumed exponential tightness and that certain sequences of functions converge. We need to modify these two concepts to allow for a sequence of spaces.

- We want to establish convergence of functions that are defined on different spaces. We therefore need a new notion of bounded and uniform convergence on compact sets. The key step in this definition will be to assign to each compact set  $K \subseteq E$  a sequence of compact sets  $K_n \subseteq E_n$  so that  $\eta_n(K_n)$  ‘converge’ to  $K$ . In fact, to have a little bit more flexibility in our assignment of compact sets, we will work below with an large index set  $\mathcal{Q}$  so that to each  $q \in \mathcal{Q}$  we associate compact sets  $K_n^q \subseteq E_n$  and  $K^q \subseteq E$ .
- Exponential tightness and buc convergence can be exploited together to make sure we get proper limiting statements. As our notion of buc convergence changes, we have to adapt our notion of exponential tightness to take into account the index set  $\mathcal{Q}$ .

We make two additional generalizations that are useful in practice.

- Often, it is hard to find an operator  $H \subseteq C_b(E) \times C_b(E)$  that is a limit of  $H_n \subseteq C_b(E_n) \times C_b(E_n)$ . Rather one finds upper and lower bounds  $H_{\dagger}$  and  $H_{\ddagger}$  for the sequence  $H_n$ . See also Question 3.10 on whether at the pre-limit level one is able to work with upper and lower bounds.
- In the context of averaging or homogenisation, the natural limiting operator  $H$  is a subset of  $C_b(E) \times C_b(F)$ , where  $F$  is some space that takes into account additional information. For example  $F = E \times \mathbb{R}$ , where the additional component  $\mathbb{R}$  takes into account the information of a fast process or a microscopic scale.

We thus start with a section on preliminaries that allows us to talk about these four extensions.

### 7.1 Preliminary definitions

Recall that  $M(E)$  is the space of bounded measurable functions  $f : E \rightarrow [-\infty, \infty]$ . Denote

$$USC_u(E) := \left\{ f \in M(E) \mid f \text{ upper semi-continuous, } \sup_x f(x) < \infty \right\},$$

$$LSC_l(E) := \left\{ f \in M(E) \mid f \text{ lower semi-continuous, } \inf_x f(x) > -\infty \right\}.$$

**Definition 7.1** (Kuratowski convergence). *Let  $\{A_n\}_{n \geq 1}$  be a sequence of subsets in a space  $E$ . We define the limit superior and limit inferior of the sequence as*

$$\limsup_{n \rightarrow \infty} A_n := \{x \in E \mid \forall U \in \mathcal{U}_x \forall N \geq 1 \exists n \geq N : A_n \cap U \neq \emptyset\},$$

$$\liminf_{n \rightarrow \infty} A_n := \{x \in E \mid \forall U \in \mathcal{U}_x \exists N \geq 1 \forall n \geq N : A_n \cap U \neq \emptyset\}.$$

where  $\mathcal{U}_x$  is the collection of open neighbourhoods of  $x$  in  $E$ . If  $A := \limsup_n A_n = \liminf_n A_n$ , we write  $A = \lim_n A_n$  and say that  $A$  is the Kuratowski limit of the sequence  $\{A_n\}_{n \geq 1}$ .

#### 7.1.1 Embedding spaces

Our main result will be based on the following setting.

**Assumption 7.2.** *We have spaces  $E_n$  and  $E, F$  and continuous maps  $\eta_n : E_n \rightarrow E$ ,  $\hat{\eta}_n : E_n \rightarrow F$  and a continuous surjective map  $\gamma : F \rightarrow E$  such that the following diagram commutes:*

$$\begin{array}{ccc} & \hat{\eta}_n & \rightarrow F \\ E_n & \searrow & \downarrow \gamma \\ & \eta_n & \rightarrow E \end{array}$$

*In addition, there is a directed set  $\mathcal{Q}$  (partially ordered set such that every two elements have an upper bound) such that for each  $q \in \mathcal{Q}$ , we have compact sets  $K_n^q \subseteq E_n$  and compact sets  $K^q \subseteq E$  and  $\hat{K}^q \subseteq F$  such that*

1. *If  $q_1 \leq q_2$ , we have  $K^{q_1} \subseteq K^{q_2}$ ,  $\hat{K}^{q_1} \subseteq \hat{K}^{q_2}$  and for all  $n$  we have  $K_n^{q_1} \subseteq K_n^{q_2}$ .*
2. *For all  $q \in \mathcal{Q}$  we have  $\bigcup_n \hat{\eta}_n(K_n^q) \subseteq \hat{K}^q$ .*

3. For each compact set  $K \subseteq E$ , there is a  $q \in \mathcal{Q}$  such that

$$K \subseteq \liminf_n \eta_n(K_n^q).$$

4. We have  $\gamma(\widehat{K}^q) \subseteq K^q$ .

**Remark 7.3.** Note that 2 implies that  $\limsup_n \widehat{\eta}_n(K_n^q) \subseteq \widehat{K}^q$  and together with 4 that  $\limsup_n \eta_n(K_n^q) \subseteq K^q$ .

Thus, the final three conditions imply that the sequences  $\eta_n(K_n^q)$  for various  $q \in \mathcal{Q}$  covers all compact sets in  $E$ , and also are covered by compact sets in  $E$  (in fact this final statement holds on the larger space  $F$ ). This implies that the index set  $\mathcal{Q}$  connects the structure of compact sets in  $E$  and  $F$  in a suitable way to (a subset) of the compact sets of the sequence  $E_n$ .

We use our index set  $\mathcal{Q}$  to extend our notion of bounded and uniform convergence on compacts sets.

**Definition 7.4.** Let Assumption 7.2 be satisfied. For each  $n$  let  $f_n \in M_b(E_n)$  and  $f \in M_b(E)$ . We say that  $\text{LIM } f_n = f$  if

- $\sup_n \|f_n\| < \infty$ ,
- for all  $q \in \mathcal{Q}$  and  $x_n \in K_n^q$  and  $x \in K^q$  such that  $\eta_n(x_n) \rightarrow x$  we have

$$\lim_{n \rightarrow \infty} |f_n(x_n) - f(x)| = 0.$$

**Remark 7.5.** Note that if  $f \in C_b(E)$  and  $f_n \in M_b(E_n)$ , we have that  $\text{LIM } f_n = f$  if and only if

- $\sup_n \|f_n\| < \infty$ ,
- if for all  $q \in \mathcal{Q}$

$$\lim_{n \rightarrow \infty} \sup_{x \in K_n^q} |f_n(x) - f(\eta_n(x))| = 0.$$

### 7.1.2 Viscosity solutions of Hamilton-Jacobi equations

Below we will introduce a more general version of viscosity solutions compared to Section 2. One recovers the old definition by taking  $B_{\dagger} = B_{\ddagger} = B$ ,  $F = E$  and  $\gamma(x) = x$ .

