Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Large Deviations for Stochastic Processes

by

Stefan Adams

Lecture note no.: 42

2013

LARGE DEVIATIONS FOR STOCHASTIC PROCESSES

By Stefan Adams[1]

Abstract: The notes are devoted to results on large deviations for sequences of Markov processes following closely the book by Feng and Kurtz ([FK06]). We outline how convergence of Fleming's nonlinear semigroups (logarithmically transformed nonlinear semigroups) implies large deviation principles analogous to the use of convergence of linear semigroups in weak convergence. The latter method is based on the range condition for the corresponding generator. Viscosity solution methods however provide applicable conditions for the necessary nonlinear semigroup convergence. Once having established the validity of the large deviation principle one is concerned to obtain more tractable representations for the corresponding rate function. This in turn can be achieved once a variational representation of the limiting generator of Fleming's semigroup can be established. The obtained variational representation of the generator allows for a suitable control representation of the rate function. The notes conclude with a couple of examples to show how the methodology via Fleming's semigroups works. These notes are based on the mini-course '*Large deviations for stochastic processes*' the author held during the workshop '*Dynamical Gibs-non-Gibbs transitions*' at EURANDOM in Eindhoven, December 2011, and at the Max-Planck institute for mathematics in the sciences in Leipzig, July 2012.

Keywords and phrases. large deviation principle, Fleming's semigroup, viscosity solution, range condition

Date: 28th November 2012.

[1]Mathematics Institute, University of Warwick, Coventry CV4 7AL, United Kingdom, S.Adams@warwick.ac.uk

The aim of these lectures is to give a very brief and concise introduction to the theory of large deviations for stochastic processes. We are following closely the book by Feng & Kurtz [FK06], and it is our subsequent aim in these notes to provide an instructive overview such that the reader can easily use the book [FK06] for particular examples or more detailed questions afterwards. In Section 1 we will first recall basic definitions and facts about large deviation principles and techniques. At the end of that section we provide a roadmap as a guidance through the notes and the book [FK06]. This is followed by an introduction to the so-called nonlinear semigroup method originally going back to Fleming [Flem85] using the so-called range condition in Section 2. In Section 3 this range condition is replaced by the weaker notion of a comparison principle. In the last Section 4 we provide control representations of the rate functions and conclude with a couple of examples.

1. Introduction

1.1. **Large deviation Principle.** We recall basic definitions of the theory of large deviations. For that we let (E,d) denote a complete, separable metric space. The theory of large deviations is concerned with the asymptotic estimation of probabilities of rare events. In its basic form, the theory considers the limit of normalisations of $\log P(A_n)$ for a sequence $(A_n)_{n\ge 1}$ of events with asymptotically vanishing probability. To be precise, for a sequence of random variables $(X_n)_{n\ge1}$ taking values in E the *large deviation principle* is formulated as follows.

Definition 1.1 (**Large Deviation Principle**). *The sequence $(X_n)_{n\in\mathbb{N}}$ satisfies a large deviation principle (LDP) if there exists a lower semicontinuous function $I\colon E\to[0,\infty]$ such that*

$$\begin{aligned} \forall\ A \text{ open} \qquad & \liminf_{n\to\infty}\frac{1}{n}\log P(X_n\in A)\ge -\inf_{x\in A} I(x),\\ \forall\ B \text{ closed} \qquad & \limsup_{n\to\infty}\frac{1}{n}\log P(X_n\in B)\le -\inf_{x\in B} I(x). \end{aligned} \tag{1.1}$$

The function I is called the rate function for the large deviation principle. A rate function I is called a good rate function if for each $a\in[0,\infty)$ the level set $\{x\colon I(x)\le a\}$ is compact.

The traditional approach to large deviation principles is via the so-called change of measure method. Indeed, beginning with the work of Cramér [Cra38] and including the fundamental work on large deviations for stochastic processes by Freidlin and Wentzell [FW98] and Donsker and Varadhan [DV76], much of the analysis has been based on change of measure techniques. In this approach, a tilted or reference measure is identified under which the events of interest have high probability, and the probability of the event under the original measure is estimated in terms of the Radon-Nikodym density relating the two measures.

Another approach to large deviation principles is analogous to the Prohorov compactness approach to weak convergence of probability measures. This has been been recently established by Puhalskii [Puh94], O'Brien and Vervaat [OV95], de Acosta [DeA97] and others. This approach is the starting point of our approach to large deviations for stochastic processes. Hence, we shall recall the basic definitions and facts about weak convergence of probability measures.

Definition 1.2. *A sequence $(X_n)_{n\geq 1}$ of E-valued random variables converges in distribution to the random variable X (that is, the distributions $P(X_n \in \cdot)$ converge weakly to $P(X \in \cdot)$) if and only if $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ for each $f \in \mathcal{C}_{\rm b}(E)$, where $\mathcal{C}_{\rm b}(E)$ is the space of all continuous bounded functions on E equipped with the supremum norm.*

The analogy between large deviations and weak convergence becomes much clearer if we recall the following equivalent formulation of convergence in distribution (weak convergence).

Proposition 1.3. *A sequence $(X_n)_{n\geq 1}$ of E-valued random variables converges in distribution to the random variable X if and only if for all open $A \subset E$,*

$$\liminf_{n\to\infty} P(X_n \in A) \geq P(X \in A), \tag{1.2}$$

or equivalently, if for all closed $B \subset E$,

$$\limsup_{n\to\infty} P(X_n \in B) \leq P(X \in B). \tag{1.3}$$

Our main aim in these notes is the development of the weak convergence approach to large deviation theory applied to stochastic processes. In the next subsection we will outline Fleming's idea [Flem85]. Before we settle some notations to be used throughout the lecture.

Notations:

We are using the following notations. Throughout, (E,d) will be a complete, separable metric space, $M(E)$ will denote the space of real-valued Borel measurable functions on E, $B(E) \subset M(E)$, the space of bounded, Borel measurable functions, $\mathcal{C}(E)$, the space of continuous functions, $\mathcal{C}_{\rm b}(E) \subset B(E)$, the space of bounded continuous functions, and $\mathcal{M}_1(E)$, the space of probability measures on E. By $\mathcal{B}(E)$ we denote the Borel σ-algebra on E. We identity an operator A with its graph and write $A \subset \mathcal{C}_{\rm b}(E) \times \mathcal{C}_{\rm b}(E)$ if the domain $\mathcal{D}(A)$ and range $\mathcal{R}(A)$ are contained in $\mathcal{C}_{\rm b}(E)$. The space of E-valued, cadlag functions on $[0,\infty)$ with the Skorohod topology will be denoted by $\mathcal{D}_E[0,\infty)$. We briefly recall the Skorohod topology and its corresponding metric. For that define $q(x,y) = 1 \wedge d(x,y)$, $x, y \in E$, and note that q is a metric on E that is equivalent to the metric d. We define a metric on $\mathcal{D}_E[0,\infty)$ which gives the Skorohod topology. Let Θ be the collection of strictly increasing functions mapping $[0,\infty)$ onto $[0,\infty)$ having the property that

$$\gamma(\theta) := \sup_{0\leq s<t} \Big\{ \Big| \log \frac{\theta(t)-\theta(s)}{t-s} \Big| \Big\} < \infty, \quad \theta \in \Theta.$$

For $x, y \in \mathcal{D}_E[0,\infty)$, define

$$d(x,y) = \inf_{\theta\in\Theta} \Big\{ \gamma(\theta) \vee \int_0^\infty {\rm e}^{-u} \sup_{t\geq 0} \{ q(x(\theta(t)\wedge u), y(t\wedge u))\, {\rm d}u\} \Big\}. \tag{1.4}$$

We denote by Δ_x the set of discontinuities of $x \in \mathcal{D}_E[0,\infty)$.

An E-valued Markov process $X = (X(t))_{t\geq 0}$ is given by its linear generator $A \subset B(E)\times B(E)$ and linear contraction semigroup $(T(t))_{t\geq 0}$,

$$T(t)f(x) = \mathbb{E}[f(X(t))|X(0) = x] = \int_E f(y)P(t,x,{\rm d}y).$$

Here, the generator is formally given as

$$\frac{\mathrm{d}}{\mathrm{d}t}T(t)f = AT(t)f, T(0)f = f;$$

or, for any $f \in \mathcal{D}(A)$, $Af = \lim_{t\downarrow 0} \frac{1}{t}(T(t)f - f)$, where this limit exists for each $f \in \mathcal{D}(A)$.

1.2. **Fleming's approach - main idea for LDP for stochastic processes.** We outline the ideas by Fleming [Flem85] and show how these can be further developed implying pathwise large deviation principles for stochastic processes. Results of Varadhan and Bryc (see Theorem 1.8) relate large deviations for sequences of random variables to the asymptotic behaviour of functionals (logarithmic moment generating functionals) of the form $1/n \log \mathbb{E}[\mathrm{e}^{nf(X_n)}]$. If we now let $(X_n)_{n\ge 1}$ be a sequence of Markov processes taking values in E and having generator A_n, one defines the linear Markov semigroup $(T_n(t))_{t\ge 0}$,

$$T_n(t)f(x) = \mathbb{E}[f(X_n(t))|X_n(0) = x], \quad t \ge 0, x \in E, \tag{1.5}$$

which, at least formally, satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}T_n(t)f = A_nT_n(t)f; \quad T_n(0)f = f. \tag{1.6}$$

Fleming (cf. [Flem85]) introduced the following nonlinear contraction (in the supremum norm) semigroup $(V_n(t))_{t\ge 0}$ for Markov processes X_n,

$$V_n(t)f(x) = \frac{1}{n}\log \mathbb{E}[\mathrm{e}^{nf(X_n(t))}|X_n(0) = x], \quad t \ge 0, x \in E, \tag{1.7}$$

and large deviations for sequences $(X_n)_{n\ge 1}$ of Markov processes can be studied using the asymptotic behaviour of the corresponding nonlinear semigroups. Again, at least formally, V_n should satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}V_n(t)f = \frac{1}{n}\mathcal{H}_n(nV_n(t)f), \tag{1.8}$$

where we define

$$H_nf := \frac{1}{n}\mathcal{H}_n(nf) = \frac{1}{n}\mathrm{e}^{-nf}A_n\mathrm{e}^{nf}. \tag{1.9}$$

Fleming and others have used this approach to prove large deviation results for sequences $(X_n)_{n\ge 1}$ of Markov processes X_n at single time points and exit times (henceforth for random sequences in E). The book by [FK06] extends this approach further, showing how convergence of the nonlinear semigroups and their generators H_n can be used to obtain both, exponential tightness and the large deviation principle for the finite dimensional distributions of the processes. Showing the large deviation principle for finite dimensional distributions of the sequence of Markov processes and using the exponential tightness gives then the full pathwise large deviation principle for the sequence of Markov processes. Before we embark on details of the single steps in this programme we will discuss basic facts of large deviation theory in the next subsection. In section 1.4 we draw up a roadmap for proving large deviations for sequences of Markov processes using Fleming's methodology. This roadmap will be the backbone of these notes and will provide guidance in studying the book [FK06] or different research papers on that subject.

1.3. **Basic facts of Large Deviation Theory.** We discuss basic properties of the large deviation principle for random variables taking values in some metric space (E, d). We start with some remarks on the definition of the large deviation principle before proving Varadhan's lemma and Bryc's formula. These two statements are the key facts in developing the weak convergence approach to large deviation principles.

Remark 1.4 (**Large deviation principle**). (a) *The large deviation principle (LDP) with rate function I and rate n implies that*

$$-I(x) = \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log P(X_n \in B_\varepsilon(x)) = \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P(X_n \in \overline{B}_\varepsilon(x)), \tag{1.10}$$

and it follows that the rate function I is uniquely determined.

(b) *If* (1.10) *holds for all $x \in E$, then the lower bound in* (1.1) *holds for all open sets $A \subset E$ and the upper bound in* (1.1) *holds for all compact sets $B \subset E$. This statement is called the weak large deviation principle,i.e., the statement where in* (1.1) *closed is replaced by compact.*

(c) *The large deviation principle is equivalent to the assertion that for each measurable $A \in \mathcal{B}(E)$,*

$$-\inf_{x\in A^\circ} \le \liminf_{n\to\infty}\frac{1}{n}\log P(X_n \in A) \le \limsup_{n\to\infty}\frac{1}{n}\log P(X_n \in A) \le -\inf_{x\in\overline{A}} I(x),$$

where A° is the interior of A and $\overline{A}$ the closure of A.

(d) *The lower semi-continuity of I is equivalent to the level set $\{x \in E \colon I(x) \le c\}$ being closed for each $c \in \mathbb{R}$. If all the level sets of I are compact, we say that the function I is good.*

In the theory of weak convergence, a sequence $(\mu_n)_{n\ge 1}$ of probability measures is *tight* if for each $\varepsilon > 0$ there exists a compact set $K \subset E$ such that $\inf_{n\in\mathbb{N}} \mu_n(K) \ge 1 - \varepsilon$. The analogous concept in large deviation theory is *exponential tightness*.

