Delft University of Technology

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

Kraaij, Richard C.; Schlottke, Mikola C.

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

Check for
updates

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

Richard C. Kraaij and Mikola C. Schlottke

Abstract. We study the well-posedness of Hamilton–Jacobi–Bellman equations on subsets of $\mathbb{R}^d$ in a context without boundary conditions. The Hamiltonian is given as the supremum over two parts: an internal Hamiltonian depending on an external control variable and a cost functional penalizing the control. The key feature in this paper is that the control function can be unbounded and discontinuous. This way we can treat functionals that appear e.g. in the Donsker–Varadhan theory of large deviations for occupation-time measures. To allow for this flexibility, we assume that the internal Hamiltonian and cost functional have controlled growth, and that they satisfy an equi-continuity estimate uniformly over compact sets in the space of controls. In addition to establishing the comparison principle for the Hamilton–Jacobi–Bellman equation, we also prove existence, the viscosity solution being the value function with exponentially discounted running costs. As an application, we verify the conditions on the internal Hamiltonian and cost functional in two examples.

Mathematics Subject Classification. 49L25, 35F21.

Keywords. Hamilton–Jacobi–Bellman equations, Comparison principle, Viscosity solutions, Optimal control theory.

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 $\mathbb{R}^d$,

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

in the context without boundary conditions and where the Hamiltonian flow generated by $\mathcal{H}$ remains inside E. In (HJB), $\lambda > 0$ is a scalar and h is a

ⓑ Birkhäuser

continuous and bounded function. The Hamiltonian $\mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R}$ is given by

$$\mathcal{H}(x, p) = \sup_{\theta \in \Theta} \left[\Lambda(x, p, \theta) - \mathcal{I}(x, \theta) \right], \tag{1.1}$$

where $\theta \in \Theta$ plays the role of a control variable. For fixed θ, the function Λ can be interpreted as an Hamiltonian itself. We call it the *internal* Hamiltonian. The function $\mathcal{I}$ can be interpreted as the cost of applying the control θ.

The main result of this paper is the *comparison principle* for (HJB) in order to establish uniqueness of viscosity solutions. The standard assumption in the literature that allows one to obtain the comparison principle in the context of optimal control problems (e.g. [2] for the first order case and [10] for the second order case) is that either there is a modulus of continuity ω such that

$$|\mathcal{H}(x, p) - \mathcal{H}(y, p)| \le \omega \left(|x - y|(1 + |p|) \right), \tag{1.2}$$

or that $\mathcal{H}$ is uniformly coercive:

$$\lim_{|p| \to \infty} \inf_x \mathcal{H}(x, p) = \infty. \tag{1.3}$$

More generally, the two estimates (1.2) and (1.3) can be combined in a single estimate, called pseudo-coercivity, see [4, (H4), Page 34], that uses the fact that the sub- and supersolution properties roughly imply that the estimate (1.2) only needs to hold for appropriately chosen x, y and p such that $\mathcal{H}$ is finite uniformly over these chosen x, y, p.

In the Hamilton–Jacobi–Bellman context, the comparison principle is typically obtained by translating (1.2) into conditions for Λ and $\mathcal{I}$ of (1.1), which include (e.g. [2, Chapter III])

(I) $|\Lambda(x, p, \theta) - \Lambda(y, p, \theta)| \le \omega_\Lambda(|x - y|(1 + |p|))$, uniformly in θ, and
(II) $\mathcal{I}$ is bounded, continuous and $|\mathcal{I}(x, \theta) - \mathcal{I}(y, \theta)| \le \omega_\mathcal{I}(|x - y|)$ for all θ.

The pseudo-coercivity property is harder to translate as in this way the control on $\mathcal{H}$ does not necessarily imply the same control on Λ, in particular in the case when $\mathcal{I}$ is unbounded. We return on this issue below.

The estimates (I) and (II) are not satisfied for Hamiltonians arising from natural examples in the theory of large deviations [12,13] for Markov processes with two scales (see e.g. [6,18,27,29] for PDE's arising from large deviations with two scales, see [3,16,17,20,21] for other works connection PDE's with large deviations). Indeed, in [6] the authors mention that well-posedness of the Hamilton–Jacobi–Bellman equation for examples arising from large deviation theory is an open problem. Recent generalizations of the coercivity condition, see e.g. [9], also do not cover these examples.

In the large deviation context, however, we typically know that we have the comparison principle for the Hamilton–Jacobi equation in terms of Λ. In addition, even though $\mathcal{I}$ might be discontinuous, we do have other types of regularity for the functional $\mathcal{I}$, see e.g. [32]. Thus, we aim to prove a comparison

principle for (HJB) on the basis of the *assumption* that we have the following natural relaxations of (or the pseudo-coercive version of) (I) and (II).

(i) For $\theta \in \Theta$, define the Hamiltonian $\mathcal{H}_\theta(x, p) := \Lambda(x, p, \theta)$. We have an estimate on $\mathcal{H}_\theta$ that is uniform over θ in compact sets $K \subseteq \Theta$. This estimate, for one fixed θ, is in spirit similar to the pseudo-coercivity estimate of [4] and is morally equivalent to the comparison principle for $\mathcal{H}_\theta$. The uniformity is made rigorous as the *continuity estimate* in Assumption 2.14 ($\Lambda 5$) below.

(ii) The cost functional $\mathcal{I}(x, \theta)$ satisfies an *equi-continuity estimate* of the type $|\mathcal{I}(x, \theta) - \mathcal{I}(y, \theta)| \leq \omega_{\mathcal{I}, C}(|x - y|)$ on sublevel sets $\{\mathcal{I} \leq C\}$ which we assume to be compact. This estimate is made rigorous in Assumption 2.15 ($\mathcal{I}5$) below.

To work with these relaxations, we introduce a procedure that allows us to restrict our analysis to compact sets in the space of controls. In the proof of the comparison principle, the sub- and supersolution properties give boundedness of $\mathcal{H}$ when evaluated in optimizing points. We then translate this boundedness to boundedness of $\mathcal{I}$, which implies that the controls lie in a compact set.

The transfer of control builds upon (i) for $\Lambda(x, p, \theta_x^0)$ when we use a control θ_x^0 that satisfies $\mathcal{I}(x, \theta_x^0) = 0$. This we call the *bootstrap procedure*: we use the comparison principle for the Hamilton–Jacobi equation in terms of $\Lambda(x, p, \theta_x^0)$ to shift the control on $\mathcal{H}$ to control on Λ and $\mathcal{I}$ for general θ. That way the comparison principle for the internal Hamiltonian Λ bootstraps to the comparison principle for the full Hamiltonian $\mathcal{H}$.

Clearly, this bootstrap argument does not come for free. We pose four additional assumptions:

(iii) The function Λ grows roughly equally fast in p: For all compact sets $\widehat{K} \subseteq E$, there are constants M, C_1, C_2 such that

$$\Lambda(x, p, \theta_1) \leq \max\left\{M, C_1 \Lambda(x, p, \theta_2) + C_2\right\},$$

for all $x \in \widehat{K}$, $p \in \mathbb{R}^d$, $\theta_1, \theta_2 \in \Theta$.

(iv) The function $\mathcal{I}$ grows roughly equally fast in x: For all $x \in E$ and $M \geq 0$ there exists an open neighbourhood U of x and constants M', C_1', C_2' such that

$$\mathcal{I}(y_1, \theta) \leq \max\{M', C_1' \mathcal{I}(y_2, \theta) + C_2'\}$$

for all $y_1, y_2 \in U$ and for all θ such that $\mathcal{I}(x, \theta) \leq M$.

(v) $\mathcal{I} \geq 0$ and for each $x \in E$, there exists θ_x^0 such that $\mathcal{I}(x, \theta_x^0) = 0$.

(vi) The functional $\mathcal{I}$ is equi-coercive in x: for any compact set $\hat{K} \subseteq E$ the set $\bigcup_{x \in \hat{K}} \{\theta \,|\, \mathcal{I}(x, \theta) \leq C\}$ is compact.

These four assumptions are stated below as Assumptions 2.14 ($\Lambda 4$), 2.15 ($\mathcal{I}4$), 2.15 ($\mathcal{I}2$), and 2.15 ($\mathcal{I}3$). To explain in more detail our argument, we give a sketch of the bootstrap procedure, which can be skipped on first reading. In

this sketch, we refrain from performing localization arguments that are needed for non-compact E.

Sketch of the bootstrap argument

Let u and v be a sub- and supersolution to $f - \lambda H f = h$ respectively. We estimate $\sup_x u(x) - v(x)$ by the classical doubling of variables by means of penalizing the distance between x and y by some penalization $\alpha \Psi(x - y)$ and aim to send $\alpha \to \infty$. Let x_α, y_α denote the optimizers, and denote by p_α the corresponding momentum $p_\alpha = \alpha \partial_x \Psi(x_\alpha - y_\alpha)$. Let θ_α be the control such that $\mathcal{H}(x_\alpha, p_\alpha) = \Lambda(x_\alpha, p_\alpha, \theta_\alpha) - \mathcal{I}(x_\alpha, \theta_\alpha)$ and let θ_α^0 be a control such that $\mathcal{I}(y_\alpha, \theta_\alpha^0) = 0$, which exists due to (v).

The supersolution property for v yields the following estimate that is uniform in $\alpha > 0$

$$\infty > \frac{\|v - h\|}{\lambda} \geq \mathcal{H}(y_\alpha, p_\alpha) \geq \Lambda(y_\alpha, p_\alpha, \theta_\alpha^0) - \mathcal{I}(y_\alpha, \theta_\alpha^0) = \Lambda(y_\alpha, p_\alpha, \theta_\alpha^0). \quad (1.4)$$

Using (iii), we obtain a uniform estimate in α:

$$\sup_\alpha \Lambda(y_\alpha, p_\alpha, \theta_\alpha) < \infty. \quad (1.5)$$

which will allow us to use (i) if we can show that the controls θ_α take their value in a compact set $K \subseteq \Theta$. For this, it suffices by (vi) to establish

$$\sup_\alpha \mathcal{I}(x_\alpha, \theta_\alpha) < \infty. \quad (1.6)$$

This, in fact, implies by (iv) that

$$\sup \mathcal{I}(y_\alpha, \theta_\alpha) \vee \mathcal{I}(x_\alpha, \theta_\alpha) < \infty$$

so that we can also apply (ii). This, in combination with the application of (i) establishes the comparison principle for $f - \lambda H f = h$.

We are thus left to prove (1.6), which is where our bootstrap comes into play. The subsolution property for u yields the following estimate that is uniform in $\alpha > 0$

$$-\infty < \frac{\|u - h\|}{\lambda} \leq \mathcal{H}(x_\alpha, p_\alpha) = \Lambda(x_\alpha, p_\alpha, \theta_\alpha) - \mathcal{I}(x_\alpha, \theta_\alpha).$$

Thus, (1.6) follows if we can establish

$$\sup_\alpha \Lambda(x_\alpha, p_\alpha, \theta_\alpha) < \infty,$$

which in turn (by (iii)) follows from

$$\sup_\alpha \Lambda(x_\alpha, p_\alpha, \theta_\alpha^0) < \infty.$$

To establish this final estimate, note that

$$\Lambda(x_\alpha, p_\alpha, \theta_\alpha^0) = \Lambda(y_\alpha, p_\alpha, \theta_\alpha^0) + \left[\Lambda(x_\alpha, p_\alpha, \theta_\alpha^0) - \Lambda(y_\alpha, p_\alpha, \theta_\alpha^0) \right]$$

and that we have control on the first term by means of (1.4) and on the second term by the pseudo-coercivity estimate of (i) on Λ for the controls θ_α^0 which lie in a compact set due to (vi). □

Thus, to summarize, we use the growth conditions posed on Λ and $\mathcal{I}$ and the pseudo-coercivity estimate for Λ to transfer the control on the full Hamiltonian $\mathcal{H}$ to the functions Λ and the cost function $\mathcal{I}$. Then the control on Λ and $\mathcal{I}$ allows us to apply the estimates (i) and (ii) to obtain the comparison principle.

Next to our main result, we also state for completeness an existence result in Theorem 2.8. The viscosity solution will be given in terms of a discounted control problem as is typical in the literature, see e.g. [2, Chapter 3]. Minor difficulties arise from working with $\mathcal{H}$ that arise from irregular $\mathcal{I}$.

Finally, we show that the conditions (i) to (vi) are satisfied in two examples that arise from large deviation theory for two-scale processes. In our companion paper [26], we will use existence and uniqueness for (HJB) for these examples to obtain large deviation principles.

Illustration in the context of an example

As an illustrating example, we consider a Hamilton–Jacobi–Bellman equation that arises from the large deviations of the empirical measure-flux pair of weakly coupled Markov jump processes that are coupled to fast Brownian motion on the torus. We skip the probabilistic background of this problem (See [26]), and come to the set-up relevant for this paper.

Let $G := \{1, \ldots, q\}$ be some finite set, and let $\Gamma = \{(a, b) \in G^2 \,|\, a \neq b\}$ be the set of directed bonds. Let $E := \mathcal{P}(G) \times [0, \infty)^\Gamma$, where $\mathcal{P}(G)$ is the set of probability measures on G. Let $F = \mathcal{P}(S^1)$ be the set of probability measures on the one-dimensional torus. We introduce Λ and $\mathcal{I}$.

- Let $r : G \times G \times \mathcal{P}(E) \times \mathcal{P}(S^1) \to [0, \infty)$ be some function that codes the $\mathcal{P}(E) \times \mathcal{P}(S^1)$ dependent jump rate of the Markov jump process over each bond $(a, b) \in \Gamma$. The internal Hamiltonian Λ is given by

$$\Lambda(\mu, p, \theta) = \sum_{(a,b) \in \Gamma} \mu_a r(a, b, \mu, \theta) \left[e^{p_b - p_a + p_{a,b}} - 1 \right].$$

- Let $\sigma^2 : S^1 \times \mathcal{P}(G) \to (0, \infty)$ be a bounded and strictly positive function. The cost function $\mathcal{I} : E \times \Theta \to [0, \infty]$ is given by

$$\mathcal{I}(\mu, w, \theta) = \mathcal{I}(\mu, \theta) = \sup_{\substack{u \in C^\infty(S^1) \\ u > 0}} \int_{S^1} \sigma^2(y, \mu) \left(-\frac{u''(y)}{u(y)} \right) \theta(\mathrm{d}y).$$

Aiming for the comparison principle, we note that classical methods do not apply. The functionals Λ are not coercive and do not satisfy (I). We show in "Appendix E" that they are also not pseudo-coercive as defined in [4]. The functional $\mathcal{I}$ is neither continuous nor bounded. Once can check e.g. that if θ is a finite combination of Dirac measures, then $\mathcal{I}(\mu, \theta) = \infty$.