**Definition 7.6.** Let  $B_{\dagger} \subseteq LSC_l(E) \times USC_u(F)$  and  $B_{\ddagger} \subseteq USC_u(E) \times LSC_l(F)$ . Fix  $h_1, h_2 \in C_b(E)$ . Consider the equations

$$f - B_{\dagger}f = h_1, \tag{7.1}$$

$$f - B_{\ddagger}f = h_2. \tag{7.2}$$

- We say that  $u : X \rightarrow \mathbb{R}$  is a subsolution of equation (7.1) if  $u \in USC_u(E)$  and if, for all  $(f, g) \in B_{\dagger}$  such that  $\sup_x u(x) - f(x) < \infty$  there is a sequence  $y_n \in F$  such that

$$\lim_{n \rightarrow \infty} u(\gamma(y_n)) - f(\gamma(y_n)) = \sup_x u(x) - f(x), \tag{7.3}$$

and

$$\limsup_{n \rightarrow \infty} u(\gamma(y_n)) - g(y_n) - h_1(\gamma(y_n)) \leq 0. \tag{7.4}$$

- We say that  $v : E \rightarrow \mathbb{R}$  is a supersolution of equation (7.2) if  $v \in LSC_l(E)$  and if, for all  $(f, g) \in B_{\dagger}$  such that  $\inf_x v(x) - f(x) > -\infty$  there is a sequence  $y_n \in Y$  such that

$$\lim_{n \rightarrow \infty} v(\gamma(y_n)) - f(\gamma(y_n)) = \inf_x v(x) - f(x), \quad (7.5)$$

and

$$\liminf_{n \rightarrow \infty} v(\gamma(y_n)) - g(y_n) - h_2(\gamma(y_n)) \geq 0. \quad (7.6)$$

- We say that  $u$  is a solution of the pair of equations (7.1) and (7.2) if it is both a subsolution for  $B_{\dagger}$  and a supersolution for  $B_{\ddagger}$ .
- We say that (7.1) and (7.2) satisfy the comparison principle if for every subsolution  $u$  to (7.1) and supersolution  $v$  to (7.2), we have

$$\sup_x u(x) - v(x) \leq \sup_x h_1(x) - h_2(x). \quad (7.7)$$

If  $B = B_{\dagger} = B_{\ddagger}$  and  $h = h_1 = h_2$ , we will say that the comparison principle holds for  $f - \lambda Bf = h$ , if for any subsolution  $u$  for  $f - \lambda Bf = h_1$  and supersolution  $v$  of  $f - \lambda Bf = h_2$  the estimate in (7.7) holds.

### 7.1.3 Notions of convergence of Hamiltonians

We now introduce our notion of upper and lower bound for the sequence  $H_n$ .

**Definition 7.7.** Consider the setting of Assumption 7.2. Suppose that for each  $n$  we have operators  $H_n \subseteq C_b(E_n) \times C_b(E_n)$ .

1. The extended  $\limsup$   $ex - \text{LIMSUP}_n H_n$  is defined by the collection  $(f, g) \in H_{\dagger} \subseteq LSC_l(E) \times USC_u(F)$  such that there exist  $(f_n, g_n) \in H_n$  satisfying

$$\text{LIM } f_n \wedge c = f \wedge c, \quad \forall c \in \mathbb{R}, \quad (7.8)$$

$$\sup_n \sup_{x \in E_n} g_n(x) < \infty, \quad (7.9)$$

and if for any  $q \in \mathcal{Q}$  and sequence  $z_{n(k)} \in K_{n(k)}^q$  (with  $k \mapsto n(k)$  strictly increasing) such that  $\lim_k \widehat{\eta}_{n(k)}(z_{n(k)}) = y$  in  $F$  with  $\lim_k f_{n(k)}(z_{n(k)}) = f(\gamma(y)) < \infty$  we have

$$\limsup_{k \rightarrow \infty} g_{n(k)}(z_{n(k)}) \leq g(y). \quad (7.10)$$

2. The extended  $\liminf$   $ex - \text{LIMINF}_n H_n$  is defined by the collection  $(f, g) \in H_{\ddagger} \subseteq USC_u(E) \times LSC_l(F)$  such that there exist  $(f_n, g_n) \in H_n$  satisfying

$$\text{LIM } f_n \vee c = f \vee c, \quad \forall c \in \mathbb{R}, \quad (7.11)$$

$$\inf_n \inf_{x \in E_n} g_n(x) > -\infty, \quad (7.12)$$

and if for any  $q \in \mathcal{Q}$  and sequence  $z_{n(k)} \in K_{n(k)}^q$  (with  $k \mapsto n(k)$  strictly increasing) such that  $\lim_k \widehat{\eta}_{n(k)}(z_{n(k)}) = y$  in  $F$  with  $\lim_k f_{n(k)}(z_{n(k)}) = f(\gamma(y)) > -\infty$  we have

$$\liminf_{k \rightarrow \infty} g_{n(k)}(z_{n(k)}) \geq g(y). \quad (7.13)$$

**Remark 7.8.** The conditions in (7.8) and (7.11) are implied by  $\text{LIM } f_n = f$ . Conditions (7.9) and (7.10) are implied by  $\text{LIM}_n g_n \leq g$  whereas conditions (7.12) and (7.13) are implied by  $\text{LIM}_n g_n \geq g$ .

Comparing this to Definition 2.3, we indeed see that the sub and super-limit can be interpreted as upper and lower bounds instead of limits.

### 7.2 Large deviations for Markov process

We proceed by stating our main large deviation result, which extends Theorem 4.2. We first give the appropriate generalization of Condition 4.1.

**Condition 7.9.** Let  $E_n, E, F, \eta_n, \hat{\eta}_n, \gamma$  be as in Assumption 7.2.

Let  $A_n \subseteq C_b(E_n) \times C_b(E_n)$  be linear operators and let  $r_n$  be positive real numbers such that  $r_n \rightarrow \infty$ . Suppose that

- The martingale problems for  $A_n$  are well-posed on  $E_n$ . Denote the processes corresponding to the operators  $A_n$  by  $t \mapsto Y_n(t)$ . Denote by  $\mathbb{P}_y^n$  the law of  $Y_n$  when  $Y_n(0) = y$ .
- For each  $n$  that  $y \mapsto \mathbb{P}_y^n$  is continuous for the weak topology on  $\mathcal{P}(D_{E_n}(\mathbb{R}^+))$ .
- [Exponential compact containment] For each  $q \in \mathcal{Q}, T > 0$  and  $a_2 > 0$  there exists  $\hat{q} = \hat{q}(q, T, a_2) \in \mathcal{Q}$  such that

$$\limsup_{n \rightarrow \infty} \sup_{y \in K_n^q} \frac{1}{r_n} \log \mathbb{P}_y^n [\exists t \leq T : Y_n(t) \notin K_n^{\hat{q}}] \leq -a_2.$$

Note that these conditions can be mapped to the ones of Condition 4.1 for the random variables  $X_n = \eta_n(Y_n)$ .