Definition 1.5. *A sequence of probability measures $(P_n)_{n\ge 1}$ on E is said to be exponentially tight if for each $\alpha > 0$, there exists a compact set $K_\alpha \subset E$ such that*

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n(K_\alpha^{\mathrm{c}}) < -\alpha.$$

A sequence $(X_n)_{n\ge 1}$ of E-valued random variables is exponentially tight if the corresponding sequence of distributions is exponentially tight.

Remark 1.6. *If $(P_n)_{n\ge 1}$ satisfies the LDP with good rate function I, then $(P_n)_{n\ge 1}$ is exponentially tight. This can be seen easily as follows. Pick $\alpha > 0$ and define $K_\alpha := \{x \in E \colon I(x) \le \alpha\}$. The set K_α is compact and we get from the large deviation upper bound the estimate*

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n(K_\alpha^{\mathrm{c}}) \le -\inf_{x\in K_\alpha^{\mathrm{c}}}\{I(x)\} \le -\alpha.$$

Exponential tightness plays the same role in large deviation theory as tightness does in weak convergence theory. The following analogue of the Prohorov compactness theorem is from [Puh94] and stated here without proof.

Theorem 1.7. *Let $(P_n)_{n\geq 1}$ be a sequence of tight probability measures on the Borel σ-algebra $\mathcal{B}(E)$ of E. Suppose in addition that $(P_n)_{n\geq 1}$ is exponentially tight. Then there exists a subsequence $(n_k)_{k\geq 1}$ along which the large deviation principle holds with a good rate function.*

The following moment characterisation of the large deviation principle due to Varadhan and Bryc is central to our study of large deviation principles for sequences of Markov processes.

Theorem 1.8. *Let $(X_n)_{n\geq 1}$ be a sequence of E-valued random variables.*

(a) *(Varadhan Lemma) Suppose that $(X_n)_{n\geq 1}$ satisfies the large deviation principle with a good rate function I. Then for each $f\in\mathcal{C}_{\rm b}(E)$,*

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[{\rm e}^{nf(X_n)}]=\sup_{x\in E}\{f(x)-I(x)\}. \tag{1.11}$$

(b) *(Bryc formula) Suppose that the sequence $(X_n)_{n\geq 1}$ is exponentially tight and that the limit*

$$\Lambda(f)=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[{\rm e}^{nf(X_n)}] \tag{1.12}$$

exists for each $f\in\mathcal{C}_{\rm b}(E)$. Then $(X_n)_{n\geq 1}$ satisfies the large deviation principle with good rate function

$$I(x)=\sup_{f\in\mathcal{C}_{\rm b}(E)}\{f(x)-\Lambda(f)\},\quad x\in E. \tag{1.13}$$

Remark 1.9. (a) *Part (a) is the natural extension of Laplace's method to infinite dimensional spaces. The limit* (1.11) *can be extended to the case of $f\colon E\to\mathbb{R}$ continuous and bounded from above. Alternatively, the limit* (1.11) *holds also under a tail or moment condition, i.e., one may assume either the tail condition*

$$\lim_{M\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}[{\rm e}^{nf(X_n)}1\!\!1\{f(X_n)\geq M\}]=-\infty,$$

or the following moment condition for some $\gamma>1$,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}[{\rm e}^{\gamma nf(X_n)}]<\infty.$$

For details we refer to Section 4.3 in [DZ98].

(b) *Suppose that the sequence $(X_n)_{n\geq 1}$ is exponentially tight and that the limit*

$$\Lambda(f)=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[{\rm e}^{nf(X_n)}]$$

exists for each $f\in\mathcal{C}_{\rm b}(E)$. Then

$$I(x)=-\inf_{f\in\mathcal{C}_{\rm b}(E)\colon f(x)=0}\Lambda(f).$$

This can be seen easily from (1.13) *and the fact that $\Lambda(f+c)=\Lambda(f)+c$ for any constant $c\in\mathbb{R}$.*

Proof. Detailed proofs are in [DZ98]. However, as the results of Theorem 1.8 are pivotal for our approach to large deviation principles on which the whole methodology in [FK06] is relying on, we decided to give an independent version of the proof. Pick $f \in \mathcal{C}_{\rm b}(E)$, $x \in E$, and for $\varepsilon > 0$, let $\delta_\varepsilon > 0$ satisfy

$$\overline{B}_\varepsilon(x) \subset \{y \in E \colon f(y) > f(x) - \delta_\varepsilon\},$$

where $\overline{B}_\varepsilon(x)$ is the closed ball of radius ε around $x \in E$. By the continuity of f, we can assume that $\lim_{\varepsilon\to 0}\delta_\varepsilon = 0$. Then the exponential version of Chebycheff's inequality gives

$$P(X_n \in \overline{B}_\varepsilon(x)) \le {\rm e}^{n(\delta_\varepsilon - f(x))}\mathbb{E}[{\rm e}^{nf(X_n)}],$$

and henceforth

$$\lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log P(X_n \in \overline{B}_\varepsilon(x)) \le -f(x) + \liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}[{\rm e}^{nf(X_n)}] \tag{1.14}$$

and

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P(X_n \in \overline{B}_\varepsilon(x)) \le -f(x) + \liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}[{\rm e}^{nf(X_n)}]. \tag{1.15}$$

The space (E,d) is regular as a metric space, hence for $x \in E$ and $\varepsilon > 0$ there is $f_{\varepsilon,x} \in \mathcal{C}_{\rm b}(E)$ and $r > 0$ satisfying $f_{\varepsilon,x}(y) \le f_{\varepsilon,x}(x) - r1\!\!1_{B^c_\varepsilon(x)}(y)$ for all $y \in E$. Thus we get

$$\mathbb{E}[{\rm e}^{nf(X_n)}] \le {\rm e}^{nf_{\varepsilon,x}(x)}{\rm e}^{-nr} + {\rm e}^{nf_{\varepsilon,x}(x)}P(X_n \in B_\varepsilon(x)),$$

and therefore

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log P(X_n \in B_\varepsilon(x)) &\ge -f_{\varepsilon,x}(x) + \limsup_{n\to\infty}\frac{1}{n}\mathbb{E}[{\rm e}^{nf_{\varepsilon,x}(X_n)}],\\ \liminf_{n\to\infty}\frac{1}{n}\log P(X_n \in B_\varepsilon(x)) &\ge -f_{\varepsilon,x}(x) + \liminf_{n\to\infty}\frac{1}{n}\mathbb{E}[{\rm e}^{nf_{\varepsilon,x}(X_n)}].\end{aligned} \tag{1.16}$$

When (1.12) holds for each $f \in \mathcal{C}_{\rm b}(E)$ we get with (1.15), (1.16) and using the fact that $f_{\varepsilon,x} \to f$ as $\varepsilon \to 0$,

$$\begin{aligned}-\sup_{f\in\mathcal{C}_{\rm b}(E)}\{f(x) - \Lambda(f)\} &\le \lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log P(X_n \in B_\varepsilon(x)) \le \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P(X_n \in B_\varepsilon(x))\\ &\le -\sup_{f\in\mathcal{C}_{\rm b}(E)}\{f(x) - \Lambda(f)\},\end{aligned}$$

and therefore (1.10) and thus a weak large deviation principle with rate function $I(x) = \sup_{f\in\mathcal{C}_{\rm b}(E)}\{f(x) - \Lambda(f)\}$, and using the exponential tightness we get the statement in (b). If (1.10) holds, then by (1.14) we have

$$\sup_{x\in E}\{f(x) - I(x)\} \le \liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}[{\rm e}^{nf(X_n)}].$$

For the corresponding upper bound note that as I is a good rate function the sequence $(X_n)_{n\ge 1}$ is exponentially tight, and henceforth there is a compact set $K \subset E$ such that

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}[{\rm e}^{nf(X_n)}] = \limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}[{\rm e}^{nf(X_n)}1\!\!1_K(X_n)].$$

For each $\varepsilon > 0$, there exists a finite cover, i.e., a finite sequence $x_1, \dots, x_m \in K$ such that $K \subset \bigcup_{i=1}^m B_\varepsilon(x_i)$ such that $\max_{y\in B_\varepsilon(x_i)}\{f(y)\} \le \varepsilon + f(x_i), i = 1, \dots, m$. Consequently,

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}[\mathrm{e}^{nf(X_n)}1\!\!1_K(X_n)] &\le \limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\big[\sum_{i=1}^m \mathrm{e}^{n(\varepsilon+f(x_i))}1\!\!1_{B_\varepsilon(x_i)}(X_n)\big]\\ &\le \max_{1\le i\le m}\big\{f(x_i)+\varepsilon+\limsup_{n\to\infty}\frac{1}{n}\log P(X_n\in B_\varepsilon(x_i))\big\},\end{aligned}$$

and hence for each $\varepsilon > 0$, there exists $x_\varepsilon \in K$ such that

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}[\mathrm{e}^{nf(X_n)}1\!\!1_K(X_n)] \le f(x_\varepsilon)+\varepsilon+\limsup_{n\to\infty}\frac{1}{n}\log P(X_n\in B_\varepsilon(x_\varepsilon)).$$

As K is compact, we can pick a subsequence along which $x_\varepsilon \to x_0 \in K$ as $\varepsilon \to 0$. If we choose $\varepsilon > 0$ such that $d(x_\varepsilon, x_0) + \varepsilon < \delta$ for a given $\delta > 0$, we arrive at

$$f(x_\varepsilon)+\varepsilon+\limsup_{n\to\infty}\frac{1}{n}\log P(X_n\in B_\varepsilon(x_\varepsilon)) \le f(x_\varepsilon)+\varepsilon+\limsup_{n\to\infty}\frac{1}{n}\log P(X_n\in B_\delta(x_0)),$$

and therefore we get that

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}[\mathrm{e}^{nf(X_n)}1\!\!1_K(X_n)] \le f(x_0)-I(x_0),$$

finishing our proof of part (a). □

To prove a large deviation principle, it is, in principle, enough to verify (1.12) for a class of functions that is smaller than $\mathcal{C}_\mathrm{b}(E)$. Let $\Gamma = \{f \in \mathcal{C}_\mathrm{b}(E)\colon (1.12) \text{ holds for } f\}$.

Definition 1.10. (a) *A collection of functions $D \subset \mathcal{C}_\mathrm{b}(E)$ is called rate function determining if whenever $D \subset \Gamma$ for an exponentially tight sequence $(X_n)_{n\ge1}$, the sequence $(X_n)_{n\ge1}$ satisfies the LDP with good rate function $I(x) = \sup_{f\in D}\{f(x) - \Lambda(f)\}$.*

(b) *A sequence $(f_n)_{n\ge1}$ of functions on E converges boundedly and uniformly on compacts (buc) to f if and only if $\sup_n \|f_n\| < \infty$ and for each compact $K \subset E$,*

$$\limsup_{n\to\infty}\sup_{x\in K}|f_n(x)-f(x)| = 0.$$

(c) *A collection of functions $D \subset \mathcal{C}_\mathrm{b}(E)$ is bounded above if $\sup_{f\in D}\sup_{y\in E}\{f(y)\} < \infty$.*

(d) *A collection of functions $D \subset \mathcal{C}_\mathrm{b}(E)$ isolates points in E, if for each $x \in E$, each $\varepsilon > 0$, and each compact $K \subset E$, there exists a function $f \in D$ satisfying*

$$|f(x)|<\varepsilon, \quad \sup_{y\in K} f(y)\le 0, \qquad \sup_{y\in K\cap B_\varepsilon^c(x)} f(y) < -\frac{1}{\varepsilon}.$$

We will later work with the following set of rate function determining functions.

Proposition 1.11. *If $D \subset \mathcal{C}_\mathrm{b}(E)$ is bounded above and isolates points, then D is rate function determining.*

Proof. We give a brief sketch of this important characterisation of rate function determining set of functions. Suppose that $(X_n)_{n\ge1}$ is exponentially tight. It suffices to show that for any sequence along which the large deviation principle holds, the corresponding rate function is given by $I(x) = \sup_{f\in D}\{f(x) - \Lambda(f)\}$. W.l.o.g. we assume that $\Gamma = \mathcal{C}_\mathrm{b}(E)$. Clearly, $\sup_{x\in D}\{f(x) - \Lambda(f)\} \le I(x)$, where $I(x)$ is given in (1.13). We shall show the reverse bound.

