Comparison Principle for Generalized Hamilton-Jacobi-Bellman Equations via a Bootstrapping Procedure

Richard C. Kraaij∗ Mikola C. Schlottke†

December 16, 2019

Abstract

We study the well-posedness of Hamilton-Jacobi-Bellman equations on subsets of \(\mathbb{R}^d\). The Hamiltonian consists of two parts: an internal Hamiltonian depending on an external control variable and a cost function penalizing the control. We show under suitable assumptions that if a comparison principle holds for the Hamilton-Jacobi equation involving only the internal Hamiltonian, then the comparison principle holds for the Hamilton-Jacobi-Bellman equation involving the full Hamiltonian. In addition to establishing uniqueness, we give sufficient conditions for existence of solutions.

Our key features are that the internal Hamiltonian is allowed to be non-Lipschitz and non-coercive in the momentum variable, and that we allow for discontinuous cost functions. To compensate for the greater generality of our approach, we assume sufficient regularity of the cost function on its sub-level sets and that the internal Hamiltonian satisfies a comparison principle uniformly in the control variable on compact sets.

As an application, we show that our established result covers both interesting examples that were posed as open problems in the literature and mean-field Hamiltonians that can not be treated with standard methods.

Keywords: Hamilton-Jacobi-Bellman equations, comparison principle, viscosity solutions, optimal control theory

MSC2010: 49L25, 35F21

1 Introduction and aim of this note

The main purpose of this note is to establish well-posedness for first-order nonlinear partial differential equations of Hamilton-Jacobi-Bellman type on subsets \(E\) of \(\mathbb{R}^d\),

\[
    u(x) - \lambda \mathcal{H}[u(x), \nabla u(x)] = h(x), \quad x \in E \subseteq \mathbb{R}^d. \tag{HJB}
\]

†Department of Mathematics and Computer Science, Eindhoven University of Technology. E-mail address: m.schlottke@tue.nl

∗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
In there, \( \lambda > 0 \) is a scalar and \( h \) is a continuous and bounded function on \( E \). The Hamiltonian \( H : E \times \mathbb{R}^d \to \mathbb{R} \) is given by

\[
H(x, p) = \sup_{\theta \in \Theta} [\Lambda(x, p, \theta) - I(x, \theta)],
\]

(1.1)

where \( \theta \in \Theta \) plays the role of a control variable. For fixed \( \theta \), the function \( \Lambda \) can frequently be interpreted as an Hamiltonian itself. We call it the *internal* Hamiltonian. The function \( I \) can be interpreted as the cost of applying the control \( \theta \).

We will establish existence of viscosity solutions (e.g. [CIL92]) in the sense of Definition 3.1 via a resolvent that is defined in terms of a standard discounted control procedure. However, the main problem we overcome in this note is to verify a comparison principle in order to establish uniqueness of viscosity solutions. The comparison principle for Hamilton-Jacobi equations is a well-studied problem in the literature. The standard assumption that allows one to obtain the comparison principle in the context of optimal control problems (e.g. [BC97]) is that either the modulus of continuity \( \omega \) such that

\[
|H(x, p) - H(y, p)| \leq \omega(|x - y| (1 + |p|)),
\]

(1.2)

or that \( H \) is uniformly coercive:

\[
\lim_{|p| \to \infty} \inf_x H(x, p) = \infty.
\]

(1.3)

The estimate (1.2) can be translated into conditions for \( f \) and \( I \) of (2.1), which include (e.g. [BC97, Chapter III])

- \(|\Lambda(x, p, \theta) - \Lambda(y, p, \theta)| \leq \omega_\Lambda(|x - y| (1 + |p|)), \text{ uniformly in } \theta, \text{ and}
- |I(x, \theta) - I(y, \theta)| \leq \omega_I(|x - y|).

However, such type of estimates are not satisfied for the examples that we are interested in. We make these examples of (HJB) more precise in Section 2. There we also explain why the standard assumptions are not satisfied and where the challenge of solving (HJB) is pointed out in the literature. Here, we focus on the motivation for our assumptions. They mainly build up on two observations:

(i) Fix a control variable \( \theta_0 \in \Theta \) and consider the Hamiltonian \( H(x, p) := \Lambda(x, p, \theta_0) \). Then in all our examples, the comparison principle is satisfied for \( u(x) - \lambda H(x, \nabla u(x)) = h(x) \).

(ii) In all our examples, the cost function \( I(x, \theta) \) satisfies an estimate of the type \(|I(x, \theta) - I(y, \theta)| \leq \omega_I C(|x - y|)\) on sublevel sets \( \{I \leq C\} \).

The main idea of this note is to take advantage of viscosity sub- and supersolution inequalities in order to work on sublevel sets of the cost function \( I \). To do so, we assume that \( H(x, p) = \Lambda(x, p, \theta_0) \) satisfies a *continuity estimate* uniformly for \( \theta_0 \) varying in a compact set. This continuity estimate captures the key information that allows to prove the comparison principle for \( H \). In the end, this is what we call the bootstrap principle: given sufficient regularity of \( I \), one can bootstrap the comparison principle for the internal Hamiltonian \( \Lambda \) to obtain a comparison principle for the full Hamiltonian \( H \). In examples, this approach proves to be a crucial improvement over known results.

In summary, the novelties of this paper are:
(i) Motivated from examples violating the standard regularity estimate (1.2) on Hamiltonians, we find weaker conditions under which the comparison principle for (HJB) is satisfied for variational Hamiltonians $\mathcal{H}$ of the type (1.1). The result is formulated in Theorem 3.4. The main bootstrapping argument is explained in simplified form in Section 4.1 and carried out in Section 6.

(ii) A proof of the comparison principle that covers a class of non-coercive Hamiltonians which typically arises in mean-field interacting particle systems that are coupled to external variables. This example has not been treated before, and we make it explicit in Proposition 3.24 of Section 3.4.

(iii) A proof of existence of a viscosity solution based on solving subdifferential inclusions in the non-compact setting. The proof relies on continuity of $\mathcal{H}$ and finding a priori estimates on the range of solutions to associated differential inclusions. The result is formulated in Theorem 3.7, and the structure of the proof is explained in Section 4.2.

With these results established, one can study large deviation problems with two time-scales from a Hamilton-Jacobi point-of-view in more generality. This is the subject of the forthcoming companion paper [KS20] which exploits the semigroup approach to large deviations by [FK06].

We believe that the bootstrap procedure we introduce in this note has the potential to also apply to second order equations or equations in infinite dimensions. For clarity of the exposition, and the already numerous applications for this setting, we stick to the finite-dimensional first-order case. The key arguments that are used in the proof in Section 6 do not depend in a crucial way on this assumption.

In Section 2, we discuss Hamiltonians that violate the standard regularity assumptions. The main results are formulated in Section 3. We proceed with a discussion of the strategy of the proofs in Section 4. In Section 5 we establish regularity properties of $\mathcal{H}$ that are used in the later proof sections. In Section 6 we establish the comparison principle. In Section 7 we establish that a resolvent operator $R(\lambda)$ in terms of an exponentially discounted control problem gives rise to viscosity solutions of the Hamilton-Jacobi-Bellman equation (HJB). Finally, in Section 8 we verify the assumptions for examples.

2 Motivation: examples violating the standard continuity assumption

Hamiltonians of the type (1.1) arise in a range of fields. In this section, we mention three examples of Hamiltonians in rising degree of complexity; one arising from optimal control theory, and two that arise in the context of stochastic systems with two time scales. We explain why examples 2 and 3 violate the standard regularity estimates. These examples illustrate the need for an alternative set of assumptions.

Example 1. In deterministic optimal control theory (e.g. [BC97]), one aims to control the dynamical system $\dot{x} = f(x, \theta)$ where the cost associated to the control $\theta$ is given by $I(x, \theta)$. In the case of minimizing an exponentially discounted cost $J$ with final pay-off $h$, this leads to a Hamilton-Jacobi-Bellman
equation (HJB) for the value function \( u(x) := \inf_{\theta} J(x, \theta) \) with a Hamiltonian \( \mathcal{H} \) that is linear in \( p \):
\[
\mathcal{H}(x, p) = \sup_{\theta \in \Theta} \left[ \langle f(x, \theta), p \rangle - \mathcal{I}(x, \theta) \right].
\] (2.1)
We include this example for completeness in Proposition 3.18. For further background on this type of optimal control problems, we refer to [BC97].

We now come to examples of Hamiltonians of the type (1.1) that frequently arise in the study of systems with multiple time-scales, e.g. geophysical flows, planetary motion, finance, weather-climate interaction models, molecular dynamics and models in statistical physics. In such systems, one can often recognize a slow and a fast component. Typically, one is interested in the behaviour of the slow component in the limit in which the separation of time scales goes to infinity. As the fast system equilibrates before the slow system has made a significant difference, the limit of such systems can be described by an ordinary or partial differential equation involving only the average behaviour of the fast component.

However, in applications an infinite separation of time scales is never achieved. Thus, the slow process still shows fluctuations around its limiting behaviour while the fast process fluctuates around its average. The effective fluctuations arise from the combination of both sources. In this two-scale context, when analysing the fluctuations by means of large-deviation techniques, one obtains Hamiltonians of the type (1.1). We refer to [KP17] for derivations in this context, and to [BGTVE16] for an extensive explanation in which the authors study ODE’s coupled to fast diffusion. In these examples, the internal Hamiltonians \( \Lambda \) capture the fluctuations of the slow component, while the cost function \( \mathcal{I} \) arises from fluctuations of averages of the fast component. The full Hamiltonian \( \mathcal{H} \) takes both contributions into account.

**Example 2.** In [BDG18], the authors study large deviations of a diffusion processes with vanishing noise on \( E = \mathbb{R}^d \) coupled to a fast jump process on a finite discrete set \( \{1, \ldots, J\} \). They identified the challenge of proving comparison principles for Hamiltonians arising in such two-scale systems, where the Hamiltonians can be casted in the form (1.1). We consider this general setting in Proposition 3.20 in Section 3.4. We illustrate the issues arising in a simpler but more concrete form. With \( d = 1 \) and \( J = 2 \), when approaching this problem from the Hamilton-Jacobi perspective, a key step (e.g. [KP17]) is to solve (HJB) with \( \mathcal{H} \) consisting of the following ingredients:

(i) The internal state space is \( E = \mathbb{R}^d \).
(ii) The set of control variables is \( \Theta = \mathcal{P}(\{1, 2\}) \).
(iii) The internal Hamiltonian \( \Lambda \) is given by
\[
\Lambda(x, p, \theta) = \frac{1}{2} a(x, 1) |p|^2 \theta_1 + \frac{1}{2} a(x, 2) |p|^2 \theta_2,
\]
where \( a(x, i) > 0 \) and \( \theta_i = \theta(\{i\}) \).
(iv) The cost function \( \mathcal{I} \) is given by
\[
\mathcal{I}(x, \theta) = \sup_{w \in \mathbb{R}^2} \left[ r_{12}(x) \theta_1 (1 - e^{w_2 - w_1}) + r_{21}(x) \theta_2 (1 - e^{w_1 - w_2}) \right],
\]
where \( r_{ij}(x) \geq 0 \).
In this example, the cost function is unbounded if \( r_{ij}(x) \) is unbounded. For instance, consider \( \theta_1 = 1 \) and \( \theta_2 = 0 \). Then by choosing \( w = (1, 0) \) in the supremum,
\[
\mathcal{I}(x, \theta) \geq C r_{12}(x),
\]
and thus \( \mathcal{I}(x, \theta) \) diverges as \( |x| \to \infty \). Establishing a general framework that also covers examples of this type is one key motivation for this note.

We now turn to another notable problem with two time-scales that motivates this paper: a system of mean-field interacting particles coupled to fast external variables.

**Example 3.** In [BCFG18], the authors prove large-deviation principles of mean-field interacting particles that are coupled to fast time-periodic variables. In this setting, the associated Hamilton-Jacobi equations are solved in [Kra17]. However, when considering a coupling to general fast random variables such as diffusions, then solving the corresponding Hamilton-Jacobi equations remained an open challenge. In full generality, we formulate this case in Proposition 3.24. For a corresponding large-deviation analysis, we refer to our companion paper [KS20]. Here we illustrate the difficulties that arise by considering the Hamiltonian in a simplified setting:

(i) The internal state space is \( E = \mathcal{P}([a, b]) \times [0, \infty) \times [0, \infty) \), embedded in \( \mathbb{R}^4 \). We denote the variables as \( x = (\mu, w) \), with \( \mu \in \mathcal{P}([a, b]) \) and \( w \in [0, \infty)^2 \).

(ii) The set of control variables is \( \Theta = \mathcal{P}(S) \), that is the probability measures on the circle \( S \).

(iii) The internal Hamiltonian \( \Lambda \) is given by
\[
\Lambda(x, p, \theta) = \mu_a r_{ab}(\mu, \theta) \left[ \exp \{ p_b - p_a + p_{ab} \} - 1 \right] \\
+ \mu_b r_{ba}(\mu, \theta) \left[ \exp \{ p_a - p_b + p_{ba} \} - 1 \right],
\]
with \( p = (p_a, p_b, p_{ab}, p_{ba}) \in \mathbb{R}^4 \) and \( \mu_i := \mu(\{i\}) \). The rates \( r_{ij} \) are non-negative.

(iv) The cost function \( \mathcal{I} : \Theta \to [0, \infty) \) is independent of \( x \) and is given by
\[
\mathcal{I}(\theta) = \sup_{u \in \mathcal{C}_c^2(S)} \int_S \left( -\frac{u''(y)}{u(y)} \right) d\theta(y)
\]

In this example, the internal Hamiltonian \( \Lambda \) is not uniformly coercive. For instance, take momenta \( p \) such that \( p_b - p_a + p_{ab} \) is constant. Then if \( |p| \to \infty \), we do not necessarily have that \( \Lambda(x, p, \theta) \to \infty \). A similar effect occurs when choosing \( p_a \to \infty \) and \( \mu_a = 0 \). Regarding the cost function, for any singular measure \( \delta_z \) with a point \( z \in S \) we have \( \mathcal{I}(\delta_z) = \infty \). This similarly holds for finite convex combinations of Dirac measures. Since this linear span is dense in \( \mathcal{P}(S) \), this implies that \( \mathcal{I} \) cannot be continuous.

### 3 Main Results

In this section, we start with preliminaries in Section 3.1 which includes the definition of viscosity solutions and that of the comparison principle.
We proceed in Section 3.2 with the main results: a comparison principle for the Hamilton-Jacobi-Bellman equation (HJB) based on variational Hamiltonians of the form (1.1), and the existence of viscosity solutions.

In Section 3.3 we collect all the assumptions that are needed for all main results in one place and discuss the applicability of our results. In Section 3.4, we verify the assumptions for the examples that motivate the problem of this note.

### 3.1 Preliminaries

For a Polish space $\mathcal{X}$, we denote by $C(\mathcal{X})$ and $C_b(\mathcal{X})$ the spaces of continuous and bounded continuous functions respectively. If $\mathcal{X} \subseteq \mathbb{R}^d$ then we denote by $C_\infty^c(\mathcal{X})$ the space of smooth functions that vanish outside a compact set. We denote by $C_\infty^cc(\mathcal{X})$ the set of smooth functions that are constant outside of a compact set, and by $\mathcal{P}(\mathcal{X})$ the space of probability measures on $\mathcal{X}$. We equip $\mathcal{P}(\mathcal{X})$ with the weak topology, that is, the one induced by convergence of integrals against bounded continuous functions.

Throughout the paper, $E$ will be the set on which we base our Hamilton-Jacobi equations. We assume that $E$ is a subset of $\mathbb{R}^d$ that is a Polish space which is contained in the $\mathbb{R}^d$ closure of its $\mathbb{R}^d$ interior. This ensures that gradients of functions are determined by their values on $E$. Note that we do not assume that $E$ is open. We assume that the space of controls $\Theta$ is Polish.

We next introduce viscosity solutions for the Hamilton-Jacobi equation with Hamiltonians like $H(x,p)$ of our introduction.

**Definition 3.1 (Viscosity solutions and comparison principle).** Let $A : C_b(E) \to C_b(E)$ be an operator with domain $\mathcal{D}(A)$, $\lambda > 0$ and $h \in C_b(E)$. Consider the Hamilton-Jacobi equation

$$f - \lambda Af = h. \quad (3.1)$$

We say that $u$ is a *(viscosity)* subsolution of equation (3.1) if $u$ is bounded, upper semi-continuous and if, for every $f \in \mathcal{D}(A)$ there exists a sequence $x_n \in E$ such that

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

$$\lim_{n \to \infty} u(x_n) - \lambda Af(x_n) - h(x_n) \leq 0.$$

We say that $v$ is a *(viscosity)* supersolution of equation (3.1) 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

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

$$\lim_{n \to \infty} v(x_n) - \lambda Af(x_n) - h(x_n) \geq 0.$$

We say that $u$ is a *(viscosity)* solution of equation (3.1) if it is both a subsolution and a supersolution to (3.1).

We say that (3.1) satisfies the comparison principle if for every subsolution $u$ and supersolution $v$ to (3.1), we have $u \leq v$. 

6
Remark 3.2. Consider the definition of subsolutions. Suppose that the test-function \( f \in D(A) \) has compact sublevel sets, then instead of working with a sequence \( x_n \), there exists \( x_0 \in E \) such that
\[
    u(x_0) - f(x_0) = \sup_x u(x) - f(x),
\]
\[
    u(x_0) - \lambda Af(x_0) - h(x_0) \leq 0.
\]
A similar simplification holds in the case of supersolutions.

Remark 3.3. For an explanatory text on the notion of viscosity solutions and fields of applications, we refer to [CIL92].

### 3.2 Hamilton-Jacobi-Bellman Equations

In this section, we state our main results, the comparison principle in Theorem 3.4, and existence of solutions in Theorem 3.7.

Consider the variational Hamiltonian \( \mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R} \) given by
\[
    \mathcal{H}(x, p) = \sup_{\theta \in \Theta} [\Lambda(x, p, \theta) - \mathcal{I}(x, \theta)].
\]

The precise assumptions on the maps \( \Lambda \) and \( \mathcal{I} \) are formulated in Section 3.3. Our first main result is that the operator \( \mathcal{H} \) constructed out of \( \mathcal{H} \) satisfies the comparison principle.