We will introduce one additional assumption in the theorem below on top of the large deviation principle for the random variables  $X_n(0)$ . In the proof below, we will argue using the Markov property for  $Y_n$  as  $X_n = \eta_n(Y_n)$  might not be Markovian. Thus, to use effectively results from semigroup theory, we need that the random variables  $Y_n(0)$  themselves concentrate on appropriate compact sets.

In the basic case of Theorem 7.14 in which  $X_n = Y_n$  this property is equivalent to exponential tightness and thus implied by the large deviation principle.

**Theorem 7.10.** Suppose that we are in the setting of Assumption 7.2 and that Condition 7.9 is satisfied. Denote  $X_n = \eta_n(Y_n)$ . Define the resolvent  $R_n(\lambda)$  and operator semigroup  $V_n(t)$  on  $C_b(E_n)$  by:

$$R_n(\lambda)h(x) := \sup_{\mathbb{Q} \in \mathcal{P}(D_{E_n}(\mathbb{R}^+))} \left\{ \int_0^\infty \lambda^{-1} e^{-\lambda^{-1}t} \left( \int h(X(t)) \mathbb{Q}(dX) - \frac{1}{r_n} S_t(\mathbb{Q} | \mathbb{P}_x^n) \right) dt \right\},$$

$$V_n(t)f(y) := \frac{1}{r_n} \log \mathbb{E} \left[ e^{r_n f(Y_n(t))} \mid Y_n(0) = y \right].$$

Let  $H_n$  be the set of pairs  $(f, g) \in C_b(E_n) \times C_b(E_n)$  such that

$$t \mapsto \exp \left\{ r_n \left( f(Y_n(t)) - f(Y_n(0)) - \int_0^t g(Y_n(s)) ds \right) \right\}$$

are martingales with respect to  $\mathcal{F}_t^n := \sigma\{Y_n(s) \mid s \leq t\}$ . Suppose furthermore that

1. For each  $a_1 > 0$  there is a  $q \in \mathcal{Q}$  such that

$$\limsup_{n \rightarrow \infty} \frac{1}{r_n} \log \mathbb{P} [Y_n(0) \notin K_n^q] \leq -a_1.$$

In addition, the large deviation principle holds for  $X_n(0) = \eta_n(Y_n(0))$  with speed  $r_n$  and good rate function  $I_0$ .

2. The processes  $X_n = \eta_n(Y_n)$  are exponentially tight on  $D_E(\mathbb{R}^+)$ .
3. There are two operators  $H_\dagger \subseteq LSC_l(E) \times USC_u(F)$  and  $H_\ddagger \subseteq USC_l(E) \times LSC_u(F)$  such that  $H_\dagger \subseteq ex - \text{LIMSUP } H_n$  and  $H_\ddagger \subseteq ex - \text{LIMINF } H_n$ .

4. Let  $D \subseteq C_b(E)$  be a quasi-dense subset of  $C_b(E)$ . Suppose that for all  $h \in D$  and  $\lambda > 0$  the comparison principle holds for viscosity subsolutions to  $f - \lambda H_{\dagger} f = h$  and supersolutions to  $f - \lambda H_{\ddagger} f = h$ .

Then there is a pseudo-resolvent  $R(\lambda)$  and a semigroup  $V(t)$  on  $C_b(E)$  such that

- If  $\lambda > 0$  and  $\text{LIM } h_n = h$ , then  $\text{LIM } R_n(\lambda)h_n = R(\lambda)h$ ;
- For  $h \in D$ , the function  $R(\lambda)h$  is the unique function that is both a subsolution to  $f - \lambda H_{\dagger} f = h$  and a supersolution to  $f - \lambda H_{\ddagger} f = h$ ;
- If  $\text{LIM } f_n = f$  and  $t_n \rightarrow t$  we have  $\text{LIM } V_n(t_n)f_n = V(t)f$ .

In addition, the processes  $X_n = \eta_n(Y_n)$  satisfy a large deviation principle on  $D_E(\mathbb{R}^+)$  with speed  $r_n$  and rate function

$$I(\gamma) = I_0(\gamma(0)) + \sup_{k \geq 1} \sup_{0=t_0 < t_1 < \dots < t_k} \sum_{i=1}^k I_{t_i - t_{i-1}}(\gamma(t_i) | \gamma(t_{i-1})). \quad (7.14)$$

Here  $\Delta_\gamma^c$  is the set of continuity points of  $\gamma$ . The conditional rate functions  $I_t$  are given by

$$I_t(y | x) = \sup_{f \in C_b(E)} \{f(y) - V(t)f(x)\}.$$

We proceed with a two remarks on how to obtain exponential tightness of the processes and the variational representation of the rate function.

We start with the exponential tightness. The verification of exponential tightness of the processes  $\eta_n(Y_n)$  comes down to verifying two statements. The first one is exponential compact containment, which has been assumed in Condition 7.9. The second one is to control the oscillations of the process, which can often be achieved by considering the exponential martingales. This has been done in the proof of Corollary 4.19 of [9]. We state it for completeness, including a definition that we need in its statement.

**Definition 7.11.** Let  $q$  be a metric that generates the topology on  $E$ . We say that  $\mathfrak{D} \subseteq C_b(E)$  approximates the metric  $q$  if for each compact  $K \subseteq E$  and  $z \in K$  there exist  $f_n \in \mathfrak{D}$  such that

$$\limsup_n \sup_{x \in K} |f_n(x) - q(x, z)| = 0.$$

**Proposition 7.12** (Corollary 4.19 [9]). Suppose that we are in the setting of Assumption 7.2 and Condition 7.9. Let  $r_n > 0$  be some sequence such that  $r_n \rightarrow \infty$ . Denote  $X_n = \eta_n(Y_n)$ . Let  $\mathfrak{D} \subseteq C_b(E)$  and  $\mathfrak{S} \subseteq \mathbb{R}$ . Suppose that

1. Either  $\mathfrak{D}$  is closed under addition and separates points in  $E$  and  $\mathfrak{S} = \mathbb{R}$  or  $\mathfrak{D}$  approximates a metric  $q$  and  $\mathfrak{S} = (0, \infty)$ .
2. For each  $\lambda \in \mathfrak{S}$  and  $f \in \mathfrak{D}$  there are  $(f_n, g_n)$  such that  $(\lambda f_n, g_n) \in \mathcal{D}(H_n)$  with  $\text{LIM } f_n = f$  and for all  $q \in \mathfrak{Q}$

$$\sup_n \sup_{x \in K_n^q} g_n(x) < \infty.$$

Then the sequence of processes  $\{X_n\}$  is exponentially tight.

