Flux large deviations of weakly interacting jump processes via well-posedness of an associated Hamilton-Jacobi equation

Kraaij, Richard C.

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

Flux large deviations of weakly interacting jump processes via well-posedness of an associated Hamilton-Jacobi equation

Richard C. Kraaij[*]

September 23, 2020

Abstract

We establish uniqueness for a class of first-order Hamilton-Jacobi equations with Hamiltonians that arise from the large deviations of the empirical measure and empirical flux pair of weakly interacting Markov jump processes. As a corollary we obtain such a large deviation principle in the context of weakly interacting processes with time-periodic rates in which the period-length converges to 0.

Keywords: Hamilton-Jacobi equation; Large deviations; weakly interacting jump processes; empirical measure; empirical flux

MSC2010 classification: 49L25; 60F10; 60J75

1 Introduction

Systems of interacting Markov jump processes appear in various contexts, e.g. in statistical physics, kinetic theory, queuing systems and communication networks. A first natural goal is to understand the limiting behaviour of appropriate observables as the number of components goes to infinity. An extension of this question is that of a large deviation principle, see e.g. [6,9,12,14,18,23,26,27] and references therein. We will consider the context of time-inhomogeneous interacting jump processes

$$(X_{n,1}(t),\dots,X_{n,n}(t))_{t\geqslant 0} \tag{1.1}$$

on a finite-state space $\{1,\dots,q\}$. We assume that the processes are fully exchangeable, jump one-by-one, and interact weakly: their jump rates depend on their empirical measure $\mu_n(t) := n^{-1}\sum_{i=1}^n \delta_{X_{n,i}(t)}$. We will study the large deviation behaviour of the trajectory of empirical measures $t \mapsto \mu_n(t)$ as n gets large.
We assume that that the interaction has the following properties.

(1) The interaction is *weak*: each of the n process in (1.1) jumps over the bond $(a,b) \in \Gamma := \big\{(a,b) \in \{1,\dots,q\}^2 \,\big|\, a \neq b\big\}$ with rate $r_n(t,a,b,\mu_n)$,

(2) The jump rates are *time-periodic with decreasing period size*. That is, there is a constant $T_0 > 0$ and a sequence of constants $\gamma_n \to \infty$ such that

$$r_n(t + \gamma_n^{-1}T_0, a, b, \mu_n) = r_n(t, a, b, \mu_n)$$

for all $t \geqslant 0$, μ_n and $(a,b) \in \Gamma$.

---

[*]Delft Institute of Applied Mathematics, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands. *E-mail address*: r.c.kraaij@tudelft.nl

(3) The rates are *converging*: there is a kernel $r(t, a, b, \mu)$ such that

$$\lim_{n \to \infty} \sup_{t \leqslant T_0} \sup_{(a,b) \in \Gamma, \mu \in \mathcal{P}_n(\{1,\ldots,n\})} \left| r_n(\gamma_n^{-1} t, a, b, \mu) - r(t, a, b, \mu) \right| = 0,$$

where $\mathcal{P}_n(\{1, \ldots, q\})$ is the set of measures of the form $n^{-1} \sum_{i=1}^{n} \delta_{x_i}$ for $x_1, \ldots, x_n \in \{1, \ldots, q\}$.

(4) The rates are *Lipschitz*: there is some $C > 0$ such that

$$\sup_{n} \sup_{t \leqslant T_0} \sup_{\mu, \nu \in \mathcal{P}_n(\{1,\ldots,q\})} \sum_{(a,b) \in \Gamma} |r_n(t, a, b, \mu) - r_n(t, a, b, \nu)| \leqslant C \, |\mu - \nu| \, .$$

The periodicity on a time-interval that is decreasing in length has the effect that the interacting particle system undergoes an effective averaging effect and this will be seen in the final large deviation result. Note that the γ_n do not model a speed-up of the process, but rather model an external factor which lives on a faster time-scale.

Recent works on path-space large deviations by [24, 26] and works in mathematical physics [3, 7], or [2, 4, 5] on the study of hydrodynamic limits or long-time (Donsker-Varadhan) large deviations, have shown that studying the process of the empirical measures together with the empirical fluxes simplifies proofs and gives greater insight in the large deviation principles. We will follow these insights and study the empirical measures of the processes in (1.1) in combination with their empirical fluxes.

This paper can thus be seen as a natural continuation of [14, 18, 24]. The papers [14, 24] are more general in the sense that they consider contexts where multiple processes can jump at the same time. If we restrict their results to the context where only a single process jumps we extend the three papers by including a time averaging effect. In addition, we extend [18] by including fluxes, [14] by allowing more general rates and include fluxes, and [24] by including more general rates. Finally, we establish the large deviation principle by using a non-standard technique using the machinery of Hamilton-Jacobi equations introduced by [16]. We give a more elaborate comparison after the statement after the introduction of our main results.

Consider the processes (1.1) and denote by $W_{n,i}(t)$ the number of jumps made by $X_{n,i}(t)$ up to time t across each directed edge $(a, b) \in \Gamma$. We will establish the large deviation principle for the trajectory of the empirical measure-flux pair

$$t \mapsto Z_n(t) := \left(\frac{1}{n} \sum_{i=1}^{n} \delta_{X_{n,i}(t)}, \frac{1}{n} \sum_{i=1}^{n} W_{n,i}(t) \right) \tag{1.2}$$

on the Skorokhod space of trajectories in $E := \mathcal{P}(\{1, \ldots, q\}) \times (\mathbb{R}^+)^{\Gamma}$. The rate function is given in Lagrangian form:

$$I(\mu, w) := \begin{cases} I_0(\mu(0), 0) + \int_0^\infty \mathcal{L}((\mu(s), w(s), (\dot{\mu}, \dot{w}(s))) ds, & \text{if } (\mu, w) \in \mathcal{AC}, \\ \infty & \text{otherwise,} \end{cases}$$

where $\mathcal{AC}$ is an appropriate space of absolutely continuous trajectories in E. The Lagrangian is given as a sum over relative entropies $S(z | v) := z \log \frac{z}{v} - z + v$:

$$\mathcal{L}((\mu(s), w(s), (\dot{\mu}, \dot{w}(s)))$$
$$:= \begin{cases} \sum_{(a,b) \in \Gamma} S\left(\dot{w}_{(a,b)} \,|\, \mu(a) \bar{r}(a, b, \mu) \right) & \text{if } \forall a : \dot{\mu}(a) = \sum_b \dot{w}_{(a,b)} - \dot{w}_{(b,a)}, \\ \infty & \text{otherwise.} \end{cases} \tag{1.3}$$

The kernel $\bar{r}$ denotes the outcome of the averaging principle from (2) and (3):

$$\bar{r}(a, b, \mu) := \frac{1}{T_0} \int_0^{T_0} r(t, a, b, \mu) dt.$$

The key step in the proof of the large deviation result in this paper, and in addition our second main result, is the establishment of the comparison principle (implying uniqueness of viscosity solutions) to a collection of associated Hamilton-Jacobi equations $f - \lambda Hf = h$, for $\lambda > 0$ and $h \in C_b(E)$. The operator H in this equation is given by $Hf(\mu, w) = \mathcal{H}((\mu, w), \nabla f(\mu, w))$ where $\mathcal{H}$ is the Legendre transform of $\mathcal{L}$ from (1.3). Its explicit representation is given by

$$\mathcal{H}((\mu, w), p) = \sum_{(a,b) \in \Gamma} \mu(a) \bar{r}(a, b, \mu) \left[\exp\left\{ p_b - p_a + p_{(a,b)} \right\} - 1 \right],$$

$$(\mu, w) \in \mathcal{P}(\{1, \ldots, q\}) \times (\mathbb{R}^+)^\Gamma, p \in \mathbb{R}^q \times \mathbb{R}^\Gamma. \quad (1.4)$$

Due to the terms of the type $\mu(a) r(a, b, \mu) \left[e^{p_b - p_a + p_{(a,b)}} - 1 \right]$ the Hamiltonian is neither Lipschitz nor uniformly coercive in p. This implies that our Hamilton-Jacobi equation can not be treated using 'standard' methods for first-order equations, see [1,11,17] and references therein. Instead, our method improves upon the method of [18] which was designed for the Hamiltonian of weakly interacting jump processes without taking into account the fluxes. The novelty of the proof of the comparison principle, compared to [18], is based on a novel 'two stage' penalization procedure, which potentially can be used to treat other types of 'non-standard' first-order Hamilton-Jacobi equations, see Sections 5.1 and 5.2.

We stress that the verification of the comparison principle is of interest beyond the large deviation statement that is proven in this paper. First of all, the comparison principle can find other applications in the field of control theory or mean-field games. Secondly, an extension of the comparison principle in this paper by the bootstrap principle introduced in [21] leads to comparison principles for more elaborate Hamilton-Jacobi (-Bellman) equations. In turn these boosted comparison principles can be used for new large deviation principles, as carried out in the forthcoming work [22] in the context of more general slow-fast systems.

We next compare our large deviation result to results in the literature.

Large deviations for weakly interacting jump processes have been studied in the past, see e.g. [9, 12, 23] in contexts with spatial structure or random fields. The methods of proof were based on direct evaluation of the asymptotics or tilting arguments based on Sanov's Theorem, Varadhan's lemma and the contraction principle.

More recent papers in the context of non-spatial processes have focused on different methods of proof [14, 18], or have included fluxes [24, 26]. Of these four papers, two [18, 26] still focus on processes with transitions of the type where one particle moves its state, whereas other two papers [14, 24] allow for more general transitions, e.g. allowing more particles to change their state at a single time or consider mass-action kinetics.

As a first remark, this paper includes an averaging effect for path-space large deviations. If we restrict ourselves to the time-homogenous case, we can compare our large deviation principle to those of [14, 18, 24, 26]. We focus our comparison to [14, 24], as this paper supersedes [18] by the inclusion of fluxes, and [24] supersedes [26] by generalizing the single-jump setting as well as letting go of the independence assumption.

In [14], the authors work in the context without fluxes. The proof of the large deviation principle is based on a variational expression for the Poisson random measure, of which it is established that the expression converges as $n \to \infty$. An approximation argument based on ergodicity is used to reduce the proof of the lower

bound to trajectories that lie in the interior of the simplex of probability measures. It is also assumed that the law of large numbers limit pushes the empirical measure into the interior of the simplex. If restricted to the context of single-jumps only, this paper covers more cases as assumptions of the type resembling these two final conditions of [14] are absent from this paper.

In [24], following [26], the empirical measure is combined with the empirical fluxes. The inclusion of the fluxes allows for a clear and direct change of measure argument leading in a straightforward way to the Lagrangrian in terms of a sum over appropriate relative entropies. In the context of single jumps, our result extends that of [24]. Two key assumptions in [24] are Assumption 2.2 (v) and (vi). The first of these two conditions is naturally reflected by the assumption that the limit of the jump rates form a *proper kernel* as in Assumption 3.1 (c). It should be noted that (v) of [24] is more restrictive and excludes for example Glauber type interactions like in Example 3.6. In addition, this paper does not assume an analogue of [24, Assumption 2.2 (vi)].

We thus see that the proof via the comparison principle in the context of systems with single jumps yields the most general results, and with additional work would allow for a generalization to the context where the rates are non-Lipschitz as in [18]. The proof of the comparison principle, however, uses a technique that is very much geared towards Hamiltonians of the type (1.4) and can not directly be adapted to the more general setting of processes with multiple simultaneous jumps of [14, 24]. More remarks on these restrictions are given in Section 5.2.

The paper is organized as follows. We start in Section 2 with basic definitions, including those of viscosity solutions of Hamilton-Jacobi equations, the comparison principle, the martingale problem, and the large deviation principle. In Section 3 we state our main results: the comparison principle and the large deviation principle. In Section 4 we give the key results that reduce the proof of the large deviation principle to the comparison principle. We then prove the comparison principle in Section 5 and we follow with the verification of the remaining assumptions for the results of Section 4 in Section 6. In Appendix A, we collect some results for the literature that are essential for the proof of the comparison principle. Their inclusion makes the paper as self-contained as possible.

2 Preliminaries

Let E be a Polish space. We denote by $\mathcal{P}(E)$ the space of Borel probability measures on E. By $\mathcal{P}_n(E)$ we denote the subset of measures that have the form $n^{-1} \sum_{i=1}^{n} \delta_{x_i}$ for some collection $\{x_i\}_{i=1}^{n} \subseteq E$.

We denote by $D_E(\mathbb{R}^+)$ the space of paths $\gamma : \mathbb{R}^+ \to E$ that are right continuous and have left limits. We endow $D_E(\mathbb{R}^+)$ with the Skorokhod topology, cf. [15, Section 3.5]. An important property is that under this topology $D_E(\mathbb{R}^+)$ is Polish if E is Polish.

We denote by $C(E)$ and $C_b(E)$ the spaces of continuous and bounded continuous functions on E. For $d \in \mathbb{N} \setminus \{0\}$ and $k \in \mathbb{N}$ let $C_b^k(\mathbb{R}^d)$ be the space of functions that have k continuous and bounded derivatives. By $C_b^\infty(\mathbb{R}^d)$ we denote the space of functions with bounded continuous derivatives of all orders.

Now consider a subset $E \subseteq \mathbb{R}^d$ that is a Polish space and that is contained in the $\mathbb{R}^d$ closure of its $\mathbb{R}^d$ interior. We denote by $C_b^k(E)$, $C_b^\infty(E)$ the spaces of functions that have an extension to $C_b^k(\mathbb{R}^d)$ and $C_b^\infty(\mathbb{R}^d)$ respectively. Finally, denote by $C_c^k(E)$ and $C_c^\infty(E)$ the subsets that have compact support in E. Note that the derivative of a continuously differentiable function on E is determined by the values of the function on E by our assumption on E.

Finally, we introduce the space $\mathcal{AC}(E)$ of absolutely continuous paths in E. A curve $\gamma : [0, T] \to E$ is absolutely continuous if there exists a function $g \in L^1([0, T], \mathbb{R}^d)$ such that for $t \in [0, T]$ we have $\gamma(t) = \gamma(0) + \int_0^t g(s)ds$. We write $g = \dot{\gamma}$.

A curve $\gamma : \mathbb{R}^+ \to E$ is absolutely continuous, i.e. $\gamma \in \mathcal{AC}(E)$, if the restriction to $[0, T]$ is absolutely continuous for every $T \geqslant 0$.

2.1 Large deviations

Let $\mathcal{X}$ be a Polish space. Later we will use both $\mathcal{X} = D_E(\mathbb{R}^+)$ and $\mathcal{X} = E$.

Definition 2.1. Let $\{X_n\}_{n \geqslant 1}$ be a sequence of random variables on $\mathcal{X}$. Furthermore, consider a function $I : \mathcal{X} \to [0, \infty]$. We say that

- the function I is a *good rate-function* if the set $\{x \,|\, I(x) \leqslant c\}$ is compact for every $c \geqslant 0$;

- the sequence $\{X_n\}_{n \geqslant 1}$ satisfies the *large deviation principle* and good rate-function I if for every closed set $A \subseteq \mathcal{X}$, we have

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}[X_n \in A] \leqslant - \inf_{x \in A} I(x),$$

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

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}[X_n \in U] \geqslant - \inf_{x \in U} I(x).$$

2.2 The martingale problem

One effective way of defining a Markov process on E is by using its infinitesimal generator, see e.g. [15]. One of the instances of this idea is that of solving the martingale problem.

We introduce the martingale problem for time-inhomogeneous processes. Note that this is a straightforward extension from the time-homogeneous case via the inclusion of time in the state-space, see for example Section 4.7.A in [15] or Proposition II.5.7 in [25].