**Theorem 3.4 (Comparison principle).** Consider the map \( \mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R} \) as in (3.2). Suppose that Assumptions 3.12 and 3.13 are satisfied for \( \Lambda \) and \( \mathcal{I} \).

Define the operator \( \mathcal{H}f(x) := \mathcal{H}(x, \nabla f(x)) \) with domain \( D(\mathcal{H}) = C^\infty_c(E) \). Then for any \( h \in C_b(E) \) and \( \lambda > 0 \), the comparison principle holds for
\[
    f - \lambda \mathcal{H}f = h.
\]

**Remark 3.5 (Uniqueness).** If \( u \) and \( v \) are two viscosity solutions of (3.3), then we have \( u \leq v \) and \( v \leq u \) by the comparison principle, giving uniqueness.

**Remark 3.6 (Domain).** The comparison principle holds with any domain that satisfies \( C^\infty_c(E) \subseteq D(\mathcal{H}) \subseteq C^1_b(E) \). We state it with \( C^\infty_c(E) \) to connect it with the existence result of Theorem 3.7, where we need to work with test functions whose gradients have compact support.

We turn to the existence of a viscosity solution for (3.3). As mentioned in the introduction, the viscosity solution is given in terms of an optimization problem with discounted cost. The Legendre dual \( \mathcal{L} : E \times \mathbb{R}^d \to [0, \infty] \) of \( \mathcal{H} \), given by
\[
    \mathcal{L}(x, v) := \sup_{p \in \mathbb{R}^d} \left( (p, v) - \mathcal{H}(x, p) \right),
\]
plays the role of a running cost. In the following Theorem, \( AC \) is the collection of absolutely continuous paths in \( E \).

**Theorem 3.7 (Existence of viscosity solution).** Consider \( \mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R} \) as in (3.2). Suppose that Assumptions 3.12 and 3.13 are satisfied for \( \Lambda \) and \( \mathcal{I} \), and that \( \mathcal{H} \) satisfies Assumption 3.16. For each \( \lambda > 0 \), let \( R(\lambda) \) be the operator
\[
    R(\lambda)h(x) = \sup_{\gamma \in AC} \int_0^\infty \lambda^{-1} e^{-\lambda^{-1}t} \left[ h(\gamma(t)) - \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s)) \right] dt.
\]

Then \( R(\lambda)h \) is the unique viscosity solution to \( f - \lambda \mathcal{H}f = h \).
3.3 Assumptions

In this section, we formulate and comment on the assumptions imposed on the Hamiltonians defined in the previous sections. We first motivate the assumptions that are required for proving the comparison principle, Theorem 3.4.

Usually, proofs of the comparison principle for a subsolution $u$ and a supersolution $v$ for the equation $f - \lambda Hf = h$ are reduced to establishing an estimate of the type

$$\liminf_{\epsilon \downarrow 0} \liminf_{\alpha \to 0} \mathcal{H}(x_{\alpha,\epsilon}, a(x_{\alpha,\epsilon} - y_{\alpha,\epsilon})) - \mathcal{H}(y_{\alpha,\epsilon}, a(x_{\alpha,\epsilon} - y_{\alpha,\epsilon})) \leq 0$$

where $(x_{\alpha,\epsilon}, y_{\alpha,\epsilon})$ are elements of $E$ such that

$$u(x_{\alpha,\epsilon}) - v(x_{\alpha,\epsilon}) - \frac{\alpha}{2}|x_{\alpha,\epsilon} - y_{\alpha,\epsilon}|^2 - \frac{\epsilon}{2}(|x_{\alpha,\epsilon}|^2 + |y_{\alpha,\epsilon}|^2) = \sup_{x,y \in E} u(x) - v(y) - \frac{\alpha}{2}|x - y|^2 - \frac{\epsilon}{2}(|x|^2 + |y|^2). \quad (3.4)$$

This final equation, and the sub- and supersolution property of $u$ and $v$ respectively, have the following consequences:

1. For all $\epsilon > 0$, the set $\{x_{\alpha,\epsilon}, y_{\alpha,\epsilon} | \alpha > 0\}$ is relatively compact in $E$;
2. For all $\epsilon > 0$, we have $|x_{\alpha,\epsilon} - y_{\alpha,\epsilon}| + |x_{\alpha,\epsilon} - y_{\alpha,\epsilon}|^2 \to 0$ as $\alpha \to \infty$;
3. We have for all $\epsilon > 0$ that

$$\inf_{\alpha} \mathcal{H}(x_{\alpha,\epsilon}, a(x_{\alpha,\epsilon} - y_{\alpha,\epsilon})) > -\infty,$$
$$\sup_{\alpha} \mathcal{H}(y_{\alpha,\epsilon}, a(x_{\alpha,\epsilon} - y_{\alpha,\epsilon})) < \infty.$$

In our bootstrap procedure, we aim to lift the comparison principle that holds for the Hamilton-Jacobi equation in terms of $\Lambda$ to that for $H$. Thus, we need to establish an estimate of the type (3.4) under assumptions of the type (1), (2) and (3) where in addition, we have to vary our control variable $\theta$. It turns out that it suffices to vary $\theta$ in a compact set in $\Theta$ that depends on $\epsilon$. In addition, to make sure that we can bootstrap, we have to relax the sup and inf in (3) to a lim sup and lim inf. A quick look at classical proofs using coercivity of $\mathcal{H}$ show that such a relaxation does not matter in the proof of the comparison principle.

A final remark. To establish the comparison principle, the quadratic distance is not special, except for being symmetric and familiar (and suited for quadratic Hamiltonians). In various examples, and at a deeper level, the theory benefits by replacing the quadratic function by a function $\Psi$ to penalize the distance between $x$ and $y$ and $\Theta$ to penalize how far for $x$ and $y$ are away from the ‘origin’.

**Definition 3.8** (Penalization function). We say that $\Psi : E^2 \to [0, \infty)$ is a penalization function if $\Psi \in C^1(E^2)$ and if $x = y$ if and only if $\Psi(x, y) = 0$.

**Definition 3.9** (Containment function). We say that a function $Y : E \to [0, \infty)$ is a containment function for $\Lambda$ if there is a constant $c_Y$ such that

- For every $c \geq 0$, the set $\{x \mid Y(x) \leq c\}$ is compact;
- We have $\sup_{\theta} \sup_{x} \Lambda(x, \nabla Y(x), \theta) \leq c_Y$. 

8
The following is our key estimate that we will use for $G = \Lambda$ which is uniform for control variables in compact sets of $\Theta$.

**Definition 3.10 (Continuity estimate).** Let $\Psi$ be a penalization function and let $G : E \times \mathbb{R}^d \times \Theta : (x, p, \theta) \mapsto G(x, p, \theta)$ be a function. Suppose that for $\epsilon > 0$ and $\alpha > 0$, we have a collection of variables $(x_{a, \epsilon}, y_{a, \epsilon})$ in $\mathbb{E}^2$ and variables $\theta_{a, \epsilon}$ in $\Theta$. We say that this collection is fundamental for $G$ with respect to $\Psi$ if:

1. For each $\epsilon$, there are compact sets $K_\epsilon \subseteq E$ and $\hat{K}_\epsilon \subseteq \Theta$ such that for all $\alpha$ we have $x_{a, \epsilon}, y_{a, \epsilon} \in K_\epsilon$ and $\theta_{a, \epsilon} \in \hat{K}_\epsilon$.

2. For each $\epsilon > 0$, we have limit points $x_\epsilon \in K_\epsilon$ and $y_\epsilon \in \hat{K}_\epsilon$ of $x_{a, \epsilon}$ and $y_{a, \epsilon}$ as $\alpha \to \infty$. For these limit points we have

   $$\lim_{\alpha \to \infty} a \Psi(x_{a, \epsilon}, y_{a, \epsilon}) = 0, \quad \Psi(x_\epsilon, y_\epsilon) = 0.$$

3. We have

   $$\limsup_{\epsilon \to 0} \limsup_{\alpha \to \infty} G(y_{a, \epsilon}, -a(\nabla \Psi(x_{a, \epsilon}, \cdot))(y_{a, \epsilon}), \theta_{a, \epsilon}) < \infty,$$

   $$\liminf_{\epsilon \to 0} \liminf_{\alpha \to \infty} G(x_{a, \epsilon}, a(\nabla \Psi(\cdot, y_{a, \epsilon}))(x_{a, \epsilon}), \theta_{a, \epsilon}) > -\infty.$$

In other words, the operator $G$ evaluated in the proper momenta is eventually bounded from above and from below.

We say that $G$ satisfies the continuity estimate if for every fundamental collection of variables we have

$$\liminf_{\alpha \to \infty} \liminf_{\epsilon \to 0} G(x_{a, \epsilon}, a(\nabla \Psi(\cdot, y_{a, \epsilon}))(x_{a, \epsilon}), \theta_{a, \epsilon}) - G(y_{a, \epsilon}, -a(\nabla \Psi(x_{a, \epsilon}, \cdot))(y_{a, \epsilon}), \theta_{a, \epsilon}) \leq 0.$$

In Section 8.3 we verify the continuity estimate in three contexts that show that the estimate is a sensible notion that is satisfied in a wide range of contexts.

**Remark 3.11.** In Appendix B, we state a slightly more general continuity estimate on the basis of multiple penalization functions. For the first reading of the proofs below, the use of this more general setting would be confusing. We want to mention, however, that all arguments below can be carried out on the basis of this more elaborate continuity estimate. Following [Kra17] a continuity estimate of this more elaborate type can be established in the context of Markov jump processes and their fluxes.

Our first assumption essentially states that we can solve the comparison principle for the Hamilton-Jacobi equation for $\Lambda$ uniformly over compact sets in $\Theta$. In addition to this assumption, we assume $(\Lambda 5)$ which states the function $\Lambda$ grows roughly equally fast in $p$ for different control variables.

**Assumption 3.12.** The function $\Lambda : E \times \mathbb{R}^d \times \Theta \to \mathbb{R}$ in the Hamiltonian (3.2) satisfies the following.

1. The map $\Lambda : E \times \mathbb{R}^d \times \Theta \to \mathbb{R}$ is continuous and for any $(x, p)$, we have boundedness: $\|\Lambda(x, p, \cdot)\|_\Theta := \sup_{\theta \in \Theta} |\Lambda(x, p, \theta)| < \infty$.

2. For any $x \in E$ and $\theta \in \Theta$, the map $p \mapsto \Lambda(x, p, \theta)$ is convex. For $p_0 = 0$, we have $\Lambda(x, p_0, \theta) = 0$ for all $x \in E$ and all $\theta \in \Theta$. 

9
(A3) There exists a containment function \( Y : E \to [0, \infty) \) in the sense of Definition 3.9.

(A4) The function \( \Lambda \) satisfies the continuity estimate.

(A5) For every compact set \( K \subseteq E \), there exist constants \( M, C_1, C_2 \geq 0 \) such that for all \( x \in K, p \in \mathbb{R}^d \) and all \( \theta_1, \theta_2 \in \Theta \), we have

\[
\Lambda(x, p, \theta_1) \leq \max \{ M, C_1 \Lambda(x, p, \theta_2) + C_2 \}.
\]

Our next assumption is on the regularity of the cost functional \( \mathcal{I} \). They are satisfied for continuous and bounded \( \mathcal{I} \) and \( \Theta \) a compact space.

**Assumption 3.13.** The functional \( \mathcal{I} : E \times \Theta \to [0, \infty) \) in (3.2) satisfies the following.

(I1) The map \((x, \theta) \mapsto \mathcal{I}(x, \theta)\) is lower semi-continuous on \( E \times \Theta \).

(I2) For any \( x \in E \), there exists a point \( \theta_x \in \Theta \) such that \( \mathcal{I}(x, \theta_x) = 0 \).

(I3) For any \( x \in E \), compact set \( K \subseteq E \) and \( C \geq 0 \), the set \( \{ \theta \in \Theta \mid \mathcal{I}(x, \theta) \leq C \} \) is compact and \( \bigcup_{x \in K} \{ \theta \in \Theta \mid \mathcal{I}(x, \theta) \leq C \} \) is relatively compact.

(I4) For any converging sequence \( x_n \to x \) in \( E \) and sequence \( \theta_n \in \Theta \), if there is an \( M > 0 \) such that \( \mathcal{I}(x_n, \theta_n) \leq M \) for all \( n \in \mathbb{N} \), then there exists a neighborhood \( U_x \) of \( x \) and a constant \( M' > 0 \) such that for any \( y \in U_x \) and \( n \in \mathbb{N} \),

\[
\mathcal{I}(y, \theta_n) \leq M' < \infty.
\]

(I5) For every compact set \( K \subseteq E \) and each \( M \geq 0 \) the collection of functions \( \{ \mathcal{I}(\cdot, \theta) \}_{\theta \in \Theta_M} \) with

\[
\Theta_M := \{ \theta \in \Theta \mid \forall x \in K : \mathcal{I}(x, \theta) \leq M \}
\]

is equicontinuous. That is: for all \( \varepsilon > 0 \), there is a \( \delta > 0 \) such that for all \( \theta \in \Theta_M \) and \( x, y \in K \) such that \( d(x, y) \leq \delta \) we have \( |\mathcal{I}(x, \theta) - \mathcal{I}(y, \theta)| \leq \varepsilon \).

**Remark 3.14 (Gamma-convergence).** The assumptions on \( \mathcal{I} \) imply that for any sequence \( x_n \to x \) in \( E \), the functionals defined by \( \mathcal{I}_n(\theta) := \mathcal{I}(x_n, \theta) \) \( \Gamma \)-converge to \( \mathcal{I}_\infty \) defined by \( \mathcal{I}_\infty(\theta) := \mathcal{I}(x, \theta) \). We give a proof in Proposition 5.2 below.

We turn to Theorem 3.7. A key ingredient in establishing the existence of a viscosity solution to Hamilton-Jacobi equations is the existence of ‘optimally’ controlled paths. The optimal controls can, for continuously differentiable Hamiltonians, be found from the Hamiltonian flow. In our context, \( \mathcal{H} \) is not continuously differentiable. We will show in Proposition 5.1, however, that \( \mathcal{H} \) is convex in \( p \). We can therefore define the subdifferential set

\[
\partial_p \mathcal{H}(x_0, p_0) := \left\{ \xi \in \mathbb{R}^d : \mathcal{H}(x_0, p) \geq \mathcal{H}(x_0, p_0) + \xi \cdot (p - p_0) \quad (\forall p \in \mathbb{R}^d) \right\}.
\]

Instead using solutions arising from the differential solution arising from the gradient of \( \mathcal{H} \), we will use solutions to differential inclusions arising from \( \partial_p \mathcal{H} \). As our set \( E \) is not necessarily equal to \( \mathbb{R}^d \), but could be, e.g. a domain with corners like \([0, \infty)^d\), we need some conditions to make sure
that the solutions to our differential inclusions remain within $E$. Assumption 3.16 below will make sure that the Hamiltonian vector field points ‘inside’ $E$.

**Definition 3.15.** The tangent cone (sometimes also called Bouligand cotangent cone) to $E$ in $\mathbb{R}^d$ at $x$ is

$$T_E(x) := \left\{ z \in \mathbb{R}^d \mid \liminf_{\lambda \downarrow 0} \frac{d(y + \lambda z, E)}{\lambda} = 0 \right\}.$$ 

**Assumption 3.16.** The map $H : E \times \mathbb{R}^d \to \mathbb{R}$ defined in (3.2) is such that $\partial_p H(x, p) \subseteq T_E(x)$ for all $p$.

**Remark 3.17.** Assumption 3.16 is intuitively implied by the comparison principle for $H$. We therefore expect Assumption 3.16 to be satisfied in any situation in which Theorem 3.4 holds.

We argue in a simple case why this is to be expected. First of all, note that the comparison principle for $H$ builds upon the maximum principle. Suppose that $E = [0, 1]$, $f, g \in C^1_b(E)$ and suppose that $f(0) - g(0) = \sup_x f(x) - g(x)$. As 0 is a boundary point, we conclude that $f'(0) \leq g'(0)$. If indeed the maximum principle holds, we must have

$$H(0, f'(0)) = Hf(0) \leq Hg(0) = H(0, g'(0))$$

implying that $p \mapsto H(0, p)$ is increasing, in other words

$$\partial_p H(x, p) \subseteq [0, \infty) = T_{[0,1]}(0).$$

### 3.4 Examples of Hamiltonians

In this section, we verify Assumptions 3.12 and 3.13 for Hamiltonians of the type

$$\mathcal{H}(x, p) = \sup_{\theta \in \Theta} [\Lambda(x, p, \theta) - \mathcal{I}(x, \theta)],$$

that include and generalize the examples from the introduction. The purpose of this section is to showcase via different examples that the method introduced in this paper is versatile enough to capture a variety of interesting examples, including Hamiltonians that could not be treated before, as for instance Proposition 3.24.

In Proposition 3.18 we consider the classical Hamilton-Jacobi-Bellman equation arising in optimal control theory. Propositions 3.20 and 3.22 correspond to the Hamiltonian that one encounters in two-scale systems as studied in [BDG18] and [KP17]. The example of Proposition 3.24 arises in models of mean-field interacting particles that are coupled to fast external variables.

Each definition below corresponds to a specification of the elements involved in (3.9). Except for the first result, all propositions follow from verifying the general Assumptions 3.12 and 3.13 on the functions $\Lambda$ and $\mathcal{I}$. We verify these in Section 8.

A simplification of our arguments recovers the comparison principle of Theorem 3 of [BC97].