We show in Sect. 5, however, that (i) to (vi) hold, implying the comparison principle for the Hamilton–Jacobi–Bellman equations. The verification of these properties is based in part on results from [23,32].

Summary and overview of the paper

To summarize, our novel bootstrap procedure allows to treat Hamilton–Jacobi–Bellman equations where:

- We assume that the cost function $\mathcal{I}$ satisfies some regularity conditions on its sub-levelsets, but allow $\mathcal{I}$ to be possibly unbounded and discontinuous.
- We assume that Λ satisfies the continuity estimate uniformly for controls in compact sets, which in spirit extends the pseudo-coercivity estimate of [4]. This implies that Λ can be possibly non-coercive, non-pseudo-coercive and non-Lipschitz as exhibited in our example above.

In particular, allowing discontinuity in $\mathcal{I}$ allows us to treat the comparison principle for examples like the one we considered above, which so far has been out of reach. 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. Of interest would be, for example, an extension of the results of [10] who work with continuous $\mathcal{I}$. For clarity of the exposition, and the already numerous applications for this setting, we stick to the finite-dimensional first-order case. We think that the key arguments that are used in the proof in Sect. 3 do not depend in a crucial way on this assumption.

The paper is organized as follows. The main results are formulated in Sect. 2. In Sect. 3 we establish the comparison principle. In Sect. 4 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 Sect. 5 we treat two examples including the one mentioned in the introduction.

2. Main results

In this section, we start with preliminaries in Sect. 2.1, which includes the definition of viscosity solutions and that of the comparison principle.

We proceed in Sect. 2.2 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 Sect. 2.3 we collect all assumptions that are needed for the main results.

2.1. 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 $C_c^\infty(\mathcal{X})$ the space of smooth functions that vanish outside a compact set. We denote by $C_{cc}^\infty(\mathcal{X})$ the set of smooth functions that are constant outside of a compact set in $\mathcal{X}$, and by $\mathcal{P}(\mathcal{X})$ the space of probability measures on $\mathcal{X}$. We equip $\mathcal{P}(\mathcal{X})$ with the weak topology 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 necessarily assume that E is open. We assume that the space of controls Θ is Polish.

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

Definition 2.1. (*Viscosity solutions and comparison principle*) Let $A : \mathcal{D}(A) \subseteq C_b(E) \to C_b(E)$ be an operator with domain $\mathcal{D}(A)$, $\lambda > 0$ and $h \in C_b(E)$. Consider the Hamilton–Jacobi equation

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

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

$$\lim_{n \uparrow \infty} u(x_n) - f(x_n) = \sup_x u(x) - f(x),$$
$$\lim_{n \uparrow \infty} u(x_n) - \lambda A f(x_n) - h(x_n) \leq 0.$$

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

$$\lim_{n \uparrow \infty} v(x_n) - f(x_n) = \inf_x v(x) - f(x),$$
$$\lim_{n \uparrow \infty} v(x_n) - \lambda A f(x_n) - h(x_n) \geq 0.$$

We say that u is a *(viscosity) solution* of Eq. (2.1) if it is both a subsolution and a supersolution to (2.1). We say that (2.1) satisfies the *comparison principle* if for every subsolution u and supersolution v to (2.1), we have $u \leq v$.

Remark 2.2. (Uniqueness) If u and v are two viscosity solutions of 2.3, then we have $u \leq v$ and $v \leq u$ by the comparison principle, giving uniqueness.

Remark 2.3. Consider the definition of subsolutions. Suppose that the test-function $f \in \mathcal{D}(A)$ has compact sublevel sets, then instead of working with a sequence x_n, there exists $x_0 \in E$ such that

$$u(x_0) - f(x_0) = \sup_x u(x) - f(x),$$
$$u(x_0) - \lambda A f(x_0) - h(x_0) \leq 0.$$

A similar simplification holds in the case of supersolutions.

Remark 2.4. For an explanatory text on the notion of viscosity solutions and fields of applications, we refer to [8].

Remark 2.5. At present, we refrain from working with unbounded viscosity solutions as we use the upper bound on subsolutions and the lower bound on supersolutions in the proof of Theorem 2.6. We can, however, imagine that the methods presented in this paper can be generalized if u and v grow slower than the containment function Υ that will be defined below in Definition 2.13.

2.2. Main results: comparison and existence

In this section, we state our main results: the comparison principle in Theorem 2.6, and existence of solutions in Theorem 2.8.

Consider the variational Hamiltonian $\mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R}$ given by

$$\mathcal{H}(x, p) = \sup_{\theta \in \Theta} \left[\Lambda(x, p, \theta) - \mathcal{I}(x, \theta) \right]. \tag{2.2}$$

The precise assumptions on the maps Λ and $\mathcal{I}$ are formulated in Sect. 2.3.

Theorem 2.6. (Comparison principle) *Consider the map $\mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R}$ as in (2.2). Suppose that Assumptions 2.14 and 2.15 are satisfied for Λ and I. Define the operator $\mathbf{H}f(x) := \mathcal{H}(x, \nabla f(x))$ with domain $\mathcal{D}(\mathbf{H}) = C^\infty_{cc}(E)$. Then:*

  (a) For any $f \in \mathcal{D}(\mathbf{H})$ the map $x \mapsto \mathbf{H}f(x)$ is continuous.
  (b) For any $h \in C_b(E)$ and $\lambda > 0$, the comparison principle holds for

$$f - \lambda \mathbf{H}f = h. \tag{2.3}$$

Remark 2.7. (Domain) The comparison principle holds with any domain that satisfies $C^\infty_{cc}(E) \subseteq \mathcal{D}(\mathbf{H}) \subseteq C^1_b(E)$. We state it with $C^\infty_{cc}(E)$ to connect it with the existence result of Theorem 2.8, where we need to work with test functions whose gradients have compact support.

Consider the Legendre dual $\mathcal{L} : E \times \mathbb{R}^d \to [0, \infty]$ of the Hamiltonian,

$$\mathcal{L}(x, v) := \sup_{p \in \mathbb{R}^d} \left[\langle p, v \rangle - \mathcal{H}(x, p) \right],$$

and denote the collection of absolutely continuous paths in E by $\mathcal{AC}$.

Theorem 2.8. (Existence of viscosity solution) *Consider $\mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R}$ as in (2.2). Suppose that Assumptions 2.14 and 2.15 are satisfied for Λ and $\mathcal{I}$, and that $\mathcal{H}$ satisfies Assumption 2.17. For each $\lambda > 0$, let $R(\lambda)$ be the operator*

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

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

Remark 2.9. The form of the solution is typical, see for example Section III.2 in [2]. It is the value function obtained by an optimization problem with exponentially discounted cost. The difficulty of the proof of Theorem 2.8 lies in treating the irregular form of $\mathcal{H}$.

2.3. Assumptions

In this section, we formulate and comment on the assumptions imposed on the Hamiltonians defined in the previous sections. The key assumptions were already mentioned in the sketch of the bootstrap method in the introduction. To these, we add minor additional assumptions on the regularity of Λ and $\mathcal{I}$ in Assumptions 2.14 and 2.15. Finally, Assumption 2.17 will imply that even if

E has a boundary, no boundary conditions are necessary for the construction of the viscosity solution.

We start with the *continuity estimate* for Λ, which was briefly discussed in (i) in the introduction. To that end, we first introduce a function that is used in the typical argument that doubles the number of variables.

Definition 2.10. (*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 $\Psi(x, y) = 0$.

We will apply the definition below for $\mathcal{G} = \Lambda$.

Definition 2.11. (*Continuity estimate*) Let Ψ be a penalization function and let $\mathcal{G} : E \times \mathbb{R}^d \times \Theta \to \mathbb{R}$, $(x, p, \theta) \mapsto \mathcal{G}(x, p, \theta)$ be a function. Suppose that for each $\varepsilon > 0$, there is a sequence of positive real numbers $\alpha \to \infty$. For sake of readability, we suppress the dependence on ε in our notation.

Suppose that for each ε and α we have variables $(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha})$ in E^2 and variables $\theta_{\varepsilon,\alpha}$ in Θ. We say that this collection is *fundamental* for $\mathcal{G}$ with respect to Ψ if:

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

(C2) For each $\varepsilon > 0$, we have $\lim_{\alpha \to \infty} \alpha\Psi(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha}) = 0$. For any limit point $(x_\varepsilon, y_\varepsilon)$ of $(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha})$, we have $\Psi(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha}) = 0$.

(C3) We have for all $\varepsilon > 0$

$$\sup_\alpha \mathcal{G}\left(y_{\varepsilon,\alpha}, -\alpha(\nabla\Psi(x_{\varepsilon,\alpha}, \cdot))(y_{\varepsilon,\alpha}), \theta_{\varepsilon,\alpha}\right) < \infty, \tag{2.4}$$

$$\inf_\alpha \mathcal{G}\left(x_{\varepsilon,\alpha}, \alpha(\nabla\Psi(\cdot, y_{\varepsilon,\alpha}))(x_{\varepsilon,\alpha}), \theta_{\varepsilon,\alpha}\right) > -\infty. \tag{2.5}$$

We say that $\mathcal{G}$ satisfies the *continuity estimate* if for every fundamental collection of variables we have for each $\varepsilon > 0$ that

$$\liminf_{\alpha \to \infty} \mathcal{G}\left(x_{\varepsilon,\alpha}, \alpha(\nabla\Psi(\cdot, y_{\varepsilon,\alpha}))(x_{\varepsilon,\alpha}), \theta_{\varepsilon,\alpha}\right)$$
$$-\mathcal{G}\left(y_{\varepsilon,\alpha}, -\alpha(\nabla\Psi(x_{\varepsilon,\alpha}, \cdot))(y_{\varepsilon,\alpha}), \theta_{\varepsilon,\alpha}\right) \leq 0. \tag{2.6}$$

Remark 2.12. In "Appendix C", we state a slightly more general continuity estimate on the basis of two penalization functions. A proof of a comparison principle on the basis of two penalization functions was given in [23].

The continuity estimate is indeed exactly the estimate that one would perform when proving the comparison principle for the Hamilton–Jacobi equation in terms of the internal Hamiltonian (disregarding the control θ). Typically, the control on $(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha})$ that is assumed in (C1) and (C2) is obtained from choosing $(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha})$ as optimizers in the doubling of variables procedure (see Lemma 3.5), and the control that is assumed in (C3) is obtained by using the viscosity sub- and supersolution properties in the proof of the comparison principle. The required restriction to compact sets in Lemma 3.5 is obtained by including in the test functions a *containment function*.

Definition 2.13. (*Containment function*) We say that a function $\Upsilon : E \to [0, \infty]$ is a *containment function* for Λ if $\Upsilon \in C^1(E)$ and there is a constant c_Υ such that

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

To conclude, our assumption on Λ contains the continuity estimate, the controlled growth, the existence of a containment function and two regularity properties.

Assumption 2.14. The function $\Lambda : E \times \mathbb{R}^d \times \Theta \to \mathbb{R}$ in the Hamiltonian (2.2) satisfies the following.

(Λ1) The map $\Lambda : E \times \mathbb{R}^d \times \Theta \to \mathbb{R}$ is continuous.
(Λ2) For any $x \in E$ and $\theta \in \Theta$, the map $p \mapsto \Lambda(x, p, \theta)$ is convex. We have $\Lambda(x, 0, \theta) = 0$ for all $x \in E$ and all $\theta \in \Theta$.
(Λ3) There exists a containment function $\Upsilon : E \to [0, \infty)$ for Λ in the sense of Definition 2.13.
(Λ4) 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

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

(Λ5) The function Λ satisfies the continuity estimate in the sense of Definition 2.11, or in the extended sense of Definition C.2.

Our second main assumption is on the properties of $\mathcal{I}$. For a compact set $K \subseteq E$ and a constant $M \geq 0$, write

$$\Theta_{K,M} := \bigcup_{x \in K} \{\theta \in \Theta \mid \mathcal{I}(x, \theta) \leq M\}, \tag{2.7}$$

and

$$\Omega_{K,M} := \bigcap_{x \in K} \{\theta \in \Theta \mid \mathcal{I}(x, \theta) \leq M\}. \tag{2.8}$$

Assumption 2.15. The functional $\mathcal{I} : E \times \Theta \to [0, \infty]$ in (2.2) satisfies the following.

($\mathcal{I}$1) The map $(x, \theta) \mapsto \mathcal{I}(x, \theta)$ is lower semi-continuous on $E \times \Theta$.
($\mathcal{I}$2) For any $x \in E$, there exists a control $\theta_x^0 \in \Theta$ such that $\mathcal{I}(x, \theta_x^0) = 0$.
($\mathcal{I}$3) For any compact set $K \subseteq E$ and constant $M \geq 0$ the set $\Theta_{K,M}$ is compact.
($\mathcal{I}$4) For each $x \in E$ and constant $M \geq 0$, there exists an open neighbourhood $U \subseteq E$ of x and constants $M', C_1', C_2' \geq 0$ such that for all $y_1, y_2 \in U$ and $\theta \in \Theta_{\{x\}, M}$ we have

$$\mathcal{I}(y_1, \theta) \leq \max\{M', C_1'\mathcal{I}(y_2, \theta) + C_2'\}.$$

($\mathcal{I}$5) For every compact set $K \subseteq E$ and each $M \geq 0$ the collection of functions $\{\mathcal{I}(\cdot, \theta)\}_{\theta \in \Omega_{K,M}}$ is equicontinuous. That is: for all $\varepsilon > 0$, there is a $\delta > 0$ such that for all $\theta \in \Omega_{K,M}$ and $x, y \in K$ such that $d(x, y) \leq \delta$ we have $|\mathcal{I}(x, \theta) - \mathcal{I}(y, \theta)| \leq \varepsilon$.

To establish the existence of viscosity solutions, we will impose one additional assumption. For a general convex functional $p \mapsto \Phi(p)$ we denote

$$\partial_p \Phi(p_0) := \left\{ \xi \in \mathbb{R}^d \ : \ \Phi(p) \geq \Phi(p_0) + \xi \cdot (p - p_0) \quad (\forall p \in \mathbb{R}^d) \right\}. \quad (2.9)$$

Definition 2.16. The tangent cone (sometimes also called *Bouligand cotingent cone*) to E in $\mathbb{R}^d$ at x is

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

Assumption 2.17. The set E is closed and convex. The map Λ is such that $\partial_p \Lambda(x, p, \theta) \subseteq T_E(x)$ for all $x \in E$, $p \in \mathbb{R}^d$ and $\theta \in \Theta$.

