

Algorithm for overcoming the curse of dimensionality for state-dependent Hamilton-Jacobi equations

Yat Tin Chow^{a,*}, Jérôme Darbon^{b,2}, Stanley Osher^{c,3}, Wotao Yin^{c,3}

^a Department of Mathematics, UCR, Riverside, CA 92521, USA

^b Division of Applied Mathematics, Brown University, Providence RI, 02912, USA

^c Department of Mathematics, UCLA, Los Angeles, CA 90095-1555, USA

ARTICLE INFO

Article history:

Received 8 April 2017

Received in revised form 15 September 2018

Accepted 25 January 2019

Available online 5 March 2019

Keywords:

Hamilton-Jacobi equations

Viscosity solution

Hopf-Lax formula

Nonconvex Hamiltonian

Differential games

Optimal control

ABSTRACT

In this paper, we develop algorithms to overcome the curse of dimensionality in non-convex state-dependent Hamilton-Jacobi partial differential equations (HJ PDEs) arising from optimal control and differential game problems. The subproblems are independent and they can be implemented in an embarrassingly parallel fashion. This is ideal for perfect scaling in parallel computing. The algorithm is proposed to overcome the curse of dimensionality [1,2] when solving HJ PDE. The major contribution of the paper is to change either the solving of a PDE problem or an optimization problem over a space of curves to an optimization problem of a single vector, which goes beyond the work of [40]. We extend the method in [7,9,15], and conjecture a (Lax-type) minimization principle to solve *state-dependent* HJ PDE when the Hamiltonian is convex, as well as a (Hopf-type) maximization principle to solve *state-dependent* HJ PDE when the Hamiltonian is *non-convex*, as a generalization of the well-known Hopf formula in [18,25,50]. We showed the validity of the formula under restricted assumption for the sake of completeness, and would like to bring our readers to [62] which validates our conjectures in a more general setting. We conjectured the weakest assumption of our formula to hold is a pseudoconvexity assumption similar to one stated in [50]. Our method is expected to have application in control theory, differential game problems and elsewhere.

© 2019 Published by Elsevier Inc.

1. Introduction

Hamilton-Jacobi-Isaacs partial differential equations (HJ PDE) are crucial in analyzing continuous/differential dynamic games, control theory problems, and dynamical systems coming from the physical world, e.g. [20]. An important application is to compute the evolution of geometric objects [42], which was first used for reachability problems in [33,34], to our knowledge.

Numerical solutions to HJ PDE have attracted a lot of attention. Most of the methods involve the introduction of a grid and a finite difference discretization of the Hamiltonian. Some of these well-known methods using discretization include

* Corresponding author.

E-mail addresses: yattinc@ucr.edu (Y.T. Chow), jerome_darbon@brown.edu (J. Darbon), sjo@math.ucla.edu (S. Osher), wotaoyin@math.ucla.edu (W. Yin).

¹ Research supported by ONR N000141712162 and NSF DMS-1720237.

² Research supported by NSF DMS-1820821.

³ Research supported by ONR grants N000141410683, N000141210838, N000141712162 and DOE grant DE-SC00183838.

ENO/WENO-type methods [41] and Dijkstra-type [16] methods such as fast marching [59] and fast sweeping [58]. However, with their discretization nature, these numerical approaches of HJ PDE suffer from poor scaling with respect to dimension d , hence rendering them impossible to be applied to problems in high dimensions.

Research has therefore been conducted by several groups in search of possible algorithms that can scale reasonably with dimension. Some new algorithms are introduced in e.g. [14,27,28]. In [7,9,15], the authors proposed a causality-free method for solving possibly non-convex and time dependent HJ PDE based on the generalized Hopf-Lax formula. Using the Hopf-Lax formula, the PDE becomes decoupled and the solution at each point can be effectively calculated by d -dimensional minimization, with d the space dimension.

In this work, we propose to extend the method in [7,9,15]. The major contribution of the paper is to change either the solving of a PDE problem or an optimization problem over a space of curves to an optimization problem of a single vector. We conjecture the (Lax-type) minimization principle to solve *state-dependent* HJ PDE when the Hamiltonian is convex. We also conjecture a (Hopf-type) maximization principle to solve *state-dependent* HJ PDE when the Hamiltonian is *non-convex* but when the initial data is *convex*. In particular the conjectured (Hopf-type) maximization principle is a generalization of the well-known Hopf formula in [18,25,50]. We validated our conjectures under restricted assumptions and refers to [62] for the validation of our formula under a less restricted setting. The optimization problems are of the same dimension as the dimensions of the HJ PDE. A coordinate descent method is suggested for the minimization procedure in the generalized Lax/Hopf formula, and numerical differentiation is used to compute the derivatives. This method is preferable since the evaluation of the function value itself requires some computational effort, especially when we handle higher dimensional optimization problems. The use of multiple initial guesses is suggested to overcome possibly multiple local extrema since the optimization process is no longer convex. A simple numerical ODE solver is used to compute the bi-characteristics in the Hamiltonian system. A numerical quadrature rule is used for computing integrals with respect to time inside the minimization or maximization principles. Coordinate descent is also used and a numerical differentiation is performed to minimize the number of calculation procedures in each iteration steps. In this paper we illustrate the practicality of our method using the simplest ODE solvers, quadrature rules and finite difference methods, namely the forward Euler, rectangular rule, and forward difference methods. Nonetheless, all these components of the optimization can be improved by using better numerical methods, e.g. pseudo-spectral methods for ODE, Gauss-Lobatto quadrature rules, etc. These choices, together with choosing an appropriate mesh-size in time, minimize errors effectively. Our method does not use a mesh in space, and solutions can be computed at each point (x, t) in a totally decoupled and embarrassingly parallel manner. Moreover, the solution to the HJ PDE is evaluated at each point (x, t) by coordinate minimization described by the minimization or maximization principles at that point.