**Proposition 3.18 (Classical Hamilton-Jacobi-Bellman).** Let $E = \mathbb{R}^d$ and $\Theta$ be a topological space. Suppose that

$$\mathcal{H}(x, p) = \sup_{\theta \in \Theta} [f(x, \theta), p) - \mathcal{I}(x, \theta)],$$

(3.10)
Set $\Lambda$ and define the operator $Hf(x) := \mathcal{H}(x, \nabla f(x))$ with domain $D(H) = C^2_c(E)$. Suppose that the map $f : \mathbb{R}^d \times \Theta \to \mathbb{R}$ and cost function $I : \mathbb{R}^d \times \Theta \to \mathbb{R}$ have the following properties

(i) $f$ is continuous and bounded on sets $B \times \Theta$, where $B$ is any bounded subset of $\mathbb{R}^d$,

(ii) $f$ is uniformly continuous on bounded sets in $\mathbb{R}^d$ uniformly in $\Theta$, that is, for each bounded set $B \subseteq \mathbb{R}^d$ there is a modulus of continuity $\omega_{f,B}$ such that
\[ \sup_{\theta \in \Theta} |f(x, \theta) - f(y, \theta)| \leq \omega_{f,B}(|x - y|); \]

(iii) $f$ is one-sided Lipschitz, that is, there is a constant $M_f \geq 0$ such that
\[ \sup_{\theta \in \Theta} (f(x, \theta) - f(y, \theta), x - y) \leq M_f |x - y|^2. \]

(iv) the map $I$ is continuous and bounded;

(v) there is a modulus of continuity $\omega_I$ such that
\[ \sup_{\theta \in \Theta} |I(x, \theta) - I(y, \theta)| \leq \omega_I(|x - y|). \]

Set $\Lambda(x, p, \theta) = (f(x, \theta), p)$. Then the comparison principle holds for $f - \lambda Hf = h$ for all $h \in C_b(E)$ and $\lambda > 0$.

In this proposition, we lack the compact level sets as required by Assumption 3.13 (I3). This is compensated by the uniform estimates over $\theta \in \Theta$. This is a significant simplification over the results that follow. As compared to the proof of Proposition 6.4 that applies to the cases below, the proof can be adjusted by choosing an approximate optimizer in (6.12) and then using the uniform estimate over $\theta \in \Theta$ for $f$ and $I$ in (6.13).

**Remark 3.19.** Note that in this example, we do not assume that the minimal value of $I$ equals 0. Due to the uniform estimates, this does not pose any problem for the comparison principle, even though the particular proof of Theorem 3.7 that we are using uses this property.

A generalization of our proof of the comparison principle to a context without uniform estimates over $\theta \in \Theta$ in which we allow $I_{\min}(x) := \inf_{\theta} I(x, \theta) \neq 0$ is possible by working with a recalibrated internal Hamiltonian and cost function: $\tilde{\Lambda}(x, p, \theta) = \Lambda(x, p, \theta) - I_{\min}(x)$ and $\tilde{I}(x, \theta) = I(x, \theta) - I_{\min}(x)$.

An adaptation of Proposition 5.3 using $\Lambda = 0$ yields that $x \mapsto I_{\min}(x)$ is continuous. As a consequence, Assumptions 3.12 and 3.13 for the maps $I$ and $\Lambda$ translate into those for $\tilde{I}$ and $\tilde{\Lambda}$ unchanged, except for (A3) and the fact that $\tilde{\Lambda}(x, 0, \theta) = 0$ is no longer satisfied. The fact that $\Lambda(x, 0, \theta) = 0$ is not relevant for the proof of the comparison principle. Regarding (A3), we need to make the additional assumption that there exists $c_Y$ such that
\[ \sup_{\theta} \sup_{x} \Lambda(x, \nabla Y(x), \theta) - I_{\min}(x) \leq c_Y. \]

As noted above, we can not use Theorem 3.7 in this context to establish the existence of viscosity solutions.

We proceed with the Hamiltonian that arises from a diffusion process coupled to a fast jump process.
Proposition 3.20 (Diffusion coupled to jumps). Let $E = \mathbb{R}^d$ and $F = \{1, \ldots, j\}$ be a finite set. Suppose the following.

(i) The set of control variables is $\Theta := \mathcal{P}(\{1, \ldots, j\})$, that is probability measures over the finite set $F$.

(ii) The function $\Lambda$ is given by

\[ \Lambda(x, p, \theta) := \sum_{i \in F} [(a(x, i)p, p) + (b(x, i), p)] \theta_i, \]

where $a : E \times F \to \mathbb{R}^{d \times d}$ and $b : E \times F \to \mathbb{R}^d$, and $\theta_i := \theta(i)$.

(iii) The cost function $\mathcal{I} : E \times \Theta \to [0, \infty)$ is given by

\[ \mathcal{I}(x, \theta) := \sup_{u \in \mathcal{D}(L)} \int_F \left[ \sum_{i \neq j} r_{ij}(x) \theta_i \left[ 1 - e^{p_i - p_j} \right] \right] \, d\theta(z), \]

with non-negative rates $r : F^2 \times E \to [0, \infty)$.

Suppose that the cost function $\mathcal{I}$ satisfies the assumptions of Proposition 8.1 and the function $\Lambda$ satisfies the assumptions of Proposition 8.3. Then Theorems 3.4 and 3.7 apply to the Hamiltonian (3.9).

Remark 3.21 (Principal eigenvalue). Under irreducibility conditions on the rates, as assumed below in Proposition 8.1, by [DV75b] the Hamiltonian $\mathcal{H}(x, p)$ is the principal eigenvalue of the matrix $A_{x, p} \in \text{Mat}_{j \times j} (\mathbb{R})$ given by

\[ A_{x, p} = \text{diag} \left( (a(x, 1)p, p) + (b(x, 1), p), \ldots, (a(x, j)p, p) + (b(x, j), p) \right) + R_x, \]

where $x, p \in \mathbb{R}^d$ and $R_x$ is the matrix

\[
\begin{pmatrix}
-\sum_{j \neq 1} r_{1j}(x) & r_{12}(x) & \ldots & r_{1j}(x) \\
r_{21}(x) & -\sum_{j \neq 2} r_{2j}(x) & \ldots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
r_{jj}(x) & \ldots & r_{j, j-1}(x) & -\sum_{j \neq j} r_{jj}(x)
\end{pmatrix},
\]

that is $(R_x)_{ii} = -\sum_{j \neq i} r_{ij}(x)$ on the diagonal and $(R_x)_{ij} = r_{ij}(x)$ for $i \neq j$.

We proceed with the Hamiltonian that arises from a diffusion process coupled to a diffusion.

Proposition 3.22 (Diffusion coupled to diffusion). Let $E = \mathbb{R}^d$ and $F$ be a smooth compact Riemannian manifold without boundary. Suppose the following.

(i) The set of control variables $\Theta$ equals the space $\mathcal{P}(F)$.

(ii) The function $\Lambda$ is given by

\[ \Lambda(x, p, \theta) := \int_F [(a(x, z)p, p) + (b(x, z), p)] \, d\theta(z), \]

with $a : E \times F \to \mathbb{R}^{d \times d}$ and $b : E \times F \to \mathbb{R}^d$.

(iii) The cost function $\mathcal{I} : E \times \Theta \to [0, \infty)$ is given by

\[ \mathcal{I}(x, \theta) := \sup_{u \in \mathcal{D}(L)} \left[ -\int_F \frac{L_x u}{u} \, d\theta \right], \]
where $L_x$ is a second-order elliptic operator locally of the form

$$L_x = \frac{1}{2} \nabla \cdot (a_x \nabla) + b_x \cdot \nabla,$$

on the domain $\mathcal{D}(L_x) := C^2(F)$, with positive-definite matrix $a_x$ and co-vectors $b_x$.

Suppose that the cost function $I$ satisfies the assumptions of Proposition 8.2 and the function $\Lambda$ satisfies the assumptions of Proposition 8.3. Then Theorems 3.4 and 3.7 apply to the Hamiltonian (3.9).

In the context of weakly interacting jump processes on a collection of states $\{1, \ldots, q\}$ the dynamics of the empirical measures takes place on $\mathcal{P}(\{1, \ldots, q\})$. Transitions occur over the bonds $(a, b) \in E^2$ with $a \neq b$. We denote the set of bonds with $\Gamma$. In the limit the total rate of transitions over the bond $(a, b)$ if the empirical measure equals $\mu$ and the background measure is $\theta$ is denoted by $v(a, b, \mu, \theta)$

**Definition 3.23 (Proper kernel).** Let $v : \Gamma \times \mathcal{P}(\{1, \ldots, q\}) \times \Theta \to \mathbb{R}^+$. We say that $v$ is a proper kernel if $v$ is continuous and if for each $(a, b) \in \Gamma$, the map $(\mu, \theta) \mapsto v(a, b, \mu, \theta)$ is either identically equal to zero or satisfies the following two properties:

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

(b) There exists a decomposition $v(a, b, \mu, \theta) = v_1(a, b, \mu(a))v_2(a, b, \mu, \theta)$ such that $v_1$ is increasing in the third coordinate and such that $v_1(a, b, \cdot, \cdot)$ is continuous and satisfies $v_1(a, b, \mu, \theta) > 0$.

A typical example of a proper kernel is given by

$$v(a, b, \mu, \theta) = \mu(a)r(a, b, \theta)e^{\partial_b V(\mu)} - \partial_b V(\mu),$$

with $r > 0$ continuous and $V \in C^1_b(\mathcal{P}(\{1, \ldots, q\}))$.

**Proposition 3.24 (Mean-field coupled to diffusion).** Let the space $E$ be given by the embedding of $E := \mathcal{P}(\{1, \ldots, q\}) \times [0, \infty)^F \subseteq \mathbb{R}^d$ and $F$ be a smooth compact Riemannian manifold without boundary. Suppose the following.

(i) The set of control variables $\Theta$ equals $\mathcal{P}(F)$.

(ii) The function $\Lambda$ is given by

$$\Lambda((\mu, \varrho), p, \theta) := \sum_{(a, b) \in \Gamma} v(a, b, \mu, \theta) \left[ \exp \left\{ \left\{ p_b - p_x + p_{(a,b)} \right\} - 1 \right\} \right]$$

with a proper kernel $v$ in the sense of Definition 3.23.

(iii) The cost function $I : E \times \Theta \to [0, \infty]$ is given by

$$I(x, \theta) := \sup_{u \in \mathcal{D}(L_x)} \left[ - \int_F \frac{L_x u}{u} \, d\theta \right],$$

where $L_x$ is a second-order elliptic operator locally of the form

$$L_x = \frac{1}{2} \nabla \cdot (a_x \nabla) + b_x \cdot \nabla,$$

on the domain $\mathcal{D}(L_x) := C^2(F)$, with positive-definite matrix $a_x$ and co-vectors $b_x$. 

14
Suppose that the cost function $I$ satisfies the assumptions of Proposition 8.2 and the function $\Lambda$ satisfies the assumptions of Proposition 8.4. Then Theorems 3.4 and 3.7 apply to the Hamiltonian (3.9).

An analogous proposition can be formulated for mean-field particles coupled to jumps as in Proposition 3.20.

4 Strategy of the proofs

We comment on the overall strategy of proofs. In Section 4.1, we explain informally without the details how the bootstrap argument works in a simple setting in which $E$ is taken to be compact. This allows us to focus on the bootstrapping argument without having to bother with the reduction to compact sets. We proceed with a discussion on the existence of a viscosity solution in Section 4.2.

4.1 The bootstrap argument in a nutshell

In this section, we explain informally the main bootstrapping idea that is behind proving the comparison principle with Hamiltonians of the type (3.2) for

\[ u(x) - H(x, \nabla u(x)) = 0, \]

assuming compactness of $E$ and $\Psi(x, y) = \frac{1}{2}|x - y|^2$. In what follows, $u_1$ is a subsolution and $u_2$ is a supersolution. Recall that for smooth functions $f$, if $(u_1 - f)$ is maximal at a point $x$, then

\[ u_1(x) - H(x, \nabla f(x)) \leq 0. \]

Similarly for the supersolution $u_2$: If $(f - u_2)$ is maximal at a point $y$, then

\[ u_2(y) - H(y, \nabla f(y)) \geq 0. \]

We sketch how to prove $u_1 \leq u_2$ in several steps.

(i) By the classical doubling of variables procedure, see e.g. [CIL92], choosing for each $\alpha > 0$ points $x_\alpha, y_\alpha$ such that

\[ u_1(x_\alpha) - u_2(y_\alpha) - \alpha \Psi(x_\alpha, y_\alpha) = \sup_{x, y \in E} u_1(x) - u_2(y) - \alpha \Psi(x, y), \]

then by the properties of $\Psi$, we have

\[ \alpha \Psi(x_\alpha, y_\alpha) \to 0 \quad (4.1) \]

and the difference $\sup_x u_1(x) - u_2(x)$ can be approximated as

\[ \sup(x_1 - u_2) \leq \lim inf_{\alpha \to \infty} u_1(x_\alpha) - u_2(y_\alpha). \]

Set $p_\alpha := \alpha(x_\alpha - y_\alpha)$. Using the subsolution inequality $u_1(x_\alpha) \leq \mathcal{H}(x_\alpha, p_\alpha)$ and the supersolution inequality $u_2(y_\alpha) \geq \mathcal{H}(y_\alpha, p_\alpha)$, one arrives at the estimate

\[ \sup(u_1 - u_2) \leq \lim inf_{\alpha \to \infty} \mathcal{H}(x_\alpha, p_\alpha) - \mathcal{H}(y_\alpha, p_\alpha). \]
(ii) Recall that the Hamiltonian is given by

\[ \mathcal{H}(x, p) = \sup_{\theta \in \Theta} [\Lambda(x, p, \theta) - \mathcal{I}(x, \theta)] . \]

Taking the optimizer \( \theta_n \) for \( \mathcal{H}(x_n, p_n) \) and estimating the Hamiltonian at \( y_n \) with this optimizer, we obtain

\[ \sup (u_1 - u_2) \leq \liminf_{\alpha \to \infty} [\Lambda(x_n, p_n, \theta_n) - \Lambda(y_n, p_n, \theta_n)] + [\mathcal{I}(y_n, \theta_n) - \mathcal{I}(x_n, \theta_n)] . \]

(iii) We assume the continuity estimate on \( \Lambda \). That means that if we have

\[ \alpha \Psi(x_n, y_n) \to 0 \quad (4.2) \]
\[ \liminf_{\alpha \to \infty} \Lambda(x_n, p_n, \theta_n) > -\infty, \quad (4.3) \]
\[ \limsup_{\alpha \to \infty} \Lambda(y_n, p_n, \theta_n) < \infty, \quad (4.4) \]

and that \( \theta_n \) are in a compact set, then the difference of \( \Lambda \)'s is controlled as

\[ \liminf_{\alpha \to \infty} [\Lambda(x_n, p_n, \theta_n) - \Lambda(y_n, p_n, \theta_n)] \leq 0. \]

We postpone the verification that \( \theta_n \) are in a compact set to the next step (iv) below. Clearly, (4.2) is immediate from 4.1. We show how the other two bounds follow from the sub- and supersolution inequalities. By the subsolution inequality,

\[ u_1(x_n) \leq \mathcal{H}(x_n, p_n) = \Lambda(x_n, p_n, \theta_n) - \mathcal{I}(x_n, \theta_n) \leq \Lambda(x_n, p_n, \theta_n), \]

and (4.3) follows since \( u_1 \) is bounded. Letting \( \theta_n^0 \) be the control variable such that \( \mathcal{I}(y_n, \theta_n^0) = 0 \), we obtain from the supersolution inequality that

\[ u_2(y_n) \leq \mathcal{H}(y_n, p_n) \geq \Lambda(y_n, p_n, \theta_n^0), \quad (4.5) \]

and therefore \( \Lambda(y_n, p_n, \theta_n^0) \) is bounded above. Assuming that

\[ \Lambda(y_n, p_n, \theta_n) \leq C_1 \Lambda(y_n, p_n, \theta_n^0) + C_2, \]

(4.4) follows. In summary, if indeed \( \theta_n \) are in a compact set, taking the \( \lim inf_{\alpha \to \infty} \) in the last estimate on \( (u_1 - u_2) \), we obtain

\[ \sup (u_1 - u_2) \leq 0 + \liminf_{\alpha \to \infty} [\mathcal{I}(y_n, \theta_n) - \mathcal{I}(x_n, \theta_n)] . \]

(iv) We assume that if the cost functions are uniformly bounded,

\[ \mathcal{I}(x_n, \theta_n) \leq M \quad \text{and} \quad \mathcal{I}(y_n, \theta_n) \leq M, \quad (4.6) \]

then (1) the control variables \( \theta_n \) are in a compact set, implying that we can carry out the argument of step (iii) above, and (2) the cost functions are continuous as a function of the internal variables \( x \), giving

\[ \limsup_{\alpha \to \infty} [\mathcal{I}(y_n, \theta_n) - \mathcal{I}(x_n, \theta_n)] = 0. \]

The required bounds on \( \mathcal{I} \) in (4.6) follow as well from the sub- and supersolution inequalities. From the subsolution inequality, we have

\[ u_1(x_n) \leq \mathcal{H}(x_n, p_n) = \Lambda(x_n, p_n, \theta_n) - \mathcal{I}(x_n, \theta_n). \]
Thus the bound on $I(x,\theta)$ follows if we establish an upper bound on $\Lambda(x, p, \theta)$. Note that

$$\Lambda(x, p, \theta) \leq C_1 \Lambda(x, p, \theta^0) + C_2$$

and

$$\Lambda(x, p, \theta^0) = \Lambda(y, p, \theta^0) + \left[ \Lambda(x, p, \theta^0) - \Lambda(y, p, \theta^0) \right].$$

We have an upper bound for the first term on the right-hand side by (4.5). The second term is bounded above by the continuity estimate, which can be carried out as we know that the $\theta^0$ are in a compact set because they satisfy $I(y, \theta_0) = 0$. Since $y$ is close to $x$ and $I$ is continuous as a function of $x$ when bounded, the bound on $I(x, \theta)$ carries over to $I(y, \theta)$.

In summary, by using the information contained in the sub- and supersolution inequalities, the continuity estimate of the functions $\Lambda$ bootstraps to a continuity estimate of $H$, giving the comparison principle.

### 4.2 Proof of the existence of a viscosity solution

For the existence of a viscosity solution to $f - \lambda Hf = h$, we will use the results of Chapter 8 of [FK06]. We will briefly discuss the method to obtain this result.