Note that $\Lambda(f)$ is a monotone function of f. Let $c > 0$ be such that $\sup_{f\in D}\sup_{y\in E} f(y) \le c$, and fix $f_0 \in \mathcal{C}_\mathrm{b}(E)$ satisfying $f_0(x) = 0$, and pick $\delta > 0$. Let $\alpha > c + |\Lambda(f_0)|$, and let K_α be a compact set (due to exponential tightness) satisfying

$$\limsup_{n\to\infty} \frac{1}{n} \log P(X_n \in K_\alpha^\mathrm{c}) \le -\alpha.$$

The continuity of f_0 gives $\varepsilon \in (0, \delta)$ such that $\sup_{y\in B_\varepsilon(x)}\{|f_0(y)|\} \le \delta$ and $\inf_{y\in K_\alpha}\{f_0(y)\} \ge -\varepsilon^{-1}$. By the property of isolating points, there exists a function $f \in D$ satisfying $f \le c\mathbb{1}_{K_\alpha^\mathrm{c}}$, $f\mathbb{1}_{K_\alpha} \le f_0\mathbb{1}_{K_\alpha} + \delta$, and $|f(x)| < \varepsilon$. It follows immediately that

$$\mathrm{e}^{nf} \le \mathrm{e}^{nc}\mathbb{1}_{K_\alpha^\mathrm{c}} + \mathrm{e}^{n(f_0+\delta)},$$

and hence that

$$\Lambda(f) \le \Lambda(f_0) + \delta.$$

Thus with $f(x) - \Lambda(f) \ge -\Lambda(f_0) - 2\delta$, we conclude with the reversed bound. □

The following conclusion of the preceding will be used frequently.

Corollary 1.12. *Fix $x \in E$, and suppose that the sequence $(f_m)_{m\ge1}$ of functions $f_m \in D \subset \mathcal{C}_\mathrm{b}(E)$ satisfies the following.*

(a) $\sup_{m\ge1}\sup_{y\in E}\{f_m(y)\} \le c < \infty$.
(b) *For each compact $K \subset E$ and each $\varepsilon > 0$,*

$$\lim_{m\to\infty} \sup_{y\in K}\{f_m(y)\} \le 0 \textit{ and } \lim_{m\to\infty} \sup_{y\in K\cap B_\varepsilon(x)^\mathrm{c}}\{f_m(y)\} = -\infty.$$

(c) $\lim_{m\to\infty} f_m(x) = 0$.

Then $I(x) = -\lim_{m\to\infty}\Lambda(f_m)$.

Proof. Pick again $x \in E$, and $f_0 \in \mathcal{C}_\mathrm{b}(E)$ with $f_0(x) = 0$. The assumptions on the sequence imply that for each compact $K \subset E$ and $\delta > 0$ there is an index $m_0 \in \mathbb{N}$ such that for $m > m_0$

$$\mathrm{e}^{nf_m} \le \mathrm{e}^{nc}\mathbb{1}_{K^\mathrm{c}} + \mathrm{e}^{n(f_0-\delta)},$$

and henceforth $\liminf_{m\to\infty}(f_m(x) - \Lambda(f_m)) \ge -\Lambda(f_0) - \delta$. Finally, assumption (c) implies that $I(x) = -\lim_{m\to\infty}\Lambda(f_m)$. □

We finish this section with some obvious generalisation to the situation when the state space E is given as a product space of complete separable metric spaces. This is in particular useful for stochastic processes as the finite dimensional distributions, i.e., the distributions of the vector $(X_n(t_1), \dots, X_n(t_k))$ for some $k \in \mathbb{N}$ and fixed times $t_i \le t_{i+1}, i = 1, \dots, k-1$, are probability measures on the corresponding product E^k of the state space E. First of all it is obvious that a sequence $(X_n)_{n\ge1}$ of random variables taking values in a product space is exponentially tight if and only if each of its marginal distributions is exponentially tight. Furthermore, one can show that if $D_1 \subset \mathcal{C}(E_1)$ and $D_2 \subset \mathcal{C}_\mathrm{b}(E_2)$ are both rate function determining in their complete separable metric spaces E_1 and E_2 respectively, that $D = \{f_1 + f_1 \colon f_1 \in D_1, f_2 \in D_2\}$ is a rate function determining set for the product space $E = E_1 \times E_2$. For details we refer to chapter 3.3 in [FK06]. We finish with the following particularly useful result for large deviation principles for sequences of Markov processes.

Proposition 1.13. *Let $(E_i, d_i), i = 1, 2$, be complete separable metric spaces. Suppose that the sequence $((X_n, Y_n))_{n\geq 1}$ of pairs of random variables taking values in $(E_1 \times E_1, d_1 + d_2)$ is exponentially tight in the product space. Let $\mu_n \in \mathcal{M}_1(E_1 \times E_2)$ be the distribution of (X_n, Y_n) and let $\mu_n(\mathrm{d}x \otimes \mathrm{d}y) = \eta_n(\mathrm{d}y|x)\mu_n^{(1)}(\mathrm{d}x)$. That is, $\mu_n^{(1)}$ is the E_1-marginal of μ_n and η_n gives the conditional distribution of Y_n given X_n. Suppose that for each $f \in \mathcal{C}_{\mathrm{b}}(E_2)$*

$$\Lambda_2(f|x) = \lim_{n\to\infty} \frac{1}{n} \log \int_{E_2} \mathrm{e}^{nf(y)} \eta_n(\mathrm{d}y|x) \tag{1.17}$$

exists, and furthermore that this convergence is uniform for x in compact subsets of E_1, and define

$$I_2(y|x) := \sup_{f\in\mathcal{C}_{\mathrm{b}}(E_2)} \{f(y) - \Lambda_2(f|x)\}. \tag{1.18}$$

If $(X_n)_{n\geq 1}$ satisfies the LDP with good rate function I_1, then $((X_n, Y_n))_{n\geq 1}$ satisfies the LDP with good rate function

$$I(x, y) = I_1(x) + I_2(y|x), \quad (x, y) \in E_1 \times E_2. \tag{1.19}$$

Proof. We give a brief sketch of the proof. First note that due to the assumption (1.17) we get for $f \in \mathcal{C}_{\mathrm{b}}(E), f_i \in \mathcal{C}_{\mathrm{b}}(E_i), i = 1, 2$, with $f(x, y) = f_1(x) + f_2(y)$, that

$$\Lambda(f) = \Lambda_1(f_1 + \Lambda_2(f_2|\cdot)),$$

and we note that $\Lambda_2(f_2|\cdot) \in \mathcal{C}_{\mathrm{b}}(E_1)$. We shall use Corollary 1.12 for the following functions

$$f_{m_1,m_2} = f^{(1)}_{m_1} + f^{(2)}_{m_2} := -(m_1^2 d_1(x_0, x)) \wedge m_1 - (m_2^2 d_2(y_0, y)) \wedge m_2.$$

Using the continuity of $\Lambda_2(f^{(2)}_{m_2}|\cdot)$, we easily see that

$$\widehat{f}^{(1)}_{m_1} = f^{(1)}_{m_1} + \Lambda_2(f^{(2)}_{m_2}|\cdot) - \Lambda_2(f^{(1)}_{m_1}|x_0)$$

satisfies all the assumptions (a)-(c) of Corollary 1.12 with $E = E_1$ and $x = x_0$. Using Corollary 1.12 two times we thus get

$$\begin{aligned} I(x_0, y_0) &= \lim_{m_1,m_2\to\infty} \Lambda(f_{m_1,m_2}) = \lim_{m_2\to\infty} \lim_{m_1\to\infty} \big\{\Lambda_1\big(f^{(1)}_{m_1} + \Lambda_2(f^{(2)}_{m_2}|\cdot) - \Lambda_2(f^{(1)}_{m_1}|x_0)\big)\big) + \Lambda_2(f^{(2)}_{m_2}|x_0)\big\} \\ &= I_1(x_0) + I_2(y_0|x_0), \end{aligned}$$

and therefore the desired result. □

The main objective in these notes is to introduce a strategy to prove large deviation principles for sequences of Markov processes. For that we let $(X_n)_{n\geq 1}$ be a sequence of Markov processes X_n with generator A_n taking values in E, that is, each X_n is a random element in the space $\mathcal{D}_E[0, \infty)$ of cadlag functions on E. In order to prove large deviation principles we need to show exponential tightness for sequences in the cadlag space $\mathcal{D}_E[0, \infty)$. We do not enter a detailed discussion of this enterprise here and simply refer to chapter 4 in [FK06] which is solely devoted to exponential tightness criteria for stochastic processes. Exponential tightness is usually easier to verify if the state space E is compact. Frequently, results can be obtained by first compactifying the state space, verifying the large deviation principle in the compact space, and inferring the large deviation principle in the original space from the result in the compact space. We leave the exponential tightness question by citing the following criterion, which is frequently used in proving exponential tightness.

Proposition 1.14. *A sequence $(X_n)_{n\geq 1}$ is exponentially tight in $\mathcal{D}_E[0,\infty)$ if and only if*

(a) *for each $T>0$ and $\alpha>0$, there exists a compact $K_{\alpha,T}\subset E$ such that*

$$\limsup_{n\to\infty}\frac{1}{n}\log P(\exists\, t\leq T\colon X_n(t)\notin K_{\alpha,T})\leq -\alpha; \tag{1.20}$$

(b) *there exists a family of functions $F\subset\mathcal{C}(E)$ that is closed under addition and separates points in E such that for each $f\in F$, $(f(X_n))_{n\geq 1}$ is exponentially tight in $\mathcal{D}_{\mathbb{R}}[0,\infty)$.*

We will refer to condition (1.20) as the *exponential compact containment condition.*

The next theorem shows that it suffices to prove large deviation principles for finite tuples of time points of the Markov processes, that is, one picks finitely many time points $0\leq t_1<t_2<\cdots<t_m, m\in\mathbb{N}$, and shows the large deviation principle on the product space E^m. For the proof we refer to chapter 4 of [FK06].

Theorem 1.15. *Assume that $(X_n)_{n\geq 1}$ is exponentially tight in $\mathcal{D}_E[0,\infty)$ and for all times $0\leq t_1<t_2<\cdots<t_m$, that the m-tuple $(X_n(t_1),\dots,X_n(t_m))_{n\geq 1}$ satisfies the LDP in E^m with rate the function $I_{t_1,\dots,t_m}$. Then $(X_n)_{n\geq 1}$ satisfies the LDP in $\mathcal{D}_E[0,\infty)$ with good rate function*

$$I(x)=\sup_{\{t_i\}\subset\Delta_x^{\mathrm{c}}}\big\{I_{t_1,\dots,t_m}(x(t_1),\dots,x(t_m))\big\},\quad x\in\mathcal{D}_E[0,\infty),$$

where Δ_x denotes the set of discontinuities of x.

Proof. We give a brief sketch. Pick $x\in\mathcal{D}_E[0,\infty)$ and $m\in\mathbb{N}$. For $\varepsilon>0$ and time points $t_1,\dots,t_m\notin\Delta_x$ we can find $\varepsilon'>0$ such that

$$B_{\varepsilon'}(x)\subset\big\{y\in\mathcal{D}_E[0,\infty)\colon (y(t_1),\dots,y(t_m))\in B_\varepsilon^m((x(t_1),\dots,x(t_m))\subset E^m\big\}.$$

Using (1.10) we easily arrive at

$$I(x)\geq I_{t_1,\dots,t_m}(x(t_1),\dots,x(t_m)).$$

For $\varepsilon>0$ and a compact set $K\subset\mathcal{D}_E[0,\infty)$ we find time points $t_1,\dots,t_m$, which are elements of a set $\mathcal{T}$ dense in $[0,\infty)$, such that

$$B_{\varepsilon'}(x)\supset\big\{y\in\mathcal{D}_E[0,\infty)\colon (y(t_1),\dots,y(t_m))\in B_\varepsilon^m((x(t_1),\dots,x(t_m))\subset E^m\big\}\cap K.$$

Using exponential tightness we obtain the reversed bound,

$$I(x)\leq\sup_{\{t_i\}\subset\mathcal{T}}\big\{I_{t_1,\dots,t_m}(x(t_1),\dots,x(t_m))\big\},$$

and choosing $\mathcal{T}=\Delta_x^{\mathrm{c}}$ we conclude with our statement. □

We can easily draw the following conclusion of the theorem.