Note that Condition 2 often follows from the convergence  $H_{\dagger} \subseteq ex - \text{LIM SUP}_n H_n$ . We proceed with a remark on the variational representation of the rate function.

**Remark 7.13.** For an expression of the large deviation rate-functional in a Lagrangian form, one can show that a variational resolvent, similar to the one in this paper, but with a Lagrangian instead of an entropy as a penalization, solves the limiting Hamilton-Jacobi equation. This has been carried out in Chapter 8 of [9]. In simple cases, this leads to an expression

$$I(\gamma) = \begin{cases} I_0(\gamma(0)) + \int \mathcal{L}(\gamma(s), \dot{\gamma}(s)) ds & \text{if } \gamma \text{ is absolutely continuous} \\ \infty & \text{otherwise.} \end{cases}$$

$\mathcal{L}$  can usually be obtained from the operators  $H_{\dagger}$  and  $H_{\ddagger}$  by a (Legendre) transformation. Often one formally has  $H_{\dagger}f(x) = \mathcal{H}(x, \nabla f(x))$ <sup>2</sup> and  $H_{\ddagger}f(x) = \mathcal{H}(x, \nabla f(x))$ .  $\mathcal{L}$  is then obtained as  $\mathcal{L}(x, \dot{x}) = \sup_p \langle p, \dot{x} \rangle - \mathcal{H}(x, p)$ .

We refrain from carrying out this step as it would follow [9, Chapter 8] exactly.

### 7.3 Strategy of the proof and discussion on the method of proof

Feng and Kurtz [9] showed in their extensive monograph that path-space large deviations of the processes  $X_n = \eta_n(Y_n)$  on  $D_E(\mathbb{R}^+)$  can be obtained by establishing exponential tightness and the convergence of the non-linear semigroups  $V_n(t)$ .

We repeat the important steps in this approach.

- (1) A projective limit theorem (rather a special version of the projective limit theorem and the inverse contraction principle, [9, Theorem 4.28]) for the Skorokhod space establishes that, given exponential tightness, it suffices to establish large deviations for the finite dimensional distributions of  $X_n = \eta_n(Y_n)$ .
- (2) By Bryc's theorem, the large deviations for finite dimensional distributions follow from the convergence of the rescaled log-Laplace transforms.
- (3) Using the Markov property, one can reduce the convergence of the log-Laplace transforms to the large deviation principle at time 0 and the convergence of semigroups.

We will give a new proof of the path-space large deviation principle on the basis of this strategy. However, the key component of establishing the convergence of semigroup will be based on the explicit identification of the resolvents of the non-linear semigroups and the semigroup convergence result of [14].

At this point we remark two differences with the main result of [9].

Throughout we assume that the maps  $\eta_n, \hat{\eta}_n$  and  $x \mapsto \mathbb{P}_x^n$  are continuous, whereas in [9] they are allowed to be measurable only. The results in [14] allow one to work with measurable resolvents also, but the methods of the first part of this paper are based on properties of continuous functions. It would be of interest to see whether these methods can be extended to the context of measurable functions also.

The key point why [9] can work with measurable maps is the approximation of the processes  $X_n$  by their Yosida approximants. This approximation does introduce an extra condition into the notions of  $ex - \text{LIMSUP}$  and  $ex - \text{LIMINF}$ . Compare our (7.9) and (7.12) to Equations (7.19) and (7.22) of [9].

### 7.4 Proof of Theorem 7.10

The following result is based on the variant of the projective limit theorem and Bryc's theorem. See Theorem 5.15, Remark 5.16 and Corollary 5.17 in [9].

<sup>2</sup>Formally  $\mathcal{H}$  should be understood as a map on the co-tangent bundle of  $E$ , so it is rather  $\mathcal{H}(x, df(x))$  than  $\mathcal{H}(x, \nabla f(x))$  in the sense of differential geometry.

**Theorem 7.14.** *Suppose that we are in the setting of Assumption 7.2 and that Condition 7.9 is satisfied. Denote  $X_n = \eta_n(Y_n)$ . Define the operator semigroup  $V_n(t)$  on  $C_b(E_n)$ :*

$$V_n(t)f(y) := \frac{1}{r_n} \log \mathbb{E} \left[ e^{r_n f(Y_n(t))} \mid Y_n(0) = y \right].$$

Suppose furthermore that

1. For each  $a_1 > 0$  there is a  $q \in \mathcal{Q}$  such that

$$\limsup_{n \rightarrow \infty} \frac{1}{r_n} \log \mathbb{P} [Y_n(0) \notin K_n^q] \leq -a_1.$$

In addition, the large deviation principle holds for  $X_n(0) = \eta_n(Y_n(0))$  with speed  $r_n$  and good rate function  $I_0$ .

2. The processes  $X_n = \eta_n(Y_n)$  are exponentially tight on  $D_E(\mathbb{R}^+)$ .
3. There is a semigroup  $V(t)$  on  $C_b(E)$  such that if  $\text{LIM } f_n = f$  and  $t_n \rightarrow t$  we have  $\text{LIM } V_n(t_n)f_n = V(t)f$ .

Then the processes  $X_n = \eta_n(Y_n)$  satisfy a large deviation principle on  $D_E(\mathbb{R}^+)$  with speed  $r_n$  and rate function

$$I(\gamma) = I_0(\gamma(0)) + \sup_{k \geq 1} \sup_{0=t_0 < t_1 < \dots < t_k} \sum_{i=1}^k I_{t_i - t_{i-1}}(\gamma(t_i) \mid \gamma(t_{i-1})). \quad (7.15)$$

Here  $\Delta_\gamma^c$  is the set of continuity points of  $\gamma$ . The conditional rate functions  $I_t$  are given by

$$I_t(y \mid x) = \sup_{f \in C_b(E)} \{f(y) - V(t)f(x)\}.$$

We will not prove this result, but refer to [9, pages 93 and 94] as it follows from essentially the projective limit theorem and Brycs result. The new contribution of this paper is a new method to obtain the convergence of semigroups based on the explicit identification of the resolvent corresponding to the semigroups  $V_n(t)$ .

*Proof of Theorem 7.10.* The large deviation statement follows from Theorem 7.14 if we can establish the convergence of semigroups, and obtain a limiting semigroup that is defined on all of  $C_b(E)$ .