In Lemma 4.1 we will show that the assumption implies that $\partial_p \mathcal{H}(x, p) \subseteq T_E(x)$, which in turn implies that the solutions of the differential inclusion in terms of $\partial_p \mathcal{H}(x, p)$ remain inside E. Motivated by our examples, we work with closed convex domains E. While in this context we can apply results from e.g. Deimling [11], we believe that similar results can be obtained in different contexts.

Remark 2.18. The statement that $\partial_p \mathcal{H}(x, p) \subseteq T_E(x)$ is intuitively implied by the comparison principle for **H** and therefore, we expect it to hold in any setting for which Theorem 2.6 holds. Here, 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 $E = [0, 1]$, $f, g \in C_b^1(E)$ and suppose that $f(0) - g(0) = \sup_x f(x) - g(x)$. As $x = 0$ is a boundary point, we conclude that $f'(0) \leq g'(0)$. If indeed the maximum principle holds, we must have

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

implying that $p \mapsto \mathcal{H}(0, p)$ is increasing, in other words
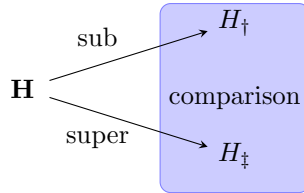
$$\partial_p \mathcal{H}(x, p)) \subseteq [0, \infty) = T_{[0,1]}(0).$$

3. The comparison principle

In this section, we establish Theorem 2.6. To establish the comparison principle for $f - \lambda \mathbf{H} f = h$ we use the bootstrap method explained in the introduction. We start by a classical localization argument.

We carry out the localization argument by absorbing the containment function Υ from Assumption 2.14 (Λ3) into the test functions. This leads to two new operators, $H_\dagger$ and $H_\ddagger$ that serve as an upper bound and a lower bound for the true **H**. We will then show the comparison principle for the Hamilton–Jacobi equation in terms of these two new operators. We therefore have to extend our notion of Hamilton–Jacobi equations and the comparison principle. This extension of the definition is standard, but we included it for completeness in the appendix as Definition A.1.

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 $f - \lambda Af = h$ are also viscosity subsolutions of $f - \lambda Bf = h$. Similarly for arrows with a subscript 'super'.

We introduce the operators $H_\dagger$ and $H_\ddagger$ in Sect. 3.1. The arrows will be established in Sect. 3.2. Finally, we will establish the comparison principle for $H_\dagger$ and $H_\ddagger$ in Sect. 3.3. Combined these two results imply the comparison principle for **H**.

Proof of Theorem 2.6. We start with the proof of (a). Let $f \in \mathcal{D}(\mathbf{H})$. Then $\mathbf{H}f$ is continuous since by Proposition B.3 in "Appendix B", the Hamiltonian $\mathcal{H}$ is continuous.

We proceed with the proof of (b). Fix $h_1, h_2 \in C_b(E)$ and $\lambda > 0$.

Let u_1, u_2 be a viscosity sub- and supersolution to $f - \lambda \mathbf{H}f = h_1$ and $f - \lambda \mathbf{H}f = h_2$ respectively. By Lemma 3.3 proven in Sect. 3.2, 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 3.4 of Sect. 3.3. Specialising to $h_1 = h_2$ gives Theorem 2.6.   □

3.1. Definition of auxiliary operators

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

Definition 3.1. The operator $\mathbf{H} \subseteq C_b^1(E) \times C_b(E)$ has domain $\mathcal{D}(\mathbf{H}) = C_{cc}^\infty(E)$ and satisfies $\mathbf{H}f(x) = \mathcal{H}(x, \nabla f(x))$, where $\mathcal{H}$ is the map

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

We proceed by introducing $H_\dagger$ and $H_\ddagger$. Recall Assumption (Λ3) and the constant $C_\Upsilon := \sup_\theta \sup_x \Lambda(x, \nabla \Upsilon(x), \theta)$ therein. Denote by $C_\ell^\infty(E)$ the set of smooth functions on E that have a lower bound and by $C_u^\infty(E)$ the set of smooth functions on E that have an upper bound.

Definition 3.2. (*The operators $H_\dagger$ and $H_\ddagger$*) For $f \in C_\ell^\infty(E)$ and $\varepsilon \in (0,1)$ set

$$f_\dagger^\varepsilon := (1 - \varepsilon)f + \varepsilon \Upsilon$$
$$H_{\dagger, f}^\varepsilon(x) := (1 - \varepsilon)\mathcal{H}(x, \nabla f(x)) + \varepsilon C_\Upsilon.$$

and set

$$H_\dagger := \left\{ (f_\dagger^\varepsilon, H_{\dagger,f}^\varepsilon) \,\middle|\, f \in C_\ell^\infty(E), \varepsilon \in (0,1) \right\}.$$

For $f \in C_u^\infty(E)$ and $\varepsilon \in (0,1)$ set

$$f_\ddagger^\varepsilon := (1+\varepsilon)f - \varepsilon \Upsilon$$
$$H_{\ddagger,f}^\varepsilon(x) := (1+\varepsilon)\mathcal{H}(x, \nabla f(x)) - \varepsilon C_\Upsilon.$$

and set

$$H_\ddagger := \left\{ (f_\ddagger^\varepsilon, H_{\ddagger,f}^\varepsilon) \,\middle|\, f \in C_u^\infty(E), \varepsilon \in (0,1) \right\}.$$

3.2. Preliminary results

The operator $\mathbf{H}$ is related to $H_\dagger, H_\ddagger$ by the following Lemma.

Lemma 3.3. *Fix $\lambda > 0$ and $h \in C_b(E)$.*

(a) *Every subsolution to $f - \lambda\mathbf{H}f = h$ is also a subsolution to $f - \lambda H_\dagger f = h$.*
(b) *Every supersolution to $f - \lambda\mathbf{H}f = h$ is also a supersolution to $f - \lambda H_\ddagger f = h$.*

We only prove (a) of Lemma 3.3, as (b) can be carried out analogously.

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

Fix $\varepsilon > 0$ and $f \in C_\ell^\infty(E)$ and let $(f_\dagger^\varepsilon, H_{\dagger,f}^\varepsilon) \in H_\dagger$ as in Definition 3.2. We will prove that there are $x_n \in E$ such that

$$\lim_{n \to \infty} \left(u - f_\dagger^\varepsilon \right)(x_n) = \sup_{x \in E} \left(u(x) - f_\dagger^\varepsilon(x) \right), \tag{3.1}$$

$$\limsup_{n \to \infty} \left[u(x_n) - \lambda H_{\dagger,f}^\varepsilon(x_n) - h(x_n) \right] \leq 0. \tag{3.2}$$

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

$$K := \left\{ x \,\middle|\, \Upsilon(x) \leq \varepsilon^{-1} \sup_x \left(u(x) - (1-\varepsilon)f(x) \right) \right\}.$$

Set $M = \varepsilon^{-1} \sup_x \left(u(x) - (1-\varepsilon)f(x) \right)$. 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_ε the function on E defined by

$$f_\varepsilon(x) := \gamma\left((1-\varepsilon)f(x) + \varepsilon\Upsilon(x) \right).$$

By construction f_ε is smooth and constant outside of a compact set and thus lies in $\mathcal{D}(H) = C_{cc}^\infty(E)$. 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

$$\lim_n (u - f_\varepsilon)(x_n) = \sup_x (u(x) - f_\varepsilon(x)), \qquad (3.3)$$

$$\limsup_n [u(x_n) - \lambda \mathbf{H} f_\varepsilon(x_n) - h(x_n)] \leq 0. \qquad (3.4)$$

As f_ε equals $f_\dagger^\varepsilon$ on K, we have from (3.3) that also

$$\lim_n (u - f_\dagger^\varepsilon)(x_n) = \sup_{x \in E} (u(x) - f_\dagger^\varepsilon(x)),$$

establishing (3.1). Convexity of $p \mapsto \mathcal{H}(x, p)$ yields for arbitrary points $x \in K$ the estimate

$$\begin{aligned}
\mathbf{H} f_\varepsilon(x) &= \mathcal{H}(x, \nabla f_\varepsilon(x)) \\
&\leq (1-\varepsilon)\mathcal{H}(x, \nabla f(x)) + \varepsilon \mathcal{H}(x, \nabla \Upsilon(x)) \\
&\leq (1-\varepsilon)\mathcal{H}(x, \nabla f(x)) + \varepsilon C_\Upsilon = H_{\dagger, f}^\varepsilon(x).
\end{aligned}$$

Combining this inequality with (3.4) yields

$$\begin{aligned}
\limsup_n &[u(x_n) - \lambda H_{\dagger, f}^\varepsilon(x_n) - h(x_n)] \\
&\leq \limsup_n [u(x_n) - \lambda \mathbf{H} f_\varepsilon(x_n) - h(x_n)] \leq 0,
\end{aligned}$$

establishing (3.2). This concludes the proof. $\qquad \square$

3.3. The comparison principle

In this section, we prove the comparison principle for the operators $H_\dagger$ and $H_\ddagger$.

Proposition 3.4. *Fix $\lambda > 0$ and $h_1, h_2 \in C_b(E)$. Let u_1 be a viscosity subsolution to $f - \lambda H_\dagger f = h_1$ and let u_2 be a viscosity supersolution to $f - \lambda H_\ddagger f = h_2$. Then we have $\sup_x u_1(x) - u_2(x) \leq \sup_x h_1(x) - h_2(x)$.*

The proof uses a variant of a classical estimate that was proven e.g. in [8, Proposition 3.7] or in the present form in Proposition A.11 of [7].

Lemma 3.5. *Let u be bounded and upper semi-continuous, let v be bounded and lower semi-continuous, let $\Psi : E^2 \to \mathbb{R}^+$ be penalization functions and let Υ be a containment function.*

Fix $\varepsilon > 0$. For every $\alpha > 0$ there exist $x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon} \in E$ such that

$$\begin{aligned}
\frac{u(x_{\alpha,\varepsilon})}{1-\varepsilon} &- \frac{v(y_{\alpha,\varepsilon})}{1+\varepsilon} - \alpha \Psi(x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon}) - \frac{\varepsilon}{1-\varepsilon}\Upsilon(x_{\alpha,\varepsilon}) - \frac{\varepsilon}{1+\varepsilon}\Upsilon(y_{\alpha,\varepsilon}) \\
&= \sup_{x,y \in E} \left\{ \frac{u(x)}{1-\varepsilon} - \frac{v(y)}{1+\varepsilon} - \alpha \Psi(x,y) - \frac{\varepsilon}{1-\varepsilon}\Upsilon(x) - \frac{\varepsilon}{1+\varepsilon}\Upsilon(y) \right\}.
\end{aligned}$$
$$(3.5)$$

Additionally, for every $\varepsilon > 0$ we have that

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

(b) All limit points of $\{(x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon})\}_{\alpha>0}$ as $\alpha \to \infty$ are of the form (z,z) and for these limit points we have $u(z) - v(z) = \sup_{x \in E}\{u(x) - v(x)\}$.

(c) We have

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

Proof of Proposition 3.4. Fix $\lambda > 0$ and $h_1, h_2 \in C_b(E)$. Let u_1 be a viscosity subsolution and u_2 be a viscosity supersolution of $f - \lambda H_\dagger f = h_1$ and $f - \lambda H_\ddagger f = h_2$ respectively. We prove Theorem 3.4 in five steps of which the first two are classical.

We sketch the steps, before giving full proofs.

Step 1: We prove that for $\varepsilon > 0$ and $\alpha > 0$, there exist points $x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha} \in E$ satisfying the properties listed in Lemma 3.5 and momenta $p^1_{\varepsilon,\alpha}, p^2_{\varepsilon,\alpha} \in \mathbb{R}^d$ such that

$$p^1_{\varepsilon,\alpha} = \alpha \nabla \Psi(\cdot, y_{\varepsilon,\alpha})(x_{\varepsilon,\alpha}), \qquad p^2_{\varepsilon,\alpha} = -\alpha \nabla \Psi(x_{\varepsilon,\alpha}, \cdot)(y_{\varepsilon,\alpha}),$$

and

$$\sup_E (u_1 - u_2) \le \lambda \liminf_{\varepsilon \to 0} \liminf_{\alpha \to \infty} \left[\mathcal{H}(x_{\varepsilon,\alpha}, p^1_{\varepsilon,\alpha}) - \mathcal{H}(y_{\varepsilon,\alpha}, p^2_{\varepsilon,\alpha}) \right]$$
$$+ \sup_E (h_1 - h_2). \tag{3.6}$$

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 Υ, and convexity of $p \mapsto \mathcal{H}(x,p)$. We conclude it suffices to establish for each $\varepsilon > 0$ that

$$\liminf_{\alpha \to \infty} \mathcal{H}(x_{\varepsilon,\alpha}, p^1_{\varepsilon,\alpha}) - \mathcal{H}(y_{\varepsilon,\alpha}, p^2_{\varepsilon,\alpha}) \le 0. \tag{3.7}$$

Step 2: We will show that there are controls $\theta_{\varepsilon,\alpha}$ such that

$$\mathcal{H}(x_{\varepsilon,\alpha}, p^1_{\varepsilon,\alpha}) = \Lambda(x_{\varepsilon,\alpha}, p^1_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) - \mathcal{I}(x_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}). \tag{3.8}$$

As a consequence we have

$$\mathcal{H}(x_{\varepsilon,\alpha}, p^1_{\varepsilon,\alpha}) - \mathcal{H}(y_{\varepsilon,\alpha}, p^2_{\varepsilon,\alpha}) \le \Lambda(x_{\varepsilon,\alpha}, p^1_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) - \Lambda(y_{\varepsilon,\alpha}, p^2_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha})$$
$$+ \mathcal{I}(y_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) - \mathcal{I}(x_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}). \tag{3.9}$$

For establishing (3.7), it is sufficient to bound the differences in (3.9) by using Assumptions 2.14 (Λ5) and 2.15 ($\mathcal{I}$5).

Step 3: We verify the conditions to apply the continuity estimate, Assumption 2.14 (Λ5).

The bootstrap argument allows us to find for each ε a subsequence $\alpha = \alpha(\varepsilon) \to \infty$ such that the variables $(x_{\varepsilon,\alpha}, x_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha})$ are fundamental for Λ with respect to Ψ (See Definition 2.11).

Step 4: We verify the conditions to apply the estimate on $\mathcal{I}$, Assumption 2.15 ($\mathcal{I}$5).