Corollary 1.16. *Suppose that $D\subset\mathcal{C}_{\mathrm{b}}(E)$ is bounded above and isolates points. Assume that $(X_n)_{n\geq 1}$ is exponentially tight in $\mathcal{D}_E[0,\infty)$ and that for each $0\leq t_1\leq\cdots\leq t_m$ and $f_1,\dots,f_m\in D$,*

$$\Lambda(t_1,\dots,t_m;f_1,\dots,f_m):=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[\mathrm{e}^{n(f_1(X_n(t_1))+\cdots+f(X_n(t_m)))}]$$

exists. Then $(X_n)_{n\ge1}$ *satisfies the LDP in* $\mathcal{D}_E[0,\infty)$ *with good rate function*

$$I(x)=\sup_{m\ge1}\sup_{\{t_1,\dots,t_m\}\subset\Delta_x}\sup_{f_1,\dots,f_m\in D}\Big\{f_1(x(t_1))+\cdots+f_m(x(t_m))-\Lambda(t_1,\dots,t_m;f_1,\dots,f_m)\Big\}.$$

1.4. **Roadmap.** We shall provide a rough "road map" for studying large deviation principles for sequences of stochastic processes. This can be used as a guide along which one can prove a large deviation principle for any given sequence of processes taking values in the cadlag space $\mathcal{D}_E[0,\infty)$ using the nonlinear semigroup $((V_n(t))_{t\ge0}$ introduced earlier. A typical application requires the following steps:

Step 1. Convergence of $(H_n)_{n\ge1}$:
Verify the convergence of $(H_n)_{n\ge1}$ to a limit operator H in the sense that for each $f\in\mathcal{D}(H)$, there exists $f_n\in\mathcal{D}(H_n)$ such that

$$f_n\to f,\qquad H_nf_n\to Hf,\ \text{as } n\to\infty,$$

where the type of convergence depends on the particular problem. However, in general convergence will be in the extended limit or graph sense. This is necessary as in several cases the generators H_n are defined on a different state spaces than the limiting operator H. Let E_n be metric spaces and let $t_n\colon E_n\to E$ be a Borel measurable and define $p_n\colon B(E)\to B(E_n)$ by $p_nf=f\circ t_n$. Let $H_n\subset B(E_n)\times B(E_n)$. Then the extended limit is the collection of $(f,g)\in B(E)\times B(E)$ such that there exist $(f_n,g_n)\in H_n$ satisfying

$$\lim_{n\to\infty}\big(\|f_n-p_nf\|+\|g_n-p_ng\|\big)=0.$$

We will denote the convergence by $ex-\lim_{n\to\infty}H_n$.

Step 2. Exponential tightness: The convergence of H_n typically gives exponential tightness, provided one can verify the exponential compact containment condition. Alternatively, one can avoid verifying the compact containment condition by compactifying the state space and verifying the large deviation principle in the compactified space. In these notes we decided not to address the problem of showing exponential tightness in detail but instead refer to chapter 4 in [FK06].

Step 3. Properties of the limiting operator: Having established the convergence to a limiting operator H we need to know more about the limit operator H such that we can prove large deviation principles using operator and nonlinear semigroup convergence. There are two possible ways to proceed in this programme. If the limit operator H satisfies a *range condition*, that is, we need existence of solutions of

$$(1\!\!1-\alpha H)f=h, \tag{1.21}$$

for all sufficiently small $\alpha>0$ and a large enough collection of functions h. If we can prove the range condition we proceed with step 3 (a) below. Unfortunately, there are very few examples for which the range condition can actually be verified in the classical sense. We overcome this difficulty by using the weak solution theory developed by Crandall and Lions. The range condition is replaced by the requirement that a *comparison principle* holds for (1.21). We shall than proceed with step 3 (b).

Step 3 (a): The main result of Section 2 is Theorem (2.5), which shows that the convergence of Fleming's log-exponential nonlinear semigroup implies the large deviation principle for Markov processes taking values in metric spaces. The convergence of the semigroup immediately implies the large deviation principle for the one dimensional distributions by Bryc's formula, see Theorem 1.8 (b). The semigroup property then gives the large deviation principle for the finite dimensional distributions, which, after verifying exponential tightness, implies the pathwise large deviation principle in Theorem 1.15.

Step 3 (b): To show that an operator satisfies the range condition is in most cases difficult. Another approach is to extend the class of test functions and to introduce the so-called viscosity solution for $(\mathbb{1}-\alpha H)f=h$ for a suitable class of functions h and f. As it turns out viscosity solutions are well suited for our setting as our generators H_n and their limiting operator H have the property that $(f,g)\in H$ and $c\in\mathbb{R}$ implies $(f+c,g)\in H$. If the limit operator satisfies the comparison principle, roughly speaking the above equation has an upper and lower solution in a certain sense, one can derive the convergence of the nonlinear semigroup, and henceforth the large deviation principle afterwards in an analogous way to step 3 (a).

Step 4. Variational representation for $\boldsymbol{H}$: In many situations one can derive a variational representation of the limiting operator H. To obtain a more convenient representation of the rate function of the derived large deviation principle one is seeking a control space U, a lower semicontinuous function L on $E\times U$, and an operator $A\colon E\to E\times U$ such that $Hf(\cdot)=\sup_{u\in U}\{Af(\cdot,u)-L(\cdot,u)\}$. These objects in turn define a control problem with reward function $-L$, and it is this control problem, or more formally, the Nisio semigroup of that control problem, which allows for a control representation of the rate function. That is, we obtain the rate function as an optimisation of an integral with the reward function.

2. Main results

In this section we will show how the convergence of Fleming's nonlinear semigroup implies large deviation principles for any finite dimensional distributions of the sequences of Markov processes taking values in E. Here we shall assume that the state space is E for all generators A_n and H_n respectively. Extension to the case of n-dependent state spaces E_n are straightforward and we refer to [FK06], chapter 5.

Definition 2.1. *Let $\boldsymbol{X}=(\boldsymbol{X},\|\cdot\|)$ be a Banach space and $H\subset\boldsymbol{X}\times\boldsymbol{X}$ an operator.*

(a) *A (possibly nonlinear) operator $H\subset\boldsymbol{X}\times\boldsymbol{X}$ is dissipative if for each $\alpha>0$ and $(f_1,g_1),(f_2,g_2)\in\boldsymbol{X}$, $\|f_1-f_2-\alpha(g_1-g_2)\|\geq\|f_1-f_2\|$. $\overline{H}$ denotes the closure of H as a graph in $\boldsymbol{X}\times\boldsymbol{X}$. If H is dissipative, then $\overline{H}$ is dissipative, and for $\alpha>0$,*

$$\overline{\mathcal{R}(\mathbb{1}-\alpha H)}=\mathcal{R}(\mathbb{1}-\alpha\overline{H})$$

and $J_\alpha=(\mathbb{1}-\alpha\overline{H})^{-1}$ is a contraction operator, that is for $f_1,f_2\in\overline{\mathcal{R}(\mathbb{1}-\alpha H)}$

$$\|J_\alpha f_1-J_\alpha f_2\|\leq\|f_1-f_2\|.$$

Alternatively, H is dissipative if $\|\alpha f-Hf\|\geq\alpha\|f\|$ for all $\alpha>0$.

(b) *The following condition will be referred to as the range condition on H. There exists $\alpha_0 > 0$ such that*

$$\mathcal{D}(H) \subset \overline{\mathcal{R}(\mathbb{1} - \alpha H)} \qquad \text{for all } 0 < \alpha < \alpha_0.$$

The primary consequence of the range condition for a dissipative operator is the following theorem by Crandall and Liggett which shows that we obtain a contraction semigroup in an obvious way.

Theorem 2.2 (**Crandall-Liggett Theorem**). *Let $H \subset \boldsymbol{X} \times \boldsymbol{X}$ be a dissipative operator satisfying the range condition. Then for each $f \in \overline{\mathcal{D}(H)}$,*

$$S(t)f = \lim_{m\to\infty} (\mathbb{1} - \frac{t}{m}\overline{H})^{-1} f$$

exists and defines a contraction semigroup $(S(t))_{t\ge 0}$ on $\overline{\mathcal{D}(H)}$.

For a proof we refer to [CL71]. The following lemma provides a useful criterion showing the range condition.

Lemma 2.3. *Let V be a contraction operator on a Banach space $\boldsymbol{X}$ (that is, $\|Vf - Vg\| \le \|f - g\|, f, g \in \boldsymbol{X}$). For $\varepsilon > 0$, define*

$$Hf = \varepsilon^{-1}(Vf - f).$$

Then H is dissipative and satisfies the range condition.

For the proof of this see [Miy91]. As outlined above, Fleming's approach is to prove first a semigroup convergence. The next result provides this convergence whenever our generators satisfy the range condition with the same constant $\alpha_0 > 0$. We summarise this as follows.

Proposition 2.4. *Let H_n be as defined and $H \subset B(E) \times B(E)$ be dissipative, and suppose that each satisfies the range condition with the same constant $\alpha_0 > 0$. Let $(S_n(t))_{t\ge 0}$ and $(S(t))_{t\ge 0}$ denote the semigroups generated by H_n and H respectively (see Theorem 2.2). Suppose $H \subset ex - \lim_{n\to\infty} H_n$. Then, for each $f \in \overline{\mathcal{D}(H)}$ and $f_n \in \overline{\mathcal{D}(H_n)}$ satisfying $\|f_n - p_n f\| \to 0$ as $n \to \infty$ and $T > 0$,*

$$\lim_{n\to\infty} \sup_{0\le t\le T} \{\|V_n(t)f_n - p_n V(t) f\| = 0.$$

For the proof we refer to [Kur73]. This result extends the linear semigroup convergence theory to the setting of the Crandall-Liggett theorem. Recall that convergence of linear semigroups can be used to prove weak convergence results in much the same way as we will use convergence of nonlinear semigroups to prove large deviation theorems. The next theorem gives the first main result, showing that convergence of the nonlinear semigroups leads to large deviation principles for stochastic processes as long we are able to show the range condition for the corresponding generators. We give a version where the state spaces of all Markov processes is E and note that one can easily generalise this to having Markov processes of metric spaces E_n (see [FK06]).

Theorem 2.5 (**Large deviation principle**). *Let A_n be generator for the Markov processes $(X_n(t))_{t\ge 0}$ and define $(V_n(t))_{t\ge 0}$ on $B(E)$ by*

$$V_n(t)f(x) = \frac{1}{n}\log \mathbb{E}[\mathrm{e}^{nf(X_n(t))}|X_n(0) = x]. \tag{2.1}$$

Let $D \subset \mathcal{C}_{\mathrm{b}}(E)$ be closed under addition, and suppose that there exists an operator semigroup $(V(t))_{t\geq 0}$ on D such that

$$\lim_{n\to\infty} \|V(t)f - V_n(t)f\| = 0 \quad \text{whenever } f \in D. \tag{2.2}$$

Let $\mu_n = P(X_n(0) \in \cdot)$ and let $(\mu_n)_{n\geq 1}$ satisfy the LDP on E with good rate function I_0, which is equivalent to the existence of $\Lambda_0(f) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{\mu_n}[\mathrm{e}^{nf}]$ for each $f \in \mathcal{C}_{\mathrm{b}}(E)$, and $I_0(x_0) = \sup_{f\in\mathcal{C}_{\mathrm{b}}(E)}\{f(x_0) - \Lambda_0(f)\}$.

Then,

(a) *For each $0 \leq t_1 < \cdots < t_k$ and $f_1, \ldots, f_k \in D$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[\mathrm{e}^{nf_1(X_n(t_1))+\cdots+nf_k(X_n(t_k))}] \\ = \Lambda_0(V(t_1)(f_1 + V(t_2 - t_1)(f_2 + \cdots + V(t_k - t_{k-1})f_k)\cdots)).$$

(b) *Let $0 \leq t_1 < \cdots < t_k$ and assume that the sequence $((X_n(t_1), \ldots, X_n(t_k))_{n\geq 1}$ of random vectors $(X_n(t_1), \ldots, X_n(t_k))$ is exponentially tight in E^m. If D contains a set that is bounded above and isolates points, then the sequence $((X_n(t_1), \ldots, X_n(t_k))_{n\geq 1}$ satisfies the LDP with rate function*

$$I_{t_1,\ldots,t_k}(x_1, \ldots, x_k) = \sup_{f_i\in\mathcal{C}_{\mathrm{b}}(E)} \Big\{ \sum_{i=1}^{k} f_i(x_i) - \Lambda_0(V(t_1)(f_1 + V(t_2 - t_1)(f_2 + \cdots))) \Big\}.$$

(c) *If $(X_n)_{n\geq 1}$ is exponentially tight in $\mathcal{D}_E[0,\infty)$ and D contains a set that is bounded above and isolates points, then $(X_n)_{n\geq 1}$ satisfies the LDP in $\mathcal{D}_E[0,\infty)$ with rate function*

$$I(x) = \sup_{\{t_i\}\subset\Delta_x^c} \big\{I_{t_1,\ldots,t_k}(x(t_1), \ldots, x(t_k))\big\}. \tag{2.3}$$

Proof. We outline important steps of the proof. Note that since D is closed under addition and $V(t)\colon D \to D$, if $f_1, f_2 \in D$, then $f_1 + V(t)f_2 \in D$. We first pick $k = 2$ and $0 \leq t_1 < t_2$. Note that X_n is a Markov process on E with filtration $\mathcal{F}_t^{(X_n)}$ at time t. Using the Markov property and the definition of the nonlinear semigroup V_n it follows for $f_1, f_2 \in D$,

$$\begin{aligned}\mathbb{E}\big[\mathrm{e}^{nf_1(X_n(t_1))+nf_2(X_n(t_2))}\big] &= \mathbb{E}\big[\mathbb{E}\big[\mathrm{e}^{nf_1(X_n(t_1))+nf_2(X_n(t_2))}\big|\mathcal{F}_{t_1}^{(X_n)}\big]\big] \\ &= \mathbb{E}\big[\mathrm{e}^{nf_1(X_n(t_1))+nV_n(t_2-t_1)f_2(X_n(t_1))}\big] \\ &= \int_E P_n \circ X_n^{-1}(0)(\mathrm{d}y)\mathrm{e}^{nV_n(t_1)(f_1+V_n(t_2-t_1)f_2)}(y),\end{aligned}$$

where the integrand is

$$\mathrm{e}^{nV_n(t_1)(f_1+V_n(t_2-t_1)f_2)}(y) = \mathbb{E}\big[\mathrm{e}^{n\{f_1(X_n(t_1)+V_n(t_2-t_1)f_2(X_n(t_1))\}}\big].$$