We thus turn to the proof of the first three statements of the theorem. To obtain these results, we apply Theorems 5.7 and 6.1 in [14]. The semigroups  $V_n$  are of the type as in Remark 3.9, whose resolvents and generators we have identified in Theorem 3.6 and Proposition 3.8.

The conditions on convergence of Hamiltonians for [14, Theorems 5.7 and 6.1] have been assumed in Theorem 7.10 and we can work with  $B_n = C_b(E_n)$  due to Proposition 3.7.

The following two ingredients for the application of [14, Theorems 5.7 and 6.1] are missing

- joint local equi-continuity of the semigroups  $\{V_n(t)\}_{n \geq 1}$ ,
- joint local equi-continuity of the resolvents  $\{R_n(\lambda)\}_{n \geq 1}$ ,

We check these properties in Lemmas 7.15 and 7.16 below.

As a consequence [14, Theorems 5.7 and 6.1] can be applied, and we obtain convergence of the resolvents  $R_n(\lambda)$  to  $R(\lambda)$ . We also obtain that  $R(\lambda)h$  yields viscosity sub and supersolutions.

In addition, we obtain the convergence of  $V_n(t)$  to a semigroup  $V(t)$  which is defined on the quasi-closure of the set

$$\bigcup_{\lambda>0} \{R(\lambda)h \mid h \in C_b(E)\}.$$

Thus, if for all  $h \in C_b(E)$  we have  $\lim_{\lambda \rightarrow 0} R(\lambda)h = h$  for the strict topology, then indeed the semigroup  $V(t)$  is defined on all of  $C_b(E)$ . We prove this in Lemma 7.17 below.

This establishes the final result.  $\square$

The estimates below will be similar in spirit to estimates carried out in Section 5. There we were able to use tightness of sets of measures that have bounded relative entropy (see Proposition A.4). Here, however, we need an argument that allows us to obtain tightness in the sense of estimates with the index set  $\mathcal{Q}$  from exponential compact containment condition and rescaled boundedness of relative entropies. A basic estimate of this type is included as Proposition B.1 and will serve as the key replacement of Proposition A.4.

**Lemma 7.15.** *Consider the setting of Theorem 7.10. The semigroups  $V_n(t)$  are locally strictly equi-continuous on bounded sets: for all  $q \in \mathcal{Q}$ ,  $\delta > 0$  and  $T > 0$ , there is a  $\hat{q} \in \mathcal{Q}$  such that for all  $n \geq 1$ ,  $h_1, h_2 \in C_b(E_n)$  and  $0 \leq t \leq T$  we have*

$$\sup_{y \in K_n^q} \{V_n(t)h_1(y) - V_n(t)h_2(y)\} \leq \delta \sup_{x \in E_n} \{h_1(x) - h_2(x)\} + \sup_{y \in K_n^{\hat{q}}} \{h_1(y) - h_2(y)\}.$$

*Proof.* Fix  $h_1, h_2 \in C_b(E_n)$ ,  $\delta > 0$ ,  $q \in \mathcal{Q}$ , and  $T > 0$ . By exponential compact containment, see Condition 7.9, there is  $\hat{q}$  such that

$$\limsup_{n \rightarrow \infty} \sup_{y \in K_n^q} \frac{1}{r_n} \log \mathbb{P} [\exists t \leq T : Y_n(t) \notin K_n^{\hat{q}} \mid Y_n(0) = y] \leq -a. \tag{7.16}$$

Fix  $n, y \in K_n^q$  and  $t \leq T$ . Consider the definition of  $V(t)h_2$ :

$$V_n(t)h_2(y) = \sup_{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+))} \left\{ \int h_2(Y_n(t)) \mathbb{Q}(dY_n) - \frac{1}{r_n} S(\mathbb{Q} \mid \mathbb{P}_n) \right\}.$$

As  $h_2$  is bounded, the optimizer  $\mathbb{Q}_n$  must satisfy  $\frac{1}{r_n} S(\mathbb{Q}_n \mid \mathbb{P}_n) \leq 2 \|h_2\|$ . Thus by Proposition B.1 and (7.16) applied to  $\mathbb{Q}_n$  restricted to the marginal at time  $t$ , we have that for each  $\delta = 2\varepsilon > 0$ , there is a  $\bar{q}$  such that

$$\sup_n \mathbb{Q}_n[Y_n(t) \notin K_n^{\bar{q}}] \leq \delta.$$

It follows that

$$\sup_{y \in K_n^q} V(t)h_1(y) - V(t)h_2(y) \leq \delta \sup_{x \in E_n} h_1(x) - h_2(x) + \sup_{y \in K_n^{\bar{q}}} h_1(y) - h_2(y). \quad \square$$

**Lemma 7.16.** *Consider the setting of Theorem 7.10. The resolvents  $R_n(\lambda)$  are locally strictly equi-continuous on bounded sets: for all  $q \in \mathcal{Q}$ ,  $\delta > 0$  and  $\lambda_0 > 0$ , there is a  $\hat{q} \in \mathcal{Q}$  such that for all  $n$  and  $h_1, h_2 \in C_b(E_n)$  and  $0 < \lambda \leq \lambda_0$  it holds*

$$\sup_{y \in K_n^q} \{R_n(\lambda)h_1(y) - R_n(\lambda)h_2(y)\} \leq \delta \sup_{x \in E_n} \{h_1(x) - h_2(x)\} + \sup_{y \in K_n^{\hat{q}}} \{h_1(y) - h_2(y)\}.$$

*Proof.* If we work for a single  $\lambda$  instead of uniformly over  $0 < \lambda \leq \lambda_0$ , we can proceed as in the proof above. We first cut-off the tail of the exponential random variable which introduces a small error. Then we use the exponential compact containment condition and Proposition B.1 to find an appropriate  $\hat{q}$  that can be used to finish the argument as in the proof of Lemma 7.15 above.

If we work with a uniform estimate over  $0 < \lambda \leq \lambda_0$ , the argument needs to be adapted as in Lemma 5.7. We carry out a similar adaptation in the proof of Lemma 7.17 below.  $\square$

**Lemma 7.17.** *Consider the setting of Theorem 7.10. For all  $h \in C_b(E)$ , we have  $\lim_{\lambda \rightarrow 0} R(\lambda)h = h$  for the strict topology.*

**Remark 7.18.** Below we give a proof based on an approximation argument and a variant of Lemma 5.8. A less direct proof can be given also by using that the domains  $\mathcal{D}(H_{\dagger})$  and  $\mathcal{D}(H_{\ddagger})$  are sufficiently rich. See Proposition 7.1 of [14]

*Proof.* Fix  $h \in C_b(E)$ . First of all,  $\sup_{\lambda} \|R(\lambda)h\| \leq \|h\|$ , so it suffices to establish uniform convergence on compact sets. Thus, fix a compact set  $K \subseteq E$ . We prove

$$\limsup_{\lambda \downarrow 0} \sup_{x \in K} |R(\lambda)h(x) - h(x)| = 0.$$

Fix  $q \in \mathcal{Q}$  such that  $K \subseteq K^q$  and set  $h_n = h \circ \eta_n$ . Then we have by construction that  $\text{LIM } h_n = h$  and by Theorem 6.1 of [14] we have  $\text{LIM } R_n(\lambda)h_n = R(\lambda)h$  for any  $\lambda > 0$ .