Let $A : \mathcal{D}(A) \subseteq C_b(E) \to C_b(\mathbb{R}^+ \times E)$ be a linear operator. For each time $s \in \mathbb{R}^+$, we denote by $A[s] : \mathcal{D}(A) \subseteq C_b(E) \to C_b(E)$ the linear operator obtained by fixing s. $A[s]$ can be interpreted as the generator at time s. In addition, we construct out of the operators $A[s]$ an operator $\vec{A}$ on $C_b(\mathbb{R}^+ \times E)$:

- $\mathcal{D}(\vec{A})$ satisfies

  $$\mathcal{D}(\vec{A}) \subseteq \left\{ f \in C_b(\mathbb{R}^+ \times E) \,\middle|\, \forall x \in E \colon f(\cdot, x) \in C_b^1(\mathbb{R}^+), \forall s \in \mathbb{R}^+ \colon f(s, \cdot) \in \mathcal{D}(A) \right\},$$

- for $f \in \mathcal{D}(\vec{A})$ we have $\vec{A}f(s, x) = \partial_s f(s, x) + (A[s]f(s, \cdot))(x)$.

Definition 2.2. Let $\mu \in \mathcal{P}(E)$. We say that the process $t \mapsto X(t)$ on $D_E(\mathbb{R}^+)$ solves *the (time-inhomogeneous) martingale problem* for $(\vec{A}, \mu)$ if for all $f \in \mathcal{D}(\vec{A})$ the process

$$M_f(t) := f(t, X(t)) - f(0, X(0)) - \int_0^t \vec{A}f(s, X(s))ds$$

$$= f(t, X(t)) - f(0, X(0)) - \int_0^t \partial_s f(s, X(s)) + (A[s]f(s, \cdot))(X(s))ds$$

is a martingale and if the projection of $\mathbb{P}$ on the time 0 coordinate equals μ.

By slight abuse of notation, we will also say that the measure of the process $t \mapsto (t, X(t))$ solves the martingale for $\vec{A}$.

2.3 Viscosity solutions to Hamilton-Jacobi equations

Definition 2.3 (Viscosity solutions)**.** Let $H : \mathcal{D}(H) \subseteq C_b(E) \to C_b(E)$, $\lambda > 0$ and $h \in C_b(E)$. Consider the Hamilton-Jacobi equation

$$f - \lambda H f = h. \tag{2.1}$$

We say that u is a *(viscosity) subsolution* of equation (2.1) if u is bounded, upper semi-continuous and if, for every $f \in \mathcal{D}(H)$ there exists a sequence $x_n \in E$ such that

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

$$\lim_{n\uparrow\infty} u(x_n) - \lambda H f(x_n) - h(x_n) \leqslant 0.$$

We say that v is a *(viscosity) supersolution* of equation (2.1) if v is bounded, lower semi-continuous and if, for every $f \in \mathcal{D}(H)$ there exists a sequence $x_n \in E$ such that

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

$$\lim_{n\uparrow\infty} v(x_n) - \lambda H f(x_n) - h(x_n) \geqslant 0.$$

We say that u is a *(viscosity) solution* of equation (2.1) if it is both a subsolution and a supersolution to (2.1).

We say that (2.1) satisfies the *comparison principle* if for every subsolution u and supersolution v to (2.1), we have $u \leqslant v$.

Remark 2.4. The comparison principle implies uniqueness of viscosity solutions. Suppose that u and v are both viscosity solutions, then the comparison principle yields that $u \leqslant v$ and $v \leqslant u$, implying that $u = v$.

Remark 2.5. Consider the definition of subsolutions. Suppose that the testfunction $f \in \mathcal{D}(H)$ has compact sublevel sets, then instead of working with a sequence x_n, we can pick a x_0 such that

$$u(x_0) - f(x_0) = \sup_x u(x) - f(x),$$

$$u(x_0) - \lambda H f(x_0) - h(x_0) \leqslant 0.$$

A similar simplification holds in the case of supersolutions.

3 Main results

In this section we give our two main results: the large deviation principle and the comparison principle. We give a short recap of some of the definitions informally given in the introduction. Let $\{1, \ldots, q\}$, $q \in \mathbb{N} \setminus \{0, 1\}$ be some finite set. Write $\Gamma = \left\{ (a, b) \in \{1, \ldots, q\}^2 \,\middle|\, a \neq b \right\}$ for the directed edge-set in $\{1, \ldots, q\}$. Let $E = \mathcal{P}(\{1, \ldots, q\}) \times (\mathbb{R}^+)^\Gamma$ be the Polish space of probability measures on $\{1, \ldots, q\}$ combined with a space in which we can keep track of the fluxes over the directed bonds in Γ.

For n points $\vec{x} = (x_1, \ldots, x_n) \in \{1, \ldots, q\}$ denote by $\mu_n[\vec{x}]$ the empirical measure $\mu_n[\vec{x}] = n^{-1} \sum_{i=1}^n \delta_{x_i}$.

We consider a collection of weakly-interacting jump processes

$$\vec{X}_n(t) = (X_{n,1}(t), \ldots, X_{n,n}(t))_{t \geqslant 0} \tag{3.1}$$

on the space $\{1, \ldots, q\}$ and $\mu_n(t) := \mu_n[\vec{X}_n(t)]$ the empirical measure of the process at time t. For any given n, we will assume that each of the n processes, if at state a, jumps to state b with rate

$$r_n(t, a, b, \mu_n(t)),$$

i.e. the processes interact weakly through their empirical measure.

We are interested in the large deviation behaviour of the trajectory $t \mapsto \mu_n(t)$ on the space $\mathcal{P}(\{1, \ldots, q\})$. Following [24, 26], it turns out that a description of the large deviation principle simplifies if we take into account also the fluxes across the bonds in Γ. Therefore, denote by

$$t \mapsto W_{n,i}(t) \in \mathbb{N}^\Gamma$$

the process that satisfies

$$W_{n,i}(t)(a, b) := \#\{s \leqslant t \mid (X_{n,i}(s-), X_{n,i}(s)) = (a, b)\}.$$

Our first result establishes the large deviation principle for the pair of processes

$$t \mapsto Z_n(t) := \left(\frac{1}{n} \sum_{i=1}^n \delta_{X_{n,i}(t)}, \frac{1}{n} \sum_{i=1}^n W_{n,i}(t) \right), \tag{3.2}$$

on the set $D_E(\mathbb{R}^+)$.

In Section 3.1, we state our large deviation principle. In Section 3.2 we give an example in the context of Glauber dynamics. We end our section of main results in Section 3.3 with the uniqueness result for the associated Hamilton-Jacobi equations.

3.1 Flux large deviations for time-periodic jump rates

Assumption 3.1. The rates $r_n(t, a, b, \mu)$ at which each of the processes in (3.1) jumps over the bond $(a, b) \in \Gamma$ at time t while the empirical measure is given by μ satisfies the following properties.

(a) The jump rates are *time-periodic with decreasing period size*. That is, there is a constant $T_0 > 0$ and a sequence of constants $\gamma_n \to \infty$ such that $r_n(t + \gamma_n^{-1} T_0, a, b, \mu_n) = r_n(t, a, b, \mu_n)$ for all $t \geqslant 0$, μ_n and $(a, b) \in \Gamma$.

(b) The rates are *converging*: there is a kernel $r(t, a, b, \mu)$ such that

$$\lim_{n \to \infty} \sup_{t \leqslant T_0} \sup_{(a,b) \in \Gamma, \mu \in \mathcal{P}_n(\{1,\ldots,n\})} \left| r_n(\gamma_n^{-1} t, a, b, \mu) - r(t, a, b, \mu) \right| = 0,$$

where $\mathcal{P}_n(\{1, \ldots, q\})$ is the set of measures of the form $n^{-1} \sum_{i=1}^n \delta_{x_i}$ for $x_1, \ldots, x_n \in \{1, \ldots, q\}$.

(c) *The averaged kernel is 'proper'.* Denote by $\bar{r}$ the kernel

$$\bar{r}(a, b, \mu) := \frac{1}{T_0} \int_0^{T_0} r(t, a, b, \mu) dt.$$

For each $(a, b) \in \Gamma$, we have either (i) or (ii):

   (i) $\bar{r}(a, b, \mu) = 0$ for all μ,
   (ii) $\inf_\mu \bar{r}(a, b, \mu) > 0$.

(d) The rates are *Lipschitz*: there is some $C > 0$ such that

$$\sup_n \sup_{t \leqslant T_0} \sup_{\mu, \nu \in \mathcal{P}_n(\{1,\ldots,q\})} \sum_{(a,b) \in \Gamma} |r_n(t, a, b, \mu) - r_n(t, a, b, \nu)| \leqslant C |\mu - \nu|.$$

Remark 3.2. Note that the time-periodic context includes the time-homogeneous contexts. Namely, if the rates do not depend on t than any choice of $T_0 > 0$ and $\gamma_n \to \infty$ will induce time-periodicity of the type above.

Our first main result is the large deviation principle for the pair of processes (3.2) in the context of time periodic rates.

Theorem 3.3. *For each n consider the process of state-flux pairs*

$$((X_{n,1}(t), W_{n,1}(t)), \dots, (X_{n,n}(t), W_{n,n}(t)))_{t \geqslant 0},$$

where the jump rates r_n satisfy Assumption 3.1.
Consider the processes $t \mapsto Z_n(t) := \left(\frac{1}{n}\sum_{i=1}^{n}\delta_{X_{n,i}(t)}, \frac{1}{n}\sum_{i=1}^{n} W_{n,i}(t)\right)$. Suppose that $Z_n(0)$ satisfies a large deviation principle on $E = \mathcal{P}(\{1,\dots,q\}) \times (\mathbb{R}^+)^\Gamma$ with good rate function I_0, then $\{Z_n\}_{n \geqslant 1}$ satisfies the large deviation principle on $D_E(\mathbb{R}^+)$ with good rate function I given by

$$I(\mu, w) = \begin{cases} I_0(\mu(0), w(0)) + \int_0^\infty \mathcal{L}((\mu(s), w(s)), (\dot{\mu}(s), \dot{w}(s)))ds & \textit{if } (\mu, w) \in \mathcal{AC}(E), \\ \infty & \textit{otherwise,} \end{cases}$$

where $\mathcal{L} : E \times \mathbb{R}^{q+|\Gamma|} \to \mathbb{R}^+$ is given by

$$\mathcal{L}\left((\mu, w), (\dot{\mu}, \dot{w})\right) = \begin{cases} \sum_{(a,b) \in \Gamma} S(\dot{w}_{(a,b)} \,|\, \mu(a)\bar{r}(a, b, \mu)) & \textit{if } \forall a : \dot{\mu}_a = \sum_b \dot{w}_{(b,a)} - \dot{w}_{(a,b)}, \\ \infty & \textit{otherwise,} \end{cases}$$

with

$$S(z \,|\, v) := \begin{cases} v & \textit{if } z = 0, \\ z \log \frac{z}{v} - z + v & \textit{if } z \neq 0, v \neq 0, \\ \infty & \textit{if } z \neq 0, v = 0. \end{cases}$$

Note that the Lagrangian is given in terms of the entropic cost of changing the flux across each bond. Indeed, S is the relative entropy corresponding to a tilt of the intensity of a Poisson measure.

Remark 3.4. Theorem 3.3 gives as a corollary the large deviation principle for the trajectory of the empirical measures only. This recovers e.g. the result of [18] but now in contracted form. The rate function J is given by

$$J(\mu) = I_0(\mu(0)) + \inf \left\{ \int_0^\infty \sum_{(a,b) \in \Gamma} S(\dot{w}_{(a,b)}(s) \,|\, \mu(a)\bar{r}(a, b, \mu(s)))ds \,\middle|\, w \in \mathcal{AC}(E), \forall a : \dot{\mu}(a) = \sum_b \dot{w}_{(b,a)} - \dot{w}_{(a,b)} \right\}$$

if μ is absolutely continuous and $J(\mu) = \infty$ otherwise.

As a second remark, a comment on the Lipschitz property in Assumption 3.1 (d).

Remark 3.5. The Lipschitz assumption can be dropped in the context that the rates are time-homogeneous.
The uniqueness of solutions of the Hamilton-Jacobi equation in Theorem 3.8 below does not depend on this statement. Thus, a large deviation principle for the time-homogeneous case without the Lipschitz condition can be obtained by dropping the $F_{f,n}$ or h_n term in Proposition 6.1 and Lemma 6.2. Alternatively, one can adapt the operator convergence proof in [18].

Even in the context of time-inhomogeneous rates, the Lipschitz property is not essential. In a work in progress, [22], this is explored in the more general context of Markov processes with two time-scales. A proof of convergence of operators and how to deal with the viscosity solutions for the limiting operators in this general two-scale context is technically more challenging. We refrain from carrying this out in this context and keep a simpler and independent proof in this paper.

3.2 Example: the Curie-Weiss-Potts model

Next, we establish the path-space large deviation principle in the context of the dynamic Curie-Weiss-Potts model.

Example 3.6 (Time-dependent potential functions). Let $V : \mathbb{R}^+ \times \mathbb{R}^q \to \mathbb{R}$ be a twice continuously differentiable function in the second component and such that $V(t+1, \cdot) = V(t, \cdot)$ for all $t \geqslant 0$. Fix $r_0 : \{1, \ldots, q\} \times \{1, \ldots, q\} \to \mathbb{R}^+$. Finally, let $\gamma_n \to \infty$ and set

$$r_n(t, a, b, \mu) := r_0(a, b) \exp \left\{ -n2^{-1} \left(V \left(\gamma_n t, \mu - n^{-1}\delta_a + n^{-1}\delta_b \right) - V(\gamma_n t, \mu) \right) \right\}.$$

As n goes to infinity, the limiting kernel r becomes

$$r(t, a, b, \mu) := r_0(a, b) \exp \left\{ \frac{1}{2} \nabla_a V(t, \mu) - \frac{1}{2} \nabla_b V(t, \mu) \right\},$$

so that

$$\bar{r}(a, b, \mu) := r_0(a, b) \int_0^1 \exp \left\{ \frac{1}{2} \nabla_a V(t, \mu) - \frac{1}{2} \nabla_b V(t, \mu) \right\} dt.$$

3.3 The comparison principle

We close our section of main results with the uniqueness of viscosity solutions for the Hamilton-Jacobi equation.

The motivation for this well-posedness result comes from a connection between large deviation theory of Markov processes and non-linear semigroup theory that by a chain of arguments leads to Theorem 3.3. This chain of arguments was first introduced in [16] and later reproved in [19, 20]. The key technical steps in this method are collected in Section 4.

This reduction, even though at first sight technical, is fully analogous to that of establishing weak convergence of Markov processes and is carried out via the convergence of their a transformed version of their infinitesimal generators. The statement that the martingale problem for the limiting operator is well posed is naturally replaced by uniqueness of solutions for the Hamilton-Jacobi equation (see [10] for the result that the uniqueness of the martingale problem for a linear operator follows from uniqueness of solutions to the Hamilton-Jacobi equation in terms of this operator).

In the proof of Theorem 3.3, we see that the natural limiting operator H is of the form $Hf(\mu, w) = \mathcal{H}((\mu, w), \nabla f(\mu, w))$ where

$$\mathcal{H}((\mu, w), p) = \sum_{(a,b) \in \Gamma} \mu(a) \bar{r}(a, b, \mu) \left[\exp \left\{ p_b - p_a + p_{(a,b)} \right\} - 1 \right]. \qquad (3.3)$$

Our second main result is the well-posedness of the Hamilton-Jacobi equation $f - \lambda Hf = h$. We state it separately as it is of use in a context that goes beyond large deviation theory. It also holds in a slightly more general setting than for the Hamiltonian obtained in Theorem 3.3. We give the proper context.

Definition 3.7. Let $\nu : \Gamma \times \mathcal{P}(\{1, \ldots, q\}) \to \mathbb{R}^+$. We say that ν is a *proper kernel* if ν is continuous and if for each $(a, b) \in \Gamma$, the map $\mu \mapsto \nu(a, b, \mu)$ is either identically equal to zero or satisfies the following two properties:

(a) $\nu(a, b, \mu) = 0$ if $\mu(a) = 0$ and $\nu(a, b, \mu) > 0$ for all μ such that $\mu(a) > 0$.