This can be easily extended to any $f_1, \ldots, f_k \in D$ such that

$$\mathbb{E}\big[\mathrm{e}^{nf_1(X_n(t_1))+\cdots+nf_k(X_n(t_k))}\big] = \int_E P_n \circ X_n^{-1}(0)(\mathrm{d}y) \exp(ng(y))$$

where we define the integrand as

$$g_n(y) = V_n(t_1)(f_1 + V_n(t_2 - t_1)(f_2 + \cdots + V_n(t_{k-1} - t_{k-2})(f_{k-1} + V_n(t_k - t_{k-1})f_k)\cdots)(y).$$

Hence, one can show using the semigroup convergence that

$$\lim_{n\to\infty}\frac{1}{n}\log E\big[\mathrm{e}^{nf_1(X_n(t_1))+\cdots+nf_k(X_n(t_k))}\big]=\Lambda_0(V(t_1)(f_1+V(t_2-t_1)(f_2+\cdots+V(tk-t_{k-1})f_k)\cdots)).$$

Using the fact that D is closed under addition and is bounded from above and isolates points we know by Proposition 1.11 that D is a rate function determining set and henceforth we obtain statement (b). Part (c) is a direct consequences of Theorem 1.15 respectively Corollary 1.16. □

In the next Corollary we obtain a different representation of the rate function in (2.3).

Corollary 2.6. *Under the assumptions of Theorem 2.5, define*

$$I_t(y|x)=\sup_{f\in D}\{f(y)-V(t)f(x)\}. \tag{2.4}$$

Then for $0\le t_1<\cdots<t_k$,

$$I_{t_1,\dots,t_k}(x_1,\dots,x_k)=\inf_{x_0\in E}\big\{I_0(x_0)+\sum_{i=1}^k I_{t_i-t_{i-1}}(x_i|x_{i-1})\big\}$$

and

$$I(x)=\sup_{\{t_i\}\subset\Delta_x^{\mathrm{c}}}\big\{I_0(x(0))+\sum_{i=1}^k I_{t_i-t_{i-1}}(x(t_i)|x(t_{i-1}))\big\}. \tag{2.5}$$

Proof. We first note that

$$\begin{aligned}\frac{1}{n}\log\mathbb{E}\big[\mathrm{e}^{nf_k(X_n(t_k))}|X_n(t_1),\dots,X_n(t_{k-1})\big]&=\frac{1}{n}\mathbb{E}\big[V_n(t_k-t_{k-1})f_k(X_n(t_{k-1})|X_n(t_1),\dots,X_n(t_{k-1})\big]\\&=V(t_k-t_{k-1})f_k(X_n(t_{k-1}))+O(\|V(t_k-t_{k-1})f_k-V_n(t_k-t_{k-1})f_k\|).\end{aligned}$$

Furthermore note that the uniform convergence in Proposition 1.13 can be replaced by the assumption that there exists a sequence $K_n\subset E_1$ such that

$$\lim_{n\to\infty}\sup_{x\in K_n}\Big|\Lambda_2(f|x)-\frac{1}{n}\log\int_{E_2}\mathrm{e}^{nf(y)}\eta_n(\mathrm{d}y|x)\Big|=0$$

and

$$\lim_{n\to\infty}\frac{1}{n}\log\mu_n^{(1)}(K_n^{\mathrm{c}})=-\infty.$$

From this and Proposition 1.13 it follows that

$$I_{t_1,\dots,t_k}(x_1,\dots,x_k)=I_{t_1,\dots,t_k}(x_1,\dots,x_{k-1})+I_{t_k-t_{k-1}}(x_k|x_{k-1}).$$

In the same way, one can show that the rate function for $(X_n(0),X_n(t_1))_{n\ge1}$ is $I_{0,t_1}(x_0,x_1)=I_0(x_0)+I_{t_1}(x_1|x_0)$, and we conclude the proof using induction and the contraction principle of large deviations (see e.g. [DZ98]).

□

These results finish step 3 (a) of our roadmap once we have shown the semigroup convergence. This is done in the following Corollary.

Corollary 2.7. *Under the assumptions of Theorem 2.5, define*

$$H_n f = \frac{1}{n}\mathrm{e}^{-nf} A_n \mathrm{e}^{nf}$$

on the domain $\mathcal{D}(H_n) = \{f \colon \mathrm{e}^{nf} \in \mathcal{D}(A_n)\}$, *and let*

$$H \subset \ ex - \lim_{n\to\infty} H_n.$$

Suppose that H *is dissipative,* $\mathcal{D}(H) \subset \mathcal{C}_\mathrm{b}(E)$, H *satisfies the range condition, and that the (norm) closure* $\overline{\mathcal{D}(H)}$ *is closed under addition and contains a set that is bounded above and isolates points. Suppose further that the sequence* $(X_n)_{n\ge 1}$ *satisfies the exponential compact containment condition* (1.20) *and that* $(X_n(0))_{n\ge1}$ *satisfies the LDP with good rate function* I_0*. Then*

(a) *The operator* H *generates a semigroup* $(V(t))_{t\ge0}$ *on* $\overline{\mathcal{D}(H)}$, *and for* $f \in \overline{\mathcal{D}(H)}$ *and* $f_n \in E$ *satisfying* $\|f - f_n\| \to 0$ *as* $n\to\infty$,

$$\lim_{n\to\infty} \|V(t)f - V_n(t)f_n\| = 0.$$

(b) $(X_n)_{n\ge1}$ *is exponentially tight and satisfies the LDP with good rate function* I *with* $D = \overline{\mathcal{D}(H)}$.

Proof. We give only a brief sketch of the proof and refer otherwise to Lemma 5.13 in [FK06]. The semigroup convergence will be shown via the Yosida approximations for the generators, that is, for $\varepsilon_n > 0$ define

$$A_n^{\varepsilon_n} = \frac{1}{\varepsilon_n}((\mathbb{1} - \varepsilon_n A_n)^{-1} - \mathbb{1}) = A_n(\mathbb{1} - \varepsilon_n A_n)^{-1}.$$

The Yosida approximation is a bounded, linear dissipative operator. One shows the semigroup convergence for the Yosida approximations, and then using for $(f_n, g_n) \in H_n$ the estimate

$$\sup_{y\in E}\{\|V_n(t)f_n(y) - V_n^{\varepsilon_n}(t)f_n\|\} \le \sqrt{2\varepsilon_n t}\,\mathrm{e}^{2n\|f_n\|}\|g_n\|,$$

and the fact that we can choose ε_n such that $\lim_{n\to\infty}\varepsilon_n \mathrm{e}^{nc} = 0$ for all $c > 0$, we obtain the desired semigroup convergence. □

3. Large deviations using viscosity semigroup convergence

In the previous section, we proved a large deviation theorem using semigroup convergence by imposing a range condition on the limit operator. A range condition essentially requires that

$$(\mathbb{1} - \alpha H)f = h \tag{3.1}$$

have a solution $f \in \mathcal{D}(H)$, for each $h \in \mathcal{D}(H)$ and small $\alpha > 0$. In this section, we show that uniqueness of f in a weak sense (i.e., the viscosity sense) is sufficient for the semigroup convergence. This approach is quite effective in applications. However, it requires a couple of new definitions and notions. First, recall the notion of upper and lower semicontinuous regularisation. For $g \in B(E)$, g^* will denote the *upper semicontinuous regularisation* of g, that is,

$$g^*(x) = \lim_{\varepsilon\to0} \sup_{y\in B_\varepsilon(x)} \{g(y)\}, \tag{3.2}$$

and g_* will denote the *lower semicontinuous regularisation*,

$$g_*(x) = \lim_{\varepsilon\to 0} \inf_{y\in B_\varepsilon(x)} \{g(y)\}. \tag{3.3}$$

Using these notions we introduce the notion of viscosity solution and define the comparison principle afterwards. The intention here is not to enter in detail the construction and properties of viscosity solutions (we shall do that later when dealing with specific examples). Here, we are only interested to see that viscosity solutions are often easier to obtain than the range condition and that it turn suffices to prove large deviation theorems.

Definition 3.1. *Let E be a compact metric space, and $H \subset \mathcal{C}(E) \times B(E)$. Pick $h \in \mathcal{C}(E)$ and $\alpha > 0$. Let $f \in B(E)$ and define $g := \alpha^{-1}(f - h)$, that is, $f - \alpha g = h$. Then*

(a) *f is a viscosity subsolution of (3.1) if and only if f is upper semicontinuous and for each $(f_0, g_0) \in H$ such that $\sup_x\{f(x) - f_0(x)\} = \|f - f_0\|$, there exists $x_0 \in E$ satisfying*

$$\begin{aligned} (f - f_0)(x_0) &= \|f - f_0\|, \text{ and} \\ \alpha^{-1}(f(x_0) - h(x_0)) &= g(x_0) \le (g_0)^*(x_0). \end{aligned} \tag{3.4}$$

(b) *f is a viscosity supersolution of (3.1) if and only if f is lower semicontinuous and for each $(f_0, g_0) \in H$ such that $\sup_{x\in E}\{f_0(x) - f(x)\} = \|f_0 - f\|$, there exists $x_0 \in E$ satisfying*

$$\begin{aligned} (f_0 - f)(x_0) &= \|f\_f\|, \text{ and} \\ \alpha^{-1}(f(x_0) - h(x_0)) &= g(x_0) \ge (g_0)_*(x_0). \end{aligned} \tag{3.5}$$

(c) *A function $f \in \mathcal{C}(E)$ is said to be a viscosity solution of (3.1) if it is both a subsolution and a supersolution.*

The basic idea of the definition of a viscosity solution is to extend the operator H so that the desired solution is included in an extended domain while at the same time keeping the dissipativity of the operator. Viscosity solutions have been introduced by Crandall and Lions (1983) to study certain nonlinear partial differential equations.

Definition 3.2 (Comparison Principle). *We say that*

$$(\mathbb{1} - \alpha H)f = h$$

satisfies a comparison principle, if $\overline{f}$ a viscosity subsolution and $\underline{f}$ a viscosity supersolution implies $\overline{f} \le \underline{f}$ on E.

The operator H we are going to analyse has the following property that $(f, g) \in H$ and $c \in \mathbb{R}$ implies $(f + c, g) \in H$. This allows for the following simplification of the definition of viscosity solution.

Lemma 3.3. *Suppose $(f, g) \in H$ and $c \in \mathbb{R}$ implies that $(f + c, g) \in H$. Then an upper semicontinuous function f is a viscosity subsolution of (3.1) if and only if for each $(f_0, g_0) \in H$, there exists $x_0 \in E$ such that*

$$\begin{aligned} f(x_0) - f_0(x_0) &= \sup_{x\in E}\{f(x) - f_0(x)\}, \quad \text{and} \\ \alpha^{-1}(f(x_0) - h(x_0)) &= g(x_0) \le (g_0)^*(x_0). \end{aligned} \tag{3.6}$$

Similarly, a lower semicontinuous function f is a viscosity supersolution of (3.1) *if and only if for each $(f_0, g_0) \in H$, there exists $x_0 \in E$ such that*

$$\begin{aligned} f_0(x_0) - f(x_0) &= \sup_{x\in E}\{f_0(x) - f(x)\}, \quad \text{and} \\ \alpha^{-1}(f(x_0) - h(x_0)) &= g(x_0) \geq (g_0)_*(x_0). \end{aligned} \tag{3.7}$$

Our main aim is to replaced the range condition by the comparison principle. The following lemma shows that this can be achieved once our limiting operator H has the property that $(f, g) \in H$ and $c \in \mathbb{R}$ implies $(f + c, g) \in H$. The generators of our nonlinear semigroups, i.e., each H_n and the limiting operator H, all share this property.

Lemma 3.4. *Let (E, d) be a compact metric space and $H \subset \mathcal{C}(E) \times B(E)$, and assume the property that $(f, g) \in H$ and $c \in \mathbb{R}$ implies $(f+c, g) \in H$. Then for all $h \in \mathcal{C}(E)\cap\overline{\mathcal{R}(\mathbb{1} - \alpha H)}$, the comparison principle holds for $f - \alpha H f = h$ for all $\alpha > 0$.*

Extensions of operators are frequently used in the study of viscosity solutions. We denote the closure of an operator H by $\overline{H}$. It is obvious that the closure $\overline{H}$ is a viscosity extension of the operator H in the following sense.