Step 5 : Using the outcomes of Steps 3 and 4, we can apply the continuity estimate of Assumption 2.14 (Λ4) and the equi-continuity of Assumption 2.15 ($\mathcal{I}$5) to estimate (3.9) for any ε:

$$
\liminf_{\alpha \to \infty} \mathcal{H}(x_{\varepsilon,\alpha}, p^1_{\varepsilon,\alpha}) - \mathcal{H}(y_{\varepsilon,\alpha}, p^2_{\varepsilon,\alpha})
$$
$$
\leq \liminf_{\alpha \to \infty} \Lambda(x_{\varepsilon,\alpha}, p^1_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) - \Lambda(y_{\varepsilon,\alpha}, p^2_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha})
$$
$$
+ \mathcal{I}(y_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) - \mathcal{I}(x_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) \leq 0, \tag{3.10}
$$

which establishes (3.7) and thus also the comparison principle.

We proceed with the proofs of the first four steps, as the fifth step is immediate.

Proof of Step 1: The proof of this first step is classical. We include it for completeness. For any $\varepsilon > 0$ and any $\alpha > 0$, define the map $\Phi_{\varepsilon,\alpha} : E \times E \to \mathbb{R}$ by

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

Let $\varepsilon > 0$. By Lemma 3.5, there is a compact set $K_\varepsilon \subseteq E$ and there exist points $x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha} \in K_\varepsilon$ such that

$$
\Phi_{\varepsilon,\alpha}(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha}) = \sup_{x,y \in E} \Phi_{\varepsilon,\alpha}(x, y), \tag{3.11}
$$

and

$$
\lim_{\alpha \to \infty} \alpha \Psi(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha}) = 0. \tag{3.12}
$$

As in the proof of Proposition A.11 of [23], it follows that

$$
\sup_E (u_1 - u_2) \leq \liminf_{\varepsilon \to 0} \liminf_{\alpha \to \infty} \left[\frac{u_1(x_{\varepsilon,\alpha})}{1 - \varepsilon} - \frac{u_2(y_{\varepsilon,\alpha})}{1 + \varepsilon} \right]. \tag{3.13}
$$

At this point, we want to use the sub- and supersolution properties of u_1 and u_2. Define the test functions $\varphi^{\varepsilon,\alpha}_1 \in \mathcal{D}(H_\dagger), \varphi^{\varepsilon,\alpha}_2 \in \mathcal{D}(H_\ddagger)$ by

$$
\varphi^{\varepsilon,\alpha}_1(x) := (1 - \varepsilon) \left[\frac{u_2(y_{\varepsilon,\alpha})}{1 + \varepsilon} + \alpha \Psi(x, y_{\varepsilon,\alpha}) + \frac{\varepsilon}{1 - \varepsilon} \Upsilon(x) + \frac{\varepsilon}{1 + \varepsilon} \Upsilon(y_{\varepsilon,\alpha}) \right]
$$
$$
+ (1 - \varepsilon)(x - x_{\varepsilon,\alpha})^2,
$$
$$
\varphi^{\varepsilon,\alpha}_2(y) := (1 + \varepsilon) \left[\frac{u_1(x_{\varepsilon,\alpha})}{1 - \varepsilon} - \alpha \Psi(x_{\varepsilon,\alpha}, y) - \frac{\varepsilon}{1 - \varepsilon} \Upsilon(x_{\varepsilon,\alpha}) - \frac{\varepsilon}{1 + \varepsilon} \Upsilon(y) \right]
$$
$$
- (1 + \varepsilon)(y - y_{\varepsilon,\alpha})^2.
$$

Using (3.11), we find that $u_1 - \varphi^{\varepsilon,\alpha}_1$ attains its supremum at $x = x_{\varepsilon,\alpha}$, and thus

$$
\sup_E (u_1 - \varphi^{\varepsilon,\alpha}_1) = (u_1 - \varphi^{\varepsilon,\alpha}_1)(x_{\varepsilon,\alpha}).
$$

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

$$u_1(x_{\varepsilon,\alpha}) - \lambda \left[(1 - \varepsilon) \mathcal{H}\left(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1\right) + \varepsilon C_\Upsilon \right] \leq h_1(x_{\varepsilon,\alpha}). \qquad (3.14)$$

With a similar argument for u_2 and $\varphi_2^{\varepsilon,\alpha}$, we obtain by the supersolution inequality that

$$u_2(y_{\varepsilon,\alpha}) - \lambda \left[(1 + \varepsilon) \mathcal{H}\left(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2\right) - \varepsilon C_\Upsilon \right] \geq h_2(y_{\varepsilon,\alpha}), \qquad (3.15)$$

where $p_{\varepsilon,\alpha}^2 := -\alpha \nabla_y \Psi(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha})$. With that, estimating further in (3.13) leads to

$$\sup_E (u_1 - u_2) \leq \liminf_{\varepsilon \to 0} \liminf_{\alpha \to \infty} \left[\frac{h_1(x_{\varepsilon,\alpha})}{1 - \varepsilon} - \frac{h_2(y_{\varepsilon,\alpha})}{1 + \varepsilon} + \frac{\varepsilon}{1 - \varepsilon} C_\Upsilon \right.$$
$$\left. + \frac{\varepsilon}{1 + \varepsilon} C_\Upsilon + \lambda \left[\mathcal{H}(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1) - \mathcal{H}(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2) \right] \right].$$

Thus, (3.6) in Step 1 follows.

Proof of Step 2: Recall that $\mathcal{H}(x, p)$ is given by

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

Since $\Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \cdot) : \Theta \to \mathbb{R}$ is bounded and continuous by ($\Lambda 1$) and ($\Lambda 4$), and $\mathcal{I}(x_{\varepsilon,\alpha}, \cdot) : \Theta \to [0, \infty]$ has compact sub-level sets in Θ by ($\mathcal{I}3$), there exists an optimizer $\theta_{\varepsilon,\alpha} \in \Theta$ such that

$$\mathcal{H}(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1) = \Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \theta_{\varepsilon,\alpha}) - \mathcal{I}(x_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}). \qquad (3.16)$$

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

$$\mathcal{H}(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1) - \mathcal{H}(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2) \leq \Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \theta_{\varepsilon,\alpha}) - \Lambda(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2, \theta_{\varepsilon,\alpha})$$
$$+ \mathcal{I}(y_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) - \mathcal{I}(x_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}). \qquad (3.17)$$

Proof of Step 3: We will construct for each $\varepsilon > 0$ a sequence $\alpha = \alpha(\varepsilon) \to \infty$ such that the collection $(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha})$ is fundamental for Λ with respect to Ψ in the sense of Definition 2.11. We thus need to verify for each $\varepsilon > 0$

(i)

$$\inf_\alpha \Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \theta_{\varepsilon,\alpha}) > -\infty, \qquad (3.18)$$

(ii)

$$\sup_\alpha \Lambda(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2, \theta_{\varepsilon,\alpha}) < \infty \qquad (3.19)$$

(iii) The set of controls $\theta_{\varepsilon,\alpha}$ is relatively compact.

To prove (i), (ii) and (iii), we introduce auxiliary controls $\theta_{\varepsilon,\alpha}^0$, obtained by $(\mathcal{I}2)$, satisfying

$$\mathcal{I}(y_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}^0) = 0. \tag{3.20}$$

We will first establish (i) and (ii) for all α. Then, for the proof of (iii), we will construct for each $\varepsilon > 0$ a suitable subsequence $\alpha \to \infty$.

Proof of Step 3, (i) and (ii):

We first establish (i). By the subsolution inequality (3.14),

$$\begin{aligned}
\frac{1}{\lambda} \inf_E (u_1 - h) &\le (1 - \varepsilon)\mathcal{H}(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1) + \varepsilon C_\Upsilon \\
&\le (1 - \varepsilon)\Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \theta_{\varepsilon,\alpha}) + \varepsilon C_\Upsilon, \tag{3.21}
\end{aligned}$$

and the lower bound (3.18) follows.

We next establish (ii). By the supersolution inequality (3.15), we can estimate

$$\begin{aligned}
(1 + \varepsilon)\Lambda(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2, \theta_{\varepsilon,\alpha}^0) &= (1 + \varepsilon)\left[\Lambda(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2, \theta_{\varepsilon,\alpha}^0) - \mathcal{I}(y_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}^0)\right] \\
&\le \left((1 + \varepsilon)\mathcal{H}\left(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2\right) - \varepsilon C_\Upsilon\right) + \varepsilon C_\Upsilon \\
&\le \frac{1}{\lambda} \sup_E (u_2 - h) + \varepsilon C_\Upsilon < \infty,
\end{aligned}$$

and the upper bound (3.19) follows by Assumption 2.14 $(\Lambda 4)$.

Proof of Step 3, (iii): To prove (iii), it suffices by Assumption 2.15 $(\mathcal{I}3)$ to find for each $\varepsilon > 0$ a subsequence α such that

$$\sup_\alpha \mathcal{I}(x_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) < \infty. \tag{3.22}$$

By (3.21), we have

$$\begin{aligned}
\frac{1}{\lambda} \inf_E (u_1 - h) &\le (1 - \varepsilon)\mathcal{H}(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1) + \varepsilon C_\Upsilon \\
&= (1 - \varepsilon)\left[\Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \theta_{\varepsilon,\alpha}) - \mathcal{I}(x_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha})\right] + \varepsilon C_\Upsilon.
\end{aligned}$$

We conclude that $\sup_\alpha \mathcal{I}(x_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) < \infty$ is implied by

$$\sup_\alpha \Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \theta_{\varepsilon,\alpha}) < \infty$$

which by $(\Lambda 4)$ is equivalent to

$$\sup_\alpha \Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \theta_{\varepsilon,\alpha}^0) < \infty. \tag{3.23}$$

To perform this estimate, we first write

$$\begin{aligned}
&\Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \theta_{\varepsilon,\alpha}^0) \\
&= \Lambda(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2, \theta_{\varepsilon,\alpha}^0) + \left[\Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \theta_{\varepsilon,\alpha}^0) - \Lambda(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2, \theta_{\varepsilon,\alpha}^0)\right].
\end{aligned} \tag{3.24}$$

To estimate the second term, we aim to apply the continuity estimate for the controls $\theta_{\varepsilon,\alpha}^0$. To do so, must establish that $(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}^0)$ is fundamental for

Λ with respect to Ψ. By Assumption 2.15 ($\mathcal{I}3$), for each ε the set of controls $\theta_{\varepsilon,\alpha}^0$ is relatively compact. Thus it suffices to establish

$$\inf_{\alpha} \Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \theta_{\varepsilon,\alpha}^0) > -\infty, \tag{3.25}$$

$$\sup_{\alpha} \Lambda(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2, \theta_{\varepsilon,\alpha}^0) < \infty. \tag{3.26}$$

These two estimates follow by Assumption 2.14 ($\Lambda4$) and (3.18) and (3.19).

The continuity estimate of Assumption 2.14 ($\Lambda5$) yields that

$$\liminf_{\alpha\to\infty} \Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \theta_{\varepsilon,\alpha}^0) - \Lambda(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2, \theta_{\varepsilon,\alpha}^0) \le 0.$$

This means that there exists a subsequence, also denoted by α such that

$$\sup_{\alpha} \Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^1, \theta_{\varepsilon,\alpha}^0) - \Lambda(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}^2, \theta_{\varepsilon,\alpha}^0) < \infty. \tag{3.27}$$

Thus, we can estimate (3.24) by (3.27) and (3.26). This implies that (3.22) holds for the chosen subsequences α and that for these the collection $(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha})$ is fundamental for Λ with respect to Ψ establishing Step 3.

Proof of Step 4:

For the subsequences constructed in Step 3, we have by (3.22) that

$$\sup_{\alpha} \mathcal{I}(x_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) < \infty. \tag{3.28}$$

As established in Step 1, following Lemma 3.5, for each $\varepsilon > 0$ the set $\{(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha})\}$ is relatively compact where α varies over the subsequences selected in Step 3. In addition, for each $\varepsilon > 0$ there exists $z \in E$ such that $(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha}) \to (z, z)$. It follows by (3.28) and Assumption 2.15 ($\mathcal{I}4$) that also

$$\sup_{\alpha} \mathcal{I}(y_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) < \infty. \tag{3.29}$$

With the bounds (3.28) and (3.29), the estimate ($\mathcal{I}5$) is satisfied for the subsequences $(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha})$. $\qquad\square$

4. Existence of viscosity solutions

In this section, we will prove Theorem 2.8. In other words, we show that for $h \in C_b(E)$ and $\lambda > 0$, the function $R(\lambda)h$ given by

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

is indeed a viscosity solution to $f - \lambda \mathbf{H}f = h$. To do so, we will use the methods of Chapter 8 of [19]. For this strategy, one needs to check three properties of $R(\lambda)$:

(a) For all $(f, g) \in \mathbf{H}$, 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.

Thus, if $R(\lambda)$ serves as a classical left-inverse to $\mathbb{1} - \lambda\mathbf{H}$ and is also a pseudo-resolvent, then it is a viscosity right-inverse of $(\mathbb{1} - \lambda\mathbf{H})$. For a second proof of this statement, outside of the control theory context, see Proposition 3.4 of [24].

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

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

$$\lambda \int_0^\infty z(t)\,\tau_\lambda(\mathrm{d}t) = \int_0^\infty \int_0^t z(s)\,\mathrm{d}s\,\tau_\lambda(\mathrm{d}t).$$

- (b) is related to a more involved integral property of exponential random variables. For $0 < \alpha < \beta$, we have

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

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 [19] on the basis of three assumptions, namely [19, Assumptions 8.9, 8.10 and 8.11]. We verify these below.

Verification of Conditions 8.9, 8.10 and 8.11. In the notation of [19], we use $U = \mathbb{R}^d$, $\Gamma = E \times U$, one operator $\mathbf{H} = \mathbf{H}_\dagger = \mathbf{H}_\ddagger$ and $Af(x,u) = \langle \nabla f(x), u\rangle$ for $f \in \mathcal{D}(\mathbf{H}) = C_{cc}^\infty(E)$.

   Regarding Condition 8.9, by continuity and convexity of $\mathcal{H}$ obtained in Propositions B.1 and B.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 [19, 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 [7]. 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

$$\int_0^T \mathcal{L}(\gamma(t), \dot{\gamma}(t))\,dt \leq M, \tag{4.1}$$

then $\gamma(t) \in K'$ for any $0 \leq t \leq T$.