(b) There exists a decomposition $\nu(a, b, \mu) = \nu_\dagger(a, b, \mu(a))\nu_\ddagger(a, b, \mu)$ such that $\nu_\dagger : \Gamma \times [0,1] \to \mathbb{R}^+$ is increasing in the third coordinate (that is, in $\mu(a)$) and such that $\nu_\ddagger : \Gamma \times \mathcal{P}(\{1, \ldots, q\}) \to \mathbb{R}^+$ is continuous and satisfies $\inf_\mu \nu_\ddagger(a, b, \mu) > 0$.

Note that the Hamiltonian in (3.3) features a proper kernel ν. Choose $\nu_\dagger(a, b, \mu) = \mu(a)$ and $\nu_\ddagger(a, b, \mu) = \bar{r}(a, b, \mu)$ and argue using Assumption 3.1 (c).

Theorem 3.8. *Consider the Hamiltonian* $H : \mathcal{D}(H) \subseteq C_b(E) \to C_b(E)$ *with domain* $C_c^\infty(E) \subseteq \mathcal{D}(H) \subseteq C_b^1(E)$ *satisfying* $Hf(\mu, w) = \mathcal{H}((\mu, w), \nabla f(\mu, w))$ *where* $\mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R}$ *is given by*

$$\mathcal{H}((\mu, w), p) = \sum_{(a,b) \in \Gamma} \nu(a, b, \mu) \left[\exp\left\{ p_b - p_a + p_{(a,b)} \right\} - 1 \right]. \qquad (3.4)$$

and where $\nu : \Gamma \times \mathcal{P}(\{1, \ldots, q\}) \to \mathbb{R}^+$ *is a proper kernel. Then for each* $\lambda > 0$ *and* $h \in C_b(E)$ *the comparison principle holds for*

$$f - \lambda H f = h.$$

The proof of the theorem can be found in Section 5.2.

Remark 3.9. Note that a natural interpretation of E is that as a subset of $\mathbb{R}^{q+|\Gamma|}$. However, due to the fact we work with probability measures, we can also interpret E as a subset of $\mathbb{R}^{q-1+|\Gamma|}$. The first interpretation is more natural to write down equations or Hamiltonians and we will do so in the subsequent sections stating the main results. Only in the second interpretation our set is a subset of $\mathbb{R}^{q-1+|\Gamma|}$ that is contained in the closure of its interior as as will be needed in the proofs of Section 5.

4 Large deviations via well-posedness of the Hamilton-Jacobi equation

The key tool that allows us to establish the path-space large deviation principle is the well-posedness of the Hamilton-Jacobi equation. In this section, we give an outline of this reduction. We base ourselves on the work [16] in which this connection was first put on strong footing. We will also refer to [19] in which a new proof of the functional analytic content of the reduction is given.

The method can best be compared to a similar situation in the context of the weak convergence of Markov processes. There one establishes: tightness, convergence of generators, and the uniqueness (and existence) of solutions of the limiting martingale problem. In the large deviation context, we follow the same strategy [1]: exponential tightness, convergence of non-linearly transformed generators, and well-posedness of the Hamilton-Jacobi equation for the limiting operator.

To properly describe the reduction, we start by introducing an appropriate martingale problem for the empirical measures and their fluxes. Afterwards, we introduce the appropriate notion of convergence of operators. After that we give the framework that reduces the large deviation principle to the comparison principle.

---

[1] In fact, the similarity is more than simply an analogy. It was established in [10] that well-posedness for the Hamilton-Jacobi equation with a linear operator establishes uniqueness for the corresponding martingale problem.

4.1 An appropriate martingale problem

We next introduce the martingale problem that corresponds to the evolution of the empirical measure and empirical fluxes of n weakly interacting Markov processes. Denote by

$$
E_n := \left\{ (\mu, w) \in E \,\middle|\, \exists (x_1, \dots, x_n) \in \{1, \dots, q\}, W \in \mathbb{N}^\Gamma : \mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}, w = \frac{1}{n} W \right\}
$$

the state space of the process of empirical measures and fluxes when working with n interacting processes.

Following the setting in Section 3, we see that at moment that one of the processes $X_{n,i}$ makes a jump from site a to b, the empirical measure makes a change from $\mu_n[\vec{X}]$ to $\mu_n[\vec{X}] + \frac{1}{n}(\delta_b - \delta_a)$, whereas the empirical flux w is increased by n^{-1} on the bond $(a, b) \in \Gamma$.

It follows that the corresponding generator for the process

$$
t \mapsto Y_n(t) := (t_0 + t, \mu_n(t), w_n(t))
$$

on $D_{\mathbb{R}^+ \times E_n}(\mathbb{R}^+)$ is given by

$$
\begin{aligned}
\vec{A}_n f(t, \mu, w) = {}& \partial_t f(t, \mu, x) \\
& + \sum_{(a,b) \in \Gamma} \mu(a) r_n(t, a, b, \mu) \left[f\left(\mu + \frac{1}{n}(\delta_b - \delta_a), w + \frac{1}{n}\delta_{(a,b)} \right) - f(\mu, w) \right].
\end{aligned}
$$

Starting from this generator, we introduce the machinery to establish the large deviation principle.

4.2 Convergence of operators

To study the convergence of operators, we first need a notion of convergence of functions on a sequence of spaces.

Definition 4.1. Let $f_n \in C_b(\mathbb{R}^+ \times E_n)$ and $f \in C_b(E)$. We say that $\operatorname{LIM} f_n = f$ if

- $\sup_n |f_n| < \infty$,

- for all $T \geqslant 0$ and $M \geqslant 0$, we have

$$
\lim_{n \to \infty} \sup_{\substack{(t,\mu,w) \in \mathbb{R}^+ \times E_n: \\ t \leqslant T, |w| \leqslant M}} |f_n(t, \mu, w) - f(\mu, w)| = 0.
$$

The second condition of the definition of LIM is equivalent to the statement that for all $(\mu, w) \in E$ and any sequence $(t_n, \mu_n, w_n) \in \mathbb{R}^+ \times E_n$ such that $(\mu_n, w_n) \to (\mu, w)$ and $\sup_n t_n < \infty$, we have $\lim_n f_n(t_n, \mu_n, w_n) = f(\mu, w)$. Note that for any function $f \in C_b(E)$, we have (by interpreting it as a function on $\mathbb{R}^+ \times E_n$ by $f(t, x) = f(x)$) the natural statement $\operatorname{LIM} f = f$.

Using the notion of convergence of functions, we can define the extended limit of operators.

Definition 4.2. Let $B_n \subseteq C_b(\mathbb{R}^+ \times E_n) \times C_b(\mathbb{R}^+ \times E_n)$. Define $ex - \operatorname{LIM} B_n \subseteq C_b(E) \times C_b(E)$ as the set

$$
\begin{aligned}
ex - \operatorname{LIM} B_n & \\
= \{ (f, g) \in C_b(E) \times C_b(E) \,|\, & \exists (f_n, g_n) \in B_n : \operatorname{LIM} f_n = f, \operatorname{LIM} g_n = g \}.
\end{aligned}
$$

4.3 The main reduction step

As announced at the beginning of the section, the large deviation principle can be derived from three main inputs:

1. exponential tightness,

2. convergence of operators,

3. well-posedness of the Hamilton-Jacobi equation.

This framework was first established in [16, Theorem 7.18], but we will work in the context of the reworked version of [19, Theorem 7.10]. The key insight is the following reduction:

- Exponential tightness allows one to reduce the large deviation principle on the Skorokhod space to that of the large deviations of the finite dimensional distributions.

- Brycs theorem, in combination with the Markov property, allows one to reduce the large deviation principle for the finite dimensional convergence to the large deviations at time 0 and the convergence of conditional generating functions (4.1).

- The conditional generating function forms a semigroup. The convergence of semigroups can be treated via a functional analytic framework.

Before giving the large deviation result, we introduce the involved operators for the functional analytic framework, and a weakened version of exponential tightness. Thus, we start by introducing a semigroup, a resolvent and an infinitesimal generator. Denote by

$$V_n(t)f(s, \mu, w) := \frac{1}{n} \log \mathbb{E}\left[e^{nf(Y_n(t))} \,\middle|\, Y_n(0) = (s, w, \mu) \right] \tag{4.1}$$

the conditional log-generating function. As a function of t the operators $V_n(t)$ form a non-linear semigroup. The formal infinitesimal generator of the semigroup is given by the operator H_n

$$\begin{aligned} \mathcal{D}(H_n) &:= \left\{ f \in C_b(\mathbb{R}^+ \times E_n) \,\middle|\, e^{nf} \in \mathcal{D}(\vec{A}_n) \right\}, \\ H_n f &:= \frac{1}{n} e^{-nf} \vec{A}_n e^{nf}, \end{aligned} \tag{4.2}$$

and the corresponding resolvent $R_n(\lambda) = (\mathbb{1} - \lambda H_n)^{-1}$ is given by

$$R_n(\lambda)h(s, \mu, w)$$
$$:= \sup_{\mathbb{Q} \in \mathcal{P}(D_{\mathbb{R}^+ \times E_n}(\mathbb{R}^+))} \left\{ \int_0^\infty \lambda^{-1} e^{-\lambda^{-1}t} \left(\int h(Y_n(t)) \mathbb{Q}(dY_n) - \frac{1}{n} S_t(\mathbb{Q} \,|\, \mathbb{P}_{n,s,\mu,w}) \right) dt \right\}$$

where S_t is the relative entropy on the set $D_{\mathbb{R}^+ \times E_n}([0, T])$ and where $\mathbb{P}_{n,s,\mu,w}$ is the law of $t \mapsto Y_n(t)$ when started in (s, μ, w). A full analysis of this triplet is given in [19].

Next, we introduce the exponential compact containment condition. This condition is a weakened version of exponential tightness on the path space. The weakening consists of only requiring exponential tightness for the time marginals. In our context, the empirical measures live on a compact space, so the condition translates to a control on the number of jumps that the process makes.

Definition 4.3. We say that the processes $t \mapsto Y_n(t) = (t_0 + t, \mu_n(t), w_n(t))$ satisfy the *exponential compact containment condition* if for each $T_0 > 0$, $M_0 > 0$, $T > 0$ and $a > 0$, there is a $M = M(T_0, M_0, T, a)$ such that

$$\limsup_{n \to \infty} \sup_{(t_0, \mu, w) \in [0, T_0] \times E_n : |w| \leqslant M} \frac{1}{n} \log P_{t_0, \mu, w} \left[|w_n(t)| > M \text{ for some } t \in [0, T] \right] \leqslant -a.$$

Theorem 4.4 (Adaptation of Theorem 7.10 of [19] to our context). *Suppose that that the exponential compact containment condition holds.*
Denote $Z_n = (\mu_n(t), w_n(t))$. Suppose that

(a) *The large deviation principle holds for $Z_n(0)$ on E with speed n and good rate function I_0.*

(b) *The processes $Z_n(t) = (\mu_n(t), w_n(t))$ are exponentially tight on $D_E(\mathbb{R}^+)$ with speed n.*

(c) *There is an operator $H \subseteq C_b(E) \times C_b(E)$ such that $H \subseteq ex - \mathrm{LIM}\, H_n$.*

(d) *For all $h \in C_b(E)$ and $\lambda > 0$ the comparison principle holds for $f - \lambda H f = h$.*

Then there is a family of operators $R(\lambda)$, $\lambda > 0$ and a semigroup $V(t)$, $t \geqslant 0$ on $C_b(E)$ such that for all $f \in C_b(E)$, $x \in E$ and $t > 0$, we have

$$V(t)f(x) = \lim_{m \to \infty} R^m\left(\frac{t}{m}\right)f(x). \tag{4.3}$$

$V(t)$ *and* $R(\lambda)$ *satisfy*

- *If $\lambda > 0$ and $\mathrm{LIM}\, h_n = h$, then $\mathrm{LIM}\, R_n(\lambda)h_n = R(\lambda)h$;*

- *For $h \in C_b(E)$, the function $R(\lambda)h$ is the unique function that is a viscosity solution to $f - \lambda H f = h$;*

- *If $\mathrm{LIM}\, f_n = f$ and $t_n \to t$ we have $\mathrm{LIM}\, V_n(t_n)f_n = V(t)f$*

In addition, the processes $Z_n = (\mu_n, w_n)$ satisfy a large deviation principle on $D_E(\mathbb{R}^+)$ with speed n and rate function

$$I(\gamma) = I_0(\gamma(0)) + \sup_{k \geqslant 1} \sup_{\substack{0 = t_0 < t_1 < \ldots, t_k \\ t_i \in \Delta_\gamma^c}} \sum_{i=1}^{k} I_{t_i - t_{i-1}}(\gamma(t_i) \,|\, \gamma(t_{i-1})). \tag{4.4}$$

Here Δ_γ^c is the set of continuity points of γ. The conditional rate functions I_t are given by

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

On the basis of this abstract result, we derive our main result.

Proof of Theorem 3.3. To apply Theorem 4.4, we have to verify (a) to (d). Assumption (a) follows by the assumption on the large deviation principle at time 0 of Theorem 3.3. The other three properties will be checked below. We give their respective statements.

We verify (b) in Proposition 6.4 and we verify (c) in Proposition 6.1 below.
(d) is the result of Theorem 3.8 which will be proven in Section 5.

At this point, we have the large deviation principle with a rate function in projective limit form. That this rate function equals the rate-function in Lagrangian form follows from Proposition 6.6. $\qquad\square$

5 Verification of the comparison principle

5.1 A general method to verify the comparison principle

Throughout this section, we assume that $E = \mathcal{P}(\{1, \ldots, q\}) \times (\mathbb{R}^+)^{|\Gamma|}$ is parametrized as a subset of $\mathbb{R}^{q-1+|\Gamma|}$ to make sure that it is contained in the closure of its interior. We correspondingly write $d = q - 1 + |\Gamma|$.

In this section, we give the main technical results that can be used to verify the comparison principle. These methods are based on those used in [8, 11, 13, 16, 18]. The key component in this method is the method of 'doubling variables'. To obtain the comparison principle, one aims to give an upper bound on

$$\sup_x u(x) - v(x). \tag{5.1}$$

A direct estimate is hard to obtain, so instead one doubles the amount of variables and a penalization to large discrepancies between the variables. Thus one tries to find an upper bound for

$$\sup_{x,y} u(x) - v(y) - \alpha \Psi(x, y)$$

which converges to 0 as $\alpha \to \infty$. If $\Psi(x, y) = 0$ if and only if $x = y$, one obtains as a consequence an upper bound on (5.1) .

This technique works in the setting of compact spaces. In the context of non-compact spaces, one also has to penalize x, y that are 'far away'. Thus later works introduce the use of 'containment' or Lyapunov functions. We introduce both these concepts below.

In this section, a novel aspect in comparison to the aforementioned papers, is the use of two 'penalization' functions instead of one.

Definition 5.1. We say that $\{\Psi_1, \Psi_2\}$, with $\Psi_i : E^2 \to \mathbb{R}^+$ is a *good pair of penalization functions* if $\Psi_i \in C^1(E^2)$ and if $x = y$ if and only if $\Psi_i(x, y) = 0$ for $i \in \{1, 2\}$.

In the proof of Theorem 3.8, we work with a penalization

$$\alpha_1 \Psi_1 + \alpha_2 \Psi_2,$$

then send $\alpha_1 \to \infty$ and afterwards $\alpha_2 \to \infty$. To be able to treat both steps in a similar fashion, we introduce an auxiliary penalization function in which α_1 is already sent to infinity:

$$\widehat{\Psi}_2(x, y) := \begin{cases} \Psi_2(x, y) & \text{if } \Psi_1(x, y) = 0, \\ \infty & \text{if } \Psi_1(x, y) \neq 0. \end{cases}$$

Finally, we introduce containment functions that allow us to restrict our analysis to compact sets.