To establish that $R(\lambda)h$ given by

$$R(\lambda)h(x) = \sup_{\gamma \in AC, \gamma(0)=x} \int_0^\infty \lambda^{-1} e^{-\lambda^{-1}t} \left[ h(\gamma(t)) - \int_0^t L(\gamma(s), \dot{\gamma}(s)) \right] dt,$$

yields a viscosity solution to $f - \lambda Hf = h$, we follow a general strategy, first used in [FK06] and summarized more general context in Proposition 3.4 of [Kra19]. For this strategy, we check three properties of $R(\lambda)$:

(a) For all $(f, g) \in H$, we have $f = R(\lambda)(f - \lambda g)$;

(b) The operator $R(\lambda)$ is a pseudo-resolvent: for all $h \in C_b(E)$ and $0 < \alpha < \beta$ we have

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

(c) The operator $R(\lambda)$ is contractive.

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

Establishing (c) is straightforward. The proof of (a) and (b) stems from two main properties of exponential random variable. Let $\tau_\lambda$ be the measure on $\mathbb{R}^+$ corresponding to the exponential random variable with mean $\lambda^{-1}$.

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

$$\lambda \int_0^\infty z(t) \tau_\lambda(dt) = \int_0^\infty \int_0^t z(s) ds \tau_\lambda(dt).$$
(b) is related to a more involved integral property of exponential random variables. For $0 < \alpha < \beta$, we have

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

Establishing (a) and (b) can then be reduced by a careful analysis of optimizers in the definition of $R(\lambda)$, and concatenation or splittings thereof. This was carried out in Chapter 8 of [FK06] on the basis of three assumptions, namely [FK06, Assumptions 8.9, 8.10 and 8.11]. We verify these conditions in Section 7.

5 Regularity of the Hamiltonian

In this section, we establish continuity, convexity and the existence of a containment function for the Hamiltonian $H$ of 3.2. We repeat its definition for convenience:

$$H(x, p) = \sup_{\theta \in \Theta} \left[\Lambda(x, p, \theta) - I(x, \theta)\right].$$

Proposition 5.1 (Regularity of the Hamiltonian). Let $H : E \times \mathbb{R}^d \to \mathbb{R}$ be the Hamiltonian as in (5.1), and suppose that Assumptions 3.12 and 3.13 are satisfied. Then:

(i) For any $x \in E$, the map $p \mapsto H(x, p)$ is convex and $H(x, 0) = 0$.

(ii) With the containment function $\Upsilon : E \to \mathbb{R}$ of (\Lambda3), we have

$$\sup_{x \in E} H(x, \nabla \Upsilon(x)) \leq C_\Upsilon < \infty.$$

Proof. The map $p \mapsto H(x, p)$ is convex as it is the supremum over convex functions.

For proving $H(x, 0) = 0$, let $x \in E$. Then by (\Lambda2) of Assumption 3.12, we have $\Lambda(x, 0, \theta) = 0$, and therefore

$$H(x, 0) = -\inf_{\theta \in \Theta} I(x, \theta) = 0,$$

since $I \geq 0$ and $I(x, \theta_x) = 0$ for some $\theta_x$ by (\Lambda2) of Assumption 3.13. Regarding (ii), we note that by (\Lambda3),

$$H(x, \nabla \Upsilon(x)) \leq \sup_{\theta \in \Theta} \Lambda(x, \nabla \Upsilon(x), \theta) \leq \sup_{\theta \in \Theta} \sup_{x \in E} \Lambda(x, \nabla \Upsilon(x), \theta) \leq C_\Upsilon.$$

To prove that $H$ is continuous, we use Assumption 3.13. What we truly need, however, is that $I$ Gamma converges as a function of $x$. We establish this result first.

Proposition 5.2 (Gamma convergence of the cost functions). Let a cost function $I : E \times \Theta \to [0, \infty]$ satisfy Assumption 3.13. Then if $x_n \to x$ in $E$, the functionals $I_n$ defined by

$$I_n(\theta) := I(x_n, \theta)$$

converge in the $\Gamma$-sense to $I_\infty(\theta) := I(x, \theta)$. That is:
1. If \( x_n \to x \) and \( \theta_n \to \theta \), then 
\[
\liminf_{n \to \infty} \mathcal{I}(x_n, \theta_n) \geq \mathcal{I}(x, \theta).
\]

2. For \( x_n \to x \) and all \( \theta \in \Theta \) there are \( \theta_n \in \Theta \) such that \( \theta_n \to \theta \) and 
\[
\limsup_{n \to \infty} \mathcal{I}(x_n, \theta_n) \leq \mathcal{I}(x, \theta).
\]

Proof. Let \( x_n \to x \). If \( \theta_n \to \theta \), then by lower semi-continuity (I1),
\[
\liminf_{n \to \infty} \mathcal{I}(x_n, \theta_n) \geq \mathcal{I}(x, \theta).
\]
For the lim-sup bound, let \( \theta \in \Theta \). If \( \mathcal{I}(x, \theta) = \infty \), there is nothing to prove. Thus suppose that \( \mathcal{I}(x, \theta) \) is finite. Then by (I4), there is a neighborhood \( U_x \) of \( x \) and a constant \( M < \infty \) such that for any \( y \in U_x \),
\[
\mathcal{I}(y, \theta) \leq M.
\]
Since \( x_n \to x \), the \( x_n \) are eventually contained in \( U_x \). Taking the constant sequence \( \theta_n := \theta \), we thus get that \( \mathcal{I}(x_n, \theta_n) \leq M \) for all \( n \) large enough. By (I5),
\[
\lim_{n \to \infty} |\mathcal{I}(x_n, \theta_n) - \mathcal{I}(x, \theta)| = 0,
\]
and the lim-sup bound follows. \( \Box \)

**Proposition 5.3** (Continuity of the Hamiltonian). Let \( \mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R} \) be the Hamiltonian defined in (3.2), and suppose that Assumptions 3.12 and 3.13 are satisfied. Then the map \( (x, p) \mapsto \mathcal{H}(x, p) \) is continuous and the Lagrangian \( (x, v) \mapsto \mathcal{L}(x, v) := \sup_p (p, v) - \mathcal{H}(x, p) \) is lower semi-continuous.

Before we start with the proof, we give a remark on the generality of its statement and on the assumption that \( \Theta \) is Polish.

**Remark 5.4.** The proof of upper semi-continuity of \( \mathcal{H} \) works in general, using continuity properties of \( \Lambda \), lower semi-continuity of \( (x, \theta) \mapsto \mathcal{I}(x, \theta) \) and the compact sublevel sets of \( \mathcal{I}(x, \cdot) \). To establish lower semi-continuity, we need that the functionals \( \mathcal{I} \) Gamma converge as a function of \( x \). This was established in Proposition 5.2.

**Remark 5.5.** In the lemma we use a sequential characterization of upper hemi-continuity. This is inspired by the natural formulation of Gamma convergence in terms of sequences. An extension of our results to spaces \( \Theta \) beyond the Polish context should take care of this issue. Without introducing the complicated matter, an extension is possible to Hausdorff \( \Theta \) that are k-spaces in which all compact sets are metrizable.

We will use the following technical result to establish upper semi-continuity of \( \mathcal{H} \).

**Lemma 5.6** (Lemma 17.30 in [AB06]). Let \( X \) and \( Y \) be two Polish spaces. Let \( \phi : X \to K(Y) \), where \( K(Y) \) is the space of non-empty compact subsets of \( Y \). Suppose that \( \phi \) is upper hemi-continuous, that is if \( x_n \to x \) and \( y_n \in \phi(x_n) \), then \( y \in \phi(x) \).

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

**Proof of Proposition 5.3.** We start by establishing upper semi-continuity of \( \mathcal{H} \). We argue on the basis of Lemma 5.6. Recall the representation of \( \mathcal{H} \) of (5.1). Set \( X = E \times \mathbb{R}^d \) for the \((x, p)\) variables, \( Y = \Theta \), and \( f(x, p, \theta) = \Lambda(x, p, \theta) - \mathcal{I}(x, \theta) \) and note that this function is upper semi-continuous by Assumption 3.13 (I1) and by Assumption 3.12 (A1).
By Assumption 3.13 (I2), we have $\mathcal{H}(x, p) \geq \Lambda(x, p, \theta)$. Thus, it suffices to restrict the supremum over $\theta \in \Theta$ to $\theta \in \phi(x, p)$ where

$$\phi(x, p) := \{ \theta \in \Theta \mid I(x, \theta) \leq 2 \| \Lambda(x, p, \cdot) \|_{\Theta} \},$$

in the sense that we have

$$\mathcal{H}(x, p) = \sup_{\theta \in \phi(x, p)} [\Lambda(x, p, \theta) - I(x, \theta)].$$

$\phi(x, p)$ is non-empty as $\theta \in \phi(x, p)$ and it is compact due to Assumption 3.13 (I3). We are left to show that $\phi$ is upper hemi-continuous.

Thus, let $(x_n, p_n, \theta_n) \to (x, p, \theta)$ with $\theta_n \in \phi(x_n, p_n)$. We establish that $\theta \in \phi(x, p)$. By (I1) and the definition of $\phi$ we find

$$I(x, \theta) \leq \liminf_n I(x_n, \theta_n) \leq \liminf_n \Lambda(x_n, p_n, \cdot) \|_{\Theta} = 2 \| \Lambda(x, p, \cdot) \|_{\Theta}$$

which implies indeed that $\theta \in \phi(x, p)$. Thus, upper semi-continuity follows by an application of Lemma 5.6.

We proceed with proving lower semi-continuity of $\mathcal{H}$. Suppose that $(x_n, p_n) \to (x, p)$, we prove that $\liminf_n \mathcal{H}(x_n, p_n) \geq \mathcal{H}(x, p)$.

Let $\theta$ be the measure such that $\mathcal{H}(x, p) = \Lambda(x, p, \theta) - I(x, \theta)$. We have

- By Proposition 5.2 there are $\theta_0$ such that $\theta_n \to \theta$ and $\limsup_n I(x_n, \theta_n) \leq I(x, \theta)$.
- $\Lambda(x_n, p_n, \theta_n)$ converges to $\Lambda(x, p, \theta)$ by Assumption (A1).

Therefore,

$$\liminf_{n \to \infty} \mathcal{H}(x_n, p_n) \geq \liminf_{n \to \infty} [\Lambda(x_n, p_n, \theta_n) - I(x_n, \theta_n)] \geq \liminf_n \Lambda(x_n, p_n, \theta_n) - \limsup_n I(x_n, \theta_n) \geq \Lambda(x, p, \theta) - I(x, \theta) = \mathcal{H}(x, p),$$

establishing that $\mathcal{H}$ is lower semi-continuous.

The Lagrangian $\mathcal{L}$ is obtained as the supremum over continuous functions. This implies $\mathcal{L}$ is lower semi-continuous.  

\[\square\]

6 The comparison principle

In this section, we establish the comparison principle for $f - \lambda \mathcal{H}f = h$ in the context of Theorem 3.4, using the general strategy of Section 4.1. Before being able to use this strategy, we need to restrict our analysis to compact sets in $E$. We will use a classical penalization technique that we will write down in operator form.

We thus introduce two new operators $H_t$ and $H_t^\perp$, which are defined in terms of $\mathcal{H}$ and the containment function $Y$ from Assumption 3.12 (A3). We will then show that the comparison principle holds for a pair of Hamilton-Jacobi equations in terms of $H_t$ and $H_t^\perp$. We therefore have to extend our notion of Hamilton-Jacobi equations and the comparison principle. This extension is standard, but we included it for completeness in Definition A.1 in the appendix.
This procedure allows us to clearly separate the reduction to compact sets on one hand, and the proof of the comparison principle on the basis of the bootstrap procedure on the other. Schematically, we will establish the following diagram:

In this diagram, an arrow connecting an operator $A$ with operator $B$ with subscript ‘sub’ means that viscosity subsolutions of $f - \lambda A f = h$ are also viscosity subsolutions of $f - \lambda B f = h$. Similarly for arrows with a subscript ‘super’.

We introduce the operators $H_\dagger$ and $H_\ddagger$ in Section 6.1. The arrows will be established in Section 6.2. Finally, we will establish the comparison principle for $H_\dagger$ and $H_\ddagger$ in Section 6.3, which by the arrows implies the comparison principle for $H$.

**Proof of Theorem 3.4.** Fix $h_1, h_2 \in C_b(E)$ and $\lambda > 0$.

Let $u_1, u_2$ be a viscosity sub- and supersolution to $f - \lambda H f = h_1$ and $f - \lambda H f = h_2$ respectively. By Lemma 6.3 proven in Section 6.2, $u_1$ and $u_2$ are a sub- and supersolution to $f - \lambda H_\dagger f = h_1$ and $f - \lambda H_\ddagger f = h_2$ respectively. Thus $\sup_E u_1 - u_2 \leq \sup_E h_1 - h_2$ by Proposition 6.4 of Section 6.3. Specialising to $h_1 = h_2$ gives Theorem 3.4.

### 6.1 Definition of auxiliary operators

In this section, we repeat the definition of $H$, and introduce the operators $H_\dagger$ and $H_\ddagger$.

**Definition 6.1.** The operator $H \subseteq C^1_b(E) \times C_b(E)$ has domain $D(H) = C^\infty_c(E)$ and satisfies $H f(x) = \mathcal{H}(x, df(x))$, where $\mathcal{H}$ is the map

$$\mathcal{H}(x, p) = \sup_{\theta \in \Theta} [\Lambda(x, p, \theta) - I(x, \theta)].$$

We proceed by introducing $H_\dagger$ and $H_\ddagger$. These new Hamiltonians will serve as natural upper and lower bound for $H$. They are defined in terms of the containment function $\Upsilon$, and essentially allow us to restrict our analysis to compact sets.

For the following definition, recall Assumption (A3) and the constant $C_\Upsilon := \sup_p \sup_x \Lambda(x, \nabla \Upsilon(x), \theta)$ therein. Denote by $C^\infty_c(E)$ the set of smooth functions on $E$ that have a lower bound and by $C^\infty_c(E)$ the set of smooth functions on $E$ that have an upper bound.

**Definition 6.2 (The operators $H_\dagger$ and $H_\ddagger$).** For $f \in C^\infty_c(E)$ and $\epsilon \in (0, 1)$ set

$$f_\epsilon^\dagger := (1 - \epsilon)f + \epsilon \Upsilon$$

$$H_\dagger f^\epsilon(x) := (1 - \epsilon)\mathcal{H}(x, \nabla f(x)) + \epsilon C_\Upsilon.$$
Lemma 6.3. Fix a creasing function such that \( f \) is assumed to lie in the compact set \( E \). Set \( M \) the operator defined by

\[
\lim_{n \to \infty} \sup_{x \in E} \left[ u(x_n) - \lambda H_{\epsilon,f}^E(x_n) - h(x_n) \right] \leq 0.
\]

and set

\[
H_\epsilon := \left\{ (f^\epsilon, H_{\epsilon,f}^E) \mid f \in C_0^\infty(E), \epsilon \in (0,1) \right\}.
\]

For \( f \in C_0^\infty(E) \) and \( \epsilon \in (0,1) \) set

\[
f^\epsilon := (1+\epsilon)f - \epsilon Y
\]

\[
H_{\epsilon,f}^E(x) := (1+\epsilon)H(x, \nabla f(x)) - \epsilon C_f.
\]

and set

\[
H_\epsilon := \left\{ (f^\epsilon, H_{\epsilon,f}^E) \mid f \in C_0^\infty(E), \epsilon \in (0,1) \right\}.
\]

6.2 Implications based on compact containment

The operator \( H \) is related to \( H_\epsilon, H_\epsilon \) by the following Lemma.

**Lemma 6.3.** Fix \( \lambda > 0 \) and \( h \in C_b(E) \).

(a) Every subsolution to \( f - \lambda H f = h \) is also a subsolution to \( f - \lambda H_\epsilon f = h \).

(b) Every supersolution to \( f - \lambda H f = h \) is also a supersolution to \( f - \lambda H_\epsilon f = h \).

We only prove (a) of Lemma 6.3, as (b) can be carried out analogously.

**Proof.** Fix \( \lambda > 0 \) and \( h \in C_b(E) \). Let \( u \) be a subsolution to \( f - \lambda H f = h \). We prove it is also a subsolution to \( f - \lambda H_\epsilon f = h \).

Fix \( \epsilon > 0 \) and \( f \in C_0^\infty(E) \) such that \( (f^\epsilon, H_{\epsilon,f}^E, \phi) \in H_\epsilon \). We will prove that there are \( x_n \in E \) such that

\[
\lim_{n \to \infty} (u - f^\epsilon_n)(x_n) = \sup_{x \in E} (u - f^\epsilon_n), \quad (6.1)
\]

\[
\limsup_{n \to \infty} \left[ u(x_n) - \lambda H_{\epsilon,f}^E(x_n) - h(x_n) \right] \leq 0. \quad (6.2)
\]

As the function \( u - (1 - \epsilon)f \) is bounded from above and \( \epsilon Y \) has compact sublevel-sets, the sequence \( x_n \) along which the first limit is attained can be assumed to lie in the compact set

\[
K := \left\{ x \mid Y(x) \leq \epsilon^{-1} \sup_x (u(x) - (1 - \epsilon)f(x)) \right\}.
\]

Set \( M = \epsilon^{-1} \sup_x (u(x) - (1 - \epsilon)f(x)) \). Let \( \gamma : \mathbb{R} \to \mathbb{R} \) be a smooth increasing function such that

\[
\gamma(r) = \begin{cases} 
  r & \text{if } r \leq M, \\
  M + 1 & \text{if } r \geq M + 2.
\end{cases}
\]

Denote by \( f_t \) the function on \( E \) defined by

\[
f_t(x) := \gamma ((1 - \epsilon)f(x) + \epsilon Y(x)).
\]

By construction \( f_t \) is smooth and constant outside of a compact set and thus lies in \( D(H) = C_0^\infty(E) \). As \( u \) is a viscosity subsolution for \( f - \lambda H f = h \) there exists a sequence \( x_n \in K \subseteq E \) (by our choice of \( K \)) with

\[
\lim_{n \to \infty} (u - f_t)(x_n) = \sup_{x} (u - f_t)(x), \quad (6.3)
\]

\[
\limsup_{n \to \infty} \left[ u(x_n) - \lambda H f_t(x_n) - h(x_n) \right] \leq 0. \quad (6.4)
\]
As $f_\varepsilon$ equals $f_\varepsilon^+\big|_K$ on $K$, we have from (6.3) that also
\[
\lim_n (u - f_\varepsilon^+) (x_n) = \sup_{x \in E} (u - f_\varepsilon^+),
\]
establishing (6.1). Convexity of $p \mapsto H(x, p)$ yields for arbitrary points $x \in K$ the estimate
\[
H f_\varepsilon(x) = H(x, \nabla f_\varepsilon(x)) \\
\leq (1 - \varepsilon) H(x, \nabla f(x)) + \varepsilon H(x, \nabla Y(x)) \\
\leq (1 - \varepsilon) H(x, \nabla f(x)) + \varepsilon C_Y = H_{\varepsilon}^f(x).
\]
Combining this inequality with (6.4) yields
\[
\limsup_n \left[ u(x_n) - \lambda H_{\varepsilon}^f(x_n) - h(x_n) \right] \\
\leq \limsup_n [u(x_n) - \lambda H f_\varepsilon(x_n) - h(x_n)] \leq 0,
\]
establishing (6.2). This concludes the proof.