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

$$\Upsilon(\gamma(\tau)) = \Upsilon(\gamma(0)) + \int_0^\tau \nabla\Upsilon(\gamma(t))\dot{\gamma}(t)\,dt$$
$$\leq \Upsilon(\gamma(0)) + \int_0^\tau \left[\mathcal{L}(\gamma(t), \dot{\gamma}(t))) + \mathcal{H}(x(t), \nabla\Upsilon(\gamma(t)))\right]\,dt$$

$$\leq \sup_K \Upsilon + M + T \sup_{x\in E} \mathcal{H}(x, \nabla\Upsilon(x)) =: C < \infty, \tag{4.2}$$

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

We proceed with the verification of Conditions 8.10 and 8.11 of [19]. By Proposition B.1, we have $\mathcal{H}(x, 0) = 0$ and hence the application of $\mathbf{H}$ to constant 1 function $\mathbb{1}$ satisfies $\mathbf{H}\mathbb{1} = 0$. Thus, Condition 8.10 is implied by Condition 8.11 (see Remark 8.12 (e) in [19]).

We establish that Condition 8.11 is satisfied: for any function $f \in \mathcal{D}(\mathbf{H}) = C^\infty_{cc}(E)$ 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 $t \geq 0$,

$$\int_0^t \mathcal{H}(x(s), \nabla f(x(s)))\, ds = \int_0^t [\dot{x}(s) \cdot \nabla f(x(s)) - \mathcal{L}(x(s), \dot{x}(s))]\, ds. \tag{4.3}$$

To do so, we solve the differential inclusion

$$\dot{x}(t) \in \partial_p \mathcal{H}(x(t), \nabla f(x(t))), \qquad x(0) = x_0, \tag{4.4}$$

where the subdifferential of $\mathcal{H}$ was defined in (2.9) 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 [11]. We have included this result as Lemma D.5, and corresponding preliminary definitions in "Appendix D". We use this result for $F(x) := \partial_p \mathcal{H}(x, \nabla f(x))$. To apply Lemma D.5, 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 2.16 on page 10).

Part (F1) follows from the properties of subdifferential sets of convex and continuous functionals. $\mathcal{H}$ is continuous in (x, p) and convex in p by Proposition B.1. Part (F3) is a consequence of Lemma 4.1, which yields that $F(x) \subseteq T_E(x)$. 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 (4.4). We establish that there is some M such that (4.1) is satisfied. By (4.4) we obtain for all $p \in \mathbb{R}^d$,

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

and as a consequence

$$\dot{x}(t)\nabla f(x(t)) - \mathcal{H}(x(t), \nabla f(x(t))) \geq \mathcal{L}(x(t), \dot{x}(t)).$$

Since f has compact support and $\mathcal{H}(y, 0) = 0$ for any $y \in E$, we estimate

$$\int_0^T \mathcal{L}(x(t), \dot{x}(t))\, ds \leq \int_0^T \dot{x}(t)\nabla f(x(t))\, dt - T \inf_{y\in \mathrm{supp}(f)} \mathcal{H}(y, \nabla f(y)).$$

By continuity of $\mathcal{H}$ 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 $M = M(T)$, with

$$M := T \sup_{y\in\text{supp}(f)} \|F(y)\| \sup_{z\in\text{supp}(f)} |\nabla f(z)| - \inf_{y\in\text{supp}(f)} \mathcal{H}(y, \nabla f(y)).$$

From the first part of the, see the argument concluding after (4.2), we find that the solution $x(t)$ remains in the compact set

$$K' := \left\{ z \in E \,\middle|\, \Upsilon(z) \leq C \right\}, \quad C := \Upsilon(x_0) + M + T \sup_x \mathcal{H}(x, \nabla\Upsilon(x)), \tag{4.5}$$

for all $t \in [0, T]$.

Step 2: We prove that there exists a solution $x(t)$ of (4.4) on $[0, T]$.

Using F, we define a new multi-valued vector-field $F'(z)$ that equals $F(z) = \partial_p \mathcal{H}(z, \nabla f(z))$ 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 $t \geq 0$,

$$\dot{y}(t) \in F'(y(t)).$$

By the estimate established in step 1 and the fact that $\Upsilon(\gamma(t)) \leq C$ for any $0 \leq t \leq T$, it follows from the argument as shown above in (4.2) 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 (4.4) on the time interval $[0, T]$. $\qquad\square$

Lemma 4.1. *Let Assumption* 2.17 *be satisfied. Then the map* $\mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R}$ *defined in* (2.2) *is such that* $\partial_p \mathcal{H}(x, p) \subseteq T_E(x)$ *for all* p *and* $x \in E$.

Proof. Fix $x \in E$ and $p_0 \in \mathbb{R}^d$. We aim to prove that $\partial_p \mathcal{H}(x, p_0) \subseteq T_E(x)$. Recall the definition of $\mathcal{H}$:

$$\mathcal{H}(x, p) = \sup_{\theta\in\Theta} \left\{ \Lambda(x, p, \theta) - \mathcal{I}(x, \theta) \right\}. \tag{4.6}$$

Let $\Omega(p) \subseteq \Theta$ be the set of controls that optimize $\mathcal{H}$: thus if $\theta \in \Omega(p)$ then $\mathcal{H}(x, p) = \Lambda(x, p, \theta) - \mathcal{I}(x, \theta)$.

The result will follow from the following claim,

$$\partial_p \mathcal{H}(x, p_0) = ch \bigcup_{\theta \in \Omega(p_0)} \partial_p \Lambda(x, p_0, \theta), \tag{4.7}$$

where ch denotes the convex hull. Having established this claim, the result follows from Assumption 2.17 and the fact that $T_E(x)$ is a convex set by Lemma D.4.

We start with the proof of (4.7). For this we will use [22, Theorem D.4.4.2]. To study the subdifferential of the function $\partial_p \mathcal{H}(x, p_0)$, it suffices to restrict the domain of the map $p \mapsto \mathcal{H}(x, p)$ to the closed ball $B_1(p_0)$ around p_0 with radius 1.

To apply [22, Theorem D.4.4.2] for this restricted map, first recall that Λ is continuous by Assumption 2.14 (Λ1) and that $\mathcal{I}$ is lower semi-continuous by Assumption 2.15 ($\mathcal{I}$1). Secondly, we need to find a compact set $\Omega \subseteq \Theta$ such that we can restrict the supremum (for any $p \in B_1(p_0)$) in (4.6) to Ω:

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

In particular, we show that we can take for Ω a sublevelset of $\mathcal{I}(x, \cdot)$ which is compact by Assumption 2.15 ($\mathcal{I}$3).

Let θ_x^0 be the control such that $\mathcal{I}(x, \theta_x^0) = 0$, which exists due to Assumption 2.15 ($\mathcal{I}$2). Let M^* be such that (with the constants M, C_1, C_2 as in Assumption 2.14 (Λ4))

$$M^* = \sup_{p \in B_1(p_0)} \left\{ \max \left\{ M, C_1 \Lambda(x, p, \theta_x^0) + C_2 \right\} - \Lambda(x, p, \theta_x^0) \right\} < \infty.$$

Note that M^* is finite as $p \mapsto \Lambda(x, p, \theta_x^0)$ is continuous on the closed unit ball $B_1(p_0)$. Then we find, due to Assumption 2.14 (Λ4), that if θ satisfies $\mathcal{I}(x, \theta) > M^*$, then for any $p \in B_1(p_0)$ we have

$$\Lambda(x, p, \theta) - \mathcal{I}(x, \theta) < \Lambda(x, p, \theta_x^0) \leq \mathcal{H}(x_0, p).$$

We obtain that if $p \in B_1(p_0)$, then we can restrict our supremum in (4.6) to the compact set $\Omega := \Theta_{\{x\}, M^*}$, see Assumption 2.15 ($\mathcal{I}$3).

Thus, it follows by [22, Theorem D.4.4.2] that

$$\partial_p \mathcal{H}(x, p_0) = ch \left(\bigcup_{\theta \in \Theta_{\{x\}, \bar{M}^*}} \partial_p \left(\Lambda(x, p_0, \theta) - \mathcal{I}(x, \theta) \right) \right),$$

where ch denotes the convex hull. Now (4.7) follows by noting that $\mathcal{I}(x, \theta)$ does not depend on p. □

5. Examples of Hamiltonians

In this section we specify our general results to two concrete examples of Hamiltonians of the type

$$\mathcal{H}(x, p) = \sup_{\theta \in \Theta} \left[\Lambda(x, p, \theta) - \mathcal{I}(x, \theta) \right]. \tag{5.1}$$

The purpose of this section is to showcase that the method introduced in this paper is versatile enough to capture interesting examples that could not be treated before.

First, we consider in Proposition 5.1 Hamiltonians that one encounters in the large deviation analysis of two-scale systems as studied in [6] and [27] when considering a diffusion process coupled to a fast jump process. Second, we consider in Proposition 5.7 the example treated in our introduction that arises from models of mean-field interacting particles that are coupled to fast external variables. This example will be further analyzed in [26].

Proposition 5.1. (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 Λ is given by*

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

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*

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

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

Suppose that the cost function $\mathcal{I}$ satisfies the assumptions of Proposition 5.9 below and the function Λ satisfies the assumptions of Proposition 5.11 below. Then Theorems 2.6 and 2.8 apply to the Hamiltonian (5.1).

Proof. To apply Theorems 2.6 and 2.8, we need to verify Assumptions 2.14, 2.15 and 2.17. Assumption 2.14 follows from Proposition 5.11, Assumption 2.15 follows from Proposition 5.9 and Assumption 2.17 is satisfied as $E = \mathbb{R}^d$. $\square$

Remark 5.2. We assume uniform ellipticity of a, which we use to establish (Λ4). This leaves our comparison principle slightly lacking to prove a large deviation principle as general as in [5]. In contrast, we do not need a Lipschitz condition on r in terms of x.

While we believe that the conditions on a can be relaxed by performing a finer analysis of the estimates in terms of a, we do not pursue this relaxation here.

Remark 5.3. The cost function is the large deviation rate function for the occupation time measures of a jump process taking values in a finite set $\{1, \ldots, J\}$, see e.g. [13,14].

Remark 5.4. In the context with $a = 0$ and $\mathcal{I}$ as general as Assumption 2.15, we improve upon the results of Chapter III of [2] by allowing a more general class of functionals $\mathcal{I}$, that are e.g. discontinuous as for example in Proposition 5.7 below.

In [10] the authors consider a second order Hamilton–Jacobi–Bellman equation, with the quadratic part replaced by a second order part. They work, however, with continuous cost functional $\mathcal{I}$. An extension of [10] that allows for a similar flexibility in the choice of $\mathcal{I}$ would therefore be of interest.

Remark 5.5. Under irreducibility conditions on the rates, as we shall assume below in Proposition 5.9, by [15] the Hamiltonian $\mathcal{H}(x,p)$ is the principal eigenvalue of the matrix $A_{x,p} \in \mathrm{Mat}_{J \times J}(\mathbb{R})$ given by

$$A_{x,p} = \mathrm{diag}\left[\langle a(x,1)p,p\rangle + \langle b(x,1),p\rangle, \ldots, \langle a(x,J)p,p\rangle + \langle b(x,J),p\rangle\right] + R_x,$$

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

$$\begin{pmatrix} -\sum_{j\neq 1} r(1,j,x) & r(1,2,x) & \cdots & r(1,J,x) \\ r(2,1,x) & -\sum_{j\neq 2} r(2,j,x) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ r(J,1,x) & \cdots & r(J,J-1,x) & -\sum_{j\neq J} r(J,j,x) \end{pmatrix},$$

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

Next we consider Hamiltonians arising in the context of weakly interacting jump processes on a collection of states $\{1, \ldots, q\}$ as described in our introduction. We analyze and motivate this example in more detail in our companion paper [26]. We give the terminology as needed for the results in this paper.

The empirical measure of the interacting processes takes its values in the set of measures $\mathcal{P}(\{1, \ldots, q\})$. The dynamics arises from mass moving over the bonds $(a,b) \in \Gamma = \{(i,j) \in \{1, \ldots, q\}^2 \mid i \neq j\}$. As the number of processes is send to infinity, there is a type of limiting result for the total mass moving over the bonds.

We will denote by $v(a,b,\mu,\theta)$ the total mass that moves from a to b if the empirical measure equals μ and the control is given by θ. We will make the following assumption on the kernel v.

Definition 5.6. (*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 μ such that $\mu(a) > 0$.
(b) There exists a decomposition $v(a,b,\mu,\theta) = v_\dagger(a,b,\mu(a))v_\ddagger(a,b,\mu,\theta)$ such that $v_\dagger$ is increasing in the third coordinate and such that $v_\ddagger(a,b,\cdot,\cdot)$ is continuous and satisfies $v_\ddagger(a,b,\mu,\theta) > 0$.

A typical example of a proper kernel is given by
$$v(a,b,\mu,\theta) = \mu(a)r(a,b,\theta)e^{\partial_a V(\mu) - \partial_b V(\mu)},$$

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

Proposition 5.7. (Mean-field coupled to diffusion) *Let the space E be given by the embedding of $E := \mathcal{P}(\{1, \ldots, J\}) \times [0, \infty)^\Gamma \subseteq \mathbb{R}^d$ and F be a smooth compact Riemannian manifold without boundary. Suppose the following.*

(i) *The set of control variables Θ equals $\mathcal{P}(F)$.*
(ii) *The function Λ is given by*

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

with a proper kernel v in the sense of Definition 5.6.
(iii) *The cost function $\mathcal{I} : E \times \Theta \to [0, \infty]$ is given by*

$$\mathcal{I}(x, \theta) := \sup_{\substack{u \in \mathcal{D}(L_x) \\ \inf u > 0}} \left[-\int_F \frac{L_x u}{u} \, d\theta \right],$$

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 $\mathcal{I}$ satisfies the assumptions of Proposition 5.10 and the function Λ satisfies the assumptions of Proposition 5.13. Then Theorems 2.6 and 2.8 apply to the Hamiltonian (5.1).

Proof. To apply Theorems 2.6 and 2.8, we need to verify Assumptions 2.14, 2.15 and 2.17. Assumption 2.14 follows from Proposition 5.13 and Assumption 2.15 follows from Proposition 5.10. We verify Assumption 2.17 in Proposition 5.19. □

Remark 5.8. The cost function stems from occupation-time large deviations of a drift-diffusion process on a compact manifold, see e.g. [15,32]. We expect Proposition 5.7 to extend also to non-compact spaces F, but we feel this technical extension is better suited for a separate paper.

5.1. Verifying assumptions for cost functions $\mathcal{I}$

Here we verify Assumption 2.15 for two types of cost functions $\mathcal{I}(x, \theta)$ appearing in the examples of Propositions 5.1 and 5.7.

Proposition 5.9. (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*

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

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

$$\mathcal{I}(x,\theta) := - \inf_{\substack{\phi \in C_b(F) \\ \inf \phi > 0}} \int \frac{L_x \phi(z)}{\phi(z)} \, d\theta$$

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

satisfies Assumption 2.15.