Definition 3.5. *$\widehat{H} \subset B(E) \times B(E)$ is a viscosity extension of H if $\widehat{H} \supset H$ and for each $h \in \mathcal{C}(E)$, f is a viscosity subsolution (supersolution) of $(\mathbb{1} - \alpha H)f = h$ if and only if f is a viscosity subsolution (supersolution) for $(\mathbb{1} - \alpha\widehat{H})f = h$.*

In our applications we will often use a viscosity extension of the limiting operator. For this we shall need a criterion when an extension of the operator is actually a viscosity extension. Using again the property that $(f, g) \in H$ and $c \in R$ implies $(f + c, g) \in H$ we arrive at the following criterion for viscosity extension.

Lemma 3.6. *Let (E, d) be a compact metric space and $H \subset \widehat{H} \subset \mathcal{C}(E) \times B(E)$. Assume that for each $(f, g) \in H$ we have a sequence $(f_n, g_n) \subset H$ such that $\|f_n - f\| \to 0$ as $n \to \infty$ and $x_n, x \in E$ with $x_n \to x$ as $n \to \infty$,*

$$g_*(x) \leq \liminf_{n\to\infty} g_n(x_n) \leq \limsup_{n\to\infty} g_n(x_n) \leq g^*(x).$$

In addition, suppose that $(f, g) \in H$ and $c \in R$ implies $(f + c, g) \in H$. Then the extension $\widehat{H}$ is a viscosity extension.

The idea is to replace the range condition of the limiting operator H by the comparison principle. The next result summarises the semigroup convergence when the limiting operator satisfies the comparison principle.

Proposition 3.7 (**Semigroup convergence using viscosity solution**). *Let (E, d) be a compact metric space. Assume for each $n \geq 1$ that $H_n \subset B(E) \times B(E)$ is dissipative and satisfies the range condition for $0 < \alpha < \alpha_0$ with $\alpha_0 > 0$ independent of n, and assume that $(f_n, g_n) \in H_n$ and $c \in \mathbb{R}$ implies $(f_n + c, g_n) \in H_n$. Let $H \subset \mathcal{C}(E) \times B(E)$. Suppose that for each $(f_0, g_0) \in H$, there exists a sequences $(f_{n,0}, g_{n,0}) \in H_n$ such that $\|f_{n,0} - f_0\| \to 0$ as $n \to \infty$ and for each $x \in E$ and sequence $z_n \to x$ as $n \to \infty$,*

$$(g_0)_*(x) \leq \liminf_{n\to\infty}(g_{n,0})(z_n) \leq \limsup_{n\to\infty}(g_{n,0})(z_n) \leq (g_0)^*(x). \tag{3.8}$$

Assume that for each $0 < \alpha < \alpha_0$, $\mathcal{C}(E) \subset \mathcal{R}(\mathbb{1} - \alpha H_n)$, and that there exists a dense subset $D_\alpha \subset \mathcal{C}(E)$ such that for each $h \in D_\alpha$, the comparison principle holds for

$$(\mathbb{1} - \alpha H)f = h. \tag{3.9}$$

Then

(a) *For each $h \in D_\alpha$, there exists a unique viscosity solution $f := R_\alpha h$ of (3.9).*

(b) *R_α is a contraction and extends to $\mathcal{C}(E)$.*

(c) *The operator*

$$\widehat{H} = \bigcup_\alpha \Big\{ (R_\alpha h, \frac{R_\alpha h - h}{\alpha}) \colon h \in \mathcal{C}(E) \Big\} \tag{3.10}$$

is dissipative and satisfies the range condition, $\overline{\mathcal{D}(\widehat{H})} \supset \mathcal{D}(H)$, and $\widehat{H}$ generates a strongly continuous semigroup $(V(t))_{t \geq 0}$ on $\overline{\mathcal{D}(\widehat{H})} = \mathcal{C}(E) =: D$ given by $V(t)f = \lim_{n\to\infty} R^n_{t/n} f$.

(d) *Let $(V_n(t))_{t \geq 0}$ denote the semigroup on $\overline{\mathcal{D}(H_n)}$ generated by H_n. For each $f \in \overline{\mathcal{D}(\widehat{H})}$ and $f_n \in \overline{\mathcal{D}(H_n)}$ with $\|f_n - f\| \to 0$ as $n \to \infty$,*

$$\lim_{n\to\infty} \|V_n(t) f_n - V(t) f\| = 0.$$

We do not give a proof here. See chapter 6 and 7 in [FK06]. Also note that the assumption (E, d) being a compact metric space can be relaxed, see chapter 7 and 9 in [FK06]. The semigroup convergence uses again Theorem 2.2 of Crandall-Liggett. We arrive at our second main large deviation result.

Theorem 3.8 (**Large deviations using semigroup convergence**). *Let (E, d) be a compact metric space and for any $n \geq 1$ let $A_n \subset B(E) \times B(E)$ be the generator for the Markov process X_n taking values in E. Define*

$$H_n f = \frac{1}{n} \mathrm{e}^{-nf} A_n \mathrm{e}^{nf} \quad \text{whenever } \mathrm{e}^{nf} \in \mathcal{D}(A_n),$$

and on $B(E)$ define the semigroup $(V_n(t))_{t\geq 0}$ by $V_n(t)f(x) = \frac{1}{n} \log \mathbb{E}[\mathrm{e}^{nf(X_n(t))} | X_n(0) = x]$. Let $H \subset \mathcal{D}(H) \times B(E)$ with $\mathcal{D}(H)$ dense in $\mathcal{C}(E)$. Suppose that for each $(f, g) \in H$ there exists $(f_n, g_n) \in H_n$ such that $\|f - f_n\| \to 0$ as $n \to \infty$ and $\sup_n \|g_n\| < \infty$, and for each $x \in E$ and $z_n \to x$ as $n \to \infty$, we have

$$g_*(x) \leq \liminf_{n\to\infty} g_n(z_n) \leq \limsup_{n\to\infty} g_n(z_n) \leq (g^*)(x). \tag{3.11}$$

Pick $\alpha_0 > 0$. Suppose that for each $0 < \alpha < \alpha_0$, there exists a dense subset $D_\alpha \subset \mathcal{C}(E)$ such that for each $h \in D_\alpha$, the comparison principle holds for (3.1). Suppose that the sequence $(X_n(0))_{n\geq 1}$ satisfies the LDP in E with the good rate function I_0.

Then

(a) *The operator $\widehat{H}$ defined in (3.10) generates a semigroup $((V(t))_{t\geq 0}$ on $\mathcal{C}(E)$ such that $\lim_{n\to\infty} \|V(t)f - V_n(t)f_n\| = 0$ whenever $f \in \mathcal{C}(E)$,$f_n \in B(E)$, and $\|f - f_n\| \to 0$ as $n \to \infty$.*

(b) $(X_n)_{n\ge 1}$ *is exponentially tight and satisfies the LDP on* $\mathcal{D}_E[0,\infty)$ *with the rate function given in Theorem 2.5, i.e.,*

$$I(x) = \sup_{\{t_i\}\subset\Delta_x^c} \Big\{ I_0(x_0) + \sum_{i=1}^k I_{t_i-t_{i-1}}(x(t_i)|x(t_{i-1}))\Big\},$$

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

Proof. We only give a brief sketch of the major steps. Having the generator A_n one defines a viscosity extension $\widehat{A}_n$ of A_n, see Lemma 3.6 and Proposition 3.7. Picking $\varepsilon_n > 0$ we consider the corresponding Yosida approximations $\widehat{A}_n^{\varepsilon_n}$, i.e.,

$$\widehat{A}_n^{\varepsilon_n} = \frac{1}{\varepsilon_n}((\mathbb{1}-\varepsilon_n\widehat{A}_n)^{-1} - \mathbb{1}) = \widehat{A}_n(\mathbb{1}-\varepsilon_n\widehat{A}_n)^{-1}.$$

The Yosida approximation is a bounded, linear dissipative operator. Then one define the generator $\widehat{H}_n^{\varepsilon_n}$ as in the statement of the theorem. Then $\widehat{H}_n^{\varepsilon_n}$ satisfies the range condition, in fact $\mathcal{R}(1-\alpha\widehat{H}_n^{\varepsilon_n}) = B(E)$ for all $\alpha > 0$, and it generates the semigroup $((V_n^{\varepsilon_n}(t))_{t\ge 0}$. We also have $D_\alpha \subset \mathcal{R}(\mathbb{1}-\alpha\widehat{H}_n^{\varepsilon_n})$ for all $\alpha > 0$. By Proposition 3.7, using (3.11), we get

$$\|V_n^{\varepsilon_n}(t)f_n - V(t)f\| \to 0 \;\text{ as } n\to\infty,$$

whenever $f_n \in B(E), f \in \mathcal{C}(E)$, and $\|f_n - f\| \to 0$ as $n\to\infty$. The remainder of the proof follows closely the proof of Theorem 2.5 and Corollary 2.6 □

One can extend the previous results in several directions, First, we can allow the state space E to be a non-compact metric space. Second, one can relax the boundedness assumptions on the domain and range of the limiting operator H This may appear odd, but it allows for greater flexibility in the choice of test functions needed to verify the comparison principle. We refer to chapter 7 and 9 in [FK06].

4. Control representation of the rate function

Having established large deviation theorems in the two previous sections we are now interested to obtain more useful representations of the rate functions. The form of the rate functions in Theorem 2.5, Corollary 2.6 and Theorem 3.8 may have limited use in applications. The main idea is to obtain a variational representation of the limiting generator H of the nonlinear Fleming semigroup which allows for a control representation of the rate function of the large deviation principles. In many situations the generators A_n of the Markov processes satisfy the maximum principle which ensures a variational representation for H. As it will turn out this representation has the form of the generator of a Nisio semigroup in control theory. Henceforth we can formally interpret the limiting Fleming semigroup $(V(t))_{t\ge 0}$ in the large deviation theorems as Nisio semigroups. We will omit technical details in this section and focus on the main ideas and will conclude later with a couple of examples.

We pick a control space U, and let q be a metric such that (U,q) is a metric space. For that control space we are seeking for an operator $A \subset B(E)\times B(E\times U)$ and a lower semicontinuous function $L\colon E\times U\mathbb{R}$ such $\mathcal{H}f = Hf$, where we define

$$\mathcal{H}f(x) = \sup_{u\in U}\{Af(x,u) - L(x,u)\}. \tag{4.1}$$

If we can establish that $Hf = \mathcal{H}f$ for all f in the domain of the generator H of Fleming's nonlinear semigroup $((V(t))_{t\ge 0}$ we are in the position to get derive a control representation of the rate function. Before we show that we introduce some basic notions and facts from control theory.

Definition 4.1. (a) *Let $\mathcal{M}_m$ denote the set of Borel measures μ on $U\times[0,\infty)$ such that $\mu(U\times[0,t]) = t$. Let $A\colon \mathcal{D}(A)\to M(E\times U)$ with $\mathcal{D}(A)\subset B(E)$ be a linear operator. The pair $(x,\lambda)\in\mathcal{D}_E[0,\infty)\times\mathcal{M}_m$ satisfies the control equation for A, if and only if $\int_{U\times[0,t]} Af(x(s),u)\lambda(\mathrm{d}u\otimes\mathrm{d}s)<\infty$, and*

$$f(x(t)) - f(x(0)) = \int_{U\times[0,t]} Af(x(s),u)\lambda(\mathrm{d}u\otimes\mathrm{d}s), \qquad f\in\mathcal{D}(A), t\ge 0. \tag{4.2}$$

We denote the set of those pairs by $\mathcal{J}$, and for $x_0\in E$, we define

$$\mathcal{J}_{x_0} = \{(x,\lambda)\in\mathcal{J}\colon x(0)=x_0\}.$$

(b) *We let $((\mathcal{V}(t))_{t\ge0}$ denote the Nisio semigroup corresponding to the control problem with dynamics given by (4.2) and reward function $-L$, that is,*

$$\mathcal{V}(t)g(x_0) = \sup_{(x,\lambda)\in\mathcal{J}^\Gamma_{x_0}} \Big\{ g(x(t)) - \int_{U\times[0,t]} L(x(s),u)\lambda(\mathrm{d}u\otimes\mathrm{d}s)\Big\}, \qquad x_0\in E, \tag{4.3}$$

where $\Gamma\subset E\times U$ is some closed set and

$$\mathcal{J}^\Gamma = \{(x,\lambda)\in\mathcal{J}\colon \int_{U\times[0,t]} 1_\gamma(x(s),u)\lambda(du\otimes ds) = t, t\ge 0\}$$

is the set of admissible controls. This restriction of the controls is equivalent to the requirement that $\lambda_s(\Gamma_{x(s)})=1$, where for any $z\in E$, $\Gamma_z := \{u\colon (z,u)\in\Gamma\}$.