6.3 The comparison principle

In this section, we prove the comparison principle for the operators $H_1$ and $H_4$.

Proposition 6.4. Fix $\lambda > 0$ and $h_1, h_2 \in C_b(E)$. Let $u_1$ be a viscosity subsolution to $f - \lambda H_4 f = h_1$ and let $u_2$ be a viscosity supersolution to $f - \lambda H_4 f = h_2$. Then we have $\sup u_1(x) - u_2(x) \leq \sup h_1(x) - h_2(x)$.

The proof uses an estimate that was proven in the proof of Proposition A.11 of [CK17] for one penalization function $\Psi$ or in the context of the more general continuity estimate of Appendix B in the proof of Proposition 4.5 of [Kra17] for two penalization functions $\{\Psi_1, \Psi_2\}$. In both contexts we use the containment function $Y$ of Assumption 3.12, (A3). We start with a key result that allows us to find optimizing points that generalize the argument of Section 4.1 to the non compact setting.

The result is a copy of Lemma A.11 of [CK17], which is in turn a variant of Lemma 9.2 in [FK06] and Proposition 3.7 in [CIL92]. We have included it for completeness.

Lemma 6.5. Let $u$ be bounded and upper semi-continuous, let $v$ be bounded and lower semi-continuous, let $\Psi : E^2 \to R^+$ be penalization functions and let $Y$ be a containment function.

Fix $\varepsilon > 0$. For every $\alpha > 0$ there exist $x_{\alpha, \varepsilon}, y_{\alpha, \varepsilon} \in E$ such that
\[
\frac{u(x_{\alpha, \varepsilon})}{1 - \varepsilon} - \frac{v(y_{\alpha, \varepsilon})}{1 + \varepsilon} - a \Psi (x_{\alpha, \varepsilon}, y_{\alpha, \varepsilon}) - \frac{\varepsilon}{1 - \varepsilon} Y(x_{\alpha, \varepsilon}) - \frac{\varepsilon}{1 + \varepsilon} Y(y_{\alpha, \varepsilon}) \\
= \sup_{x, y \in E} \left\{ \frac{u(x)}{1 - \varepsilon} - \frac{v(y)}{1 + \varepsilon} - a \Psi (x, y) - \frac{\varepsilon}{1 - \varepsilon} Y(x) - \frac{\varepsilon}{1 + \varepsilon} Y(y) \right\}. \quad (6.5)
\]
Additionally, for every $\varepsilon > 0$ we have that
(a) The set $\{x_{\alpha, \varepsilon}, y_{\alpha, \varepsilon} \mid \alpha > 0\}$ is relatively compact in $E$.
(b) All limit points of $\{(x_{\alpha, \varepsilon}, y_{\alpha, \varepsilon}) \mid \alpha > 0\}$ as $\alpha \to \infty$ are of the form $(z, z)$ and for these limit points we have $u(z) - v(z) = \sup_{x \in E} \{u(x) - v(x)\}$. 23
We prove that for $\lambda > 0$ and $h_1, h_2 \in C_b(E)$. Let $u_1$ be a viscosity subsolution and $u_2$ be a viscosity supersolution of $f - \lambda H \phi = h_1$ and $f - \lambda H \phi = h_2$ respectively. We prove Theorem 6.4 in two steps.

**Step 1:** We prove that for $\epsilon > 0$ and $\alpha > 0$, there exist points $x_{\epsilon, \alpha}, y_{\epsilon, \alpha} \in E$ and momenta $p_{\epsilon, \alpha}^1, p_{\epsilon, \alpha}^2 \in \mathbb{R}^d$ such that

$$
\sup_{\epsilon} (u_1 - u_2) \leq \lambda \liminf_{\epsilon \to 0} \liminf_{\alpha \to \infty} \left[ H(x_{\epsilon, \alpha}, p_{\epsilon, \alpha}^1) - H(y_{\epsilon, \alpha}, p_{\epsilon, \alpha}^2) \right] + \sup_{\epsilon} (h_1 - h_2). \quad (6.6)
$$

This step is solely based on the sub- and supersolution properties of $u_1, u_2$, the continuous differentiability of the penalization function $\Psi(x, y)$, the containment function $\Phi$, and convexity of $p \mapsto H(x, p)$.

**Step 2:** Using Assumptions 3.12 and 3.13, we prove that

$$
\liminf_{\epsilon \to 0} \liminf_{\alpha \to \infty} \left[ H(x_{\epsilon, \alpha}, p_{\epsilon, \alpha}^1) - H(y_{\epsilon, \alpha}, p_{\epsilon, \alpha}^2) \right] \leq 0.
$$

**Proof of Step 1:** For any $\alpha > 0$ and any $\alpha > 0$, define the map $\Phi_{\epsilon, \alpha} : E \times E \to \mathbb{R}$ by

$$
\Phi_{\epsilon, \alpha}(x, y) := \frac{u_1(x)}{1 - \epsilon} - \frac{u_2(y)}{1 + \epsilon} - \frac{\alpha}{1 - \epsilon} \Psi(x) - \frac{\epsilon}{1 + \epsilon} \Phi (x) - \frac{\epsilon}{1 + \epsilon} \Phi (y).
$$

Let $\epsilon > 0$. By Lemma 6.5, there is a compact set $K_{\epsilon} \subseteq E$ and there exist points $x_{\epsilon, \alpha}, y_{\epsilon, \alpha} \in K_{\epsilon}$ such that

$$
\Phi_{\epsilon, \alpha}(x_{\epsilon, \alpha}, y_{\epsilon, \alpha}) = \sup_{x, y \in E} \Phi_{\epsilon, \alpha}(x, y), \quad (6.7)
$$

and

$$
\lim_{\alpha \to \infty} \alpha \Psi(x_{\epsilon, \alpha}, y_{\epsilon, \alpha}) = 0. \quad (6.8)
$$

As in the proof of Proposition A.11 of [Kra17], it follows that

$$
\sup_{E} (u_1 - u_2) \leq \liminf_{\epsilon \to 0} \liminf_{\alpha \to \infty} \left[ \frac{u_1(x_{\epsilon, \alpha})}{1 - \epsilon} - \frac{u_2(y_{\epsilon, \alpha})}{1 + \epsilon} \right]. \quad (6.9)
$$

At this point, we want to use the sub- and supersolution properties of $u_1$ and $u_2$. Define the test functions $\phi_1^{\epsilon, \alpha} \in D(H_1)$, $\phi_2^{\epsilon, \alpha} \in D(H_1)$ by

$$
\phi_1^{\epsilon, \alpha}(x) := (1 - \epsilon) \left[ \frac{u_2(y_{\epsilon, \alpha})}{1 + \epsilon} + \alpha \Psi(x, y_{\epsilon, \alpha}) + \frac{\epsilon}{1 - \epsilon} \Phi (x) + \frac{\epsilon}{1 + \epsilon} \Phi (y_{\epsilon, \alpha}) \right] + (1 - \epsilon)(x - x_{\epsilon, \alpha})^2,
$$

$$
\phi_2^{\epsilon, \alpha}(y) := (1 + \epsilon) \left[ \frac{u_1(x_{\epsilon, \alpha})}{1 - \epsilon} - \alpha \Psi(x_{\epsilon, \alpha}, y) - \frac{\epsilon}{1 - \epsilon} \Phi (x_{\epsilon, \alpha}) - \frac{\epsilon}{1 + \epsilon} \Phi (y) \right] - (1 + \epsilon)(y - y_{\epsilon, \alpha})^2.
$$

Using (6.7), we find that $u_1 - \phi_1^{\epsilon, \alpha}$ attains its supremum at $x = x_{\epsilon, \alpha}$, and thus

$$
\sup_{E} (u_1 - \phi_1^{\epsilon, \alpha}) = (u_1 - \phi_1^{\epsilon, \alpha})(x_{\epsilon, \alpha}).
$$
Denote $p_{\epsilon, a}^1 := a \nabla_x \psi(x_{\epsilon, a}, y_{\epsilon, a})$. By our addition of the penalization $(x - x_{\epsilon, a})^2$ to the test function, the point $x_{\epsilon, a}$ is in fact the unique optimizer, and we obtain from the subsolution inequality that

$$u_1(x_{\epsilon, a}) - \lambda \left[ (1 - \epsilon) H(x_{\epsilon, a}, p_{\epsilon, a}^1) + \epsilon C_Y \right] \leq h_1(x_{\epsilon, a}). \quad (6.10)$$

With a similar argument for $u_2$ and $p_{\epsilon, a}^2$, we obtain by the supersolution inequality that

$$u_2(y_{\epsilon, a}) - \lambda \left[ (1 + \epsilon) H(y_{\epsilon, a}, p_{\epsilon, a}^2) - \epsilon C_Y \right] \geq h_2(y_{\epsilon, a}). \quad (6.11)$$

where $p_{\epsilon, a}^2 := -a \nabla_y \psi(x_{\epsilon, a}, y_{\epsilon, a})$. With that, estimating further in (6.9) leads to

$$\sup(\epsilon u_1 - u_2) \leq \liminf_{\epsilon \to 0} \liminf_{a \to 0} \left[ \frac{h_1(x_{\epsilon, a})}{1 - \epsilon} - \frac{h_2(y_{\epsilon, a})}{1 + \epsilon} + \frac{\epsilon}{1 - \epsilon} C_Y \right] + \frac{\epsilon}{1 + \epsilon} C_Y + \lambda \left[ H(x_{\epsilon, a}, p_{\epsilon, a}^1) - H(y_{\epsilon, a}, p_{\epsilon, a}^2) \right].$$

Thus, (6.6) in Step 1 follows.

**Proof of Step 2:** Recall that $H(x, p)$ is given by

$$H(x, p) = \sup_{\Theta \in \Theta} [\Lambda(x, p, \theta) - I(x, \theta)].$$

Since $\Lambda(x_{\epsilon, a}, p_{\epsilon, a}^1) : \Theta \to \mathbb{R}$ is bounded and continuous by (A1) and $I(x_{\epsilon, a}, \cdot) : \Theta \to [0, \infty]$ has compact sub-level sets in $\Theta$ by (I3), there exists an optimizer $\theta_{\epsilon, a} \in \Theta$ such that

$$H(x_{\epsilon, a}, p_{\epsilon, a}^1) = \Lambda(x_{\epsilon, a}, p_{\epsilon, a}^1, \theta_{\epsilon, a}) - I(x_{\epsilon, a}, \theta_{\epsilon, a}). \quad (6.12)$$

Choosing the same point in the supremum of the second term $H(y_{\epsilon, a}, p_{\epsilon, a}^2)$, we obtain for all $\epsilon > 0$ and $a > 0$ the estimate

$$H(x_{\epsilon, a}, p_{\epsilon, a}^1) - H(y_{\epsilon, a}, p_{\epsilon, a}^2) \leq \Lambda(x_{\epsilon, a}, p_{\epsilon, a}^1, \theta_{\epsilon, a}) - \Lambda(y_{\epsilon, a}, p_{\epsilon, a}^2, \theta_{\epsilon, a}) + I(y_{\epsilon, a}, \theta_{\epsilon, a}) - I(x_{\epsilon, a}, \theta_{\epsilon, a}). \quad (6.13)$$

We will establish an upper bound for this difference using the continuity estimate (A4) and equi-continuity (I5).

To apply the continuity estimate (A4), we need to verify (3.5) and (3.6) (see (6.15) and (6.17) below) for the variables $\theta_{\epsilon, a}$. In addition, we need to establish that $\theta_{\epsilon, \beta}$ are contained in a compact set.

To apply (I5), we need to control the size of $I(x_{\epsilon, a}, \theta_{\epsilon, a})$ and $I(y_{\epsilon, a}, \theta_{\epsilon, a})$ along subsequences, which by Assumption (I3) implies the above requirement that along these subsequences $\theta_{\epsilon, \beta}$ are contained in a compact set. To obtain control on the size of $I$, we employ an auxiliary argument based on the continuity estimate for the measures $\theta_{\epsilon, a}^0$, obtained by (I2), satisfying

$$I(y_{\epsilon, a}, \theta_{\epsilon, a}^0) = 0. \quad (6.14)$$

The application of the continuity estimate for $\theta_{\epsilon, a}^0$ only requires to check (3.5) and (3.6) as the measures $\theta_{\epsilon, a}^0$ are contained in a compact set by (6.14)
and (I3). Thus, we will first establish

\[
\liminf_{\varepsilon \to 0} \inf_{a \to \infty} \Lambda(x_{t,a}, p_{t,a}^1, \theta_{t,a}) > -\infty, \quad (6.15)
\]

\[
\liminf_{\varepsilon \to 0} \inf_{a \to \infty} \Lambda(x_{t,a}, p_{t,a}^1, \theta_{t,a}^0) > -\infty, \quad (6.16)
\]

\[
\limsup_{\varepsilon \to 0} \sup_{a \to \infty} \Lambda(y_{t,a}, p_{t,a}^2, \theta_{t,a}) < \infty, \quad (6.17)
\]

\[
\limsup_{\varepsilon \to 0} \sup_{a \to \infty} \Lambda(y_{t,a}, p_{t,a}^2, \theta_{t,a}^0) < \infty. \quad (6.18)
\]

Note that by (A5) the bounds in (6.15) and (6.16) are equivalent. Similarly (6.17) and (6.18) are equivalent.

By the subsolution inequality (6.10),

\[
\frac{1}{\Lambda} \inf_{E} (u_1 - h) \leq (1 - \varepsilon) \mathcal{H}(x_{t,a}, p_{t,a}^1) + \varepsilon C_Y \quad (6.19)
\]

\[
\leq (1 - \varepsilon) \Lambda(x_{t,a}, p_{t,a}^1, \theta_{t,a}) + \varepsilon C_Y,
\]

and the lower bounds (6.15) and (6.16) follow.

By the supersolution inequality (6.11), we can estimate

\[
(1 + \varepsilon) \Lambda(y_{t,a}, p_{t,a}^2, \theta_{t,a}^0) = (1 + \varepsilon) \left[ \Lambda(y_{t,a}, p_{t,a}^2, \theta_{t,a}^0) - \mathcal{I}(y_{t,a}, \theta_{t,a}^0) \right]
\]

\[
\leq \left( (1 + \varepsilon) \mathcal{H}(y_{t,a}, p_{t,a}^2) - \varepsilon C_Y \right) + \varepsilon C_Y
\]

\[
\leq \frac{1}{\Lambda} \sup_{E} (u_2 - h) + \varepsilon C_Y < \infty,
\]

and the upper bounds (6.17) and (6.18) follow.

Since the \( \theta_{t,a}^0 \) are contained in a compact set by (I3), we conclude by the continuity estimate (A4) that

\[
\liminf_{\varepsilon \to 0} \inf_{a \to \infty} \left[ \Lambda \left( x_{t,a}, p_{t,a}^1, \theta_{t,a}^0 \right) - \Lambda \left( y_{t,a}, p_{t,a}^2, \theta_{t,a}^0 \right) \right] \leq 0.
\]

Without loss of generality, we can choose for all small \( \varepsilon \) subsequences \( (x_{t,a}, y_{t,a}) \) (denoted the same) such that also

\[
\liminf_{\varepsilon \to 0} \inf_{a \to \infty} \left[ \Lambda \left( x_{t,a}, p_{t,a}^1, \theta_{t,a}^0 \right) - \Lambda \left( y_{t,a}, p_{t,a}^2, \theta_{t,a}^0 \right) \right] \leq 0. \quad (6.20)
\]

We proceed to establish that along this collection of subsequences we have \( \limsup_{a \to \infty} \mathcal{I}(x_{t,a}, \theta_{t,a}) < \infty. \) We return to the first inequality of (6.19), combined with (6.12), to obtain

\[
\frac{1}{\Lambda} \inf_{E} (u_1 - h) \leq (1 - \varepsilon) \mathcal{H}(x_{t,a}, p_{t,a}^1) + \varepsilon C_Y
\]

\[
= (1 - \varepsilon) \left[ \Lambda(x_{t,a}, p_{t,a}^1, \theta_{t,a}) - \mathcal{I}(x_{t,a}, \theta_{t,a}) \right] + \varepsilon C_Y.
\]

We conclude that \( \limsup_{a \to \infty} \mathcal{I}(x_{t,a}, \theta_{t,a}) < \infty \) is implied by

\[
\limsup_{a \to \infty} \Lambda(x_{t,a}, p_{t,a}^1, \theta_{t,a}) < \infty
\]

which by (A5) is equivalent to

\[
\limsup_{a \to \infty} \Lambda(x_{t,a}, p_{t,a}^1, \theta_{t,a}^0) < \infty.
\]

26
This, however, yields what we want by (6.18) and (6.20):
\[
\limsup_{a \to \infty} \Lambda(x_{t,a}, p_{t,a}^1, \theta_{t,a}^0) \leq \limsup_{a \to \infty} \Lambda(y_{t,a}, p_{t,a}^2, \theta_{t,a}^0) + \limsup_{a \to \infty} \left[ \Lambda \left( x_{t,a}, p_{t,a}^1, \theta_{t,a}^0 \right) \right] - \Lambda \left( y_{t,a}, p_{t,a}^2, \theta_{t,a}^0 \right) < \infty.
\]
We thus obtain
\[
\limsup_{a \to \infty} I(x_{t,a}, \theta_{t,a}) < \infty.
\]
Therefore, by (7.3), for each \( \epsilon > 0 \) the \( \theta_{t,a} \) are contained in a compact set. With the bounds (6.15) and (6.17), we conclude by the continuity estimate (A4) that
\[
\liminf_{\epsilon \to 0} \liminf_{a \to \infty} \left[ \Lambda \left( x_{t,a}, p_{t,a}^1, \theta_{t,a}^0 \right) \right] - \Lambda \left( y_{t,a}, p_{t,a}^2, \theta_{t,a}^0 \right) \leq 0. \tag{6.21}
\]
By (6.8), we have along a subsequence \( (x_{t,a}, y_{t,a}) \to (z_t, z_t) \in K_t \times K_t \) as \( a \to \infty \). Therefore by (7.4) there exists a subsequence of \( (x_{t,a}, y_{t,a}) \) (denoted the same) and a constant \( M'_t < \infty \) such that for all \( a > 0 \) large enough,
\[
I(x_{t,a}, \theta_{t,a}) \leq M'_t \quad \text{and} \quad I(y_{t,a}, \theta_{t,a}) \leq M'_t.
\]
Hence by (7.5), for any \( \epsilon > 0 \),
\[
\limsup_{a \to \infty} |I(y_{t,a}, \theta_{t,a}) - I(x_{t,a}, \theta_{t,a})| = 0. \tag{6.22}
\]
Then combining (6.21) with (6.22) gives an estimate on (6.13) which completes Step 2.