Proof. ($\mathcal{I}1$): For a fixed vector $w \in \mathbb{R}^J$, the map

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

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

($\mathcal{I}2$): Let $x \in E$. First note that for all θ, 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 $\theta_x^0 \in \Theta$ such that for any $f : \{1,\ldots,J\} \to \mathbb{R}$,

$$\sum_i L_x f(i)\theta_x^0(i) = 0. \tag{5.2}$$

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

$$\left(1 - e^{b-a}\right) \leq -(b-a) \quad \text{for all } a,b > 0,$$

we obtain that

$$\sum_{ij} r(i,j,x)\theta_x^0(i)\left(1 - e^{w_j - w_i}\right) \leq \sum_{ij} r(i,j,x)\theta_x^0(i)\left(w_j - w_i\right)$$

$$= \sum_i (L_x w)(i)\theta_x^0(i) = 0$$

by (5.2). Since $\mathcal{I} \geq 0$, this implies $\mathcal{I}(x,\theta_x^0) = 0$.

($\mathcal{I}3$): Any closed subset of Θ is compact.

($\mathcal{I}4$): Let $x_n \to x$ 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 $y \in K$,

$$\mathcal{I}(y,\theta) \leq \sum_{ij,i \neq j} r(i,j,y)\theta_i \leq \sum_{ij,i \neq j} r(i,j,y) \leq \sum_{ij,i \neq j} \bar{r}_{ij}, \quad \bar{r}_{ij} := \sup_{y \in K} r(i,j,y).$$

Hence $\mathcal{I}$ is uniformly bounded on $K \times \Theta$, and ($\mathcal{I}4$) follows with U_x the interior of K.

($\mathcal{I}5$) : 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 $d(x, y) \le \delta$ and for all $\theta \in \mathcal{P}(F)$, we have

$$|\mathcal{I}(x, \theta) - \mathcal{I}(y, \theta)| \le \varepsilon. \tag{5.3}$$

Let $x, y \in K$. By continuity of the rates the $\mathcal{I}(x, \cdot)$ are uniformly bounded for $x \in K$:

$$0 \le \mathcal{I}(x, \theta) \le \sum_{ij, i \neq j} r(i, j, x)\theta_i \le \sum_{ij, i \neq j} r(i, j, x) \le \sum_{ij, i \neq j} \bar{r}_{ij}, \quad \bar{r}_{ij} := \sup_{x \in K} r(i, j, x).$$

For any $n \in \mathbb{N}$, there exists $w^n \in \mathbb{R}^J$ such that

$$0 \le \mathcal{I}(x, \theta) \le \sum_{ij, i \neq j} r_{ij}(x)\theta_i(1 - e^{w_j^n - w_i^n}) + \frac{1}{n}.$$

By reorganizing, we find for all bonds (a, b) the bound

$$\theta_a e^{w_b^n - w_a^n} \le \frac{1}{r_{K,a,b}} \left[\sum_{ij, i \neq j} r(i, j, x)\theta_i + \frac{1}{n} \right] \le \frac{1}{r_{K,a,b}} \left[\sum_{ij, i \neq j} \bar{r}_{ij} + \frac{1}{n} \right],$$

where $r_{K,a,b} := \inf_{x \in K} r(a, b, x)$. Thereby, evaluating in $\mathcal{I}(y, \theta)$ the same vector w^n to estimate the supremum,

$$\mathcal{I}(x, \theta) - \mathcal{I}(y, \theta)$$

$$\le \frac{1}{n} + \sum_{ab, a \neq b} r(a, b, x)\theta_a(1 - e^{w_b^n - w_a^n}) - \sum_{ab, a \neq b} r(a, b, y)\theta_a(1 - e^{w_b^n - w_a^n})$$

$$\le \frac{1}{n} + \sum_{ab, a \neq b} |r(a, b, x) - r(a, b, y)|\theta_a + \sum_{ab, a \neq b} |r(a, b, y) - r(a, b, x)|\theta_a e^{w_b^n - w_a^n}$$

$$\le \frac{1}{n} + \sum_{ab, a \neq b} |r(a, b, x) - r(a, b, y)| \left(1 + \frac{1}{r_{K,a,b}} \left[\sum_{ij, i \neq j} \bar{r}_{ij} + 1 \right] \right).$$

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 (5.3). $\qquad \square$

Proposition 5.10. (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_x^{ij} and b_x^i on F. Suppose that for all i, j the maps

$$x \mapsto a_x^{i,j}(\cdot), \qquad x \mapsto b_x^i(\cdot) \tag{5.4}$$

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

$$\mathcal{I}(x, \theta) := \sup_{\substack{u \in \mathcal{D}(L_x) \\ u > 0}} \left[-\int_F \frac{L_x u}{u} \, d\theta \right]$$

satisfies Assumption 2.15.

Proof. $(\mathcal{I}1)$: For any fixed function $u \in \mathcal{D}(L_x)$ such that $u > 0$, the function $(-L_x u / u)$ is continuous on F. Note that by definition of $\mathcal{I}$ it suffices to only consider $u > 0$. Thus, for any such fixed $u > 0$ it follows by (5.4) 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 $\mathcal{I}(x, \theta)$ is lower semicontinuous as the supremum over continuous functions.

$(\mathcal{I}2)$: Let $x \in E$. The stationary measure $\theta_x^0 \in \Theta$ satisfying

$$\int_F L_x g(z) \, \theta_x^0(\mathrm{d}z) = 0 \quad \text{for all } g \in \mathcal{D}(L_x) \tag{5.5}$$

is the minimizer of $\mathcal{I}(x, \cdot)$, that is $\mathcal{I}(x, \theta_x^0) = 0$. This follows by considering the Hille-Yosida approximation L_x^ε of L_x and using the same argument (using $w = \log u$) as in Proposition 5.9 for these approximations. For any $u > 0$ and $\varepsilon > 0$,

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

$$\leq -\int_F \frac{L_x^\varepsilon u}{u} \, d\theta + \frac{1}{\inf_F u} \| (L_x^\varepsilon - L_x)u \|_F$$

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

Sending $\varepsilon \to 0$ and then using (5.5) gives $(\mathcal{I}2)$.

$(\mathcal{I}3)$: Since $\Theta = \mathcal{P}(F)$ is compact, any closed subset of Θ is compact. Hence any union of sub-level sets of $\mathcal{I}(x, \cdot)$ is relatively compact in Θ.

$(\mathcal{I}4)$: Fix $x \in E$ and $M \geq 0$. Let $\theta \in \Theta_{\{x\}, M}$. As $\mathcal{I}(x, \theta) \leq M$, we find by [31] that the density $\frac{\mathrm{d}\theta}{\mathrm{d}z}$ exists, where $\mathrm{d}z$ denotes the Riemannian volume measure.

In addition it follows from [31, Theorem 1.4] there exists constants $c_1(y)$, $c_2(y), c_3(y), c_4(y), c_1(y), c_2(y)$ being positive, depending continuously on a_y, a_y^{-1}, b_y (See the derivation of [30, Eq. (2.18), (2.19)]), but not on θ, such that

$$c_1(y) \int_F |\nabla g_\theta|^2 \, dz - c_2(y) \leq \mathcal{I}(y, \theta) \leq c_3(y) \int_F |\nabla g_\theta|^2 \, dz + c_4(y), \tag{5.6}$$

where $g_\theta = (\mathrm{d}\theta / \mathrm{d}z)^{1/2}$.

As the dependence is continuous in y, we can find a open set $U \subseteq E$ of x such that there are constants c_1, c_2, c_3, c_4, c_1, c_3 being positive, that do not depend on θ, such that for any $y \in U$:

$$c_1 \int_F |\nabla g_\theta|^2 \, dz - c_2 \leq \mathcal{I}(y, \theta) \leq c_3 \int_F |\nabla g_\theta|^2 \, dz + c_4. \tag{5.7}$$

From (5.7), $(\mathcal{I}4)$ immediately follows.

$(\underline{\mathcal{I}5})$: 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 [32]. $\qquad\square$

5.2. Verifying assumptions for functions Λ

In this section we verify Assumption 2.14 for two types of functions $\Lambda(x, p, \theta)$ appearing in the examples of Propositions 5.1 and 5.7.

Proposition 5.11. (Quadratic function Λ) *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*

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

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 $K \subseteq \mathbb{R}^d$,

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

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

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

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. Then Assumption 2.14 holds.

Remark 5.12. The above proposition is slightly more general than what we consider in Proposition 5.1, as there we assume that $F = \{1, \ldots, J\}$ is a finite set.

Proof.

$(\underline{\Lambda 1})$: Let $(x, p) \in E \times \mathbb{R}^d$. Continuity of Λ 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 $\Upsilon : E \to \mathbb{R}$ defined by

$$\Upsilon(x) := \frac{1}{2} \log\left(1 + |x|^2\right)$$

is a containment function for Λ. For any $x \in E$ and $\theta \in \mathcal{P}(F)$, we have

$$\Lambda(x, \nabla\Upsilon(x), \theta) = \int_F \langle a(x, z)\nabla\Upsilon(x), \nabla\Upsilon(x)\rangle \, d\theta(z) + \int_F \langle b(x, z), \nabla\Upsilon(x)\rangle \, d\theta(z)$$

$$\leq a_{\{x\},\max}|\nabla\Upsilon(x)|^2 + b_{\{x\},\max}|\nabla\Upsilon(x)|$$

$$\leq C(1 + |x|)\frac{x^2}{(1 + x^2)^2} + C(1 + |x|)\frac{x}{(1 + x^2)},$$

and the boundedness condition follows with the constant

$$C_\Upsilon := C \sup_x (1 + |x|)\left[\frac{x^2}{(1 + x^2)^2} + \frac{x}{(1 + x^2)}\right] < \infty.$$

(Λ4): 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

$$\Lambda(x, p, \theta_1) \leq \max\{M, C_1\Lambda(x, p, \theta_2) + C_2\}. \tag{5.8}$$

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 $K \times \times S^c \times \Theta$. Thus, we can find a constant C_2 such that (5.8) holds.

(Λ5): By the assumption on $a(x, z)$, the function Λ is uniformly coercive in the sense that for any compact set $K \subseteq E$,

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

and the continuity estimate follows by Proposition 5.15.

$\square$

We proceed with the example in which Λ depends on p through exponential functions (Proposition 5.7). Let $q \in \mathbb{N}$ be an integer and

$$\Gamma := \left\{ (a,b) \,\middle|\, a,b \in \{1,\ldots,q\},\ a \neq b \right\}$$

be the set of oriented edges in $\{1,\ldots,q\}$.

Proposition 5.13. (Exponential function Λ) *Let $E \subseteq \mathbb{R}^d$ be the embedding of $E = \mathcal{P}(\{1,\ldots,q\}) \times (\mathbb{R}^+)^{|\Gamma|}$ and Θ be a topological space. Suppose that Λ 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 5.6. 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

$$\sup_{\mu} \sup_{\theta_1,\theta_2} \frac{v(a,b,\mu,\theta_1)}{v(a,b,\mu,\theta_2)} \leq C. \tag{5.9}$$

Then Λ satisfies Assumption 2.14.

Remark 5.14. Similar to the previous proposition, the assumptions on Λ are satisfied when $\Theta = \mathcal{P}(F)$ for some Polish space F, we have $v(a,b,\mu,\theta) = \mu(a) \int r(a,b,\mu,z)\theta(\mathrm{d}z)$, and there are constants $0 < r_{min} \leq r_{max} < \infty$ such that for all $(a,b) \in \Gamma$ such that $\sup_{\mu,z} r(a,b,\mu,z) > 0$, we have

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

Regarding (5.9), for $(a,b) \in \Gamma$ for which $v(a,b,\cdot,\cdot)$ is non-trivial, we have

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

Proof of Proposition 5.13. (Λ1): The function Λ is continuous as the sum of continuous functions.

(Λ2): Convexity of Λ as a function of p follows from the fact that Λ is a finite sum of convex functions, and $\Lambda(x,0,\theta) = 0$ is evident.

(Λ3): The function $\Upsilon : E \to \mathbb{R}$ defined by

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

is a containment function for Λ. For a verification, see [23].

(Λ4): Note that

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

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

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

Thus the estimate holds with $M=0$, $C_1=C$ and $C_2=\sup_{\mu,\theta}\sum_{a,b} v(a,b,\mu,\theta)$.

$(\underline{\Lambda 5})$: The continuity estimate is the content of Proposition 5.18 below.

$\square$

5.3. Verifying the continuity estimate

With the exception of the verification of the continuity estimate in Assumption 2.14 the verification in Sect. 5.2 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 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 Θ a topological space.

Proposition 5.15. (Coercive Λ) *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*

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

Then the continuity estimate holds for Λ with respect to any penalization function Ψ.

Proof. Let $\Psi(x,y)=\frac{1}{2}(x-y)^2$. Let $(x_{\alpha,\varepsilon},y_{\alpha,\varepsilon},\theta_{\varepsilon,\alpha})$ be fundamental for Λ with respect to Ψ. Set $p_{\alpha,\varepsilon}=\alpha(x_{\varepsilon,\alpha}-y_{\varepsilon,\alpha})$. By the upper bound (2.4), we find that for sufficiently small $\varepsilon>0$ there is some $\alpha(\varepsilon)$ such that

$$\sup_{\alpha\geq\alpha(\varepsilon)} \Lambda\left(y_{\varepsilon,\alpha},p_{\varepsilon,\alpha},\theta_{\varepsilon,\alpha}\right)<\infty.$$

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

$$\liminf_{\alpha\to\infty} \Lambda\left(x_{\varepsilon,\alpha},p_{\varepsilon,\alpha},\theta_{\varepsilon,\alpha}\right) - \Lambda\left(y_{\alpha,\varepsilon},p_{\varepsilon,\alpha},\theta_{\varepsilon,\alpha}\right)$$
$$\leq \lim_{\alpha'\to\infty} \Lambda\left(x_{\varepsilon,\alpha'},p_{\varepsilon,\alpha'},\theta_{\varepsilon,\alpha'}\right) - \Lambda\left(y_{\varepsilon,\alpha'},p_{\varepsilon,\alpha'},\theta_{\varepsilon,\alpha'}\right)=0$$

establishing the continuity estimate. $\square$

Proposition 5.16. (One-sided Lipschitz Λ) *Let $\Lambda : E \times \mathbb{R}^d \times \Theta \to \mathbb{R}$ satisfy*

$$\Lambda(x, \alpha(x-y), \theta) - \Lambda(y, \alpha(x-y), \theta) \le c(\theta)\omega(|x-y| + \alpha(x-y)^2) \quad (5.10)$$

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

Then the continuity estimate holds for Λ with respect to $\Psi(x, y) = \frac{1}{2}(x - y)^2$.