Definition 5.2. Let $\mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R}$, we say that $\Upsilon : E \to \mathbb{R}$ is a *good containment function* (for $\mathcal{H}$) if

(Υa) $\Upsilon \geqslant 0$ and there exists a point z_0 such that $\Upsilon(z_0) = 0$,

(Υb) Υ is continuously differentiable,

(Υc) for every $c \geqslant 0$, the sublevel set $\{z \in E \mid \Upsilon(z) \leqslant c\}$ is compact,

(Υd) we have $\sup_{z \in E} \mathcal{H}(z, \nabla \Upsilon(z)) < \infty$.

The following result gives us the main technical input for the proof of the comparison principle.

Lemma 5.3. *Let* $u : E \to \mathbb{R}$ *be bounded and upper semi-continuous, let* $v : E \to \mathbb{R}$ *be bounded and lower semi-continuous. Let* $\{\Psi_1, \Psi_2\}$ *be a good pair of penalization functions and let* $\Upsilon : E \to \mathbb{R}^+$ *be a good containment function for* $\mathcal{H}$.

Fix $\varepsilon > 0$. *Then there is a compact set* $K_\varepsilon \subseteq E$ *such that for every* $\alpha \in (0, \infty)^2$, $\alpha = (\alpha_1, \alpha_2)$ *there exist points* $x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon} \in K_\varepsilon$, *such that*

$$\frac{u(x_{\alpha,\varepsilon})}{1-\varepsilon} - \frac{v(y_{\alpha,\varepsilon})}{1+\varepsilon} - \sum_{i=1}^{2} \alpha_i \Psi_i(x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon}) - \frac{\varepsilon}{1-\varepsilon}\Upsilon(x_{\alpha,\varepsilon}) - \frac{\varepsilon}{1+\varepsilon}\Upsilon(y_{\alpha,\varepsilon})$$

$$= \sup_{x,y \in E} \left\{ \frac{u(x)}{1-\varepsilon} - \frac{v(y)}{1+\varepsilon} - \sum_{i=1}^{2} \alpha_i \Psi_i(x,y) - \frac{\varepsilon}{1-\varepsilon}\Upsilon(x) - \frac{\varepsilon}{1+\varepsilon}\Upsilon(y) \right\}.$$

In addition, for all $\varepsilon > 0$ *and* $\alpha_2 > 0$ *there are limit points* $x_{\alpha_2,\varepsilon}, y_{\alpha_2,\varepsilon} \in K_\varepsilon$ *of* $x_{(\alpha_1,\alpha_2),\varepsilon}$ *and* $y_{(\alpha_1,\alpha_2),\varepsilon}$ *as* $\alpha_1 \to \infty$ *and we have*

$$\lim_{\alpha_1 \to \infty} \alpha_1 \Psi_1(x_{(\alpha_1,\alpha_2),\varepsilon}, y_{(\alpha_1,\alpha_2),\varepsilon}) = 0,$$

$$\Psi_1(x_{\alpha_2,\varepsilon}, y_{\alpha_2,\varepsilon}) = 0.$$

In addition

$$\frac{u(x_{\alpha_2,\varepsilon})}{1-\varepsilon} - \frac{v(y_{\alpha_2,\varepsilon})}{1+\varepsilon} - \alpha_2 \widehat{\Psi}_2(x_{\alpha_2,\varepsilon}, y_{\alpha_2,\varepsilon}) - \frac{\varepsilon}{1-\varepsilon}\Upsilon(x_{\alpha_2,\varepsilon}) - \frac{\varepsilon}{1+\varepsilon}\Upsilon(y_{\alpha_2,\varepsilon})$$

$$= \sup_{x,y \in E} \left\{ \frac{u(x)}{1-\varepsilon} - \frac{v(y)}{1+\varepsilon} - \alpha_2 \widehat{\Psi}_2(x,y) - \frac{\varepsilon}{1-\varepsilon}\Upsilon(x) - \frac{\varepsilon}{1+\varepsilon}\Upsilon(y) \right\}.$$

Remark 5.4. The result remains true for Ψ_i and Υ that are lower semi-continuous instead of C^1.

The proof of Lemma 5.3, carried out below, is based on the following standard result.

Lemma 5.5 (Proposition 3.7 of [11] or Lemma 9.2 in [16]). *Let* $F : E \to \mathbb{R} \cup \{-\infty\}$ *be bounded above, upper semi-continuous, and such that for each* $c \in \mathbb{R}$ *the set* $\{(x,y) \in E^2 \,|\, F(x,y) \geqslant c\}$ *is compact. Let* $G : E^2 \to [0, \infty]$ *be lower semi-continuous and such that* $x = y$ *implies* $G(x,y) = 0$.
Then there exist for each $\alpha > 0$ *variables* $(x_\alpha, y_\alpha) \in E^2$ *such that*

$$F(x_\alpha, y_\alpha) - \alpha G(x_\alpha, y_\alpha) = \sup_{x,y \in E} \{F(x,y) - \alpha G(x,y)\}.$$

In addition, we have

(a) *The set* $\{x_\alpha, y_\alpha \,|\, \alpha > 0\}$ *is relatively compact in* E;

(b) *Any limit point* (x_0, y_0) *of* $(x_\alpha, y_\alpha)_{\alpha > 0}$ *as* $\alpha \to \infty$ *satisfies* $G(x_0, y_0) = 0$ *and* $F(x_0, y_0) = \sup_{x,y \in E, G(x,y)=0} F(x,y)$;

(c) *We have* $\lim_{\alpha \to \infty} \alpha G(x_\alpha, y_\alpha) = 0$.

Proof of Lemma 5.3. As u, v are bounded and the Ψ_i are bounded from below, we find using the semi-continuity properties of all functions involved, and the compact level sets of Υ, that there is a compact set $K_\varepsilon \subseteq E$ and variables $x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon} \in K_\varepsilon$ as in the first claim of the lemma.

The second statement follows from Lemma 5.5 by taking for α the variable α_1, and for F and G the functions

$$F(x,y) := \left\{ \frac{u(x)}{1-\varepsilon} - \frac{v(y)}{1+\varepsilon} - \alpha_2 \Psi_2(x,y) - \frac{\varepsilon}{1-\varepsilon} \Upsilon(x) - \frac{\varepsilon}{1+\varepsilon} \Upsilon(y) \right\},$$

$$G(x,y) := \Psi_1(x,y).$$

$\square$

The following result gives us the explicit condition that can be used to verify the comparison principle.

Proposition 5.6. *Let* $H : \mathcal{D}(H) \subseteq C_b(E) \to C_b(E)$ *have domain* $\mathcal{D}(H)$ *satisfying* $C_c^\infty(E) \subseteq \mathcal{D}(H) \subseteq C_b^1(E)$ *and be of the form* $Hf(x) = \mathcal{H}(x, \nabla f(x))$. *Assume that the map* $\mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R}$ *is continuous and that for each* $x \in E$ *the map* $p \mapsto \mathcal{H}(x, p)$ *is convex.*

Fix $\lambda > 0$, $h \in C_b(E)$ *and consider* u *and* v *sub- and supersolution to* $f - \lambda Hf = h$.

Let $k \in \mathbb{N} \setminus \{0\}$ *and let* $\{\Psi_1, \Psi_2\}$ *be a pair of good penalization functions and* Υ *be a good containment function. Moreover, for every* $\alpha = (\alpha_1, \alpha_2) \in (0, \infty)^2$ *and* $\varepsilon > 0$ *let* $x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon} \in E$ *be such that*

$$\frac{u(x_{\alpha,\varepsilon})}{1-\varepsilon} - \frac{v(y_{\alpha,\varepsilon})}{1+\varepsilon} - \sum_{i=1}^{2} \alpha_i \Psi_i(x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon}) - \frac{\varepsilon}{1-\varepsilon} \Upsilon(x_{\alpha,\varepsilon}) - \frac{\varepsilon}{1+\varepsilon} \Upsilon(y_{\alpha,\varepsilon})$$

$$= \sup_{x,y \in E} \left\{ \frac{u(x)}{1-\varepsilon} - \frac{v(y)}{1+\varepsilon} - \sum_{i=1}^{2} \alpha_i \Psi_i(x,y) - \frac{\varepsilon}{1-\varepsilon} \Upsilon(x) - \frac{\varepsilon}{1+\varepsilon} \Upsilon(y) \right\}. \quad (5.2)$$

Suppose that

$$\liminf_{\varepsilon \to 0} \liminf_{\alpha_2 \to \infty} \liminf_{\alpha_1 \to \infty} \mathcal{H}\left(x_{\alpha,\varepsilon}, \sum_{i=1}^{2} \alpha_i \nabla \Psi_i(\cdot, y_{\alpha,\varepsilon})(x_{\alpha,\varepsilon}) \right)$$

$$- \mathcal{H}\left(y_{\alpha,\varepsilon}, -\sum_{i=1}^{2} \alpha_i \nabla \Psi_i(x_{\alpha,\varepsilon}, \cdot)(y_{\alpha,\varepsilon}) \right) \leqslant 0, \quad (5.3)$$

then $u \leqslant v$. *In other words:* $f - \lambda Hf = h$ *satisfies the comparison principle.*

Proof. Using the convexity of H and the definitions of sub- and supersolutions, we find by Lemma A.4 that

$$\sup_x u(x) - v(x)$$

$$\leqslant \frac{h(x_{\alpha,\varepsilon})}{1-\varepsilon} - \frac{h(y_{\alpha,\varepsilon})}{1+\varepsilon} \quad (5.4)$$

$$+ \frac{\varepsilon}{1-\varepsilon} \mathcal{H}(x_{\alpha,\varepsilon}, \nabla \Upsilon(x_{\alpha,\varepsilon})) + \frac{\varepsilon}{1+\varepsilon} \mathcal{H}(y_{\alpha,\varepsilon}, \nabla \Upsilon(y_{\alpha,\varepsilon})) \quad (5.5)$$

$$+ \lambda \left[\mathcal{H}\left(x_{\alpha,\varepsilon}, \sum_{i=1}^{2} \alpha_i \nabla \Psi_i(\cdot, y_{\alpha,\varepsilon})(x_{\alpha,\varepsilon}) \right) - \mathcal{H}\left(y_{\alpha,\varepsilon}, -\sum_{i=1}^{2} \alpha_i \nabla \Psi_i(x_{\alpha,\varepsilon}, \cdot)(y_{\alpha,\varepsilon}) \right) \right].$$

$$(5.6)$$

Consecutively taking $\liminf$ over $\alpha_1, \alpha_2, \varepsilon$, the term (5.6) vanishes by assumption. The term in (5.5) vanishes as well, due to the uniform bounds on $\mathcal{H}(z, \nabla \Upsilon(z))$ by property (Υd) of Definition 5.2. Consecutively taking limit points as in Lemma 5.3 by sending α_1, then α_2 to infinity, we find a pair $(x_\varepsilon, y_\varepsilon)$ with $\Psi_1(x_\varepsilon, y_\varepsilon) =$

16

$\Psi_2(x_\varepsilon, y_\varepsilon) = 0$. This implies $x_\varepsilon = y_\varepsilon$. Thus, taking $\liminf$ over α_1 and α_2 the term in (5.4) is bounded above by

$$\sup_{z \in K_\varepsilon} \frac{h(z)}{1 - \varepsilon} - \frac{h(z)}{1 + \varepsilon} \leqslant \frac{2\varepsilon}{1 - \varepsilon^2} |h|,$$

which vanishes if $\varepsilon \to 0$. We conclude that the comparison principle holds for $f - \lambda Hf = h$. $\qquad \square$

The next lemma establishes that the Hamiltonian applied to the pair of penalization functions is either bounded below or above. Using coercivity properties of $\mathcal{H}$, this allows us to derive properties of the sequences $(x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon})$ that can be used afterwards to help the the verification of (5.3).

Lemma 5.7. *Let* $H : \mathcal{D}(H) \subseteq C_b(E) \to C_b(E)$ *with domain* $\mathcal{D}(H)$ *satisfying* $C_c^\infty(E) \subseteq \mathcal{D}(H) \subseteq C_b^1(E)$ *and such that* $Hf(x) = \mathcal{H}(x, \nabla f(x))$. *Assume that the map* $\mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R}$ *is continuous and that for each* $x \in E$ *the map* $p \mapsto \mathcal{H}(x, p)$ *is convex.*
Let $h \in C_b(E)$ *and* $\lambda > 0$ *and let* u *be a subsolution and* v *a supersolution to* $f - \lambda H = h$.
Let $\{\Psi_1, \Psi_2\}$ *be a pair of good penalization functions and* Υ *be a good containment function. Moreover, for every* $\alpha = (\alpha_1, \alpha_2) \in (0, \infty)^2$ *and* $\varepsilon > 0$ *let* $x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon} \in E$ *be as in* (5.2). *Then we have that*

$$\sup_{\varepsilon, \alpha} \mathcal{H}\left(y_{\alpha,\varepsilon}, -\sum_{i=1}^2 \alpha_i(\nabla \Psi_i(x_{\alpha,\varepsilon}, \cdot))(y_{\alpha,\varepsilon})\right) < \infty, \tag{5.7}$$

$$\inf_{\varepsilon, \alpha} \mathcal{H}\left(x_{\alpha,\varepsilon}, \sum_{i=1}^2 \alpha_i(\nabla \Psi_i(\cdot, y_{\alpha,\varepsilon}))(x_{\alpha,\varepsilon})\right) > -\infty. \tag{5.8}$$

The proof is an adaptation of Lemma 9.3 in [16]. For a similar version in compact setting see Lemma 5 in [18].

Proof. We only prove the first statement, the second can be proven similarly. Using that v is a supersolution to $f - \lambda Hf = h$, we find that it is a supersolution to the equation $f - \lambda H_\ddagger f = h$, where $H_\ddagger$ is a super-extension of H that includes functions of the type $y \mapsto (-(1 + \varepsilon) \sum_{i=1}^2 \alpha_i \Psi_i(x, y) - \varepsilon \Upsilon(y)$ in its domain, see Lemma A.2. It follows that for the points $(x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon})$, we have

$$\mathcal{H}\left(y_{\alpha,\varepsilon}, -(1 + \varepsilon) \sum_{i=1}^2 \alpha_i \nabla \Psi_i(x_{\alpha,\varepsilon}, \cdot)(y_{\alpha,\varepsilon}) - \varepsilon \nabla \Upsilon(y_{\alpha,\varepsilon})\right)$$
$$\leqslant \frac{v(y_{\alpha,\varepsilon}) - h(y_{\alpha,\varepsilon})}{\lambda} \leqslant \frac{|v - h|}{\lambda}.$$

By the convexity of $p \mapsto \mathcal{H}(x, p)$, we find

$$\mathcal{H}\left(y_{\alpha,\varepsilon}, -\sum_{i=1}^2 \alpha_i \nabla \Psi_i(x_{\alpha,\varepsilon}, \cdot)(y_{\alpha,\varepsilon})\right)$$
$$\leqslant \frac{1}{1 + \varepsilon} \mathcal{H}\left(y_{\alpha,\varepsilon}, -(1 + \varepsilon) \sum_{i=1}^2 \alpha_i \nabla \Psi_i(x_{\alpha,\varepsilon}, \cdot)(y_{\alpha,\varepsilon}) - \varepsilon \nabla \Upsilon(y_{\alpha,\varepsilon})\right)$$
$$+ \frac{\varepsilon}{1 + \varepsilon} \mathcal{H}\left(y_{\alpha,\varepsilon}, \nabla \Upsilon(y_{\alpha,\varepsilon})\right),$$

which implies

$$\sup_\alpha \mathcal{H}\left(y_{\alpha,\varepsilon}, -\sum_{i=1}^2 \alpha_i (\nabla \Psi_i (x_{\alpha,\varepsilon}, \cdot))(y_{\alpha,\varepsilon})\right)$$

$$\leqslant \frac{1}{1+\varepsilon}\left(\frac{|\nu - h|}{\lambda} + \varepsilon \sup_z \mathcal{H}(z, \nabla \Upsilon(z))\right) < \infty.$$

Taking the supremum over ε yields the final claim. $\qquad\square$

5.2 The verification of the comparison principle for our explicit Hamiltonian

We now turn to the verification of Theorem 3.8, that is, the verification of the comparison principle for the Hamilton-Jacobi equation with Hamiltonians of the type

$$\mathcal{H}((\mu, w), p) = \sum_{(a,b)\in\Gamma} \nu(a, b, \mu)\left[\exp\left\{p_b - p_a + p_{(a,b)}\right\} - 1\right].$$

To obtain the comparison principle, we aim to apply Proposition 5.6. To do so, we first need to choose a pair of good penalization functions (Ψ_1, Ψ_2) and a good containment function Υ. This is the first thing we will do in the section. Afterwards, we verify (5.3) which is the key hypothesis of Proposition 5.6.