7 Construction of viscosity solutions

In this Section, we will show that \( R(\lambda)h \), for \( h \in C_b(E) \), \( \lambda > 0 \) of Theorem 3.7 is indeed a viscosity solution to \( f - \lambda Hf = h \). To do so, we will use the methods of Chapter 8 of [FK06] which are based on the strategy laid out in Section 4.2.

In particular, we will verify [FK06, Conditions 8.9, 8.10 and 8.11] which imply by [FK06, Theorem 8.27] and the comparison principle for \( f - \lambda Hf = h \) that \( R(\lambda)h \) is a viscosity solution to \( f - \lambda Hf = h \).

Verification of Conditions 8.9, 8.10 and 8.11. In the notation of [FK06], we use \( U = \mathbb{R}^d \), \( \Gamma = E \times U \), one operator \( H = H_T = H_T \) and \( A_f(x,u) = \langle \nabla f(x), u \rangle \) for \( f \in D(H) = C^\infty_c(E) \).

Regarding Condition 8.9, by continuity and convexity of \( H \) obtained in Propositions 5.1 and 5.3, parts 8.9.1, 8.9.2, 8.9.3 and 8.9.5 can be proven e.g. as in the proof of [FK06, Lemma 10.21] for \( \psi = 1 \). Part 8.9.4 is a consequence of the existence of a containment function, and follows as shown in the proof of Theorem A.17 of [CK17]. Since we use the argument further below, we briefly recall it here. We need to show that for any compact set \( K \subseteq E \), any finite time \( T > 0 \) and finite bound \( M \geq 0 \), there exists a compact set \( K' = K'(K, T, M) \subseteq E \) such that for any absolutely continuous path \( \gamma : [0, T] \to E \) with \( \gamma(0) \in K \), if
\[
\int_0^T \mathcal{L}(\gamma(t), \dot{\gamma}(t)) \, dt \leq M,
\]
then
then $\gamma(t) \in K'$ for any $0 \leq t \leq T.$

For $K \subseteq E, T > 0, M \geq 0$ and $\gamma$ as above, this follows by noting that

$$Y(\gamma(t)) = Y(\gamma(0)) + \int_0^T \nabla Y(\gamma(t)) \dot{\gamma}(t) \, dt$$

$$\leq Y(\gamma(0)) + \int_0^T [L(\gamma(t), \dot{\gamma}(t)) + \mathcal{H}(x(t), \nabla Y(\gamma(t)))] \, dt$$

$$\leq \sup_k Y + M + T \sup_{x \in E} \mathcal{H}(x, \nabla Y(x)) =: C < \infty,$$

(7.2)

for any $0 \leq \tau \leq T,$ so that the compact set $K' := \{z \in E : Y(z) \leq C\}$ satisfies the claim.

We proceed with the verification of Conditions 8.10 and 8.11 of [FK06]. By Proposition 5.1, we have $\mathcal{H}(x, 0) = 0$ and hence $H_1 = 0.$ Thus, Condition 8.10 is implied by Condition 8.11 (see Remark 8.12 (e) in [FK06]).

We establish that Condition 8.11 is satisfied: for any function $f \in \mathcal{D}(\mathcal{H}) = C^\infty_0(E)$ and $x_0 \in E,$ there exists an absolutely continuous path $x : [0, \infty) \to E$ such that $x(0) = x_0$ and for any $t \geq 0,$

$$\int_0^t \mathcal{H}(x(s), \nabla f(x(s))) \, ds = \int_0^t [\dot{x}(s) \cdot \nabla f(x(s)) - L(x(s), \dot{x}(s))] \, ds.$$

(7.3)

To do so, we solve the differential inclusion

$$\dot{x}(t) \in \partial_p \mathcal{H}(x(t), \nabla f(x(t))), \quad x(0) = x_0,$$

(7.4)

where the subdifferential of $\mathcal{H}$ was defined in (3.8) on page 10.

Since the addition of a constant to $f$ does not change the gradient, we may assume without loss of generality that $f$ has compact support. A general method to establish existence of differential inclusions $\dot{x} \in F(x)$ is given by Lemma 5.1 of Deimling [Dei92]. We have included this result as Lemma C.4, and corresponding preliminary definitions in Appendix C. We use this result for $F(x) := \partial_p \mathcal{H}(x, \nabla f(x)).$ To apply Lemma C.4, we need to verify that:

(F1) $F$ is upper hemi-continuous and $F(x)$ is non-empty, closed, and convex for all $x \in E.$

(F2) $\|F(x)\| \leq c(1 + |x|)$ on $E,$ for some $c > 0.$

(F3) $F(x) \cap T_E(x) \neq \emptyset$ for all $x \in E.$ (For the definition of $T_E,$ see Definition 3.15 on page 10.

While part (F1) follows from the properties of a subdifferential set and (F3) is a consequence of Assumption 3.16, part (F2) is in general not satisfied.

To circumvent this problem, we use properties of $\mathcal{H}$ to establish a-priori bounds on the range of solutions.

**Step 1:** Let $T > 0,$ and assume that $x(t)$ solves (7.4). We establish that there is some $M$ such that (7.1) is satisfied. By (7.4) we obtain for all $p \in \mathbb{R}^d,$

$$\mathcal{H}(x(t), p) \geq \mathcal{H}(x(t(\tau), \nabla f(x(t)(\tau))) + \dot{x}(t) \cdot (p - \nabla f(x(t))),$$

and as a consequence

$$\dot{x}(t) \nabla f(x(t)) - \mathcal{H}(x(t), \nabla f(x(t))) \geq L(x(t), \dot{x}(t)).$$
Since \( f \) has compact support and \( H(y,0) = 0 \) for any \( y \in E \), we estimate
\[
\int_0^T L(x(t), \dot{x}(t)) \, dt \leq \int_0^T \dot{x}(t) \nabla f(x(t)) \, dt - T \inf_{y \in \text{supp}(f)} H(y, \nabla f(y)).
\]

By continuity of \( H \) the field \( F \) is bounded on compact sets, so the first term can be bounded by
\[
\int_0^T \dot{x}(t) \nabla f(x(t)) \, dt \leq T \sup_{y \in \text{supp}(f)} \| F(y) \| \sup_{z \in \text{supp}(f)} | \nabla f(z) |.
\]

Therefore, for any \( T > 0 \), we obtain that the integral over the Lagrangian is bounded from above by
\[
M := T \sup_{y \in \text{supp}(f)} \| F(y) \| \sup_{z \in \text{supp}(f)} | \nabla f(z) | - \inf_{y \in \text{supp}(f)} H(y, \nabla f(y)).
\]

From the first part of the, see the argument concluding after (7.2), we find that the solution \( x(t) \) remains in the compact set
\[
K' := \{ z \in E \mid \Upsilon(z) \leq C \}, \quad C := \Upsilon(x_0) + M + T \sup_x H(x, \nabla \Upsilon(x)), \quad (7.5)
\]
for all \( t \in [0, T] \).

Step 2: We prove that there exists a solution \( x(t) \) of (7.4) on \([0, T]\).

Using \( F \), we define a new multi-valued vector-field \( F'(z) \) that equals \( F(z) = \partial_y H(z, \nabla f(z)) \) inside \( K' \), but equals \( \{ 0 \} \) outside a neighborhood of \( K \). This can e.g. be achieved by multiplying with a smooth cut-off function \( g_{K'} : E \to [0,1] \) that is equal to one on \( K' \) and zero outside of a neighborhood of \( K' \).

The field \( F' \) satisfies (F1), (F2) and (F3) from above, and hence there exists an absolutely continuous path \( y : [0, \infty) \to E \) such that \( y(0) = x_0 \) and for almost every \( t \geq 0 \),
\[
y(t) \in F'(y(t)).
\]

By the estimate established in step 1 and the fact that \( \Upsilon(y(t)) \leq C \) for any \( 0 \leq t \leq T \), it follows from the argument as shown above in (7.2) that the solution \( y \) stays in \( K' \) up to time \( T \). Since on \( K' \), we have \( F' = F \), this implies that setting \( x = y|_{[0,T]} \), we obtain a solution \( x(t) \) of (7.4) on the time interval \([0, T] \).

8 Verification for examples of Hamiltonians

In this section, we verify the conditions on \( \Lambda \) and \( I \) for the example Hamiltonians of Section 3.4. Since the conditions on the functions \( \Lambda \) and \( I \) are independent of each other, we verify these conditions separately. In Section 8.1, we consider Assumption 3.13 for \( I \). In Sections 8.2, we consider Assumption 3.12 for \( \Lambda \). The continuity estimates will be verified separately in Section 8.3.
8.1 Verifying assumptions for cost functions \( \mathcal{I} \)

We verify Assumption 3.13 for two types of cost functions \( \mathcal{I}(x, \theta) \), corresponding to the examples of Section 3.4.

We start by considering the case in which the cost function is the large deviation rate function for the occupation-time measures of jump process taking values in a finite set \( \{1, \ldots, J\} \), see e.g. [DV75a, DH08]. We follow this example in Proposition 8.2 in which the cost function stems from occupation-time large deviations of a drift-diffusion process on a compact manifold, see e.g. [DV75b, Pin07]. We expect these results to extend also to non-compact spaces, but we feel this is better suited for a separate paper.

**Proposition 8.1** (Donsker-Varadhan functional for jump processes). Consider a finite set \( F = \{1, \ldots, J\} \) and let \( \Theta := \mathcal{P}(\{1, \ldots, J\}) \) be the set of probability measures on \( F \). For \( x \in E \), let \( L_x : C_b(F) \to C_b(F) \) be the operator given by

\[
L_x f(i) := \sum_{j=1}^{J} r(i, j, x) [f(j) - f(i)], \quad f : \{1, \ldots, J\} \to \mathbb{R}.
\]

Suppose that the rates \( r : \{1, \ldots, J\}^2 \times E \to \mathbb{R}^+ \) are continuous as a function on \( E \) and moreover satisfy the following:

(i) For any \( x \in E \), the matrix \( R(x) \) with entries \( R(x)_{ij} := r(i, j, x) \) for \( i \neq j \) and \( R(x)_{ii} = -\sum_{j \neq i} r(i, j, x) \) is irreducible.

(ii) For each pair \((i, j)\), we either have \( r(i, j, \cdot) \equiv 0 \) or for each compact set \( K \subseteq E \), it holds that

\[
r_K(i, j) := \inf_{x \in K} r(i, j, x) > 0.
\]

Then the Donsker-Varadhan functional \( \mathcal{I} : E \times \Theta \to \mathbb{R}^+ \) defined by

\[
\mathcal{I}(x, \theta) := \sup_{w \in \mathbb{R}^J} \sum_{ij} r(i, j, x)\theta_i [1 - e^{w_j - w_i}]
\]

satisfies Assumption 3.13.

**Proof.** \((I1):\) For a fixed vector \( w \in \mathbb{R}^J \), the map

\[
(x, \theta) \mapsto \sum_{ij} r(i, j, x)\theta_i [1 - e^{w_j - w_i}]
\]

is continuous on \( E \times \Theta \). Hence \( \mathcal{I}(x, \theta) \) is lower semicontinuous as the supremum over continuous functions.

\((I2):\) Let \( x \in E \). First note that for all \( \theta \), the choice \( w = 0 \) implies that \( \mathcal{I}(x, \theta) \geq 0 \). By the irreducibility assumption on the rates \( r(i, j, x) \), there exists a unique measure \( \theta_x \in \Theta \) such that for any \( f : \{1, \ldots, J\} \to \mathbb{R} \),

\[
\sum_{i} L_x f(i)\theta_x(i) = 0. \tag{8.1}
\]

We establish that \( \mathcal{I}(x, \theta_x) = 0 \). Let \( w \in \mathbb{R}^J \). By the elementary estimate

\[
(1 - e^{b-a}) \leq b - a \quad \text{for all} \quad a, b > 0,
\]

Theorem 8.3 (Donsker-Varadhan functional for jump processes). Consider a finite set \( F = \{1, \ldots, J\} \) and let \( \Theta := \mathcal{P}(\{1, \ldots, J\}) \) be the set of probability measures on \( F \). For \( x \in E \), let \( L_x : C_b(F) \to C_b(F) \) be the operator given by

\[
L_x f(i) := \sum_{j=1}^{J} r(i, j, x) [f(j) - f(i)], \quad f : \{1, \ldots, J\} \to \mathbb{R}.
\]

Suppose that the rates \( r : \{1, \ldots, J\}^2 \times E \to \mathbb{R}^+ \) are continuous as a function on \( E \) and moreover satisfy the following:

(i) For any \( x \in E \), the matrix \( R(x) \) with entries \( R(x)_{ij} := r(i, j, x) \) for \( i \neq j \) and \( R(x)_{ii} = -\sum_{j \neq i} r(i, j, x) \) is irreducible.

(ii) For each pair \((i, j)\), we either have \( r(i, j, \cdot) \equiv 0 \) or for each compact set \( K \subseteq E \), it holds that

\[
r_K(i, j) := \inf_{x \in K} r(i, j, x) > 0.
\]

Then the Donsker-Varadhan functional \( \mathcal{I} : E \times \Theta \to \mathbb{R}^+ \) defined by

\[
\mathcal{I}(x, \theta) := \sup_{w \in \mathbb{R}^J} \sum_{ij} r(i, j, x)\theta_i [1 - e^{w_j - w_i}]
\]

satisfies Assumption 3.13.

**Proof.** \((I1):\) For a fixed vector \( w \in \mathbb{R}^J \), the map

\[
(x, \theta) \mapsto \sum_{ij} r(i, j, x)\theta_i [1 - e^{w_j - w_i}]
\]

is continuous on \( E \times \Theta \). Hence \( \mathcal{I}(x, \theta) \) is lower semicontinuous as the supremum over continuous functions.

\((I2):\) Let \( x \in E \). First note that for all \( \theta \), the choice \( w = 0 \) implies that \( \mathcal{I}(x, \theta) \geq 0 \). By the irreducibility assumption on the rates \( r(i, j, x) \), there exists a unique measure \( \theta_x \in \Theta \) such that for any \( f : \{1, \ldots, J\} \to \mathbb{R} \),

\[
\sum_{i} L_x f(i)\theta_x(i) = 0. \tag{8.1}
\]

We establish that \( \mathcal{I}(x, \theta_x) = 0 \). Let \( w \in \mathbb{R}^J \). By the elementary estimate

\[
(1 - e^{b-a}) \leq b - a \quad \text{for all} \quad a, b > 0,
\]
we obtain that
\[
\sum_{ij} r(i, j, x) \theta_z(i) (1 - e^{w_j - w_i}) \leq \sum_{ij} r(i, j, x) \theta_z(i) (w_j - w_i) = \sum_i (L_\mu w)(i) \theta_z(i) = 0
\]
by (8.1). Since \( I \geq 0 \), this implies \( I(x, \theta_z) = 0 \).

(\text{I3}): Any closed subset of \( \Theta \) is compact.

(\text{I4}): Let \( x_n \rightarrow x \) in \( E \). It follows that the sequence is contained in some compact set \( K \subseteq E \) that contains the \( x_n \) and \( x \) in its interior. For any \( y \in K \),
\[
I(y, \theta) \leq \sum_{ij, i \neq j} r(i, j, y) \theta_i \leq \sum_{ij, i \neq j} r(i, j, y) \leq \sum_{ij, i \neq j} \bar{r}_{ij}, \quad \bar{r}_{ij} := \sup_{y \in K} r(i, j, y).
\]
Hence \( I \) is uniformly bounded on \( K \times \Theta \), and (I4) follows with \( U_z \) the interior of \( K \).