Proof. Let $\Psi(x, y) = \frac{1}{2}(x-y)^2$. Let $(x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon}, \theta_{\varepsilon,\alpha})$ be fundamental for Λ with respect to Ψ. Set $p_{\alpha,\varepsilon} = \alpha(x_{\varepsilon,\alpha} - y_{\varepsilon,\alpha})$. We find

$$\liminf_{\alpha \to \infty} \Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) - \Lambda(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha})$$

$$\le \liminf_{\alpha \to \infty} c(\theta_{\varepsilon,\alpha})\omega\left(|x_{\varepsilon,\alpha} - y_{\varepsilon,\alpha}| + \alpha(x_{\varepsilon,\alpha} - y_{\varepsilon,\alpha})^2\right)$$

which equals 0 as $\sup_\theta c(\theta) < \infty$, $\lim_{\delta \downarrow 0} \omega(\delta) = 0$ and property (C2) of a fundamental collection of variables. $\qquad\square$

For the empirical measure of a collection of independent processes one obtains maps Λ 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 [23].

Proposition 5.17. *Suppose that $E = [-1, 1]$ and that $\Lambda(x, p, \theta)$ is given by*

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

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

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

We have

$$\Lambda(x_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha}) - \Lambda(y_{\varepsilon,\alpha}, p_{\varepsilon,\alpha}, \theta_{\varepsilon,\alpha})$$

$$= (y_{\varepsilon,\alpha} - x_{\varepsilon,\alpha})\, c_+(\theta_{\varepsilon,\alpha})\left[e^{p_{\varepsilon,\alpha}} - 1\right] + (x_{\varepsilon,\alpha} - y_{\varepsilon,\alpha})\, c_-(\theta_{\varepsilon,\alpha})\left[e^{-p_{\varepsilon,\alpha}} - 1\right]$$

Now note that $y_{\varepsilon,\alpha} - x_{\varepsilon,\alpha}$ is positive if and only if $e^{p_{\varepsilon,\alpha}} - 1$ 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. $\qquad\square$

Proposition 5.18. *Suppose $E = \mathcal{P}(\{1, \dots, q\} \times (\mathbb{R}^+)^\Gamma$ and suppose that Λ 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 Λ with respect to penalization functions (see Sect. C)

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

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

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

In this context, one can use coercivity like in Proposition 5.15 in combination with directional properties used in the proof of Proposition 5.17 above.

To be more specific: the proof of this proposition can be carried out exactly as the proof of Theorem 3.8 of [23]: namely at any point a converging subsequence is constructed, the variables α need to be chosen such that we also get convergence of the measures $\theta_{\varepsilon,\alpha}$ in $\mathcal{P}(F)$.

5.4. Verifying assumption 2.17 for the exponential internal Hamiltonian

Proposition 5.19. *Let Λ be as in Proposition 5.7:*

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

Then we have $\partial_p \Lambda((\mu, x), p) \subseteq T_E(\mu, w)$.

A sketch of the verification of Assumption 2.17. We sketch the proof in a simplified case, the general case being similar. Consider $E = \mathcal{P}(\{a, b\})$ (ignoring the flux for the moment), and identify E with the simplex in $\mathbb{R}^2$. Fix the control $\theta \in \Theta$. We have to show $\partial_p \Lambda(\mu, p, \theta) \subseteq T_E(\mu)$. Recall that $T_E(\mu)$ is the tangent cone at μ, that means the vectors at μ pointing inside of E. We compute the vector $\nabla_p \Lambda(\mu, p, \theta) \in \mathbb{R}^2$,

$$\nabla_p \Lambda(\mu, p, \theta) = \begin{pmatrix} -v(a, b, \mu, \theta)e^{p_b - p_a} + v(b, a, \mu, \theta)e^{p_a - p_b} \\ v(a, b, \mu, \theta)e^{p_b - p_a} - v(b, a, \mu, \theta)e^{p_a - p_b} \end{pmatrix}.$$

For $\mu = (\mu_a, \mu_b) \in E$ with $\mu_a, \mu_b > 0$, the tangent cone $T_E(\mu)$ is spanned by $(1, -1)^T$. Since $\nabla_p \Lambda(\mu, p, \theta)$ is orthogonal to $(1, 1)^T$, we indeed find that $\partial_p \Lambda(\mu, p, \theta) \subseteq T_E(\mu)$ in that case. For $\mu = (1, 0)$, the tangent cone is $T_E(1, 0) = \{\lambda(-1, 1)^T : \lambda \geq 0\}$. We have

$$\nabla_p \Lambda(\mu, p, \theta) = \begin{pmatrix} -v(a, b, \mu, \theta)e^{p_b - p_a} \\ v(a, b, \mu, \theta)e^{p_b - p_a} \end{pmatrix},$$

which is parallel to $(-1, 1)^T$, and therefore $\partial_p \Lambda(\mu, p, \theta) \subseteq T_E(\mu)$. The argument is similar for $\mu = (0, 1)$. The general case (including the fluxes) follows from a more tedious, but straightforward, computation. □

Acknowledgements

A. Viscosity solutions

In Sect. 3 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 Sect. 2.1.

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

$$f - \lambda A_1 f = h_1, \tag{A.1}$$

$$f - \lambda A_2 f = h_2. \tag{A.2}$$

We say that u is a *(viscosity) subsolution* of Eq. (A.1) if u is bounded, upper semi-continuous and if for all $(f, g) \in A_1$ there exists a sequence $x_n \in E$ such that

$$\lim_{n \uparrow \infty} u(x_n) - f(x_n) = \sup_x u(x) - f(x),$$
$$\lim_{n \uparrow \infty} u(x_n) - \lambda g(x_n) - h(x_n) \leq 0.$$

We say that v is a *(viscosity) supersolution* of Eq. (A.2) if v is bounded, lower semi-continuous and if for all $(f, g) \in A_2$ there exists a sequence $x_n \in E$ such that

$$\lim_{n \uparrow \infty} v(x_n) - f(x_n) = \inf_x v(x) - f(x),$$
$$\lim_{n \uparrow \infty} v(x_n) - \lambda g(x_n) - h(x_n) \geq 0.$$

If $h_1 = h_2$, we say that u is a *(viscosity) solution* of Eqs. (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. Regularity of the Hamiltonian

In this section, we establish continuity, convexity and the existence of a containment function for the Hamiltonian $\mathcal{H}$ of (2.2). We repeat its definition for convenience:

$$\mathcal{H}(x, p) = \sup_{\theta \in \Theta} \left[\Lambda(x, p, \theta) - \mathcal{I}(x, \theta) \right]. \tag{B.1}$$

Proposition B.1. (Regularity of the Hamiltonian) *Let $\mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R}$ be the Hamiltonian as in* (B.1)*, and suppose that Assumptions 2.14 and 2.15 are satisfied. Then:*

(i) *For any $x \in E$, the map $p \mapsto \mathcal{H}(x, p)$ is convex and $\mathcal{H}(x, 0) = 0$.*
(ii) *With the containment function $\Upsilon : E \to \mathbb{R}$ of (Λ3), we have*

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

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

For proving $\mathcal{H}(x, 0) = 0$, let $x \in E$. Then by (Λ2) of Assumption 2.14, we have $\Lambda(x, 0, \theta) = 0$, and therefore

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

since $\mathcal{I} \geq 0$ by Assumption 2.15 and $\mathcal{I}(x, \theta_x^0) = 0$ for some θ_x^0 by ($\mathcal{I}$2) of Assumption 2.15. Regarding (ii), we note that by (Λ3),

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

$\square$

To prove that $\mathcal{H}$ is continuous, we use Assumption 2.15. What we truly need, however, is that $\mathcal{I}$ Gamma converges as a function of x. We establish this result first.

Proposition B.2. (Gamma convergence of the cost functions) *Let a cost function $\mathcal{I} : E \times \Theta \to [0, \infty]$ satisfy Assumption 2.15. Then if $x_n \to x$ in E, the functionals $\mathcal{I}_n$ defined by*

$$\mathcal{I}_n(\theta) := \mathcal{I}(x_n, \theta)$$

converge in the Γ-sense to $\mathcal{I}_\infty(\theta) := \mathcal{I}(x, \theta)$. 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 controls $\theta_n \in \Theta$ 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 semicontinuity ($\mathcal{I}$1),

$$\liminf_{n \to \infty} \mathcal{I}(x_n, \theta_n) \geq \mathcal{I}(x, \theta).$$

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 $(\mathcal{I}4)$, 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 $(\mathcal{I}5)$,

$$\lim_{n \to \infty} |\mathcal{I}(x_n, \theta_n) - \mathcal{I}(x, \theta)| \leq 0,$$

and the lim-sup bound follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proposition B.3. (Continuity of the Hamiltonian) *Let $\mathcal{H} : E \times \mathbb{R}^d \to \mathbb{R}$ be the Hamiltonian defined in (2.2), and suppose that Assumptions 2.14 and 2.15 are satisfied. Then the map $(x, p) \mapsto \mathcal{H}(x, p)$ is continuous and the Lagrangian $(x, v) \mapsto \mathcal{L}(x, v) := \sup_p \langle p, v \rangle - \mathcal{H}(x, p)$ 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 Θ is Polish.

Remark B.4. The proof of upper semi-continuity of $\mathcal{H}$ works in general, using continuity properties of Λ, 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 B.2.

Remark B.5. In the lemma we use a sequential characterization of upper hemi-continuity which holds if Θ is Polish. This is inspired by the natural formulation of Gamma convergence in terms of sequences. An extension of our results to spaces Θ beyond the Polish context should be possible to Hausdorff Θ 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 B.6. (Lemma 17.30 in [1]) *Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces. Let $\phi : \mathcal{X} \to \mathcal{K}(\mathcal{Y})$, where $\mathcal{K}(\mathcal{Y})$ is the space of non-empty compact subsets of $\mathcal{Y}$. Suppose that ϕ is upper hemi-continuous, that is if $x_n \to x$ and $y_n \to y$ and $y_n \in \phi(x_n)$, then $y \in \phi(x)$.*

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

Proof of Proposition B.3. We start by establishing upper semi-continuity of $\mathcal{H}$. We argue on the basis of Lemma B.6. Recall the representation of $\mathcal{H}$ of (B.1). Set $\mathcal{X} = E \times \mathbb{R}^d$ for the (x, p) variables, $\mathcal{Y} = \Theta$, and $f(x, p, \theta) = \Lambda(x, p, \theta) - \mathcal{I}(x, \theta)$ and note that this function is upper semi-continuous by Assumption 2.15 $(\mathcal{I}1)$ and by Assumption 2.14 $(\Lambda1)$.

By Assumption 2.15 ($\mathcal{I}2$), we have $\mathcal{H}(x,p) \geq \Lambda(x,p,\theta_x^0)$, where θ_x^0 is a control such that $\mathcal{I}(x,\theta_x^0) = 0$. Thus, it suffices to restrict the supremum over $\theta \in \Theta$ to $\theta \in \phi(x,p)$ where

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

where $\left\|\cdot\right\|_\Theta$ denotes the supremum norm on Θ. Note that $\left\|\Lambda(x,p,\cdot)\right\|_\Theta < \infty$ by Assumption 2.14 ($\Lambda4$). It follows that

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

$\phi(x,p)$ is non-empty as $\theta_x^0 \in \phi(x,p)$ and it is compact due to Assumption 2.15 ($\mathcal{I}3$). We are left to show that ϕ 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 ($\mathcal{I}1$) and the definition of ϕ we find

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

which implies indeed that $\theta \in \phi(x,p)$. Thus, upper semi-continuity follows by an application of Lemma B.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 θ be the measure such that $\mathcal{H}(x,p) = \Lambda(x,p,\theta) - \mathcal{I}(x,\theta)$. We have

- By Proposition B.2 there are θ_n such that $\theta_n \to \theta$ and $\limsup_n \mathcal{I}(x_n,\theta_n) \leq \mathcal{I}(x,\theta)$.
- $\Lambda(x_n,p_n,\theta_n)$ converges to $\Lambda(x,p,\theta)$ by Assumption ($\Lambda1$).

Therefore,

$$\begin{aligned}
\liminf_{n \to \infty} \mathcal{H}(x_n,p_n) &\geq \liminf_{n \to \infty} \left[\Lambda(x_n,p_n,\theta_n) - \mathcal{I}(x_n,\theta_n) \right] \\
&\geq \liminf_{n \to \infty} \Lambda(x_n,p_n,\theta_n) - \limsup_{n \to \infty} \mathcal{I}(x_n,\theta_n) \\
&\geq \Lambda(x,p,\theta) - \mathcal{I}(x,\theta) = \mathcal{H}(x,p),
\end{aligned}$$

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. $\qquad\square$

C. 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 [16,23]. In absence of a general method to solve these issues, ad-hoc procedures can be introduced. One such ad-hoc procedure introduced in [23] 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 C.1. We say that $\{\Psi_1, \Psi_2\}$, $\Psi_i : E^2 \to \mathbb{R}^+$ is a *pair of penalization functions* if $\Psi_i \in C^1(E^2)$ and if $x = y$ if and only if $\Psi_i(x, y) = 0$ for all i.

Definition C.2. (Continuity estimate) Let $\mathcal{G} : E \times \mathbb{R}^d \times \Theta \to \mathbb{R}$, $(x, p, \theta) \mapsto \mathcal{G}(x, p, \theta)$ be a function and $\{\Psi_1, \Psi_2\}$ be a pair of penalization functions. Suppose that for each $\varepsilon > 0$ there is a sequence $\alpha_2 \to \infty$. As before, we suppress the dependence on ε. Suppose that for each ε and α_2 , there is a sequence $\alpha_1 \to \infty$. We suppress writing the dependence of the sequence α_1 on ε and α_2. We write $\alpha = (\alpha_1, \alpha_2)$.