We start by construction of a good containment function.

Lemma 5.8. *Consider* $\mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R}$ *given by*

$$\mathcal{H}((\mu, w), p) = \sum_{(a,b)\in\Gamma} \nu(a, b, \mu)\left[\exp\left\{p_b - p_a + p_{(a,b)}\right\} - 1\right]$$

where $\nu : \Gamma \times \mathcal{P}(\{1, \dots, q\})$ *is continuous and non-negative.*
Then the function $\Upsilon(\mu, w) = \sum_{(a,b)\in\Gamma} \log\left(1 + w_{(a,b)}\right)$ *is a good containment function for* $\mathcal{H}$.

Proof. As $\mathcal{P}(\{1, \dots, q\})$ is compact and $x \mapsto \log(1 + x)$ has compact level sets on $\mathbb{R}^+$ the map Υ has compact level sets in E also. Clearly Υ is smooth. Thus, it suffices to show $\sup_{\mu, w} \mathcal{H}((\mu, w), \nabla\Upsilon(\mu, w)) < \infty$:

$$\mathcal{H}((\mu, w), \nabla\Upsilon(\mu, w)) = \sum_{(a,b)\in\Gamma} \nu(a, b, \mu)\left[\exp\left\{(1 + w_{(a,b)})^{-1}\right\} - 1\right]$$

$$\leqslant \sum_{(a,b)\in\Gamma} \nu(a, b, \mu)\left[\exp\{1\} - 1\right].$$

The claim follows as ν is continuous, and, therefore, bounded. $\qquad\square$

We proceed by constructing a pair of good penalization functions. For Ψ_1 we use a version of the quadratic distance on the space of measures which was first introduced in [18]. For Ψ_2 we use a standard quadratic distance on the space of fluxes. The exact definition follows.
For $x \in \mathbb{R}$, let $x^- := x \wedge 0$ and $x^+ = x \vee 0$.

Lemma 5.9. *Define* Ψ_1, Ψ_2 *by*

$$\Psi_1(\mu, \hat{\mu}) = \frac{1}{2}\sum_a ((\mu(a) - \hat{\mu}(a))^-)^2 = \frac{1}{2}\sum_a ((\hat{\mu}(a) - \mu(a))^+)^2,$$

$$\Psi_2(w, \hat{w}) := \frac{1}{2}\sum_{(a,b)\in\Gamma} (w_{(a,b)} - \hat{w}_{(a,b)})^2.$$

The pair (Ψ_1, Ψ_2) *is a pair of good penalization functions for* $E = \mathcal{P}(\{1, \ldots, 1\}) \times (\mathbb{R}^+)^\Gamma$. *In addition, we have*

$$(\nabla \Psi_1(\cdot, \hat{\mu}))(\mu) = -(\nabla \Psi_1(\mu, \cdot))(\hat{\mu}),$$
$$(\nabla \Psi_2(\cdot, \hat{w}))(w) = -(\nabla \Psi_2(w, \cdot))(\hat{w}).$$

The use of Ψ_1 is highly specific for the space $\mathcal{P}(\{1, 2, \ldots, q\})$. The special form is motivated by the linear constraint $\sum \mu(a) = 1$. The use of the standard quadratic distance leads to 'loss of control' over the variables when applying Lemma 5.7. This issue is related to the discussion in Remark 5.10. The adaptation of the quadratic distance takes into account the form of our Hamiltonian and the linear constraint in a symmetric way and is geared towards re-establishing the control via Lemma 5.7.

Proof of Lemma 5.9. Note that as $\sum_i \mu(i) = \sum_i \hat{\mu}(i) = 1$, we find that $\Psi_1(\mu, \hat{\mu}) = 0$ implies that $\mu = \hat{\mu}$. As Ψ_2 is a quadratic distance on $(\mathbb{R}^+)^\Gamma$, we indeed have that $(\mu, w) = (\hat{\mu}, \hat{w})$ if and only if $\Psi_1(\mu, \hat{\mu}) + \Psi_2(w, \hat{w}) = 0$.
The second claim follows by direct verification. $\qquad\square$

We proceed with the verification of of the comparison principle by establishing the key estimate (5.3) of Proposition 5.6.

Proof of Theorem 3.8. The proof is an adaptation of the proof of Theorem 4 in [18]. Fix $h \in C_b(E)$ and $\lambda > 0$. Let u be a subsolution and v be a supersolution to $f - \lambda Hf = h$.
We verify (5.3) of Proposition 5.6 using containment function Υ and penalization functions Ψ_1, Ψ_2 from Lemmas 5.8 and 5.9. For $\varepsilon > 0$, $\alpha_1, \alpha_2 > 0$ let $x_{\alpha, \varepsilon} := (\mu_{\alpha, \varepsilon}, w_{\alpha, \varepsilon})$ and $y_{\alpha, \varepsilon} := (\hat{\mu}_{\alpha, \varepsilon}, \hat{w}_{\alpha, \varepsilon})$ be as in (5.2).
To establish the theorem, we will show that already after taking one liminf, the bound is satisfied. Indeed, we will show for fixed $\varepsilon > 0$ and $\alpha_2 > 0$ that

$$\liminf_{\alpha_1 \to \infty} \mathcal{H}\left(x_{\alpha, \varepsilon}, \sum_{i=1}^2 \alpha_i \nabla \Psi_i(\cdot, y_{\alpha, \varepsilon})(x_{\alpha, \varepsilon})\right)$$
$$- \mathcal{H}\left(y_{\alpha, \varepsilon}, -\sum_{i=1}^2 \alpha_i \nabla \Psi_i(x_{\alpha, \varepsilon}, \cdot)(y_{\alpha, \varepsilon})\right) \leqslant 0. \quad (5.9)$$

By Lemma 5.3 for fixed α_2, ε and sending $\alpha_1 \to \infty$, we find limit points $(x_{\alpha_2, \varepsilon}, y_{\alpha_2, \varepsilon}) = ((\mu_{\alpha_2, \varepsilon}, w_{\alpha_2, \varepsilon}), (\hat{\mu}_{\alpha_2, \varepsilon}, \hat{w}_{\alpha_2, \varepsilon}))$ of the sequence $((\mu_{\alpha, \varepsilon}, w_{\alpha, \varepsilon}), (\hat{\mu}_{\alpha, \varepsilon}, \hat{w}_{\alpha, \varepsilon}))$. Without loss of generality, going to a subsequence if necessary, we assume that these sequences converge to their respective limit point. By the definition of $\mathcal{H}$, we have

$$\mathcal{H}\left(x_{\alpha, \varepsilon}, \sum_{i=1}^2 \alpha_i \nabla \Psi_i(\cdot, y_{\alpha, \varepsilon})(x_{\alpha, \varepsilon})\right) - \mathcal{H}\left(y_{\alpha, \varepsilon}, -\sum_{i=1}^2 \alpha_i \nabla \Psi_i(x_{\alpha, \varepsilon}, \cdot)(y_{\alpha, \varepsilon})\right)$$
$$= \sum_{(a,b) \in \Gamma} [v(a, b, \mu_{\alpha, \varepsilon}) - v(a, b, \hat{\mu}_{\alpha, \varepsilon})] \times \qquad\qquad (5.10)$$
$$\left[e^{\alpha_1 \left((\mu_{\alpha, \varepsilon}(b) - \hat{\mu}_{\alpha, \varepsilon}(b))^- - (\mu_{\alpha, \varepsilon}(a) - \hat{\mu}_{\alpha, \varepsilon}(a))^-\right) + \alpha_2 (w_{\alpha, \varepsilon}(a, b) - \hat{w}_{\alpha, \varepsilon}(a, b))} - 1\right].$$

To ease the notation, and focus on the parts that matter, we will write $c_{\alpha, \varepsilon}(a, b) := \alpha_2 (w_{\alpha, \varepsilon}(a, b) - \hat{w}_{\alpha, \varepsilon}(a, b))$ as this term will not play a role in our bounds below. In fact, for fixed ε and α_2, we have for all $(a, b) \in \Gamma$ that

$$\sup_{\alpha_1} |c_{\alpha, \varepsilon}(a, b)| < \infty \qquad\qquad (5.11)$$

because by the construction of Lemma 5.3 we have

$$\sup_{\alpha_1} \alpha_2 \Psi_2(w_{\alpha,\varepsilon}, \hat{w}_{\alpha,\varepsilon}) < \infty.$$

We will show that each term in (5.10) separately is bounded from above by 0 asymptotically. Pick some ordering of the ordered pairs $(a, b) \in \Gamma$, and assume that we have some sequence α_1 such that the $\liminf_{\alpha_1 \to \infty}$ of the first l terms in equation (5.10) are bounded above by 0. We construct a subsequence so that also term $l+1$ is asymptotically bounded above by 0. The result then follows by induction. Thus, suppose that (i, j) is the pair corresponding to the $l+1$-th term of the sum in (5.10). We go through the two options of ν being a proper kernel. Clearly, if $\nu(i, j, \pi) = 0$ for all π then we are done. Therefore, we assume that $\nu(i, j, \pi) \neq 0$ for all π such that $\pi(i) > 0$ and that conditions (a) and (b) of having a proper kernel are satisfied. *Case 1:* If $\mu_{\alpha_2,\varepsilon}(i) > 0$, we know by (5.7), using that $\nu(i, j, \cdot)$ is bounded away from 0 on a neighbourhood of $\mu_{\alpha_2,\varepsilon}$ (condition (a) of having a proper kernel), that

$$\sup_{\alpha_1} e^{\alpha_1 \left((\mu_{\alpha,\varepsilon}(j) - \hat{\mu}_{\alpha,\varepsilon}(j))^- - (\mu_{\alpha,\varepsilon}(i) - \hat{\mu}_{\alpha,\varepsilon}(i))^- \right) + c_{\alpha,\varepsilon}(i,j)} - 1 < \infty.$$

As we also have that the exponential is bounded from below by 0, we can pick a subsequence $\alpha(n) = (\alpha_1(n), \alpha_2)$ and some constant c such that

$$e^{\alpha_1(n) \left((\mu_{\alpha(n),\varepsilon}(j) - \hat{\mu}_{\alpha(n),\varepsilon}(j))^- - (\mu_{\alpha(n),\varepsilon}(i) - \hat{\mu}_{\alpha(n),\varepsilon}(i))^- \right) + c_{\alpha(n),\varepsilon}(i,j)} - 1 \to c.$$

Using that $\pi \to \nu(i, j, \pi)$ is uniformly continuous on compact sets, we see

$$\liminf_{\alpha_1 \to \infty} [\nu(i, j, \mu_{\alpha,\varepsilon}) - \nu(i, j, \hat{\mu}_{\alpha,\varepsilon})] \times$$
$$\left[e^{\alpha_1 \left((\mu_{\alpha,\varepsilon}(j) - \nu_{\alpha,\varepsilon}(j))^- - (\mu_{\alpha,\varepsilon}(i) - \nu_{\alpha,\varepsilon}(i))^- \right) + c_{\alpha,\varepsilon}(i,j)} - 1 \right]$$
$$\leqslant \lim_{n \to \infty} \left[\nu(i, j, \mu_{\alpha(n),\varepsilon}) - \nu(i, j, \hat{\mu}_{\alpha(n),\varepsilon}) \right] \times$$
$$\lim_{n \to \infty} \left[e^{\alpha_1(n) \left((\mu_{\alpha(n),\varepsilon}(j) - \hat{\mu}_{\alpha(n),\varepsilon}(j))^- - (\mu_{\alpha(n),\varepsilon}(i) - \hat{\mu}_{\alpha(n),\varepsilon}(i))^- \right) + c_{\alpha,\varepsilon}(i,j)} - 1 \right]$$
$$= c \lim_{n \to \infty} \left[\nu(i, j, \mu_{\alpha(n),\varepsilon}) - \nu(i, j, \hat{\mu}_{\alpha(n),\varepsilon}) \right] = 0.$$

Case 2: Suppose that $\mu_{\alpha,\varepsilon}(i), \hat{\mu}_{\alpha,\varepsilon}(i) \to 0$ as $\alpha_1 \to \infty$. Again by (5.7), we get

$$M := \sup_{\alpha_1} \nu(i, j, \hat{\mu}_{\alpha,\varepsilon}) \left[e^{\alpha_1 \left((\mu_{\alpha,\varepsilon}(j) - \hat{\mu}_{\alpha,\varepsilon}(j))^- - (\mu_{\alpha,\varepsilon}(i) - \hat{\mu}_{\alpha,\varepsilon}(i))^- \right) + c_{\alpha,\varepsilon}(i,j)} - 1 \right] < \infty. \tag{5.12}$$

First of all, if $\sup_{\alpha_1} \alpha_1 \left((\mu_{\alpha,\varepsilon}(j) - \hat{\mu}_{\alpha,\varepsilon}(j))^- - (\mu_{\alpha,\varepsilon}(i) - \hat{\mu}_{\alpha,\varepsilon}(i))^- \right) + c_{\alpha,\varepsilon}(i,j) < \infty$, then the argument given in step 1 above also takes care of this situation. So suppose that this supremum is infinite. Clearly, the contribution $\alpha_1 (\mu_{\alpha,\varepsilon}(j) - \hat{\mu}_{\alpha,\varepsilon}(j))^-$ is negative, and the one of $c_{\alpha,\varepsilon}$ is uniformly bounded by (5.11), which implies that $\sup_{\alpha_1} \alpha_1 (\hat{\mu}_{\alpha,\varepsilon}(i) - \mu_{\alpha,\varepsilon}(i))^+ = \infty$. This means that we can assume without loss of generality that

$$\alpha_1 (\hat{\mu}_{\alpha,\varepsilon}(i) - \mu_{\alpha,\varepsilon}(i)) \to \infty, \qquad \hat{\mu}_{\alpha,\varepsilon}(i) > \mu_{\alpha,\varepsilon}(i). \tag{5.13}$$

The bound on the right, combined with property (a) of ν being a proper kernel, implies that $\nu(i, j, \hat{\mu}_{\alpha,\varepsilon}) > 0$. We rewrite the term $(a, b) = (i, j)$ in equation (5.10) as

$$\left[\frac{\nu(i, j, \mu_{\alpha,\varepsilon})}{\nu(i, j, \hat{\mu}_{\alpha,\varepsilon})} - 1 \right] \times$$
$$\nu(i, j, \hat{\mu}_{\alpha,\varepsilon}) \left[e^{\alpha_1 \left((\mu_{\alpha,\varepsilon}(j) - \hat{\mu}_{\alpha,\varepsilon}(j))^- - (\mu_{\alpha,\varepsilon}(i) - \hat{\mu}_{\alpha,\varepsilon}(i))^- \right) + c_{\alpha,\varepsilon}(i,j)} - 1 \right].$$

The term on the second line is bounded above by M introduced in (5.12) and bounded below by $-|\nu|$. Thus, we can take a subsequence of α_1, also denoted by α_1, such that the right-hand side converges. By (5.13), the right-hand side is non-negative. Therefore, it suffices to show that

$$\liminf_{\alpha_1 \to \infty} \frac{\nu(i, j, \mu_{\alpha, \varepsilon})}{\nu(i, j, \hat{\mu}_{\alpha, \varepsilon})} \leqslant 1.$$

By property (b) of ν being a proper kernel, we find $\nu(i, j, \mu) = \nu_\dagger(i, j, \mu(i))\nu_\ddagger(i, j, \mu)$ which implies that

$$\liminf_{\alpha_1 \to \infty} \frac{\nu(i, j, \mu_{\alpha, \varepsilon})}{\nu(i, j, \hat{\mu}_{\alpha, \varepsilon})} = \liminf_{\alpha_1 \to \infty} \frac{\nu_\dagger(i, j, \mu_{\alpha, \varepsilon}(i))}{\nu_\dagger(i, j, \hat{\mu}_{\alpha, \varepsilon}(i))} \frac{\nu_\ddagger(i, j, \mu_{\alpha, \varepsilon})}{\nu_\ddagger(i, j, \hat{\mu}_{\alpha, \varepsilon})}$$
$$\leqslant \left(\limsup_{\alpha_1 \to \infty} \frac{\nu_\dagger(i, j, \mu_{\alpha, \varepsilon}(i))}{\nu_\dagger(i, j, \hat{\mu}_{\alpha, \varepsilon}(i))} \right) \left(\lim_{\alpha_1 \to \infty} \frac{\nu_\ddagger(i, j, \mu_{\alpha, \varepsilon})}{\nu_\ddagger(i, j, \hat{\mu}_{\alpha, \varepsilon})} \right) \leqslant \frac{\nu_\ddagger(i, j, \mu_{\alpha_2, \varepsilon})}{\nu_\ddagger(i, j, \mu_{\alpha_2, \varepsilon})} = 1,$$

where we use that $r \mapsto \nu_\dagger(i, j, r)$ is increasing and the bound in (5.13) for the first term and that $\pi \mapsto \nu_\ddagger(i, j, \mu)$ is continuous and bounded away from zero in a neighborhood of $\mu_{\alpha_2, \varepsilon}$ for the second term.