(\text{I5}): Let \( d \) be some metric that metrizes the topology of \( E \). We will prove that for any compact set \( K \subseteq E \) and \( \varepsilon > 0 \) there is some \( \delta > 0 \) such that for all \( x, y \in K \) with \( d(x, y) \leq \delta \) and for all \( \theta \in \mathcal{P}(F) \), we have
\[
|I(x, \theta) - I(y, \theta)| \leq \varepsilon. \tag{8.2}
\]

Let \( x, y \in K \). By continuity of the rates the \( I(x, \cdot) \) are uniformly bounded for \( x \in K \):
\[
0 \leq I(x, \theta) \leq \sum_{ij, i \neq j} r(i, j, x) \theta_i \leq \sum_{ij, i \neq j} r(i, j, x) \leq \sum_{ij, i \neq j} \bar{r}_{ij}, \quad \bar{r}_{ij} := \sup_{x \in K} r(i, j, x).
\]

For any \( n \in \mathbb{N} \), there exists \( w^n \in \mathbb{R}^j \) such that
\[
0 \leq I(x, \theta) \leq \sum_{ij, i \neq j} r(i, j, x) \theta_i (1 - e^{w_j^n - w_i^n}) + \frac{1}{n}.
\]
By reorganizing, we find for all bonds \( (a, b) \) the bound
\[
\theta_a e^{w_j^n - w_i^n} \leq \frac{1}{T_{K,a,b}} \left[ \sum_{ij, i \neq j, a \neq i, b \neq j} r(i, j, x) \theta_i + \frac{1}{n} \right] \leq \frac{1}{T_{K,a,b}} \left[ \sum_{ij, i \neq j} \bar{r}_{ij} + \frac{1}{n} \right].
\]

Thereby, evaluating in \( I(y, \theta) \) the same vector \( w^n \) to estimate the supremum,
\[
I(x, \theta) - I(y, \theta) \leq \frac{1}{n} + \sum_{ab, a \neq b} r(a, b, x) \theta_a (1 - e^{w_j^n - w_i^n}) - \sum_{ab, a \neq b} r(a, b, y) \theta_a (1 - e^{w_j^n - w_i^n})
\]
\[
\leq \frac{1}{n} + \sum_{ab, a \neq b} |r(a, b, x) - r(a, b, y)| \theta_a + \sum_{ab, a \neq b} |r(a, b, y) - r(a, b, x)| \theta_a e^{w_j^n - w_i^n}
\]
\[
\leq \frac{1}{n} + \sum_{ab, a \neq b} |r(a, b, x) - r(a, b, y)| \left( 1 + \frac{1}{T_{K,a,b}} \left[ \sum_{ij, i \neq j} \bar{r}_{ij} + 1 \right] \right)
\]
We take \( n \to \infty \) and use that the rates \( x \mapsto r(a, b, x) \) are continuous, and hence uniformly continuous on compact sets, to obtain (8.2).
**Proposition 8.2** (Donsker-Varadhan functional for drift-diffusions). Let $F$ be a smooth compact Riemannian manifold without boundary and set $\Theta := \mathcal{P}(F)$, the set of probability measures on $F$. For $x \in E$, let $L_x : C^2(F) \subseteq C_b(F) \to C_b(F)$ be the second-order elliptic operator that in local coordinates is given by

$$L_x = \frac{1}{2} \nabla \cdot (a_x \nabla) + b_x \cdot \nabla,$$

where $a_x$ is a positive definite matrix and $b_x$ is a vector field having smooth entries $a_i^j$ and $b_i^j$ on $F$. Suppose that for all $i, j$ the maps

$$x \mapsto a_i^j(\cdot), \quad x \mapsto b_i^j(\cdot)$$

are continuous as functions from $E$ to $C_b(F)$, where we equip $C_b(F)$ with the supremum norm. Then the functional $I : E \times \Theta \to [0, \infty]$ defined by

$$I(x, \theta) := \sup_{u \in D(L_x)} \left[ - \int_F \frac{L_x u}{u} \ d\theta \right]$$

satisfies Assumption 3.13.

**Proof.** \(\text{(I)}\): For any fixed function $u \in D(L_x)$ that is strictly positive on $F$, the function $(-L_x u/u)$ is continuous on $F$. For any fixed $u$ it follows by (8.3) and compactness of $f$ that

$$(x, \theta) \mapsto - \int_F \frac{L_x u}{u} \ d\theta$$

is continuous on $E \times \Theta$. As a consequence $I(x, \theta)$ is lower semicontinuous as the supremum over continuous functions.

\(\text{(II)}\): Let $x \in E$. The stationary measure $\theta_x \in \Theta$ satisfying

$$\int_F L_x g(z) \ d\theta_x(z) = 0 \quad \text{for all } g \in D(L_x)$$

(8.4)

is the minimizer of $I(x, \cdot)$, that is $I(x, \theta_x) = 0$. This follows by considering the Hille-Yosida approximation $L_x^\epsilon$ of $L_x$ and using the same argument (using $w = \log u$) as in Proposition 8.1 for these approximations. For any $u > 0$ and $\epsilon > 0$,

$$- \int_F \frac{L_x u}{u} \ d\theta = - \int_F \frac{L_x^\epsilon u}{u} \ d\theta + \int_F \frac{(L_x^\epsilon - L_x) u}{u} \ d\theta$$

$$\leq - \int_F \frac{L_x^\epsilon u}{u} \ d\theta + \frac{1}{\inf_{F} u} \| (L_x^\epsilon - L_x) u \|_F$$

$$\leq - \int_F L_x^\epsilon \log(u) \ d\theta + o(1).$$

Sending $\epsilon \to 0$ and then using (8.4) gives \(\text{(II)}\).

\(\text{(III)}\): Since $\Theta = \mathcal{P}(F)$ is compact, any closed subset of $\Theta$ is compact. Hence any union of sub-level sets of $I(x, \cdot)$ is relatively compact in $\Theta$.

\(\text{(IV)}\): Let $x_n \to x$ in $E$ and $\theta_n$ be a sequence in $\Theta$, and suppose that $I(x_n, \theta_n) \leq M$ for some constant $M$ independent of $n$. Let $dz$ be the Riemannian measure on $F$. By Pinsky’s results in [Pin85, Pin07], if $I(y, \theta) < \infty$, then the density $\frac{d\pi}{dz}$ exists. In addition, there are constants $c_1, c_2, c_3, c_4$ depending only on $a_y, b_y$, and not on $\theta$, such that

$$c_1(y) \int_F |\nabla g_\theta|^2 \ dz - c_2(y) \leq I(y, \theta) \leq c_3(y) \int_F |\nabla g_\theta|^2 \ dz + c_4(y), \quad (8.5)$$
Furthermore, there exists a constant $L > 0$, in particular, as can be seen by the derivation of [P’s Eq. (2.18), (2.19)], the constants depend continuously on $y \in E$ by our continuity assumptions on $a_y$ and $b_y$.

Applying this to our sequences $x_n$ and $\theta_n$, we have

$$\int_E |\nabla g_{\theta_n}|^2 \, dz \leq M',$$

for a constant $M'$. This implies again by (8.5) that for any $y$ in some neighborhood of $x$ that

$$\mathcal{I}(y, \theta_n) \leq C < \infty,$$

with a constant independent of $n$.

$(\mathcal{I}5)$: Since the coefficients $a_x$ and $b_x$ of the operator $L_x$ depend continuously on $x$, assumption $(\mathcal{I}5)$ follows from Theorem 2 of [Pin07].

### 8.2 Verifying assumptions for functions $\Lambda$

We verify Assumption 3.12 for three types of functions $\Lambda$ corresponding to the examples of Section 3.4. We start with $\Lambda$’s that are given as integrals over quadratic polynomials in $p$.

**Proposition 8.3 (Quadratic function $\Lambda$).** Let $E = \mathbb{R}^d$ and $\Theta = \mathcal{P}(F)$ for some compact Polish space $F$. Suppose that the function $\Lambda : E \times \mathbb{R}^d \times \Theta \to \mathbb{R}$ is given by

$$\Lambda(x, p, \theta) = \int_F (a(x, z)p, p) \, d\theta(z) + \int_F (b(x, z), p) \, d\theta(z),$$

where $a : E \times F \to \mathbb{R}^{d \times d}$ and $b : E \times F \to \mathbb{R}^d$ are continuous. Suppose that for every compact set $K \subseteq \mathbb{R}^d$,

$$a_{K, \text{min}} := \inf_{x \in K, z \in F, |p| = 1} \langle a(x, z)p, p \rangle > 0,$$

$$a_{K, \text{max}} := \sup_{x \in K, z \in F, |p| = 1} \langle a(x, z)p, p \rangle < \infty,$$

$$b_{K, \text{max}} := \sup_{x \in K, z \in F, |p| = 1} |\langle b(x, z), p \rangle| < \infty.$$

Furthermore, there exists a constant $L > 0$ such that for all $x, y \in E$ and $z \in F$,

$$\|a(x, z) - a(y, z)\| \leq L|x - y|,$$

and suppose that the functions $b$ are one-sided Lipschitz continuous. Then Assumption 3.12 holds.

**Proof.** $(\Lambda 1)$: Let $(x, p) \in E \times \mathbb{R}^d$. By the boundedness assumptions on $a$ and $b$,

$$\sup_{\theta} |\Lambda(x, p, \theta)| \leq a_{\{x\}, \text{max}} + b_{\{x\}, \text{max}} < \infty,$$

and hence the function $\theta \mapsto |\Lambda(x, p, \theta)|$ is bounded on $\mathcal{P}(F)$. Continuity of $\Lambda$ is a consequence of the fact that

$$\Lambda(x, p, \theta) = \int_F V(x, p, z) \, d\theta(z)$$

is the pairing of a continuous and bounded function $V(x, p, \cdot)$ with the measure $\theta \in \mathcal{P}(F)$.
(Λ2): Let $x \in E$ and $\theta \in \mathcal{P}(F)$. Convexity of $p \mapsto \Lambda(x, p, \theta)$ follows since $a(x, z)$ is positive definite by assumption. If $p_0 = 0$, then evidently $\Lambda(x, p_0, \theta) = 0$.

(Λ3): We show that the map $Y : E \to \mathbb{R}$ defined by
\[
Y(x) := \frac{1}{2} \log \left(1 + |x|^2 \right)
\]
is a containment function for $\Lambda$. For any $x \in E$ and $\theta \in \mathcal{P}(F)$, we have
\[
\Lambda(x, \nabla Y(x), \theta) = \int_E \langle a(x, z) \nabla Y(x), \nabla Y(x) \rangle \, d\theta(z) + \int_E \langle b(x, z), \nabla Y(x) \rangle \, d\theta(z) \\
\leq a_{\{x\}, \max} |\nabla Y(x)|^2 + b_{\{x\}, \max} |\nabla Y(x)| \\
\leq C(1 + |x|) \frac{x^2}{(1 + x^2)^2} + C(1 + |x|) \frac{x}{1 + x^2},
\]
and the boundedness condition follows with the constant
\[
C_Y := C \sup_x (1 + |x|) \left[ \frac{x^2}{(1 + x^2)^2} + \frac{x}{1 + x^2} \right] < \infty.
\]

(Λ4): By the assumption on $a(x, z)$, the function $\Lambda$ is uniformly coercive in the sense that for any compact set $K \subseteq E$,
\[
\inf_{x \in K, \theta \in \Theta} \Lambda(x, p, \theta) \to \infty \quad \text{as} \quad |p| \to \infty,
\]
and the continuity estimate follows by Proposition 8.6.

(Λ5): Let $K \subseteq E$ be compact. We have to show that there exist constants $M, C_1, C_2 \geq 0$ such that for all $x \in K$, $p \in \mathbb{R}^d$ and all $\theta_1, \theta_2 \in \mathcal{P}(F)$, we have
\[
\Lambda(x, p, \theta_1) \leq \max \{M, C_1 \Lambda(x, p, \theta_2) + C_2 \}. \quad (8.6)
\]
Fix $\theta_1, \theta_2 \in \mathcal{P}(F)$. We have for $x \in K$
\[
\int \langle a(x, z) p, p \rangle \, d\theta_1(z) \leq \frac{a_{K, \max}}{a_{K, \min}} \int \langle a(x, z) p, p \rangle \, d\theta_2(z)
\]
In addition, as $a_{K, \min} > 0$ and $b_{K, \max} < \infty$ we have for any $C > 0$ and sufficiently large $|p|$ that
\[
\int \langle b(x, z), p \rangle \, d\theta_1(z) - (C + 1) \int \langle b(x, z), p \rangle \, d\theta_2(z) \leq C \int \langle a(x, z) p, p \rangle \, d\theta_2(z)
\]
Thus, for sufficiently large $|p|$ (depending on $C$) we have
\[
\Lambda(x, p, \theta_1) \leq (1 + C) \Lambda(x, p, \theta_2).
\]
Fix a $C := C_1$ and denote the set of ‘large’ $p$ by $S$. The map $(x, p, \theta) \mapsto \Lambda(x, p, \theta)$ is bounded on $K \times \times S \times \Theta$. Thus, we can find a constant $C_2$ such that (8.6) holds. \qed

We proceed with an example in which $\Lambda$ depends on $p$ through exponential functions. Let $q \in \mathbb{N}$ be an integer and
\[
\Gamma := \{(a, b) : a, b \in \{1, \ldots, q\}, a \neq b\}
\]
be the set of oriented edges in $\{1, \ldots, q\}$. 34
Proposition 8.4 (Exponential function $\Lambda$). Let $E \subseteq \mathbb{R}^d$ be the embedding of $E = \mathcal{P}(\{1, \ldots, q\}) \times (\mathbb{R}^+)^\Gamma$ and $\Theta$ be a topological space. Suppose that $\Lambda$ is given by

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

where $v$ is a proper kernel in the sense of Definition 3.23. Suppose in addition that there is a constant $C > 0$ such that for all $(a, b) \in \Gamma$ such that $v(a, b, \cdot, \cdot) \neq 0$ we have

$$\sup_{\mu} \sup_{\theta_1, \theta_2} \frac{v(a, b, \mu, \theta_1)}{v(a, b, \mu, \theta_2)} \leq C. \quad (8.7)$$

Then $\Lambda$ satisfies Assumption 3.12.

Remark 8.5. Similar to previous proposition, the assumptions on $\Lambda$ are satisfied when $\Theta = \mathcal{P}(F)$ for some Polish space $F$, and if $v(a, b, \mu, \theta) = \mu(a) \int r(a, b, \mu, z) \theta(dz)$ and there are constants $0 < r_{\min} \leq r_{\max} < \infty$ such that for all $(a, b) \in \Gamma$ such that $\sup_{\mu} r(a, b, \mu, z) > 0$, we have

$$r_{\min} \leq \inf_{\mu} \inf_{z} r(a, b, \mu, z) \leq \sup_{\mu} \sup_{z} r(a, b, \mu, z) \leq r_{\max}.$$

Regarding (8.7), for $(a, b) \in \Gamma$ for which $v(a, b, \cdot, \cdot)$ is non-trivial, we have

$$\frac{v(a, b, \mu, \theta_1)}{v(a, b, \mu, \theta_2)} = \frac{\int r(a, b, \mu, z) \theta_1(dz)}{\int r(a, b, \mu, z) \theta_2(dz)} \leq \frac{r_{\max}}{r_{\min}}.$$

Proof of Proposition 8.4. (A1): The function $\Lambda$ is continuous as the sum of continuous functions. Boundedness of $\Lambda$ as a function of $\theta$ follows from the boundedness assumption (8.7).

(A2): Convexity of $\Lambda$ as a function of $p$ follows from the fact that $\Lambda$ is a finite sum of convex functions, and $\Lambda(\mu, 0, \theta) = 0$ is evident.

(A3): The function $Y : E \to \mathbb{R}$ defined by

$$Y(\mu, w) := \sum_{(a, b) \in \Gamma} \log \left[ 1 + \frac{w_{(a,b)}}{z_{(a,b)}} \right]$$

is a containment function for $\Lambda$. For a verification, see [Kra17].

(A4): The continuity estimate is the content of Proposition 8.9 below.

(A5): Note that

$$\Lambda((\mu, w), \theta_1, p) \leq \sum_{(a, b) \in \Gamma} v(a, b, \mu, \theta_1) e^{p_{a} + p_{b} - p_{a}}$$

$$\leq C \sum_{(a, b) \in \Gamma} v(a, b, \mu, \theta_2) e^{p_{a} + p_{b} - p_{a}}$$

$$\leq C \sum_{(a, b) \in \Gamma} v(a, b, \mu, \theta_2) \left[ e^{p_{a} + p_{b} - p_{a}} - 1 \right] + C_2.$$
8.3 Verifying the continuity estimate

With the exception of the verification of the continuity estimate in Assumption 3.12 the verification in Section 8.2 is straightforward. On the other hand, the continuity estimate is an extension of the comparison principle, and is therefore more complex. We verify the continuity estimate in three contexts, which we hope, illustrates, that the continuity estimate follows from essentially the same arguments as the standard comparison principle. We will do this for:

- Coercive Hamiltonians
- One-sided Lipschitz Hamiltonians
- Hamiltonians arising from large deviations of empirical measures.

This list is not meant to be an exhaustive list, but to illustrate that the continuity estimate is a sensible extension of the comparison principle, which is satisfied in a wide range of contexts. In what follows, \( E \subseteq \mathbb{R}^d \) is a Polish subset and \( \Theta \) a topological space.

**Proposition 8.6 (Coercive \( \Lambda \)).** Let \( \Lambda : E \times \mathbb{R}^d \times \Theta \to \mathbb{R} \) be continuous and uniformly coercive: that is, for any compact \( K \subseteq E \) we have

\[
\inf_{x \in K, \theta \in \Theta} \Lambda(x, p, \theta) \to \infty \quad \text{as} \quad |p| \to \infty.
\]

Then the continuity estimate holds for \( \Lambda \) with respect to any penalization function \( \Psi \).

**Proof.** Let \( \Psi(x, y) = \frac{1}{2}(x - y)^2 \). Let \((x_{a, \epsilon}, y_{a, \epsilon}, \theta_{e, a})\) be fundamental for \( \Lambda \) with respect to \( \Psi \). Set \( p_{a, \epsilon} = a(x_{e, \epsilon} - y_{e, \epsilon}) \). By the upper bound (3.5), we find that for sufficiently small \( \epsilon > 0 \) there is some \( a(\epsilon) \) such that

\[
\sup_{a \geq a(\epsilon)} \Lambda(y_{a, \epsilon}, p_{a, \epsilon}, \theta_{e, a}) < \infty.
\]

As the variables \( y_{a, \epsilon} \) are contained in a compact set by property (C1) of fundamental collections of variables, the uniform coercivity implies that the momenta \( p_{a, \epsilon} \) for \( a \geq a(\epsilon) \) remain in a bounded set. Thus, we can extract a subsequence \( a' \) such that \((x_{e, a'}, y_{e, a'}, p_{e, a'}, \theta_{e, a'})\) converges to \((x, y, p, \theta)\) with \( x = y \) due to property (C2) of fundamental collections of variables. By continuity of \( \Lambda \) we find

\[
\lim_{a \to a'} \inf_{a \to a} \Lambda(x_{a, \epsilon}, p_{a, \epsilon}, \theta_{e, a}) - \Lambda(y_{a, \epsilon}, p_{a, \epsilon}, \theta_{e, a}) \\
\leq \lim_{a' \to a} \Lambda(x_{a', \epsilon}, p_{a', \epsilon}, \theta_{e, a'}) - \Lambda(y_{e, a'}, p_{e, a'}, \theta_{e, a'}) = 0
\]

establishing the continuity estimate. \( \square \)

**Proposition 8.7 (One-sided Lipschitz \( \Lambda \)).** Let \( \Lambda : E \times \mathbb{R}^d \times \Theta \to \mathbb{R} \) satisfy

\[
\Lambda(x, a(x - y), \theta) - \Lambda(y, a(x - y), \theta) \leq c(\theta)\omega(a(x - y)^2)
\]

(8.8)

for some collection of constants \( c(\theta) \) satisfying \( \sup_\theta c(\theta) < \infty \) and a function \( \omega : \mathbb{R}^+ \to \mathbb{R}^+ \) satisfying \( \lim_{\delta \to 0} \omega(\delta) = 0 \).