Pick  $x \in K$  and let  $x_n \in K_n^q$  such that  $\eta_n(x_n) \rightarrow x$ . We have

$$\begin{aligned} R(\lambda)h(x) - h(x) &= R(\lambda)h(x) - R_n(\lambda)h_n(x_n) \\ &\quad + R_n(\lambda)h_n(x_n) - h_n(x_n) \\ &\quad + h_n(x_n) - h(x). \end{aligned}$$

Thus the result follows if we can prove that for each  $\varepsilon > 0$  there is a  $\lambda$  such that

$$\sup_n |R_n(\lambda)h_n(x_n) - h_n(x_n)| \leq \varepsilon. \tag{7.17}$$

Denote by  $\mathbb{P}_y^n$  the law of  $Y_n$  on  $D_{E_n}(\mathbb{R}^+)$  when started in  $y \in E_n$ . We have

$$\begin{aligned} &R_n(\lambda)h_n(x_n) - h_n(x_n) \\ &= \sup_{\mathbb{Q} \in \mathcal{P}(D_{E_n}(\mathbb{R}^+))} \int_0^\infty h(\eta_n(Y_n(t))) - h(\eta_n(x_n)) \mathbb{Q}(dY_n) - \frac{1}{r_n} S_t(\mathbb{Q} | \mathbb{P}_{x_n}^n) \tau_\lambda(dt). \end{aligned} \tag{7.18}$$

As in Lemma 5.8, we argue via a lower and upper bound.

$$\begin{aligned} R_n(\lambda)h_n(x_n) - h_n(x_n) &\geq \int_0^\infty h(\eta_n(Y_n(t))) - h(\eta_n(x_n)) \mathbb{P}_{x_n}^n(dY_n) \tau_\lambda(dt) \\ &= \int_0^\infty h(X_n(t)) - h(X_n(0)) \mathbb{P}_{x_n}^n \circ \eta_n^{-1}(dX_n) \tau_\lambda(dt) \end{aligned}$$

As the processes  $X_n$  are exponentially tight by assumption, the upper bound follows as tightness gives control on the modulus of continuity of the trajectories  $X_n$ .

For the lower bound, we can find  $\mathbb{Q}_{n,\lambda} \in D_{E_n}(\mathbb{R}^+)$  such that

$$\frac{1}{r_n} \int_0^\infty S_t(\mathbb{Q}_{n,\lambda} | \mathbb{P}_{x_n}^n) \tau_\lambda(dt) \leq 2 \|h\| \tag{7.19}$$

that achieve the suprema in (7.18). As in the proof of Lemma 5.8, define  $T(\lambda) = -\lambda \log \left( \frac{\varepsilon}{4\|h\|} \right)$ . This leads to

$$R_n(\lambda)h_n(x_n) - h_n(x_n) \leq \frac{1}{2}\varepsilon + \int_0^{T(\lambda)} h(X_n(t)) - h(X_n(0))\mathbb{Q}_{n,\lambda} \circ \eta_n^{-1}(dX_n)\tau_\lambda(dt). \quad (7.20)$$

Equation (7.19) yields that the restrictions of  $\mathbb{Q}_{n,\lambda} \circ \eta_n^{-1}$  to  $\mathcal{F}_T$  satisfy

$$\sup_\lambda \sup_n \frac{1}{r_n} S_{T(\lambda)}(\mathbb{Q}_{n,\lambda} \circ \eta_n^{-1} | \mathbb{P}_{x_n}^n \circ \eta_n^{-1}) \leq 8\|h\|^2 \varepsilon^{-1}.$$

Denote as before  $\widehat{\mathbb{Q}}_{n,\lambda}$  the measure  $\mathbb{Q}_{n,\lambda}$  up to time  $T(\lambda)$ , concatenated with the Markovian kernel  $\mathbb{P}_y^n$ . Thus,

$$\sup_\lambda \sup_n \frac{1}{r_n} S_1(\widehat{\mathbb{Q}}_{n,\lambda} \circ \eta_n^{-1} | \mathbb{P}_{x_n}^n \circ \eta_n^{-1}) \leq 8\|h\|^2 \varepsilon^{-1}.$$

As the measures  $\mathbb{P}_{x_n}$  are exponentially tight, the measures  $\widehat{\mathbb{Q}}_{n,\lambda} \circ \eta_n^{-1}$  restricted to  $\mathcal{F}_1$  are tight due to Proposition B.1. Tightness implies we can control the modulus of continuity, which implies we can upper bound (7.20) uniformly in  $n$  by  $\varepsilon$  by choosing  $\lambda$  small.  $\square$

## A Properties of relative entropy

The following result by Donsker and Varadhan can be derived from Lemma's 4.5.8 and 6.2.13 of [6].

**Lemma A.1** (Donsker Varadhan). *Let  $\mathcal{X}$  be a Polish space. We have the following duality relations:*

$$\begin{aligned} \log \langle e^f, \mu \rangle &= \sup_{\nu \in \mathcal{P}(\mathcal{X})} \{ \langle f, \nu \rangle - S(\nu | \mu) \} & \forall f \in C_b(\mathcal{X}), \\ S(\nu | \mu) &= \sup_{f \in C_b(\mathcal{X})} \{ \langle f, \nu \rangle - \log \langle e^f, \mu \rangle \} & \forall \nu \in \mathcal{P}(\mathcal{X}). \end{aligned}$$

By the second property of previous lemma, we immediately obtain lower semi-continuity of  $S$ .

**Lemma A.2.** *The map  $(\nu, \mu) \rightarrow S(\nu | \mu)$  is lower semi-continuous.*

We next give an extension of Theorem D.13 in [6], given as Exercise 5.13 in [20].