Thus, cases 1 and 2 inductively establish an upper bound for (5.9), concluding the proof. $\square$

Remark 5.10. Note that the motivation for the definition of the non-standard Ψ_1 in [18], as well as the introduction of the use of two penalization functions in this paper comes from the bound obtained in (5.12). Indeed, in [18] the use of Ψ_1 allowed us to obtain (5.13), which is needed to complete the argument.

In our setting, where we work with fluxes, using a single penalization function $\Psi = \Psi_1 + \Psi_2$ multiplied by α, would not allow us to obtain (5.13) due to the interference coming from Ψ_2. Instead considering these two functions separately with separate multiplicative constants, allows us to establish the important inequality in (5.13).

Note that an argument like the one carried out in this proof does not seem directly applicable in the context of mass-action kinetics. In particular, if one allows transitions leading to a jumps in the rescaled dynamics of the type $n^{-1}(\delta_c + \delta_d - \delta_a - \delta_b)$, one gets instead of (5.13) a statement of the type

$$\alpha_1(\hat{\mu}_{\alpha, \varepsilon}(a) - \mu_{\alpha, \varepsilon}(a)) + \alpha_1(\hat{\mu}_{\alpha, \varepsilon}(b) - \mu_{\alpha, \varepsilon}(b)) \to \infty.$$

From such a statement, one cannot derive that $\hat{\mu}_{\alpha, \varepsilon}(a) > \mu_{\alpha, \varepsilon}(a)$ and $\hat{\mu}_{\alpha, \varepsilon}(a) > \mu_{\alpha, \varepsilon}(b)$ for large α. This makes it impossible to proceed with the present argument. It seems that a new type of penalization procedure is needed.

6 Convergence of operators, exponential tightness and the variational representation of the rate function

6.1 Convergence of operators

We proceed with the verification of (c) of Theorem 4.4, namely that there exists an operator H such that $H \subseteq ex - LIM H_n$.

Proposition 6.1. *Consider the setting of Theorem 3.3. Let H be the operator with domain $\mathcal{D}(H) = C_b^2(E)$ satisfying $Hf(x) = \mathcal{H}(x, \nabla f(x))$ with $\mathcal{H}$ as in (3.4):*

$$\mathcal{H}((\mu, w), p) = \sum_{(a,b) \in \Gamma} \mu(a)\bar{r}(a, b, \mu) \left[\exp\{p_b - p_a + p_{(a,b)}\} - 1 \right]. \qquad (6.1)$$

Then we have $H \subseteq ex - LIM H_n$.

To recall the definition of $\mathsf{ex}-\mathrm{LIM}$: we will prove that for each $f \in C_b^2(E)$ there are $f_n \in C_b(\mathbb{R}^+ \times E_n)$ such that

$$\mathrm{LIM}\, f_n = f, \tag{6.2}$$

$$\mathrm{LIM}\, H_n f_n = Hf. \tag{6.3}$$

The proof of this proposition will be carried out in three steps.

- In Lemma 6.2, we will establish the convergence of operators in the context where the time dependence is essentially removed by working along a time sequence $s_n = s\gamma_n^{-1}$. In this context we will show that for any function f and small perturbation h_n we have $H_n[\gamma_n^{-1}s](f + h_n) \to H_0[s]f$.

- In Lemma 6.3, we will introduce a sequence of functions that are periodic over the respective intervals on which the jump-rates r_n oscillate. These functions will satisfy the conditions for the small perturbations h_n of the previous step.

- We prove Proposition 6.1 by showing that the perturbations h_n have the effect that $H_n(f + h_n)$ are constant in time.

We first introduce some auxiliary notation.
Let $H_0[t]$ be the operators with domain $C_b^2(E)$ satisfying for $f \in C_b^2(E)$: $H_0[t]f(x) = \mathcal{H}_0[t](x, \nabla f(x))$ and where

$$\mathcal{H}_0[t]((\mu, w), p) = \sum_{(a,b) \in \Gamma} \mu(a) r(t, a, b, \mu) \left[\exp\left\{ p_b - p_a + p_{(a,b)} \right\} - 1 \right]. \tag{6.4}$$

Following (4.2), we find that for f with $e^{nf} \in \mathcal{D}(\vec{A}_n)$ that

$$H_n f(t, \mu, w) := \frac{1}{n} e^{-nf(t,\mu,w)} \cdot \vec{A}_n e^{nf}(t, \mu, w)$$
$$= \partial_t f(t, \mu, w) + H_n[t]f(t, \mu, w). \tag{6.5}$$

We similarly define the Hamiltonian $H_n[t]$ at time t in terms of the generator $A_n[t]$ at time t. For f such that $e^{nf} \in \mathcal{D}(A_n[t])$, set

$$H_n[t]f(t, \mu, w) := \frac{1}{n} e^{-nf(t,\mu,w)} \cdot A_n[t] e^{nf}(t, \mu, w), = \frac{1}{n} e^{-nf(t,\mu,w)} \cdot (A_n[t] e^{nf}(t, \cdot, \cdot))(\mu, w).$$

Finally, denote for $(\hat{a}, \hat{b}) \in \Gamma$ the measure $\mu^{\hat{a}, \hat{b}} = \mu + \frac{1}{n}\left(\delta_{\hat{b}} - \delta_{\hat{a}}\right)$ and flux $w^{\hat{a}, \hat{b}} = w + \frac{1}{n}\delta_{(\hat{a}, \hat{b})}$.

Lemma 6.2. *Consider the setting of Proposition 6.1. Let $f \in C_b^2(E)$ and let $h_n : \mathbb{R}^+ \times E_n \to \mathbb{R}$ be functions such that*

$$\lim_{n \to \infty} \sup_{s \in \mathbb{R}^+, (\mu, w) \in E_n} \sup_{(\hat{a}, \hat{b}) : \mu(\hat{a}) > 0} n \left| h_n(s, \mu^{\hat{a}, \hat{b}}, w^{\hat{a}, \hat{b}}) - h_n(s, \mu, w) \right| = 0. \tag{6.6}$$

We then have that $\sup_n \sup_s |H_n[s](f + h_n)| < \infty$ and

$$\lim_{n \to \infty} H_n[\gamma_n^{-1}s](f + h_n)(s, \mu_n, w_n) = H_0[s]f(\mu, w) \tag{6.7}$$

for any sequence $(\mu_n, w_n) \in E_n$ such that $(\mu_n, w_n) \to (\mu, w)$ uniformly in $s \geqslant 0$.

Proof. Fix $f \in C_b^2(E)$ and h_n satisfying (6.6). We will prove that $\sup_n \sup_s |H_n[s](f + h_n)| < \infty$ and that for any sequence $(\mu_n, w_n) \in E_n$ such that $(\mu_n, w_n) \to (\mu, w)$, we have

$$\lim_{n \to \infty} H_n[\gamma_n^{-1}s](f + h_n)(t_n, \mu_n, w_n) = H_0[s]f(\mu, w) \tag{6.8}$$

22

uniformly in s. We consider the left-hand side:

$$H_n[\gamma_n^{-1}s](f+h_n)(\mu_n,w_n)$$
$$= \sum_{(a,b)\in\Gamma} \mu_n(a)r_n(\gamma_n^{-1}s,a,b,\mu_n)$$
$$\times \left[e^{n\left(f(\mu_n^{a,b},w_n^{a,b}-f(\mu_n,w_n)+h_n(\mu_n^{a,b},w_n^{a,b})-h(\mu_n,w_n)\right)}-1\right]$$
$$= \sum_{(a,b)\in\Gamma} \mu_n(a)r_n(\gamma_n^{-1}s,a,b,\mu_n)\left[e^{n\left(f(\mu_n^{a,b},w_n^{a,b}-f(\mu_n,w_n))+o(1)\right)}-1\right],$$

where $o(1)$ is a term that vanishes in n uniformly in all parameters by (6.6). As $f \in C_b(E)$, a first order Taylor expansion of f around (μ_n, w_n), using Assumption 3.1 (b) and that $(\mu_n, w_n) \to (\mu, w)$, we find indeed that (6.8) holds uniformly in s. Note that the first order expansion of f in the exponent can also be used to establish that $\sup_n \sup_s |H_n[s](f+h_n)| < \infty$. $\qquad\square$

Next, we take care of the fluctuating rates. Note that to prove $H \subseteq ex - \mathrm{LIM}\, H_n$, we need to find for each $f \in \mathcal{D}(H)$ a sequence $f_n \in \mathcal{D}(H_n)$ such that $\mathrm{LIM}\, f_n = f$ and $\mathrm{LIM}\, H_n f_n = Hf$. By definition, we have each $f \in \mathcal{D}(H)$ that $\mathrm{LIM}\, f = f$. This, combined with the argument in the proof of Lemma 6.2 would be sufficient to establish the convergence of operators if there were no time periodicity. Our context, however, is more difficult. We will modify the functions $f \in \mathcal{D}(H_n)$ with a perturbative term, itself oscillating, that will cancel out the oscillatory behaviour of the jump rates.

Lemma 6.3. *Consider the setting of Proposition 6.1. Fix $f \in C_b^2(E)$. For each n, define $F_{f,n} : \mathbb{R}^+ \times E_n$ as*

$$F_{f,n}(t,\mu,w) := \int_0^t H_n[s]f(\mu,w)ds - \frac{t}{\gamma_n^{-1}T_0}\int_0^{\gamma_n^{-1}T_0} H_n[s]f(\mu,w)ds.$$

Then

(a) $F_{f,n}(t+\gamma_n^{-1}T_0,\mu,w) = F_{f,n}(t,\mu,w)$ *for all n and (t,μ,w),*

(b) *We have $\lim_n |F_{f,n}| = 0$.*

(c) *The functions $F_{f,n}$ satisfy the following Lipschitz estimate*

$$\lim_{n\to\infty} \sup_{s\in\mathbb{R}^+,(\mu,w)\in E_n} \sup_{(\hat{a},\hat{b}):\mu(\hat{a})>0} n\left|F_{f,n}(\mu^{\hat{a},\hat{b}},w^{\hat{a},\hat{b}}) - F_{f,n}(\mu,w)\right| = 0. \quad (6.9)$$

Proof. Property (a) is immediate as $F_n(\gamma_n^{-1}T_0,\mu,w) = 0$ and the $\gamma_n^{-1}t_0$ periodicity of the jump-rates of Assumption 3.1 (a). Due to this periodicity, it suffices for (b) to establish

$$\sup_{t\in[0,\gamma_n^{-1}T_0],(\mu,w)\in E_n} |F_n(t,\mu,w)| \to 0.$$

By a change of variables $u = \gamma_n s$, we find

$$F_n(t,\mu,w) = \gamma_n^{-1}\int_0^{\gamma_n t} H_n[\gamma_n^{-1}u]f(\mu,w)du - \frac{t}{T_0}\int_0^{T_0} H_n[\gamma_n^{-1}u]f(\mu,w)du.$$

Using (6.7) with $h_n = 0$, we can replace, up to uniform errors c_n that satisfy $c_n \to 0$, the integrands by $H[u]f(\mu,w)$. This implies

$$F_n(t,\mu,w) = c_n + \gamma_n^{-1}\int_0^{\gamma_n t} H[u]f(\mu,w)du - \frac{t}{T_0}\int_0^{T_0} H[u](\mu,w)du$$

which gives

$$\sup_{t\in[0,\gamma_n^{-1}T_0],(\mu,w)\in E_n} |F_n(t,\mu,w)| \leqslant c_n + \gamma_n^{-1}C$$

concluding the proof of (b). For the proof of (c), we will first establish that here is a constant C_f such that for any n, any $(\mu,w)\in E_n$, $(\hat{a},\hat{b})\in\Gamma$ such that $\mu(\hat{a}) > 0$ and $s\geqslant 0$ we have

$$n\left|H_n[s]f\left(\mu+\frac{1}{n}(\delta_{\hat{b}}-\delta_{\hat{a}}),w+\frac{1}{n}\delta_{(\hat{a},\hat{b})}\right) - H_n[s]f(\mu,w)\right| \leqslant C_f. \qquad (6.10)$$

Applying the definition of H_n yields

$$H_n[s]f\left(\mu^{\hat{a},\hat{b}},w^{\hat{a},\hat{b}}\right) - H_n[s]f(\mu,w)$$
$$= \sum_{(a,b)\in\Gamma} \mu^{\hat{a},\hat{b}}(a)r_n(s,a,b,\mu^{\hat{a},\hat{b}})\left[e^{\frac{1}{n}\left(f(\mu^{\hat{a},\hat{b}}+\frac{1}{n}(\delta_b-\delta_a),w^{\hat{a},\hat{b}}+\frac{1}{n}\delta_{(a,b)})\right)}-1\right]$$
$$- \sum_{(a,b)\in\Gamma} \mu(a)r_n(s,a,b,\mu)\left[e^{\frac{1}{n}\left(f(\mu+\frac{1}{n}(\delta_b-\delta_a),w+\frac{1}{n}\delta_{(a,b)})\right)}-1\right].$$

Standard arguments for obtaining Lipschitz estimates using Assumption 3.1 (d) and $f\in C_b^2(E)$ yield (6.10). Using (6.10) in line 5, we find

$$\sup_{t\geqslant 0} n\left|F_{f,n}(t,\mu^{\hat{a},\hat{b}},w^{\hat{a},\hat{b}}) - F_{f,n}(t,\mu,w)\right|$$
$$\sup_{t\in[0,\gamma_n^{-1}T_0]} n\left|F_{f,n}(t,\mu^{\hat{a},\hat{b}},w^{\hat{a},\hat{b}}) - F_{f,n}(t,\mu,w)\right|$$
$$= \sup_{t\in[0,\gamma_n^{-1}T_0]} n\left|\int_0^t H_n[s]f(\mu^{\hat{a},\hat{b}},w^{\hat{a},\hat{b}}) - H_n[s]f(\mu,w)\,ds\right|$$
$$\leqslant \sup_{t\in[0,\gamma_n^{-1}T_0]} n\int_0^t \left|H_n[s]f(\mu^{\hat{a},\hat{b}},w^{\hat{a},\hat{b}}) - H_n[s]f(\mu,w)\right|\,ds$$
$$\leqslant \sup_{t\in[0,\gamma_n^{-1}T_0]} \int_0^t C_f\,ds$$
$$\leqslant \gamma_n^{-1}\hat{C}_f$$

establishing (c). $\qquad\square$

Proof of Proposition 6.1. Recall that for each $f\in\mathcal{D}(H) = C_b^2(E)$, we need to establish the existence of $f_n\in C_b(\mathbb{R}^+\times E_n)$ such that

$$\text{LIM}\, f_n = f, \qquad (6.11)$$
$$\text{LIM}\, H_n f_n = Hf. \qquad (6.12)$$

Fix $f\in C_b^2(E)$. Using the functions $F_{f,n}$ from Lemma 6.3, we define a suitable collection of functions f_n that approximate f and which take care of the periodic behaviour in the Hamiltonian:

$$f_n(t,\mu,w) := f(\mu,w) - \left(\int_0^t H_n[s]f(\mu,w)ds - \frac{t}{\gamma_n^{-1}T_0}\int_0^{\gamma_n^{-1}T_0} H_n[s]f(\mu,w)ds\right)$$
$$= f(\mu,w) - F_{f,n}(t,\mu,w).$$

By Lemma 6.3, we have (6.11). We proceed with establishing (6.12). We use the form in (6.5) to establish this result.