Considering our study of the Fleming's semigroup method for proving large deviation principles we expect that $\mathcal{V}(t) = V(t)$, and that we will able to simplify the form of the rate function I using the variational structure induced by A and the reward function L. Our primary result is the representation of the rate function of our large deviation principles in the following theorem. Unfortunately that theorem requires a long list of technical conditions to be satisfied which we are not going to list in detail. Instead we provide the key conditions and refer to the complete list of condition 8.9, 8.10, and 8.11 in [FK06].

Conditions ($\diamond$):

(C1) $A\subset\mathcal{C}_\mathrm{b}(E)\times\mathcal{C}(E\times U)$ and $\mathcal{D}(A)$ separates points.

(C2) $\Gamma\subset E\times U$ closed and for each $x_0\in E$ there exists $(x,\lambda)\in\mathcal{J}^\Gamma$ with $x(0)=x_0$.

(C3) $L\colon E\times U\to[0,\infty]$ is lower semicontinuous, and for each compact $K\subset E$ the set $\{(x,u)\in\Gamma\colon L(x,u)\le c\}\cap(K\times U)$ is relatively compact.

(C4) Compact containment condition. That is for each compact set $K\subset E$ and time horizon $T>0$ there is another compact set $\widehat{K}\subset$ such that $x(0)\in K$ and the boundedness of the integral imply that the terminal point $x(T)\in\widehat{K}$.

(C5) Some growth condition for $Af(x,u)$.

(C6) For each $x_0 \in E$, there exists $(x,\lambda) \in \mathcal{J}^\Gamma$ such that $x(0) = x_0$ and $\int_{U\times[0,\infty]} L(x(s),u)\lambda(\mathrm{d}u\otimes\mathrm{d}s) = 0$.

(C7) Certain growth conditions in case there are two limiting operators instead of H (lower and upper bound). This applies to extensions described in chapter 7 of [FK06].

Theorem 4.2. *Let $A \subset \mathcal{C}_\mathrm{b}(E)\times C(E\times U)$ and $L\colon E\times U \to \mathbb{R}_+$ be lower semicontinuous satisfying conditions ($\diamond$). Suppose that all the conditions from Theorem 2.5 hold. Define $\mathcal{V}$ as in* (4.3)*, and assume that*

$$\mathcal{V} = V \text{ on } D \subset \mathcal{C}_\mathrm{b}(E),$$

where D is the domain of the nonlinear semigroup $((V(t))_{t\ge 0}$ in Theorem 2.5. Assume in addition that D contains a set that is bounded from above and isolates points. Then, for $x \in \mathcal{D}_E[0,\infty)$,

$$I(x) = I_0(x(0)) + \inf_{\lambda\colon (x,\lambda)\in\mathcal{J}^\Gamma_{x_0}} \Big\{ \int_{U\times[0,\infty)} L(x(s),u)\lambda(\mathrm{d}u\otimes\mathrm{d}s)\Big\}. \tag{4.4}$$

Proof. We outline the main steps of the proof. First note that for $f\in D$ and $x_0\in$ we have

$$V(t)f(x_0) = \mathcal{V}(t)f(x_0) = -\inf_{(x,\lambda)\in\mathcal{J}^\Gamma_{x_0}} \Big\{ \int_{U\times[0,\infty]} L(x(s),u)\lambda(\mathrm{d}u\otimes\mathrm{d}s) - f(x(t))\Big\}.$$

We are going to use Corollary 2.6 again with a sequence $(f_m)_{m\ge1}$ of functions $f_m\in D$ satisfying the conditions of Corollary 1.12. We thus get

$$\begin{aligned} I_t(x_1|x_0) &= \sup_{f\in D}\{f(x_1) - \mathcal{V}(t)f(x_0)\} \\ &= \sup_{f\in D}\inf_{(x,\lambda\in\mathcal{J}^\Gamma_{x_0}} \Big\{ f(x_1) - f(x(t)) + \int_{U\times[0,t]} L(x(s),u)\lambda(\mathrm{d}u\otimes\mathrm{d}s)\Big\} \\ &= \lim_{m\to\infty}\inf_{(x,\lambda\in\mathcal{J}^\Gamma_{x_0}} \Big\{ f_m(x_1) - f_m(x(t)) + \int_{U\times[0,t]} L(x(s),u)\lambda(\mathrm{d}u\otimes\mathrm{d}s)\Big\}. \end{aligned}$$

There are two possibilities. Either $I_t(x_1|x_0)=\infty$ or there is a minimising sequence $(x_m,\lambda_m)_{m\ge1}$ such that $x_m(t)\to x_1$ as $m\to\infty$ and $\limsup_{m\to\infty}\int_{U\times[0,t]} L(x_m(s),\lambda_m(\mathrm{d}u\otimes\mathrm{d}s) < \infty$. There exists a subsequence whose limit (x^*,λ^*) satisfies $x^*(0)=x_0$ and $x^*(t)=x_1$, and

$$\begin{aligned} I_t(x_1|x_0) &= \int_{U\times[0,t]} L(x^*(s),u)\lambda^*(\mathrm{d}u\otimes\mathrm{d}s) \\ &= \inf_{(x,\lambda)\in\mathcal{J}^\Gamma_{x_0}\colon x(0)=x_0, x(t)=x_1} \Big\{ \int_{U\times[0,t]} L(x(s),u)\lambda(\mathrm{d}u\otimes\mathrm{d}s)\Big\}. \end{aligned}$$

This concludes the proof. $\square$

For the remaining technical details we refer to chapter 8 and 9 in [FK06]. We briefly discuss the possible variational representation for H. We assume that the limiting Markov generator A (as the limit of the generators A_n of our Markov processes) has as its domain $D=\mathcal{D}(A)$ an algebra and that $f\in D$ implies $\mathrm{e}^f\in D$. We define a new set of transformed operators as follows. Define for $f,g\in D$:

$$\begin{aligned} Hf &= \mathrm{e}^{-f}A\mathrm{e}^{f},\\ A^g f &= \mathrm{e}^{-g}A(f\mathrm{e}^{g})-(\mathrm{e}^{-g}f)A\mathrm{e}^{g},\\ Lg &= A^g g - Hg. \end{aligned} \tag{4.5}$$

The following natural duality between H and L was first observed by Fleming.

$$\begin{aligned} Hf(x) &= \sup_{g\in D}\{A^g f(x)-Lg(x)\},\\ Lg(x) &= \sup_{f\in D}\{A^g f(x)-Hf(x)\}, \end{aligned} \tag{4.6}$$

with both suprema attained at $f=g$. In our strategy for proving large deviation principles we are given a sequence $(H_n)_{n\ge 1}$ of operators H_n of the form

$$H_n f = \frac{1}{n}\mathrm{e}^{-nf}A_n\mathrm{e}^{nf}.$$

We define

$$\begin{aligned} A_n^g f &= \mathrm{e}^{-ng}A_n(f\mathrm{e}^{ng})-(\mathrm{e}^{-ng}f)A_n\mathrm{e}^{ng},\\ L_n^g &= A_n^g g - H_n g. \end{aligned} \tag{4.7}$$

Then for $D\subset\mathcal{C}_{\mathrm{b}}(E)$ with the above properties we get the corresponding duality

$$\begin{aligned} H_n f(x) &= \sup_{g\in D}\{A_n^g f(x)-L_n g(x)\},\\ L_n g(x) &= \sup_{f\in D}\{A_n^g f(x)-H_n f(x)\}. \end{aligned}$$

In application we typically compute L_n and A_n^g and then directly verify that the desired duality (4.6) holds in the limit. Also note that (4.6) is a representation with control space $U=D$ which may not be the most convenient choice for a control space.

When E is a convex subset of a topological vector space, the limiting operator H can be (in most cases) written as $\mathcal{H}(x,\nabla f(x))$, where $(\mathcal{H}(x,p)$ is convex in the second entry and ∇f is the gradient of f. This allows for an alternative variational representation for the operator H, namely through the Fenchel-Legendre transform. For that let E^* be the dual of E, and define for each $\ell\in E^*$

$$L(x,\ell)=\sup_{p\in E}\{\langle p,\ell\rangle-\mathcal{H}(x,p)\}. \tag{4.8}$$

Having some sufficient regularity on $\mathcal{H}(x,\cdot)$, we obtain

$$\mathcal{H}(x,p)=\sup_{\ell\in E^*}\{\langle p,\ell\rangle-L(x,\ell)\}, \tag{4.9}$$

which implies

$$Hf(x)=\mathcal{H}(x,\nabla f(x))=\sup_{\ell\in E^*}\{\langle\nabla f(x),\ell\rangle-L(x,\ell)\}.$$

This transform approach has been employed by Freidlin and Wentzell and is commonly used when studying random perturbations of dynamical systems, see [FW98]. We leave a thorough discussion of these issues and refer to chapter 8 in [FK06] and turn to the following examples.

Example 4.3. The best-known example of large deviations for Markov processes goes back to Freidlin and Wentzell (see [FW98]) and considers diffusions with small diffusion coefficient. Let X_n be a sequence of diffusion processes taking values in $\mathbb{R}^d$ satisfying the Itô equation

$$X_n(t) = x + \frac{1}{\sqrt{n}} \int_0^t \sigma(X_n(s))\,\mathrm{d}W(s) + \int_0^t b(X_n(s))\,\mathrm{d}s.$$

Putting $a(x) = \sigma(x)\cdot\sigma^T(x)$ the generator is given as

$$A_n g(x) = \frac{1}{2n}\sum_{i,j} a_{ij}(x)\partial_i\partial_j g(x) + \sum_i b_i(x)\partial_i g(x),$$

where we take $\mathcal{D}(A_n) = \mathcal{C}_{\mathrm{c}}^2(\mathbb{R}^d)$. We compute

$$H_n f(x) = \frac{1}{2n}\sum_{i,j} a_{ij}(x)\partial_i\partial_j g(x) + \frac{1}{2}\sum_{ij} a_{ij}(x)\partial_i f(x)\partial_j f(x) + \sum_i b_i(x)\partial_i g(x).$$

Henceforth we easily get $Hf = \lim_{n\to\infty} H_n f$ with

$$Hf(x) = \frac{1}{2}(\nabla f(x))^T \cdot a(x)\cdot\nabla f(x)\rangle + b(x)\cdot\nabla f(x).$$

We shall find a variational representation of H. This is straightforward once we introduce a pair of functions on $\mathbb{R}^d\times\mathbb{R}^d$, namely

$$\begin{aligned}\mathcal{H}(x,p) &:= \frac{1}{2}|\sigma^T(x)\cdot p|^2 + b(x)\cdot p,\\ L(x,\ell) &= \sup_{p\in\mathbb{R}^d}\{\langle p,\ell\rangle - \mathcal{H}(x,p)\}.\end{aligned}$$

Thus

$$Hf(x) = \mathcal{H}(x,\nabla f(x)) = \mathcal{H}f(x) := \sup_{u\in\mathbb{R}^d}\{Af(x,u) - L(x,u)\}$$

with $Af(x,u) = u\cdot\nabla f(x)$. Hence we can use H, i.e., $\mathcal{H}$, as a Nisio semigroup generator, and using Theorem 4.2 we obtain the rate in variational form,

$$I(x) = I_0(x_0) + \inf_{u\colon (x,u)\in\mathcal{J}}\Big\{\int_0^\infty L(x(s),u(s))\,\mathrm{d}s\Big\},$$

where I_0 is the rate function for $(X_n(0))_{n\ge1}$ and $\mathcal{J}$ is the collection of solutions of

$$f(x(t)) - f(x(0)) = \int_0^t Af(x(s),u(s))\,\mathrm{d}s, \qquad f\in\mathcal{C}_{\mathrm{c}}^2(\mathbb{R}^d).$$

This leads to $\dot{x}(t) = u(t)$, and thus

$$I(x) = \begin{cases} I_0(x(0)) + \int_0^\infty L(x(s),\dot{x}(s))\,\mathrm{d}s & \text{if } x \text{ is absolutely continuous,}\\ \infty & \text{otherwise.}\end{cases}$$

There is another variational representation of the operator H leading to a different expression of the rate function I. Put

$$Af(x,u) = u\cdot(\sigma^T(x)\nabla f(x)) + b(x)\cdot\nabla f(x), \quad f\in\mathcal{C}_0^2(\mathbb{R}^d),$$

$L(x,u)=\frac{1}{2}|u|^2$, and define $\mathcal{H}f(x)=\sup_{u\in\mathbb{R}^d}\{Af(x,u)-L(x,u)\}$. Then $H=\mathcal{H}$ and we obtain the following expression of the rate function

$$I(x)=I_0(x(0))+\frac{1}{2}\inf\Big\{\int_0^\infty|u(s)|^2\,\mathrm{d}s\colon u\in L^2[0,\infty)\text{ and }x(t)=x(0)\\ +\int_0^t b(x(s))\,\mathrm{d}s+\int_0^t\sigma(x(s))u(s)\,\mathrm{d}s\Big\}.$$