**Proposition A.3.** *Let  $\mathcal{X}$  be a Polish space. Suppose  $\mathcal{F}$  is a sub- $\sigma$ -algebra of  $\mathcal{B}(\mathcal{X})$ . Let  $\mu, \nu$  be two Borel probability measures on  $\mathcal{X}$ . Let  $\mu_1, \nu_1$  denote the restrictions of  $\mu$  and  $\nu$  to  $\mathcal{F}$ . Suppose that the regular conditional probabilities  $\mu(\cdot | x), \nu(\cdot | x)$  of  $\mu$  and  $\nu$  with respect to  $\mathcal{F}$  exist. Then the map*

$$x \mapsto S(\nu(\cdot | x) | \mu(\cdot | x))$$

is measurable and we have

$$S(\nu | \mu) = S(\nu_1 | \mu_1) + \int S(\nu(\cdot | x) | \mu(\cdot | x)) \nu_1(dx).$$

The final result of this appendix is the equi-coercivity of relative entropy in the second component.

**Proposition A.4.** *Let  $\mathcal{X}$  be a Polish space. Let  $K \subseteq \mathcal{P}(\mathcal{X})$  be a weakly compact set. For all  $c \geq 0$  the set*

$$A := \bigcup_{\mu \in K} \{ \nu \in \mathcal{P}(\mathcal{X}) | S(\nu | \mu) \leq c \}$$

is compact for the weak topology on  $\mathcal{P}(\mathcal{X})$ .

*Proof.* First of all, note that by Lemma A.1 the set  $A$  is weakly closed as  $(\nu, \mu) \mapsto S(\mu | \nu)$  is lower semicontinuous. Thus, it suffices that  $A$  is contained in a weakly compact set.

We start by proving that the map  $p(f) := \sup_{\mu \in K} \log \langle e^f, \mu \rangle$  is continuous for the strict topology. First note that the map  $(f, \mu) \mapsto \log \langle e^f, \mu \rangle$  is jointly continuous for the strict and weak topology. As a consequence  $p$  is lower semi-continuous.

Let  $f_\alpha$  be a net that converges strictly to  $f$ . Let  $\mu_\alpha$  be such that  $p(f_\alpha) = \log \langle e^{f_\alpha}, \mu_\alpha \rangle$ . By compactness there is a subnet  $\mu_\beta$  that has a weak limit  $\mu_0 \in K$ . By joint continuity of  $(f, \mu) \mapsto \log \langle e^f, \mu \rangle$ , we find

$$\limsup_{\alpha \rightarrow \infty} p(f_\alpha) = \limsup_{\alpha \rightarrow \infty} \log \langle e^{f_\alpha}, \mu_\alpha \rangle = \log \langle e^f, \mu_0 \rangle \leq p(f)$$

establishing that  $p$  is upper semi-continuous and thus continuous.

Fix  $c \geq 0$ . Let  $U := \{f \in C_b(\mathcal{X}) \mid p(f) \vee p(-f) \leq 1\}$ . As  $p$  is strictly continuous, we find that  $U$  is a strict neighbourhood of 0.

Let  $\nu$  be such that there is  $\mu_\nu$  with  $S(\nu | \mu_\nu) \leq c$ . For  $f \in U$  we have by Young's inequality, or equivalently, Lemma A.1, that

$$|\langle \nu, f \rangle| \leq (\log \langle e^f, \mu_\nu \rangle \vee \log \langle e^{-f}, \mu_\nu \rangle) + S(\nu | \mu_\nu) \leq 1 + c.$$

It follows that

$$A \subseteq \{\nu \mid \forall f \in U : |\langle f, \nu \rangle| \leq 1 + c\}$$

so that  $A$  is contained in a weakly compact set by the Bourbaki-Alaoglu theorem, see e.g. [13, Theorem 20.9.(4)].

Compactness of  $A$  follows by the lower semi-continuity of  $(\nu, \mu) \mapsto S(\nu | \mu)$ , Lemma A.2.  $\square$

## B Tightness for exponentially tilted measures

The next technical result essentially states the following. Suppose that we have two sequences of measures  $\mathbb{P}_n$  and  $\mathbb{Q}_n$  such that

1. the measures  $\mathbb{P}_n$  are exponentially tight at speed  $r_n \rightarrow \infty$ ,
2. there is some  $M \geq 0$  such that  $\sup_n \frac{1}{r_n} S(\mathbb{Q}_n | \mathbb{P}_n) \leq M$ .

Then the sequence  $\mathbb{Q}_n$  is tight.

The proposition below states this result, but allows for the context of a changing sequence of spaces and collections of measures instead of sequences.

**Proposition B.1.** *Let  $\mathcal{X}_n$  be a collection of Polish spaces. For each  $n$  let  $A_n \subseteq \mathcal{P}(\mathcal{X}_n)$  and suppose that for each  $a$  there is a collection of compact sets  $K_n = K_n(a) \subseteq \mathcal{X}_n$  such that*

$$\limsup_n \frac{1}{r_n} \log \sup_{\mu \in A_n} \mu(K_n^c) \leq -a$$

*for some sequence of constants  $r_n \rightarrow \infty$ . Suppose in addition that  $B_n \subseteq \mathcal{P}(\mathcal{X}_n)$  are collections of measures such that there exists a constant  $M \geq 0$  such that*

$$\sup_n \sup_{\nu \in B_n} \inf_{\mu \in A_n} \frac{1}{r_n} S(\nu | \mu) \leq M.$$

*Then there is for each  $\varepsilon > 0$  a constant  $a = a(\varepsilon)$  such that for the compact sets  $K_n = K_n(a(\varepsilon))$  we have*

$$\limsup_n \sup_{\nu \in B_n} \nu(K_n^c) \leq 2\varepsilon.$$

*Proof.* Fix  $\varepsilon > 0$ , we will construct sets  $K_n$  such that  $\limsup_n \sup_{\nu \in B_n} \nu(K_n^c) \leq 2\varepsilon$  which suffices to establish the claim. Fix  $a$  such that  $a > M\varepsilon^{-1}$ . By assumption there are sets  $K_n \subseteq \mathcal{X}_n$  such that

$$\limsup_n \sup_{\mu \in A_n} \frac{1}{r_n} \log \mu(K_n^c) \leq -a - 1. \tag{B.1}$$

We introduce an auxiliary function in terms of which we can define relative entropy. Denote by  $G : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  the function

$$G(t) = \begin{cases} 0 & \text{if } t = 0, \\ t \log t & \text{if } t > 0. \end{cases}$$

Thus, we have by assumption that

$$\sup_n \sup_{\nu \in B_n} \inf_{\mu \in A_n} \frac{1}{r_n} \int G\left(\frac{d\nu}{d\mu}\right) d\mu \leq M. \tag{B.2}$$

Note that  $G(t)t^{-1} = \log(t)$  and that for  $t \geq e^{r_n M \varepsilon^{-1}}$  we have that

$$t \geq c_n := e^{r_n \frac{M}{\varepsilon}} \implies \frac{G(t)}{t} \geq \frac{G(c_n)}{c_n} = \log(c_n) = r_n \frac{M}{\varepsilon}.$$