Let $(t_n, \mu_n, w_n) \in \mathbb{R}^+ \times E_n$ be such that $\mu_n, w_n \to (\mu, w)$ and $\sup_n t_n < \infty$. Note that the application of $H_n[t_n]$ to $f - F_{f,n}$ and the application of the time derivative to the first integral term of $-F_{f,n}$ yield

$$H_n[t_n](f - F_{f,n})(t_n, \mu_n, w_n) - H_n[t_n]f(\mu_n, w_n)$$

which converges to 0 by Lemma 6.2 as (6.9) implies (6.6). We thus obtain the final expression

$$
\begin{aligned}
H_n(f_n - F_{f,n})(t_n, \mu_n, w_n) &= \frac{1}{\gamma_n T_0} \int_0^{\gamma_n^{-1} T_0} H_n[s] r(s, \mu_n, w_n) \, ds + o(1), \\
&= \frac{1}{T_0} \int_0^{T_0} H_n[\gamma_n^{-1} u] f(\mu_n, w_n) \, du + o(1), \quad (6.13)
\end{aligned}
$$

which does not depend on t_n. Using (6.7) of Lemma 6.2 and the dominated convergence theorem, this yields

$$\lim_n H_n(f_n - F_{f,n})(t_n, \mu_n, w_n) = Hf(\mu, w)$$

establishing (6.12).

$\qquad\square$

6.2 Verifying exponential tightness

The next condition in Theorem 4.4 is exponential tightness.

Proposition 6.4. *The processes* $t \mapsto (t_0 + t, \mu_n(t), w_n(t))$ *started in a compact set are exponentially tight on* $D_E(\mathbb{R}^+)$.

It is well known in the context of weak convergence that tightness follows from compact containment and the convergence of generators. The same holds in the context of large deviations. The following proposition is the exponential compact containment condition. This property, combined with the convergence of operators established in Proposition 6.1 yields the result by [16, Corollary 4.19] or [19, Proposition 7.12].

Proposition 6.5. *For each compact set* $K \subseteq (\mathbb{R}^+)^\Gamma$, $T_0 > 0$, $T > 0$ *and* $a > 0$, *there is a compact set* $\hat{K} \subseteq (\mathbb{R}^+)^\Gamma$ *depending on* K, T, a *such that*

$$\limsup_{n \to \infty} \sup_{(t_0, \mu_0, w_0): t_0 \leqslant T_0, w \in K} \frac{1}{n} \log \mathbb{P}\left[w_n(t) \notin \hat{K} \,\middle|\, (t(0), \mu_n(0), w_n(0)) = (t_0, \mu_0, w_0)\right] \leqslant -a.$$

The proof is based on a standard martingale argument and is given for completeness, see e.g. Section 4.6 of [16].

Proof. Recall that containment function $\Upsilon(\mu, w) = \Upsilon(w) = \sum_{(a,b) \in \Gamma} \log(1 + w_{(a,b)})$ introduced in Lemma 5.8 and that the argument in its proof also yields

$$\sup_{\mu, w} \sup_t \mathcal{H}[t]((\mu, w), \nabla \Upsilon(\mu, w)) =: c_{0,\Upsilon} < \infty.$$

Choose $\beta > 0$ such that $T c_{0,\Upsilon} + 1 - \beta \leqslant -a$. As Υ has compact sublevel sets, we can choose a c such that

$$K \subseteq \{(\mu, w) \,|\, \Upsilon(\mu, w) \leqslant c\}$$

Next, set $G := \{w \,|\, \Upsilon(w) < c + \beta\}$ and let $\hat{K}$ be the closure of G (which is a compact set). Let $f(x) := \iota \circ \Upsilon$ where ι is some smooth increasing function such that

$$
\iota(r) = \begin{cases} r & \text{if } r \leqslant c + \beta, \\ c + \beta + 1 & \text{if } r > c + \beta + 2. \end{cases}
$$

25

Set $g_n := H_n f$. By definition it follows that $LIM\, f = f$. We now bound $H_n f$ from above using that f is derived from Υ.

By (6.5), we find

$$\sup_{\mu,w,t} H_n f(t,\mu,w) = \sup_{\mu,w,t} H_n[t] f(t,\mu,w)$$
$$= \sup_{\mu,w,t} H_n[t] f(\gamma_n^{-1} t, \mu, w).$$

Noting that $g(\mu,w) = \mathcal{H}(\mu,w,\nabla\Upsilon(\mu,w)) \leqslant c_{0,\gamma}$ if $w \in \hat{K}$, we find by Lemma 6.2 that

$$\limsup_n \sup_{t,\mu,w} H_n f(t,\mu,w) \leqslant \sup_{t,\mu,w} H_0[t] f(\mu,w) \leqslant c_{0,\gamma}. \tag{6.14}$$

We now define a martingale that we will use to control the probability of leaving the set G. let

$$M_n(t) := \exp\left\{ n \left(f(\mu_n(t), w_n(t))) - f(\mu_n(0), w_n(0)) - \int_0^t g_n(s, \vec{X}_n(s), W_n(s)) ds \right) \right\}.$$

Let τ be the stopping time $\tau := \inf\{t \geqslant 0 \,|\, w_n(t) \notin G\}$.

By construction M_n is a martingale and by the optional stopping theorem $t \mapsto M_n(t \wedge \tau)$ is a martingale also. We obtain that if the process is started at (t_0, μ_0, w_0) such that $w_0 \in K$:

$$\mathbb{P}\left[w_n(t) \notin \hat{K} \text{ for some } t \in [0,T] \right]$$
$$\leqslant \mathbb{P}\left[w_n(t) \notin G \text{ for some } t \in [0,T]\right]$$
$$= \mathbb{E}\left[\mathbb{1}_{\{w_n(t) \notin G \text{ for some } t \in [0,T]\}} M_n(t \wedge \tau) M_n(t \wedge \tau)^{-1} \right]$$
$$\leqslant \exp\left\{ -n \left(\inf_{w \notin G} \Upsilon(w) - \sup_{w \in K} \Upsilon(w) \right.\right.$$
$$\left.\left. -T \sup_{(\mu,w) \in \mathcal{P}_n(1,\ldots,q) \times G} \sup_t g_n(t,\mu,w) \right) \right\}$$
$$\times \mathbb{E}\left[\mathbb{1}_{\{w_n(t) \notin G \text{ for some } t \in [0,T]\}} M_n(t \wedge \tau) \right].$$

Using (6.14), we obtain that the term in the exponential is bounded by $n\,(c_{0,\gamma} T - \beta) \leqslant -n\alpha$ for sufficiently large n. The final expectation is bounded by 1 due to the martingale property of $M_n(t \wedge \tau)$. This establishes the claim.

$\square$

6.3 Establishing the Lagrangian form of the rate function

Proposition 6.6. *The rate function of Theorem 4.4 can be re-expressed in variational form as in Theorem 3.3.*

Proof. The result follows from a combination of the outcomes of Theorem 4.4 with Theorem [16, Theorem 8.27] and [16, Theorem 8.14]. We argue in three steps.

(1) We come up with a new solution $\mathbf{R}(\lambda)h$ to the Hamilton-Jacobi equation $f - \lambda H f = h$ and a new semigroup $\mathbf{V}(t)$ using control methods.

(2) Using uniqueness of solutions, we infer that the new solution must equal $R(\lambda)h$ from Theorem 4.4. Similarly we find that the new semigroup $\mathbf{V}(t)$ must equal $V(t)$. This leads to a new representation of the rate-function in terms of a Lagrangian $\widehat{\mathcal{L}}$ given by the Legendre transform of $\mathcal{H}$.

(3) We show that $\widehat{\mathcal{L}} = \mathcal{L}$.

Step 1: We start with the application of [16, Theorem 8.27]. We use this result taking $\mathbf{H} = \mathbf{H}_\dagger = \mathbf{H}_\ddagger$ (in the terminology of [16]) all equal to the the operator H of our paper defined in (6.1). We furthermore use

$$\widehat{\mathcal{L}}((\mu, w), (\dot{\mu}, \dot{w})) = \sup_p \left\{ \sum_a p_a \dot{\mu}_a + \sum_{(a,b) \in \Gamma} p_{(a,b)} \dot{w}_{(a,b)} - \mathcal{H}((\mu, w), p) \right\}, \quad (6.15)$$

and

$$\mathcal{A}f((\mu, w), (\dot{\mu}, \dot{w})) \sum_a \nabla_a f(\mu, w) \cdot \dot{\mu}_a + \sum_{(a,b) \in \Gamma} \nabla_{(a,b)} f(\mu, w) \cdot \dot{w}_{(a,b)}.$$

Note that by convex-duality (with respect to the velocity-momentum variables) $\mathcal{H}$ is the Legendre transform of $\widehat{\mathcal{L}}$. The conditions for [16, Theorem 8.27] are Conditions 8.9, 8.10 and 8.11 in [16], which can be checked in a straightforward way, following the methods of e.g. [8,18], or Section 10.3.5 of [16] with $\psi = 1$. The final condition for [16, Theorem 8.27] is the comparison principle, which is the result of Theorem 3.8.

We obtain from [16, Theorem 8.27] that there are two families of operators $\mathbf{R}(\lambda)$, $\lambda > 0$ and $\mathbf{V}(t)$, $t \geqslant 0$ given in variational form

$$\mathbf{R}(\lambda)h(x) := \sup_{\gamma \in \mathcal{AC}(E), \gamma(0)=x} \int_0^\infty \lambda^{-1} e^{-\lambda^{-1}t} \left(h(\gamma(t)) - \int_0^t \widehat{\mathcal{L}}(\gamma(s), \dot{\gamma}(s)) ds \right) dt,$$

$$\mathbf{V}(t)f(x) := \sup_{\gamma \in \mathcal{AC}(E), \gamma(0)=x} f(\gamma(t)) - \int_0^t \widehat{\mathcal{L}}(\gamma(s), \dot{\gamma}(s)) ds,$$

where $x = (\mu, w)$. Similarly as in Theorem 4.4 the results of [16, Theorem 8.27 and Section 8] yield

$$\mathbf{V}(t)f(x) = \lim_{m \to \infty} \mathbf{R}^m \left(\frac{t}{m} \right) f(x). \quad (6.16)$$

and such that for $\lambda > 0$ and $h \in C_b(E)$, the function $\mathbf{R}(\lambda)h$ is the unique function that is a viscosity solution to $f - \lambda Hf = h$.

Step 2: We rewrite the rate function in Lagrangian form.

As both $R(\lambda)h$ and $\mathbf{R}(\lambda)h$ are viscosity solutions to $f - \lambda Hf = h$, the comparison principle of Theorem 3.8 yields that they are equal. By (4.3) and (6.16) we also find $\mathbf{V}(t) = V(t)$.

By a duality argument, performed in e.g. [16, Theorem 8.14] it follows that the rate function in Theorem 4.4 can be rewritten in Lagrangian form, with Lagrangian given in (6.15).

Step 3: Finally, we show that $\widehat{\mathcal{L}} = \mathcal{L}$. Note that

$$\sum_a p_a \dot{\mu}_a + \sum_{(a,b) \in \Gamma} p_{(a,b)} \dot{w}_{(a,b)}$$

$$= \sum_a p_a \left(\dot{\mu}_a - \sum_{b:(a,b) \in \Gamma} \left(\dot{w}_{(b,a)} - \dot{w}_{(a,b)} \right) \right) + \sum_{(a,b) \in \Gamma} \dot{w}_{(a,b)} \left(p_{(a,b)} - p_a + p_b \right).$$

The map $\mathcal{H}$ only depends on the combinations $p_{(a,b)} - p_a + p_b$. Therefore, taking the Legendre transform of $\mathcal{H}$, we find that $\widehat{\mathcal{L}}$ equals infinity if there is some a such that $\dot{\mu}_a \neq \sum_{b:(a,b) \in \Gamma} \left(\dot{w}_{(b,a)} - \dot{w}_{(a,b)} \right)$. In the case that for all a we have $\dot{\mu}_a = \sum_{b:(a,b) \in \Gamma} \left(\dot{w}_{(b,a)} - \dot{w}_{(a,b)} \right)$, the Legendre transform reduces to a supremum over the combinations $p_{(a,b)} - p_a + p_b$. By computing the straightforward Legendre transform of the function $r \mapsto a [e^r - 1]$ with $a > 0$ we find that indeed $\mathcal{L} = \widehat{\mathcal{L}}$.

Thus, in both cases $\mathcal{L} = \widehat{\mathcal{L}}$, establishing the result of Theorem 3.3 $\qquad\square$

27

A Viscosity solutions, auxiliary arguments

In Section 5.1, we refer at two points to results from [8]. We repeat these arguments here for sake completeness. The setting is as in Section 5.1 .
We start by establishing that we can replace our Hamiltonian H by a proper upper bound $H_\dagger$ and lower bound $H_\ddagger$.

Definition A.1. We say that $H_\dagger : \mathcal{D}(H_\dagger) \subseteq C(E) \to C(E)$ is a *viscosity sub-extension* of H if $H \subseteq H_\dagger$ (as a graph) and if for every $\lambda > 0$ and $h \in C_b(E)$ a viscosity subsolution to $f - \lambda H f = h$ is also a viscosity subsolution to $f - \lambda H_\dagger f = h$. Similarly, we define a *viscosity super-extension*.

The $H_\dagger, H_\ddagger$ that we will consider are constructed by introducing the unbounded containment function Υ into the domain:

$$\mathcal{D}(H_\dagger) := C_b^1(E) \cup \{x \mapsto (1-\varepsilon)\Psi_\alpha(x,y) + \varepsilon\Upsilon(x) + c \,|\, \alpha, \varepsilon > 0, c \in \mathbb{R}\},$$
$$\mathcal{D}(H_\ddagger) := C_b^1(E) \cup \{y \mapsto -(1+\varepsilon)\Psi_\alpha(x,y) - \varepsilon\Upsilon(y) + c \,|\, \alpha, \varepsilon > 0, c \in \mathbb{R}\}.$$

Here we write Ψ_α for the function $\alpha_1 \Psi_1 + \alpha_2 \Psi_2$. The introduction of the containment function in the domain will allow us to work on compact sets rather than on the full space.
For $f \in \mathcal{D}(H_\dagger)$, set $H_\dagger f(x) = \mathcal{H}(x, \nabla f(x))$ and for $f \in \mathcal{D}(H_\ddagger)$, set $H_\ddagger f(x) = \mathcal{H}(x, \nabla f(x))$.

Lemma A.2. *The operator* $(H_\dagger, \mathcal{D}(H_\dagger))$ *is a viscosity sub-extension of* H *and* $(H_\ddagger, \mathcal{D}(H_\ddagger))$ *is a viscosity super-extension of* H.

In the proof we need Lemma 7.7 from [16]. We recall it here for the sake of readability.