◇

Example 4.4. We consider a Poisson walk with small increments on $\mathbb{R}$. Let $(X_n(t))_{t\geq 0}$ be the random walk on $\mathbb{R}$ that jumps n^{-1} forward at rate bn and backward at rate dn, with $b,d\in(0,\infty)$. The generator of this Markov process is given as

$$(A_nf)(x)=bn[f(x+\frac{1}{n})-f(x)]+dn[f(x-\frac{1}{n})-f(x)].$$

If $\lim_{n\to\infty}X_n(0)=x\in\mathbb{R}$, then $\lim_{n\to\infty}X_n(t)=x+(b-d)t,t>0$. Henceforth, in the limit $n\to\infty$, the process X_n becomes a deterministic flow solving $\dot{x}=(b-d)$. For all $n\in\mathbb{N}$ we have

$$X_n(t)=\sum_{i=1}^n(X_i^{(bt)}-Y_i^{(dt)}),$$

where $X_i^t,Y_i^t,i=1,\dots,n$, are independent Poisson random variables with mean bt and dt respectively. Thus we may employ Cramér's theorem for sums of i.i.d. random variables (see [DZ98] for details) to derive the rate function

$$I(at)=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n(X_n(t)=at|X_n(0)=0)=\sup_{\lambda\in\mathbb{R}}\big\{at\lambda-\Lambda(\lambda)\big\},$$

where the logarithmic moment generating function Λ is given by

$$\Lambda(\lambda)=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}_n[\mathrm{e}^{\lambda nX_n(t)}]=b(\mathrm{e}^{\lambda}-1)+d(\mathrm{e}^{-\lambda}-1). \tag{4.10}$$

We can rewrite the rate function as $I(at)=tL(a)$ with

$$L(a)=\sup_{\lambda\in\mathbb{R}}\big\{a\lambda-b(\mathrm{e}^{\lambda}-1)-d(\mathrm{e}^{-\lambda}-1)\big\}.$$

Using that the increments of the Poisson process are independent over disjoint time intervals, one can easily show a large deviation principle. We do not do that here but instead employ our roadmap in proving the large deviation principle for the sequence $(X_n)_{n\geq 1}$ of Markov processes. We obtain easily

$$(\mathcal{H}f)(x):=\lim_{n\to\infty}\frac{1}{n}\mathrm{e}^{-nf(x)}\big(A_n\mathrm{e}^{nf(x)}\big)(x)=b(\mathrm{e}^{f'(x)}-1)+d(\mathrm{e}^{-f'(x)}-1).$$

Furthermore, we easily see that the Lagrangian L in (4.10) is the Legendre transform of the following formal Hamiltonian (coincides in this case with the logarithmic generating function)

$$H(\lambda)=b(\mathrm{e}^{\lambda}-1)+d(\mathrm{e}^{-\lambda}-1). \tag{4.11}$$

Thus,

$$H(\lambda)=(\mathcal{H}f_\lambda)(x),\qquad\text{for }f_\lambda(x)=\lambda x,$$

and by the convexity of H,

$$(\mathcal{H}f)(x) = H(f'(x)) = \sup_{a\in\mathbb{R}} \big\{af'(x) - L(a)\big\}.$$

Then

$$\mathbb{P}_n\big((X_n(t))_{t\in[0,T]} \approx (\gamma_t)_{t\in[0,T]}\big) \approx \exp\Big\{-n\int_0^T L(\gamma_t,\dot\gamma_t)\,\mathrm{d}t\Big\},$$

where the Lagrangian L is the function $(x,\lambda)\mapsto L(x,\lambda)$ with $x=\gamma_t$ and $\lambda=\dot\gamma_t$. Thus in our case (only dependence on the time derivative) we obtain

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n\big(X_n(t))_{t\in[0,T]} \approx (\gamma_t)_{t\in[0,T]}\big) = \int_0^T L(\dot\gamma_t)\,\mathrm{d}t.$$

The latter LDP can be obtained also directly using Cramér's theorem and the independence of the increments for disjoint time intervals. However, this example is to show how our roadmap may lead to LDP for sequences of Markov processes. We omitted all technical details which can easily be filled in.

◇

In the next example we study spin-flips which are frequently studied in statistical mechanics. We outline an example from the recent paper by van Enter et al. [vEFHR10]

Example 4.5. In this example we shall be concerned with large deviation behaviour of the trajectories of the mean magnetisation for spin-flip dynamics. We consider a simple case in dimension one, namely Ising spins on the one-dimensional torus $\mathbb{T}_N$ represented by the set $\{1,\dots,N\}$ subject to a rate-1 independent spin-flip dynamics. We will write $\mathbb{P}_N$ for the law of this process. We are interested in the trajectory of the mean magnetisation, i.e., $t\mapsto m_N(t) = \frac{1}{N}\sum_{=1}^N \sigma_i(t)$, where $\sigma_i(t)$ is the spin at site i at time t. A spin-flip from $+1$ to -1 (or from -1 to $+1$) corresponds to a jump of size $-2N^{-1}$ (or $+2N^{-1}$) in the magnetisation. Henceforth the generator A_N of the process $(m_N(t))_{t\ge 0}$ is given by

$$(A_Nf)(m) = \frac{1+m}{2}N\big[f(m-2N^{-1}) - f(m)\big] + \frac{1-m}{2}N\big[f(m+2N^{-1}) - f(m)\big]$$

for any $m\in\{-1,-1+2N^{-1},\dots,1-2N^{-1},1\}$. If $\lim_{N\to\infty} m_N = m$ and $f\in\mathcal{C}^1(\mathbb{R})$ with bounded derivative, then

$$\lim_{N\to\infty}(A_Nf)(m_N) = (Af)(m) \qquad \text{with } (Af)(m) = -2mf'(m).$$

This is the generator of the deterministic flow $m(t) = m(0)\mathrm{e}^{-2t}$, solving the equation $\dot m(t) = -2m(t)$. We shall prove that the magnetisation satisfies the LDP with rate function given as an integral over a finite time horizon $T\in(0,\infty)$ of the Lagrangian, that is, for every trajectory $\gamma = (\gamma_t)_{t\in[0,T]}$, we shall get that

$$\mathbb{P}_N\big((m_N(t))_{t\in[0,T]} \approx (\gamma_t)_{t\in[0,T]}\big) \approx \exp\Big\{-N\int_0^T L(\gamma_t,\dot\gamma_t)\,\mathrm{d}t\Big\},$$

where the Lagrangian can be computed using our roadmap following the scheme of Feng and Kurtz [FK06]. We obtain

$$(\mathcal{H}f)(m) := \lim_{N\to\infty}(\mathcal{H}_Nf)(m_N),$$

where

$$(\mathcal{H}_N f)(m) = \lim_{N\to\infty} \frac{1}{N} \mathrm{e}^{-Nf(m_n)} A_N(\mathrm{e}^{Nf})(m_N)$$

and $\lim_{N\to\infty} m_N = m$. This gives

$$(\mathcal{H}f)(m) = \frac{m+1}{2}\big[\mathrm{e}^{-2f'(m)} - 1\big] + \frac{1-m}{2}\big[\mathrm{e}^{2f'(m)} - 1\big].$$

We can see easily that this can be written as

$$(\mathcal{H}f)(m) = H(m, f'(m))$$

with the Hamiltonian

$$H(m,p) = \frac{m+1}{2}\big(\mathrm{e}^{-2p} - 1\big) + \frac{1-m}{2}\big(\mathrm{e}^{2p} - 1\big).$$

The convexity of $p \mapsto H(m,p)$ leads to the variational representation

$$H(m,p) = \sup_{q\in\mathbb{R}} \big\{pq - L(m,q)\big\},$$

where some computation gives us the Lagrangian

$$\begin{aligned} L(m,q) &= \sup_{p\in\mathbb{R}} \big\{pq - H(m,p)\big\} \\ &= \frac{q}{2}\log\Big(\frac{q+\sqrt{q^2+4(1-m^2)}}{2(1+m)}\Big) - \frac{1}{2}\sqrt{q^2+4(1-m^2)} + 1. \end{aligned}$$

We thus obtain the LDP for $(m_N(t))_{t\in[0,T]}$ for the magnetisation process with finite time horizon $T \in (0,\infty)$ and rate function given by the Lagrangian or reward function $L(m,q)$ where $m = \gamma_t$ and $q = \dot{\gamma}_t$, and

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}_N\big((m_N(t))_{t\in[0,T]} \approx (\gamma_t)_{t\in[0,T]}\big) = \exp\Big\{ - N \int_0^T L(\gamma_t, \dot{\gamma}_t)\,\mathrm{d}t\Big\}.$$

$\diamond$

Remark 4.6. *We conclude our list of examples with the remark that in* [vEFHR10] *the scheme of Feng and Kurtz (see our discussion above and* [FK06]*) is applied to a more general spin-flip dynamics. First, the spin-flip is now considered on the d-dimensional torus $\mathbb{T}_N^d$ and one studies Glauber dynamics of Ising spins, i.e., on the configuration space $\Omega_N = \{-1,+1\}^{\mathbb{T}_N^d}$. The second extension is that the independence of the spin-flips is dropped and the dynamics is defined via the generator A_n acting on functions $f\colon \Omega_N \to \mathbb{R}$ as*

$$(A_N f)(\sigma) = \sum_{i\in\mathbb{T}_N^d} c_i(\sigma)[f(\sigma^i) - f(\sigma)],$$

where σ^i denotes the configuration obtained from the configuration σ by flipping the spin at site i, and where $c_i(\sigma)$ are given rates. The period empirical measure is a random probability measure on Ω_N defined as the random variable

$$L_N\colon \Omega_N \to \mathcal{M}_1(\Omega_N), \quad \sigma \mapsto L_N(\sigma) = \frac{1}{|\mathbb{T}_N^d|} \sum_{i\in\mathbb{T}_N^d} \delta_{\theta_i\sigma},$$

where θ_i is the shift operator by $i \in \mathbb{T}_N^d$. Having the dynamics defined via the generator A_N one is interested in large deviations of the trajectory of the empirical measures. This LDP can *be derived following the roadmap of the scheme of Feng and Kurtz, and this has been done in* [vEFHR10]*. This shows how effective the scheme discussed in these notes can be.*

Acknowledgment. The author is very grateful to Aernout van Enter, Roberto Fernández, Frank den Hollander, and Frank Redig for the kind invitation and their support to lecture this mini-course at EU-Random during the workshop '*Dynamical Gibs-non-Gibbs transitions*' in Eindhoven, December 2011. The authors thanks Felix Otto and his group at the Max Planck institute for kind hospitality, support and their interest.

References

[Cra38] H. Cramér, Sur un nouveau théoreme-limite de la theorie des probabilitiés, Acta. Sci. et Ind. **736**, 5-23 (1938).

[CL71] M.G. Crandall and T.M. Liggett, Generation of Semigroups of Nonlinear Transformations on General Banach Spaces, Amer. J. Math. **93**, 265-298 (1971).

[vEFHR10] A.C.D. van Enter, R. Fernández, F. den Hollander and F> Redig, A large deviation view on dynamical Gibbs-non-Gibbs transitions, Mosc. Math. J., Volume **10**, Number 4, 687711(2010).

[DeA97] A. de Acosta, Exponential tightness and Projective Systems in Large Deviation Theory, Festschrift for Lucien Le Cam, Springer, New York, 143-156 (1997).

[DZ98] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd edition, Springer, Berlin (1998).

[DV76] M.D. Donsker and S.R.S. Varadhan, Asymptotic evaluation of Markov process expectations for large times, I,II,III, Comm. Pure Appl. Math. **27**, 1-47 (1975), **28**, 279-301 (1975), **29**, 389-461 (1976).

[FK06] J. Feng and T.G. Kurtz, *Large Deviations for Stochastic Processes*, American Mathematical Society (2006).

[Flem85] W.H. Fleming, A stochastic control approach to some large deviation problems, *Recent mathematical methods in dynamics programming (Rome 1984)*, Lecture Notes in Math. **1119**, Springer-Verlag Berlin-New York, 52-66 (1985).

[FW98] M.I. Freidlin and A.D. Wentzell, *Random perturbations of dynamical systems*, Second Edition, Springer-Verlag, New York (1998).

[Kur73] T.G. Kurtz, Convergence of sequences of semigroups of nonlinear operators with an application to gas kinetics, Trans. Amer. Math. Soc. **186**, 259-272 (1973).

[Miy91] I. Miyadera, *Nonlinear Semigroups*, Translations of MAthematical Monographs **108**, AMS (1991).

[OV95] G.L. O'Brien and W. Vervaat, Compactness in the theory of large deviations, Stochastic Process. Appl. **57**, 1-10 (1995).

[Puh94] A. Puhalskii, The method of stochastic exponentials for large deviations, Stochastic Process. Appl. **54**, 45-70 (1994).