Then the continuity estimate holds for \( \Lambda \) with respect to \( \Psi(x, y) = \frac{1}{2}(x - y)^2 \).
Proof. Let $\Psi(x, y) = \frac{1}{2}(x - y)^2$. Let $(x_{\alpha, t}, y_{\alpha, t}, \theta_{t, a})$ be fundamental for $\Lambda$ with respect to $\Psi$. Set $p_{a, t} = a(x_{t, a} - y_{t, a})$. We find

$$
\liminf_{a \to \infty} \Lambda(x_{t, a}, p_{t, a}, \theta_{t, a}) - \Lambda(y_{t, a}, p_{t, a}, \theta_{t, a}) \\
\leq \liminf_{a \to \infty} c(\theta) \omega(a(x - y)^2)
$$

which equals 0 as $\sup_{\theta} c(\theta) < \infty$, $\lim_{t \to 0} \omega(\theta) = 0$ and property (C1) of a fundamental collection of variables.

For the empirical measure of a collection of independent processes one obtains maps $\Lambda$ that are neither uniformly coercive nor Lipschitz. Also in this context one can establish the continuity estimate. We treat a simple 1d case and then state a more general version for which we refer to [Kra17].

**Proposition 8.8.** Suppose that $E = [-1, 1]$ and that $\Lambda(x, p, \theta)$ is given by

$$
\Lambda(x, p, \theta) = \frac{1 - x}{2} c^2(\theta) \left[ e^{2p - 1} + \frac{1 + x}{2} c_-(\theta) \left[ e^{-2p} - 1 \right] \right]
$$

with $c_-, c_+$ non-negative functions of $\theta$. Then the continuity estimate holds for $\Lambda$ with respect to $\Psi(x, y) = \frac{1}{2}(x - y)^2$.

**Proof.** Let $\Psi(x, y) = \frac{1}{2}(x - y)^2$. Let $(x_{\alpha, t}, y_{\alpha, t}, \theta_{t, a})$ be fundamental for $\Lambda$ with respect to $\Psi$. Set $p_{a, t} = a(x_{t, a} - y_{t, a})$.

We have

$$
\Lambda(x_{t, a}, p_{t, a}, \theta_{t, a}) - \Lambda(y_{t, a}, p_{t, a}, \theta_{t, a}) \\
= \frac{y_{t, a} - x_{t, a}}{2} c^2(\theta_{t, a}) \left[ e^{2p_{t, a} - 1} + \frac{1 + x_{t, a}}{2} c^2(\theta_{t, a}) \left[ e^{-2p_{t, a}} - 1 \right] \right]
$$

Now note that $y_{t, a} - x_{t, a}$ is positive if and only if $e^{2p_{t, a} - 1}$ is negative so that the first term is bounded above by 0. With a similar argument the second term is bounded above by 0. Thus the continuity estimate is satisfied. 

**Proposition 8.9.** Suppose $E = \mathcal{P}([1, \ldots, q]) \times (\mathbb{R}^+)^\Gamma$ and suppose that $\Lambda$ is given by

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

where $v$ is a proper kernel. Then the continuity estimate holds for $\Lambda$ with respect to penalization functions (see Section B)

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

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

Here we denote $r^+ = r \vee 0$ for $r \in \mathbb{R}$.

In this context, one can use coercivity like in Proposition 8.6 in combination with directional properties used in the proof of Proposition 8.8 above.

To be more specific: the proof of this proposition can be carried out exactly as the proof of Theorem 3.8 of [Kra17]: namely at any point a converging subsequence is constructed, the variables $a$ need to be chosen such that we also get convergence of the measures $\theta_{t, a}$ in $\mathcal{P}(F)$. 

37
Acknowledgment

MS acknowledges financial support through NWO grant 613.001.552.

A Viscosity solutions

In Section 6 we work with a pair of Hamilton-Jacobi equations instead of a single Hamilton-Jacobi equation. To this end, we need to extend the notion of a viscosity solution and that of the comparison principle of Section 3.1.

Definition A.1. Let $A_1 \subseteq \mathcal{C}(E) \times \mathcal{C}(E)$ and $A_2 \subseteq \mathcal{C}(E) \times \mathcal{C}(E)$. Fix $\lambda > 0$ and $h_1, h_2 \in C_b(E)$. Consider the equations

$$f - \lambda A_1 f = h_1,$$

$$f - \lambda A_2 f = h_2.$$

We say that $u$ is a (viscosity) subsolution of equation (A.1) if

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

$$\lim_{n \to \infty} u(x_n) - \lambda g(x_n) - h(x_n) \leq 0.$$

We say that $v$ is a (viscosity) supersolution of equation (A.2) if

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

$$\lim_{n \to \infty} v(x_n) - \lambda g(x_n) - h(x_n) \geq 0.$$

If $h_1 = h_2$, we say that $u$ is a (viscosity) solution of equations (A.1) and (A.2) if it is both a subsolution to (A.1) and a supersolution to (A.2).

We say that (A.1) and (A.2) satisfy the comparison principle if for every subsolution $u$ to (A.1) and supersolution $v$ to (A.2), we have $\sup_E u - v \leq \sup_E h_1 - h_2$.

As before, if test functions have compact levelsets, the existence of a sequences can be replaced by the existence of a point.

B A more general continuity estimate

In classical literature, the comparison principle for the Hamilton-Jacobi equation $f - \lambda H f = h$ is often proven using a squared distance as a penalization function. This often works well due to the quadratic structure of the Hamiltonian. In different contexts, e.g. for the Hamiltonians arising from the large deviations of jump processes, this is not natural, see the issues arising in the proofs in [DIS90, Kra17]. In absence of a general method to solve these issues, ad-hoc procedures can be introduced. One such ad-hoc procedure introduced in [Kra17] is to work with multiple penalization functions (in that context $\{\Psi_1, \Psi_2\}$) that explore different parts of the state-space.
Any argument that has been carried out in the main text can be carried out with the generalization of the continuity estimate below.

Definition B.1. We say that $\{\Psi_i\}_{i \in \{1, \ldots, k\}}$, $\Psi_i : E^2 \to \mathbb{R}^+$ is a collection 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 all $i$.

Definition B.2 (Continuity estimate). Let $G : E \times \mathbb{R}^d \times \Theta : (x, p, \theta) \mapsto G(x, p, \theta)$ be a function and $\{\Psi_i\}$ be a collection of penalization functions. Suppose that we have a collection of variables $(x_{\epsilon, \theta}, y_{\epsilon, \theta})$ in $E^2$ and $\theta_{\epsilon, \theta}$ in $\Theta$. We say that this collection is fundamental for $G$ and $\{\Psi_i\}$ if

(C1) For each $\epsilon$, there are compact sets $K_\epsilon \subseteq E$ and $\tilde{K}_\epsilon \subseteq \Theta$ such that for all $\alpha$ we have $x_{\epsilon, \theta}, y_{\epsilon, \theta} \in K_\epsilon$ and $\theta_{\epsilon, \theta} \in \tilde{K}_\epsilon$.

(C2) For each $\epsilon > 0$, we have the following inductive statement: For each $m \in \{1, \ldots, k\}$ and $\alpha_{m+1}, \ldots, \alpha_k > 0$ there are limit points $x_{(\alpha_{m+1}, \ldots, \alpha_k), \epsilon} \in K_\epsilon$ and $y_{(\alpha_{m+1}, \ldots, \alpha_k), \epsilon} \in K_\epsilon$ of $x_{(\alpha_m, \ldots, \alpha_k), \epsilon}$ and $y_{(\alpha_m, \ldots, \alpha_k), \epsilon}$ as $\alpha_m \to \infty$. For these limit points we have

$$\lim_{\alpha_m \to \infty} \sum_{i=1}^m \Psi_i(x_{(\alpha_m, \ldots, \alpha_k), \epsilon}, y_{(\alpha_m, \ldots, \alpha_k), \epsilon}) = 0,$$

and

$$\lim_{m \to \infty} \sum_{i=1}^m \Psi_i(x_{(\alpha_{m+1}, \ldots, \alpha_k), \epsilon}, y_{(\alpha_{m+1}, \ldots, \alpha_k), \epsilon}) = 0.$$

(C3) We have

\begin{align}
\limsup_{\epsilon \to 0} \limsup_{\alpha_1 \to \infty} \ldots \limsup_{\alpha_k \to \infty} G \left( y_{\alpha, \theta} - \sum_{i=1}^k a_i (\nabla \Psi_i(x_{\alpha, \theta})) (y_{\alpha, \theta}), \theta_{\alpha, \theta} \right) < \infty, \\
\liminf_{\epsilon \to 0} \liminf_{\alpha_1 \to \infty} \ldots \liminf_{\alpha_k \to \infty} \left( x_{\alpha, \theta} - \sum_{i=1}^k a_i (\nabla \Psi_i(y_{\alpha, \theta})) (y_{\alpha, \theta}), \theta_{\alpha, \theta} \right) > -\infty,
\end{align}

(B.1)

(B.2)

In other words, the operator $G$ evaluated in the proper momenta is eventually bounded from above and from below.

We say that $G$ satisfies the continuity estimate if for every fundamental collection of variables we have

\begin{align}
&\liminf_{\epsilon \to 0} \liminf_{\alpha_1 \to \infty} \ldots \liminf_{\alpha_k \to \infty} G \left( x_{\alpha, \theta} - \sum_{i=1}^k a_i \nabla \Psi_i(x_{\alpha, \theta}) (x_{\alpha, \theta}), \theta_{\alpha, \theta} \right) \\
&- G \left( y_{\alpha, \theta} - \sum_{i=1}^k a_i \nabla \Psi_i(y_{\alpha, \theta}) (y_{\alpha, \theta}), \theta_{\alpha, \theta} \right) \leq 0. 
\end{align}

(B.3)

\section{Differential inclusions}

To establish that Condition 8.11 of [FK06] is satisfied in the proof of Theorem 3.7, we need to solve a differential inclusion. The following appendix is based on [Dei92, Kun00] and is a copy of the one in [KM18]. We state it for completeness.

Let $D \subseteq \mathbb{R}^d$ be a non-empty set. A multi-valued mapping $F : D \to 2^{\mathbb{R}^d \setminus \{\emptyset\}}$ is a map that assigns to every $x \in D$ a set $F(x) \subseteq \mathbb{R}^d$, $F(x) \neq \emptyset$. 39
**Definition C.1.** Let \( I \subseteq \mathbb{R} \) be an interval with \( 0 \in I \), \( D \subseteq \mathbb{R}^d \), \( x \in D \) and \( F : D \to 2^{\mathbb{R}^d} \setminus \emptyset \) a multi-valued mapping. A function \( \gamma \) such that

(a) \( \gamma : I \to D \) is absolutely continuous,

(b) \( \gamma(0) = x \),

(c) \( \dot{\gamma}(t) \in F(\gamma(t)) \) for almost every \( t \in I \)

is called a solution of the differential inclusion \( \dot{\gamma} \in F(\gamma) \) a.e., \( \gamma(0) = x \).

If we assume sufficient regularity on the multi-valued mapping \( F \), we can ensure the existence of a solution to differential inclusions that remain inside \( D \).

**Definition C.2.** Let \( D \subseteq \mathbb{R}^d \) be a non-empty set and let \( F : D \to 2^{\mathbb{R}^d} \setminus \emptyset \) be a multi-valued mapping.

(i) We say that \( F \) is closed, compact or convex valued if each set \( F(x) \), \( x \in D \) is closed, compact or convex, respectively.

(ii) We say that \( F \) is upper hemi-continuous at \( x \in D \) if for each neighbourhood \( U \) of \( F(x) \), there is a neighbourhood \( V \) of \( x \) in \( D \) such that \( F(V) \subseteq U \). We say that \( F \) is upper hemi-continuous if it is upper hemi-continuous at every point. \( F \) is upper hemi-continuous if and only if for each sequence \( x_n \rightarrow x \) in \( D \) and \( \xi_n \in F(x_n) \) such that \( \xi_n \rightarrow \xi \) we have \( \xi \in F(x) \).

**Definition C.3.** Let \( D \subseteq \mathbb{R}^d \) be a closed non-empty set. The tangent cone to \( D \) at \( x \) is

\[
T_D(x) := \left\{ z \in \mathbb{R}^d \mid \liminf_{\lambda \downarrow 0} \frac{d(y + \lambda z, D)}{\lambda} = 0 \right\}.
\]

The set \( T_D(x) \) is sometimes called the the Bouligand cotangent cone.

**Lemma C.4** (Theorem 2.2.1 in [Kun00], Lemma 5.1 in [Dei92]). Let \( D \subseteq \mathbb{R}^d \) be closed and let \( F : D \to 2^{\mathbb{R}^d} \setminus \emptyset \) satisfy

(a) \( F \) has closed convex values and is upper hemi-continuous;

(b) for every \( x \), we have \( F(x) \cap T_D(x) \neq \emptyset \);

(c) \( F \) has bounded growth: there is some \( c > 0 \) such that \( \|F(x)\| = \sup \{|z| \mid z \in F(x)\} \leq c(1 + |x|) \) for all \( x \in D \).

Then the differential inclusion \( \dot{\gamma} \in F(\gamma) \) has a solution on \( \mathbb{R}^+ \) for every starting point \( x \in D \).

**References**

[AB06] Charalambos D. Aliprantis and Kim C. Border. *Infinite dimensional analysis*. Springer, Berlin, third edition, 2006. A hitchhiker’s guide.

[BC97] Martino Bardi and Italo Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Boston, MA: Birkhäuser, 1997.

[BCFG18] Lorenzo Bertini, Raphael Chetrite, Alessandra Faggionato, and Davide Gabrielli. Level 2.5 large deviations for continuous-time markov chains with time periodic rates. *Annales Henri Poincaré*, 19(10):3197–3238, 10 2018.
[BDG18] Amarjit Budhiraja, Paul Dupuis, and Arnab Ganguly. Large deviations for small noise diffusions in a fast markovian environment. *Electron. J. Probab.*, 23:33 pp., 2018.

[BGTVE16] Freddy Bouchet, Tobias Grafke, Tomás Tangarife, and Eric Vanden-Eijnden. Large deviations in fast–slow systems. *Journal of Statistical Physics*, 162(4):793–812, Feb 2016.

[CIL92] Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User’s guide to viscosity solutions of second order partial differential equations. *Bull. Am. Math. Soc., New Ser.*, 27(1):1–67, 1992.

[CK17] Francesca Collet and Richard C. Kraaij. Dynamical moderate deviations for the Curie-Weiss model. *Stochastic Processes and their Applications*, 127(9):2900 – 2925, 2017.

[Dei92] Klaus Deimling. *Multivalued differential equations*, volume 1 of De Gruyter Series in Nonlinear Analysis and Applications. Walter de Gruyter & Co., Berlin, 1992.

[DH08] Frank Den Hollander. *Large deviations*, volume 14. American Mathematical Soc., 2008.

[DIS90] Paul Dupuis, Hitoshi Ishii, and H. Mete Soner. A viscosity solution approach to the asymptotic analysis of queueing systems. *The Annals of Probability*, 18(1):pp. 226–255, 1990.

[DV75a] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, i. *Communications on Pure and Applied Mathematics*, 28(1):1–47, 1975.

[DV75b] Monroe D. Donsker and S. R. S. Varadhan. On a variational formula for the principal eigenvalue for operators with maximum principle. *Proceedings of the National Academy of Sciences*, 72(3):780–783, 1975.

[FK06] Jin Feng and Thomas G. Kurtz. *Large Deviations for Stochastic Processes*. American Mathematical Society, 2006.

[KM18] Richard C. Kraaij and Louis Mahé. Well-posedness of hamilton-jacobi equations in population dynamics and applications to large deviations. *preprint; ArXiv:1803.04525*, 2018.

[KP17] Rohini Kumar and Lea Popovic. Large deviations for multi-scale jump-diffusion processes. *Stochastic Processes and their Applications*, 127(4):1297 – 1320, 2017.

[Kra17] Richard C. Kraaij. Flux large deviations of weakly interacting jump processes via well-posedness of an associated Hamilton-Jacobi equation. *preprint; ArXiv:1711.00274*, 2017.

[Kra19] Richard C. Kraaij. A general convergence result for viscosity solutions of Hamilton-Jacobi equations and non-linear semigroups. *preprint; ArXiv:1903.04196*, 2019.

[KS20] Richard C. Kraaij and Mikola Schlottke. Large deviations for two-scale mean-field interacting markov processes. *Work in progress*, 2020.

[Kun00] M. Kunze. *Non-Smooth Dynamical Systems*. Number nr. 1744 in Lecture Notes in Mathematics. Springer, 2000.
[P'85] Ross Pinsky et al. The $I$-Function for Diffusion Processes with Boundaries. *The Annals of Probability*, 13(3):676–692, 1985.

[Pin85] Ross G Pinsky. On evaluating the Donsker-Varadhan $I$-function. *The Annals of Probability*, pages 342–362, 1985.

[Pin07] Ross G. Pinsky. Regularity properties of the Donsker-Varadhan rate functional for non-reversible diffusions and random evolutions. *Stoch. Dyn.*, 7(2):123–140, 2007.