Lemma A.3 (Lemma 7.7 in [16])**.** *Let* H *and* $H_\dagger : \mathcal{D}(H_\dagger) \subseteq C(E) \to C(E)$ *be two operators. Suppose that for all* $(f, g) \in H_\dagger$ *there exist* $\{(f_n, g_n)\} \subseteq H_\dagger$ *that satisfy the following conditions:*

(a) *For all* n, *the function* f_n *is lower semi-continuous.*

(b) *For all* n, *we have* $f_n \leqslant f_{n+1}$ *and* $f_n \to f$ *point-wise.*

(c) *Suppose* $x_n \in E$ *is a sequence such that* $\sup_n f_n(x_n) < \infty$ *and* $\inf_n g_n(x_n) > -\infty$, *then* $\{x_n\}_{n \geqslant 1}$ *is relatively compact and if a subsequence* $x_{n(k)}$ *converges to* $x \in E$, *then*
$$\limsup_{k \to \infty} g_{n(k)}(x_{n(k)}) \leqslant g(x).$$

Then $H_\dagger$ *is a viscosity sub-extension of* H.
An analogous result holds for super-extensions $H_\ddagger$ *by taking* f_n *a decreasing sequence of upper semi-continuous functions and by replacing requirement (c) with*

(c′) *Suppose* $x_n \in E$ *is a sequence such that* $\inf_n f_n(x_n) > -\infty$ *and* $\sup_n g_n(x_n) < \infty$, *then* $\{x_n\}_{n \geqslant 1}$ *is relatively compact and if a subsequence* $x_{n(k)}$ *converges to* $x \in E$, *then*
$$\liminf_{k \to \infty} g_{n(k)}(x_{n(k)}) \geqslant g(x).$$

Proof of Lemma A.2. We only prove the sub-extension part.
Consider a collection of smooth functions $\phi_n : \mathbb{R} \to \mathbb{R}$ defined as $\phi_n(x) = x$ if $x \leqslant n$ and $\phi_n(x) = n + 1$ for $x \geqslant n + 1$. Note that $\phi_{n+1} \geqslant \phi_n$ for all n.
Fix a function $f \in \mathcal{D}(H_\dagger)$ of the type $f(x) = (1-\varepsilon)\Psi_\alpha(x,y) + \varepsilon\Upsilon(x) + c$ and write $g = H_\dagger f$. Moreover set $f_n := \phi_n \circ f$. Since f is bounded from below, $f_n \in C_c^2(E)$

for all n and as $n \mapsto \phi_n$ is increasing also $n \mapsto f_n$ is increasing and $\lim_n f_n = f$ point-wise.

As $f_n \in C_c^2(E)$, we have $f_n \in \mathcal{D}(H)$ and we can write $g_n = H f_n$.

We verify conditions of Lemma A.3 for (f_n, g_n) and (f, g). (a) and (b) have already been verified above. For (c), let $\{x_n\}_{n \geqslant 1}$ be a sequence such that $\sup_n f_n(x_n) = M < \infty$. It follows by the compactness of the level sets of Υ and the positivity of Ψ_α that the sequence $\{x_n\}_{n \geqslant 1}$ is contained in the compact set

$$K := \{z \in E \mid f(z) \leqslant M + 1\}.$$

Note that K has non-empty interior by the assumptions on Ψ_α and Υ. In particular, if h_1, h_2 are continuously differentiable and if $h_1(z) = h_2(z)$ for $z \in K$, then $\nabla h_1(z) = \nabla h_2(z)$ for $z \in K$.

Suppose $x_{n(k)}$ is a subsequence converging to some point x. As f is bounded on K, there exists a sufficiently large N such that for all $n \geqslant N$ and $y \in K$, we have $f_n(y) = f(y)$ and

$$g_n(y) = \mathcal{H}(y, \nabla f_n(y)) = \mathcal{H}(y, \nabla f(y)) = g(y).$$

Thus, we have $\limsup_k g_{n(k)}(x_{n(k)}) \leqslant g(x)$. $\qquad \square$

We proceed with a standard argument that is needed for the proof of Proposition 5.6. It is a copy of the argument of Proposition A.9 of [8].

Lemma A.4. *Consider the setting of Proposition 5.6. Then it holds that*

$$\sup_x u(x) - v(x)$$

$$\leqslant \frac{h(x_{\alpha,\varepsilon})}{1 - \varepsilon} - \frac{h(y_{\alpha,\varepsilon})}{1 + \varepsilon}$$

$$+ \frac{\varepsilon}{1 - \varepsilon} \mathcal{H}(x_{\alpha,\varepsilon}, \nabla \Upsilon(x_{\alpha,\varepsilon})) + \frac{\varepsilon}{1 + \varepsilon} \mathcal{H}(y_{\alpha,\varepsilon}, \nabla \Upsilon(y_{\alpha,\varepsilon}))$$

$$+ \lambda \left[\mathcal{H}\left(x_{\alpha,\varepsilon}, \sum_{i=1}^{2} \alpha_i \nabla \Psi_i(\cdot, y_{\alpha,\varepsilon})(x_{\alpha,\varepsilon}) \right) - \mathcal{H}\left(y_{\alpha,\varepsilon}, -\sum_{i=1}^{2} \alpha_i \nabla \Psi_i(x_{\alpha,\varepsilon}, \cdot)(y_{\alpha,\varepsilon}) \right) \right]$$

Proof. For sake of readability, we write $\Psi_\alpha = \alpha_1 \Psi_1 + \alpha_2 \Psi_2$.

By Lemma A.2 we get immediately that u is a subsolution to $f - \lambda H_\dagger f = h$ and v is a supersolution to $f - \lambda H_\ddagger f = h$. Thus, it suffices to verify the comparison principle for the equations involving the extensions $H_\dagger$ and $H_\ddagger$.

By Remark 2.5, we can find $x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon} \in E$ such that (5.2) is satisfied and such that

$$u(x_{\alpha,\varepsilon}) - \lambda \mathcal{H}\left(x_{\alpha,\varepsilon}, (1 - \varepsilon) \nabla \Psi_\alpha(\cdot, y_{\alpha,\varepsilon})(x_{\alpha,\varepsilon}) + \varepsilon \nabla \Upsilon(x_{\alpha,\varepsilon}) \right) \leqslant h(x_{\alpha,\varepsilon}), \qquad (A.1)$$

$$v(y_{\alpha,\varepsilon}) - \lambda \mathcal{H}\left(y_{\alpha,\varepsilon}, -(1 + \varepsilon) \nabla \Psi_\alpha(x_{\alpha,\varepsilon}, \cdot)(y_{\alpha,\varepsilon}) - \varepsilon \nabla \Upsilon(y_{\alpha,\varepsilon}) \right) \geqslant h(y_{\alpha,\varepsilon}). \qquad (A.2)$$

For all α we have

$$\sup_x u(x) - v(x)$$

$$= \lim_{\varepsilon \to 0} \sup_x \frac{u(x)}{1 - \varepsilon} - \frac{v(x)}{1 + \varepsilon}$$

$$\leqslant \liminf_{\varepsilon \to 0} \sup_{x,y} \frac{u(x)}{1 - \varepsilon} - \frac{v(y)}{1 + \varepsilon} - \Psi_\alpha(x, y) - \frac{\varepsilon}{1 - \varepsilon} \Upsilon(x) - \frac{\varepsilon}{1 + \varepsilon} \Upsilon(y)$$

$$= \liminf_{\varepsilon \to 0} \frac{u(x_{\alpha,\varepsilon})}{1 - \varepsilon} - \frac{v(y_{\alpha,\varepsilon})}{1 + \varepsilon} - \Psi_\alpha(x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon}) - \frac{\varepsilon}{1 - \varepsilon} \Upsilon(x_{\alpha,\varepsilon}) - \frac{\varepsilon}{1 + \varepsilon} \Upsilon(y_{\alpha,\varepsilon})$$

$$\leqslant \liminf_{\varepsilon \to 0} \frac{u(x_{\alpha,\varepsilon})}{1 - \varepsilon} - \frac{v(y_{\alpha,\varepsilon})}{1 + \varepsilon}, \qquad (A.3)$$

as Υ and Ψ_α are non-negative functions. We now aim to use that u and v are viscosity sub- and supersolutions. For all $z \in E$, the map $p \mapsto \mathcal{H}(z, p)$ is convex. Thus, (A.1) implies that

$$u(x_{\alpha,\varepsilon}) \leqslant h(x_{\alpha,\varepsilon}) + (1-\varepsilon)\lambda H(x_{\alpha,\varepsilon}, \nabla\Psi_\alpha(\cdot, y_{\alpha,\varepsilon})(x_{\alpha,\varepsilon})) \\ + \varepsilon\lambda\mathcal{H}(x_{\alpha,\varepsilon}, \nabla\Upsilon(x_{\alpha,\varepsilon})). \quad (A.4)$$

We aim for a complementary inequality for v. First note that because Ψ_1, Ψ_2 are such that $-(\nabla\Psi_\alpha(x_{\alpha,\varepsilon}, \cdot))(y_{\alpha,\varepsilon}) = \nabla\Psi_\alpha(\cdot, y_{\alpha,\varepsilon})(x_{\alpha,\varepsilon})$. Next, we need a more sophisticated bound using the convexity of $\mathcal{H}$:

$$H(y_{\alpha,\varepsilon}, \nabla\Psi_\alpha(\cdot, y_{\alpha,\varepsilon})(x_{\alpha,\varepsilon})) \\ \leqslant \frac{1}{1+\varepsilon}H(y_{\alpha,\varepsilon}, (1+\varepsilon)\nabla\Psi_\alpha(\cdot, y_{\alpha,\varepsilon})(x_{\alpha,\varepsilon}) - \varepsilon\nabla\Upsilon(y_{\alpha,\varepsilon})) + \frac{\varepsilon}{1+\varepsilon}H(y_{\alpha,\varepsilon}, \nabla\Upsilon(y_{\alpha,\varepsilon})).$$

Thus, (A.2) gives us

$$v(y_{\alpha,\varepsilon}) \geqslant h(y_{\alpha,\varepsilon}) + \lambda(1+\varepsilon)H(y_{\alpha,\varepsilon}, \nabla\Psi_\alpha(\cdot, y_{\alpha,\varepsilon})(x_{\alpha,\varepsilon})) - \varepsilon\lambda H(y_{\alpha,\varepsilon}, \nabla\Upsilon(y_{\alpha,\varepsilon})). \quad (A.5)$$

By combining (A.3) with (A.4) and (A.5), we find

$$\sup_x u(x) - v(x)$$
$$\leqslant \liminf_{\varepsilon \to 0} \liminf_{\alpha \to \infty} \left\{ \frac{h(x_{\alpha,\varepsilon})}{1-\varepsilon} - \frac{h(y_{\alpha,\varepsilon})}{1+\varepsilon} \right.$$
$$+ \frac{\varepsilon}{1-\varepsilon}\mathcal{H}(x_{\alpha,\varepsilon}, \nabla\Upsilon(x_{\alpha,\varepsilon})) + \frac{\varepsilon}{1+\varepsilon}\mathcal{H}(y_{\alpha,\varepsilon}, \nabla\Upsilon(y_{\alpha,\varepsilon}))$$
$$\left. + \lambda\left[\mathcal{H}(x_{\alpha,\varepsilon}, \nabla\Psi_\alpha(\cdot, y_{\alpha,\varepsilon})(x_{\alpha,\varepsilon})) - \mathcal{H}(y_{\alpha,\varepsilon}, \nabla\Psi_\alpha(\cdot, y_{\alpha,\varepsilon})(x_{\alpha,\varepsilon}))\right] \right\}.$$

This establishes the claim. $\qquad\square$

References

[1] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations.* Boston, MA: Birkhäuser, 1997.

[2] L. Bertini, R. Chetrite, A. Faggionato, and D. Gabrielli. Level 2.5 large deviations for continuous-time markov chains with time periodic rates. *Annales Henri Poincaré*, 19(10):3197–3238, 10 2018.

[3] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim. Macroscopic fluctuation theory for stationary non-equilibrium states. *J. Statist. Phys.*, 107(3-4):635–675, 2002.

[4] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim. Large deviations of the empirical current in interacting particle systems. *Teor. Veroyatn. Primen.*, 51(1):144–170, 2006.

[5] L. Bertini, A. Faggionato, and D. Gabrielli. Large deviations of the empirical flow for continuous time Markov chains. *Ann. Inst. H. Poincaré Probab. Statist.*, 51(3):867–900, 08 2015.

[6] V. S. Borkar and R. Sundaresan. Asymptotics of the invariant measure in mean field models with jumps. *Stoch. Syst.*, 2(2):322–380, 2012.

[7] K. N. C. Maes and B. Wynants. On and beyond entropy production; the case of Markov jump processes. *Markov Processes and Related Fields*, 14:445–464, 2008.

[8] F. Collet and R. C. Kraaij. Dynamical moderate deviations for the Curie-Weiss model. *Stochastic Processes and their Applications*, 127(9):2900 – 2925, 2017.

[9] F. Comets. Nucleation for a long range magnetic model. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 23(2):135–178, 1987.

[10] C. Costantini and T. Kurtz. Viscosity methods giving uniqueness for martingale problems. *Electron. J. Probab.*, 20:no. 67, 1–27, 2015.

[11] M. G. Crandall, H. Ishii, and P.-L. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Am. Math. Soc., New Ser.*, 27(1):1–67, 1992.

[12] P. Dai Pra and F. den Hollander. Mckean-Vlasov limit for interacting random processes in random media. *Journal of Statistical Physics*, 84(3-4):735–772, 1996.

[13] X. Deng, J. Feng, and Y. Liu. A singular 1-D Hamilton-Jacobi equation, with application to large deviation of diffusions. *Communications in Mathematical Sciences*, 9(1), 2011.

[14] P. Dupuis, K. Ramanan, and W. Wu. Large deviation principle for finite-state mean field interacting particle systems. *preprint; ArXiv:1601.06219*, 2016.

[15] S. N. Ethier and T. G. Kurtz. *Markov processes: Characterization and Convergence.* Wiley, 1986.

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

[17] W. H. Fleming and H. M. Soner. *Controlled Markov processes and viscosity solutions. 2nd ed.* New York, NY: Springer, 2nd ed. edition, 2006.

[18] R. C. Kraaij. Large deviations for finite state Markov jump processes with mean-field interaction via the comparison principle for an associated Hamilton–Jacobi equation. *Journal of Statistical Physics*, 164(2):321–345, 2016.

[19] R. C. Kraaij. The exponential resolvent of a markov process and large deviations for markov processes via hamilton-jacobi equations. *preprint; ArXiv:1910.03946*, 2019.

[20] R. 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.

[21] R. C. Kraaij and M. C. Schlottke. Comparison Principle for Generalized Hamilton-Jacobi-Bellman Equations via a Bootstrapping Procedure. *preprint; ArXiv:1912.06579*, 2019.

[22] R. C. Kraaij and M. C. Schlottke. Large deviations for two-scale interacting markov processes. *Work in progress*, 2020.

[23] C. Léonard. Large deviations for long range interacting particle systems with jumps. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 31(2):289–323, 1995.

[24] R. I. A. Patterson and D. R. M. Renger. Large deviations of jump process fluxes. *Mathematical Physics, Analysis and Geometry*, 22(3):21, Sep 2019.

[25] E. Perkins. Dawson-Watanabe superprocesses and measure-valued diffusions. In *Lectures on probability theory and statistics. Ecole d'eté de probabilités de Saint-Flour XXIX - 1999, Saint-Flour, France, July 8–24, 1999*, pages 125–329. Berlin: Springer, 2002.

[26] D. R. M. Renger. Flux large deviations of independent and reacting particle systems, with implications for macroscopic fluctuation theory. *Journal of Statistical Physics*, 172(5):1291–1326, 9 2018.

[27] A. Shwartz and A. Weiss. Large deviations with diminishing rates. *Mathematics of Operations Research*, 30(2):281–310, 2005.