Thus, for each  $\nu \in B_n$  there is a  $\mu \in A_n$  such that we obtain by (B.2) that

$$\int_{\{\frac{d\nu}{d\mu} \geq c_n\}} \frac{d\nu}{d\mu} d\mu \leq \frac{\varepsilon}{r_n M} \int_{\{\frac{d\nu}{d\mu} \geq c_n\}} G\left(\frac{d\nu}{d\mu}\right) d\nu \leq \varepsilon. \tag{B.3}$$

We use this inequality to bound  $\nu_n[K_n^c]$ . For sufficiently large  $n$ , we find by (B.1) and (B.3) that

$$\begin{aligned} \nu(K_n^c) &= \int_{K_n^c \cap \{\frac{d\nu}{d\mu} \geq c_n\}} \frac{d\nu}{d\mu} d\mu + \int_{K_n^c \cap \{\frac{d\nu}{d\mu} < c_n\}} \frac{d\nu}{d\mu} d\mu \\ &\leq \varepsilon + \mu(K_n^c) c_n \\ &\leq \varepsilon + e^{-r_n a} c_n. \end{aligned}$$

As  $a > M\varepsilon^{-1}$ , the right-hand exponential term on the final line converges to 0 independent from the choice of  $\varepsilon$ . This establishes the claim.  $\square$

## References

- [1] Charalambos D. Aliprantis and Kim C. Border, *Infinite dimensional analysis*, third ed., Springer, Berlin, 2006. A hitchhiker’s guide. MR-2378491
- [2] Martino Bardi and Italo Capuzzo-Dolcetta, *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations.*, Boston, MA: Birkhäuser, 1997 (English). MR-1484411
- [3] Vladimir I. Bogachev, *Measure theory*, Springer-Verlag, 2007. MR-2267655
- [4] A. de Acosta, *A general non-convex large deviation result with applications to stochastic equations*, Probability Theory and Related Fields **118** (2000), no. 4, 483–521. MR-1808373
- [5] A. de Acosta, *A general nonconvex large deviation result II*, Ann. Probab. **32** (2004), no. 3, 1873–1901. MR-2073180
- [6] Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, second ed., Springer-Verlag, 1998. MR-1619036
- [7] Paul Dupuis, Hitoshi Ishii, and H. Mete Soner, *A viscosity solution approach to the asymptotic analysis of queueing systems*, The Annals of Probability **18** (1990), no. 1, pp. 226–255. MR-1043946

- [8] Stewart N. Ethier and Thomas G. Kurtz, *Markov processes: Characterization and convergence*, Wiley, 1986. MR-0838085
- [9] Jin Feng and Thomas G. Kurtz, *Large deviations for stochastic processes*, American Mathematical Society, 2006. MR-2260560
- [10] Wendell H. Fleming, *Exit probabilities and optimal stochastic control*, Applied Mathematics and Optimization **4** (1977), 329–346. MR-0512217
- [11] Wendell H. Fleming and Panagiotis E. Souganidis, *Pde-viscosity solution approach to some problems of large deviations*, Annali della Scuola Normale Superiore di Pisa – Classe di Scienze **Ser. 4, 13** (1986), no. 2, 171–192. MR-0876121
- [12] Tosio Kato, *Remarks on pseudo-resolvents and infinitesimal generators of semi-groups*, Proc. Japan Acad. **35** (1959), 467–468. MR-0117570
- [13] Gottfried Köthe, *Topological vector spaces I*, Springer-Verlag, 1969. MR-0248498
- [14] Richard C. Kraaij, *A general convergence result for viscosity solutions of Hamilton-Jacobi equations and non-linear semigroups*, To appear in Journal of Functional Analysis (2019).
- [15] Richard C. Kraaij, *Strict continuity of the transition semigroup for the solution of a well-posed martingale problem*, preprint; arXiv:1909.05113 (2019).
- [16] Thomas G. Kurtz, *Convergence of sequences of semigroups of nonlinear operators with an application to gas kinetics.*, Trans. Am. Math. Soc. **186** (1974), 259–272. MR-0336482
- [17] Mauro Mariani, *A  $\Gamma$ -convergence approach to large deviations*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **18** (2018), no. 3, 951–976. MR-3807592
- [18] A. Puhalskii, *Large deviations of semimartingales: a maxingale problem approach. I. Limits as solutions to a maxingale problem*, Stochastics Stochastics Rep. **61** (1997), no. 3-4, 141–243. MR-1488137
- [19] Anatolii Puhalskii, *Large deviations and idempotent probability*, Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics, vol. 119, Chapman & Hall/CRC, Boca Raton, FL, 2001. MR-1851048
- [20] Firas Rassoul-Agha and Timo Seppäläinen, *A course on large deviations with an introduction to Gibbs measures*, Graduate Studies in Mathematics, vol. 162, American Mathematical Society, Providence, RI, 2015. MR-3309619
- [21] F. Dennis Sentes, *Bounded continuous functions on a completely regular space*, Trans. Amer. Math. Soc. **168** (1972), 311–336. MR-0295065
- [22] S. Sheu, *Stochastic control and exit probabilities of jump processes*, SIAM Journal on Control and Optimization **23** (1985), no. 2, 306–328. MR-0777462
- [23] H. F. Trotter, *Approximation of semi-groups of operators*, Pacific J. Math. **8** (1958), 887–919. MR-0103420
- [24] Jan A. van Casteren, *Markov processes, Feller semigroups and evolution equations*, Series on Concrete and Applicable Mathematics, vol. 12, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2011. MR-2779929
- [25] S. L. Zabell, *Mosco convergence in locally convex spaces.*, J. Funct. Anal. **110** (1992), no. 1, 226–246 (English). MR-1190424

**Acknowledgments.** The author wants to thank two anonymous referees for their careful reading of the manuscript and their comments that improved the text.

---

# Electronic Journal of Probability

## Electronic Communications in Probability

---

### Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS<sup>1</sup>)
- Easy interface (EJMS<sup>2</sup>)

### Economical model of EJP-ECP

- Non profit, sponsored by IMS<sup>3</sup>, BS<sup>4</sup>, ProjectEuclid<sup>5</sup>
- Purely electronic

### Help keep the journal free and vigorous

- Donate to the IMS open access fund<sup>6</sup> (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

---

<sup>1</sup>LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

<sup>2</sup>EJMS: Electronic Journal Management System <http://www.vtex.lt/en/ejms.html>

<sup>3</sup>IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

<sup>4</sup>BS: Bernoulli Society <http://www.bernoulli-society.org/>

<sup>5</sup>Project Euclid: <https://projecteuclid.org/>

<sup>6</sup>IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>