Suppose that for each triplet $(\varepsilon, \alpha_1, \alpha_2)$ as above we have variables $(x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha})$ in E^2 and variables $\theta_{\varepsilon,\alpha}$ in Θ. We say that this collection is *fundamental* for $\mathcal{G}$ with respect to $\{\Psi_1, \Psi_2\}$ if:

(C1) For each ε, there are compact sets $K_\varepsilon \subseteq E$ and $\widehat{K}_\varepsilon \subseteq \Theta$ such that for all α we have $x_{\varepsilon,\alpha}, y_{\varepsilon,\alpha} \in K_\varepsilon$ and $\theta_{\varepsilon,\alpha} \in \widehat{K}_\varepsilon$.
(C2) For each $\varepsilon > 0$ and α_2 there are limit points $x_{\varepsilon,\alpha_2}, y_{\varepsilon,\alpha_2} \in K_\varepsilon$ of $x_{\varepsilon,\alpha}$ and $y_{\varepsilon,\alpha}$ as $\alpha_1 = \alpha_1(\varepsilon, \alpha_2) \to \infty$. For each ε there are limit points $x_\varepsilon, y_\varepsilon$ in K_ε of x_{ε,α_2} and y_{ε,α_2} as $\alpha_2 \to \infty$. We furthermore have

$$\Psi_1(x_{\varepsilon,\alpha_2}, y_{\varepsilon,\alpha_2}) = 0 \qquad\qquad \forall\, \varepsilon > 0, \forall\, \alpha_2,$$
$$\Psi_1(x_\varepsilon, y_\varepsilon) + \Psi_2(x_\varepsilon, y_\varepsilon) = 0, \qquad\qquad \forall\, \varepsilon > 0,$$
$$\lim_{\alpha_1 \to \infty} \alpha_1 \Psi_1(x_{\varepsilon,\alpha_1,\alpha_2}, x_{\varepsilon,\alpha_1,\alpha_2}) = 0, \qquad\qquad \forall\, \varepsilon > 0, \forall\, \alpha_2,$$
$$\lim_{\alpha_2 \to \infty} \alpha_2 \Psi_2(x_{\varepsilon,\alpha_1}, x_{\varepsilon,\alpha_1}) = 0, \qquad\qquad \forall\, \varepsilon > 0.$$

(C3) We have

$$\sup_{\alpha_2} \sup_{\alpha_1} \mathcal{G}\left(y_{\varepsilon,\alpha}, -\sum_{i=1}^{2} \alpha_i (\nabla \Psi_i(x_{\varepsilon,\alpha}, \cdot))(y_{\varepsilon,\alpha}), \theta_{\varepsilon,\alpha} \right) < \infty, \qquad\qquad \text{(C.1)}$$

$$\inf_{\alpha_2} \inf_{\alpha_1} \mathcal{G}\left(x_{\varepsilon,\alpha}, \sum_{i=1}^{2} \alpha_i (\nabla \Psi_i(\cdot, y_{\varepsilon,\alpha}))(y_{\varepsilon,\alpha}), \theta_{\varepsilon,\alpha} \right) > -\infty. \qquad\qquad \text{(C.2)}$$

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

We say that $\mathcal{G}$ satisfies the *continuity estimate* if for every fundamental collection of variables we have for each $\varepsilon > 0$ that

$$\liminf_{\alpha_2 \to \infty} \liminf_{\alpha_1 \to \infty} \mathcal{G}\left(x_{\varepsilon,\alpha}, \sum_{i=1}^{2} \alpha_i \nabla \Psi_i(\cdot, y_{\varepsilon,\alpha})(x_{\varepsilon,\alpha}), \theta_{\varepsilon,\alpha} \right)$$
$$- \mathcal{G}\left(y_{\varepsilon,\alpha}, -\sum_{i=1}^{2} \alpha_i \nabla \Psi_i(x_{\varepsilon,\alpha}, \cdot)(y_{\varepsilon,\alpha}), \theta_{\varepsilon,\alpha} \right) \leq 0. \qquad\qquad \text{(C.3)}$$

D. Differential inclusions

To establish that Condition 8.11 of [19] is satisfied in the proof of Theorem 2.8, we need to solve a differential inclusion. The following appendix is based on [11,28] and is a copy of the one in [25]. 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 $F(x) \subseteq \mathbb{R}^d$, $F(x) \neq \emptyset$.

Definition D.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 γ 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 D.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 $\mathcal{U}$ of $F(x)$, there is a neighbourhood $\mathcal{V}$ of x in D such that $F(\mathcal{V}) \subseteq \mathcal{U}$. 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 $x_n \to x$ in D and $\xi_n \in F(x_n)$ such that $\xi_n \to \xi$ we have $\xi \in F(x)$.

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

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

The set $T_D(x)$ is sometimes called the the *Bouligand cotingent cone*.

Lemma D.4. (Proposition 4.1 in [11]) *Let $D \subseteq \mathbb{R}^d$ be a closed, convex, non-empty set. Then the set $T_D(x)$ is convex and contains 0.*

Lemma D.5. (Theorem 2.2.1 in [28], Lemma 5.1 in [11]) *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 $\|F(x)\| = \sup \{|z| \,\big|\, z \in F(x)\} \leq c(1 + |x|)$ for all $x \in D$.*

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

E. Pseudo-coercivity for exponential Hamiltonians

In this section we consider the notions of pseudo-coercivity and the continuity estimate for the Hamiltonian-Jacobi equation $f - \Lambda f = h$ on $E = [0, \infty)$, with $h \in C_b(E)$ for the Hamiltonian $\Lambda f(x) = \Lambda(x, f'(x))$ with

$$\Lambda(x, p) = x \left[e^{-p} - 1 \right] \tag{E.1}$$

which is a simplified version of the Hamiltonian given in the introduction and Proposition 5.7.

The pseudo-coercivity estimate of [4, Pages 34 and 35] translates into the present context as

$$|\Lambda(x, p) - \Lambda(y, p)| \leq m \left(|x - y|(1 + |p|) \right) Q(x, y, p) \tag{E.2}$$

where $Q(x, y, p) = \max \left(\Psi(H(x, p)), \Psi(H(y, p)) \right)$, where $m : [0, \infty) \to [0, \infty)$ is such that $\lim_{t \downarrow 0} m(t) = 0$ and $\Psi : \mathbb{R} \to [0, \infty)$ is continuous.

We first make some general remarks on the relation between pseudo-coercivity and the continuity estimate. We then show that for E.1, pseudo-coercivity fails, whereas the continuity estimate holds (in the case of a single θ).

The continuity estimate (for a single θ) is more general than pseudo-coercivity in the sense that:

- It does not rely on the fact that you use a multiple of $|x-y|^2$ as a penalization in the comparison principle. This is of importance for Hamiltonians Λ on e.g. the set of probability measures $\mathcal{P}(\{1, \ldots, q\})$ with $q \in \{3, 4, \ldots\}$ like in Proposition 5.18.
- It removes the necessity of taking absolute values in the estimate. This last fact is important as can be seen for the Hamilton–Jacobi equation for $f - \lambda \Lambda f = h$, $h \in C_b$, $\lambda > 0$ for the Hamiltonian $\Lambda f(x) = \Lambda(x, f'(x))$ with

$$\Lambda(x, p) = x \left[e^{-p} - 1 \right], \qquad x \in [0, \infty), p \in \mathbb{R},$$

for which the comparison principle holds, whereas for $\widetilde{\Lambda}(x, p) := \Lambda(x, -p)$ it fails. Our continuity estimate holds for Λ, but not for $\widetilde{\Lambda}$. Pseudo-coercivity fails for both as we explain below.

We first show that pseudo-coercivity fails for the Hamilton–Jacobi equation in terms of Λ of (E.1).

Lemma E.1. Λ *is not pseudo-coercive.*

Proof. A counterexample suffices.

Let W be the Lambert function. That is, W is the inverse of ϕ where $\phi(x) = xe^x$. Next, let $x_\alpha = 0$, $y_\alpha = \alpha^{-1} W(\alpha)$. Then we have $y_\alpha \to 0$, $\alpha y_\alpha \to \infty$, $\alpha y_\alpha^2 = \frac{W(\alpha)^2}{\alpha} \to 0$ and

$$y_\alpha e^{\alpha y_\alpha} = 1.$$

Thus, with $p_\alpha = \alpha(x_\alpha - y_\alpha) = -\alpha y_\alpha$,

$$\Lambda(x_\alpha, p_\alpha) - \Lambda(y_\alpha, p_\alpha) = y_\alpha - 1 \to -1 \tag{E.3}$$

contradicting (E.2).   □

    Note that the comparison principle for $f - \lambda \Lambda f = h$ does in fact hold. This is due to the fact that one only needs to establish that $\liminf_{\alpha \to \infty} \Lambda(x_\alpha, p_\alpha) - \Lambda(y_\alpha, p_\alpha) \leq 0$ for appropriately chosen $x_\alpha, y_\alpha, p_\alpha$ without absolute value signs, see Proposition 5.17. The removal of absolute value signs is essential: the comparison principle for $f - \widetilde{\Lambda} f = h$ for

$$\widetilde{\Lambda}(x, p) = x \left[e^p - 1 \right]$$

fails. This is related to the statement that an associated large deviation principle fails, see Example E of [33].

References

[1] Aliprantis, C.D., Border, K.C.: Infinite Dimensional Analysis. A Hitchhiker's Guide, 3rd edn. Springer, Berlin (2006)

[2] Bardi, M., Capuzzo-Dolcetta, I.: Optimal Control and Viscosity Solutions of Hamilton–Jacobi–Bellman equations. Birkhäuser, Boston (1997)

[3] Bardi, M., Cesaroni, A., Ghilli, D.: Large deviations for some fast stochastic volatility models by viscosity methods. Discrete Contin. Dyn. Syst. **35**(9), 3965–3988 (2015)

[4] Barles, G.: Solutions de viscosité des équations de Hamilton–Jacobi. Mathématiques & Applications (Berlin) [Mathematics & Applications], Vol. 17. Springer, Paris (1994)

[5] Budhiraja, A., Dupuis, P.: Analysis and Approximation of Rare Events: Representations and Weak Convergence Methods, volume 94 of Probability Theory and Stochastic Modelling. Springer, Berlin (2019)

[6] Budhiraja, A., Dupuis, P., Ganguly, A.: Large deviations for small noise diffusions in a fast Markovian environment. Electron. J. Probab. **23**, 33 (2018)

[7] Collet, F., Kraaij, R.C.: Dynamical moderate deviations for the Curie–Weiss model. Stoch. Process. Appl. **127**(9), 2900–2925 (2017)

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

[9] Cutrì, A., Da Lio, F.: Comparison and existence results for evolutive non-coercive first-order Hamilton–Jacobi equations. ESAIM Control Optim. Calc. Var. **13**(3), 484–502 (2007)

[10] Da Lio, F., Ley, O.: Convex Hamilton–Jacobi equations under superlinear growth conditions on data. Appl. Math. Optim. **63**(3), 309–339 (2011)

[11] Deimling, K.: Multivalued Differential Equations. De Gruyter Series in Nonlinear Analysis and Applications, vol. 1. Walter de Gruyter & Co., Berlin (1992)

[12] Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications, 2nd edn. Springer, Berlin (1998)

[13] Den Hollander, F.: Large Deviations, vol. 14. American Mathematical Society, New York (2008)

[14] Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time, i. Commun. Pure Appl. Math. **28**(1), 1–47 (1975)

[15] Donsker, M.D., Varadhan, S.R.S.: On a variational formula for the principal eigenvalue for operators with maximum principle. Proc. Natl. Acad. Sci. **72**(3), 780–783 (1975)

[16] Dupuis, P., Ishii, H., Soner, H.M.: A viscosity solution approach to the asymptotic analysis of queueing systems. Ann. Probab. **18**(1), 226–255 (1990)

[17] Evans, L.C., Souganidis, P.E.: A PDE approach to certain large deviation problems for systems of parabolic equations. Ann. Inst. H. Poincaré Anal. Non Linéaire, 6(suppl.):229–258 (1989). Analyse non linéaire (Perpignan, 1987)

[18] Feng, J., Fouque, J.-P., Kumar, R.: Small-time asymptotics for fast mean-reverting stochastic volatility models. Ann. Appl. Probab. **22**(4), 1541–1575 (2012)

[19] Feng, J., Kurtz, T.G.: Large Deviations for Stochastic Processes. American Mathematical Society, Philadelphia (2006)

[20] Fleming, W.H., Souganidis, P.E.: PDE-viscosity solution approach to some problems of large deviations. Annali della Scuola Normale Superiore di Pisa - Classe di Scienze, Ser. 4 **13**(2), 171–192 (1986)

[21] Ghilli, D.: Viscosity methods for large deviations estimates of multiscale stochastic processes. ESAIM COCV **24**(2), 605–637 (2018)

[22] Hiriart-Urruty, J.-B., Lemaréchal, C.: Fundamentals of Convex Analysis. Grundlehren Text Editions. Springer, Berlin (2001). Abridged version of ıt Convex analysis and minimization algorithms. I [Springer, Berlin, 1993; MR1261420 (95m:90001)] and ıt II [ibid.; MR1295240 (95m:90002)]

[23] Kraaij, R.C.: Flux large deviations of weakly interacting jump processes via well-posedness of an associated Hamilton–Jacobi equation. Bernoulli **(to appear)** (2017)

[24] Kraaij, R.C.: A general convergence result for viscosity solutions of Hamilton–Jacobi equations and non-linear semigroups. J. Funct. Anal. **(to appear)** (2019)

[25] Kraaij, R.C., Mahé, L.: Well-posedness of Hamilton–Jacobi equations in population dynamics and applications to large deviations. Stoch. Process. Appl. (2020)

[26] Kraaij, R.C., Schlottke, M.C.: A large deviation principle for Markovian slow-fast systems. preprint; arXiv:2011.05686

[27] Kumar, R., Popovic, L.: Large deviations for multi-scale jump-diffusion processes. Stoch. Process. Appl. **127**(4), 1297–1320 (2017)

[28] Kunze, M.: Non-Smooth Dynamical Systems. Number nr. 1744 in Lecture Notes in Mathematics. Springer, Berlin (2000)

[29] Peletier, M.A., Schlottke, M.C.: Large-deviation principles of switching Markov processes via Hamilton–Jacobi equations. preprint; ArXiv:1901.08478 (2019)

[30] Pinsky, R.G.: The *I*-function for diffusion processes with boundaries. Ann. Probab. **13**(3), 676–692 (1985)

[31] Pinsky, R.G.: On evaluating the Donsker–Varadhan I-function. Ann. Probab. **66**, 342–362 (1985)

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

[33] Shwartz, A., Weiss, A.: Large deviations with diminishing rates. Math. Oper. Res. **30**(2), 281–310 (2005)

Richard C. Kraaij
Delft Institute of Applied Mathematics
Delft University of Technology
Mekelweg 4
2628CD Delft
The Netherlands
e-mail: r.c.kraaij@tudelft.nl

Mikola C. Schlottke
Department of Mathematics and Computer Science
Eindhoven University of Technology
Eindhoven
The Netherlands
e-mail: mikola.schlottke@outlook.com