As for high dimensional control, we would like to compare our work with [40], which is concerned with discrete approximations of a particular set of control, namely the discrete linear-convex control problems involving a quadratic control. In the paper, both the final parameter v as well as the whole curve representing controls are unknowns. The major contribution of our paper are to remove the optimization over the control (getting the HJ PDE as in [7,9,15]), as well as extending our method in [7,9,15] to a much more general setting, i.e. differential games/nonconvex problems and the Hamilton-Jacobi limit. Also, our algorithm is faster comparing with [40] even at the discrete level because we remove the optimization over the control. We have also proposed both the Lax and the Hopf version, while [40] focuses only on the analogue of a Lax type formula. We would like to remark that by considering and working on the dual, the algorithm can be made much faster in many cases, especially when the initial data is convex.

Our formal statements of the conjectures that are used for computation will be given in section 3. However, before we give an exact formulation of our conjectures, let us briefly provide our main conjectures. In what follows, we denote p as the co-state variable and $H(x, p, t)$ as the Hamiltonian, as well as $\varphi(x, t)$ as the value function satisfying (2.1)–(2.2). We also denote $(\gamma(t), p(t))$ as the bi-characteristic curve in the phase space that shall satisfy the constraints in the following formulae. Then we conjecture the followings:

1. **Minimization principle** (Lax Formula) Assume $H(x, p, t) \in C^2$ and is convex w.r.t. p , and (A5) is satisfied, then there exists t_0 such that the viscosity solution to (2.1)–(2.2) can be represented as for $t \leq t_0$ that:

$$\varphi(x, t) = \min_{v \in \mathbb{R}^d} \left\{ g(\gamma(x, v, 0)) + \int_0^t \left\{ \langle p(x, v, s), \partial_p H(\gamma(x, v, s), p(x, v, s), s) \rangle - H(\gamma(x, v, s), p(x, v, s), s) \right\} ds : \right. \\ \left. \begin{aligned} \dot{\gamma}(x, v, s) &= \partial_p H(\gamma(x, v, s), p(x, v, s), s), \\ \dot{p}(x, v, s) &= -\partial_x H(\gamma(x, v, s), p(x, v, s), s), \\ \gamma(x, v, t) &= x, \quad p(x, v, t) = v \end{aligned} \right\} \quad (1.1)$$

and its discrete approximation given a small δ ,

$$\varphi(x, t) \approx \min_{v \in \mathbb{R}^d} \left\{ g(x_0(x, v)) + \delta \sum_{n=1}^{N-1} \left\{ \langle p_n(x, v), \partial_p H(x_n(x, v), p_n(x, v), t_n) \rangle - H(x_n(x, v), p_n(x, v), t_n) \right\} : \right.$$

$$\left. \begin{aligned} x_{n+1}(x, v) - x_n(x, v) &= \delta \partial_p H(x_n(x, v), p_n(x, v), t_n), \\ p_{n+1}(x, v) - p_n(x, v) &= \delta \partial_x H(x_n(x, v), p_n(x, v), t_n), \\ x_N &= x, p_N = v \end{aligned} \right\}$$

We would like to remark that the bi-characteristics $(\gamma(x, v, \cdot), p(x, v, \cdot))$ depend on the two initial conditions x and v , and this dependence is emphasized using notations that include the two independent variables.

2. **Maximization principle** (Hopf Formula) Assume $H(x, p, t) \in C^2$ and $g(p) \in C^2$ is convex w.r.t. p that satisfies (A5). Assume the pseudoconvex condition stated in Conjecture 3.12 holds, then there exists t_0 such that the viscosity solution to (2.1)–(2.2) can be represented as, for $t < t_0$, that:

$$\begin{aligned} \varphi(x, t) &= \max_{v \in \mathbb{R}^d} \left\{ \langle x, v \rangle - g^*(p(x, v, 0)) - \int_0^t \left\{ H(\gamma(x, v, s), p(x, v, s), s) - \langle \partial_x H(\gamma(x, v, s), p(x, v, s), s), \gamma(x, v, s) \rangle \right\} ds : \right. \\ &\quad \left. \begin{aligned} \dot{\gamma}(x, v, s) &= \partial_p H(\gamma(x, v, s), p(x, v, s), s), \\ \dot{p}(x, v, s) &= -\partial_x H(\gamma(x, v, s), p(x, v, s), s), \\ \gamma(x, v, t) &= x, p(x, v, t) = v \end{aligned} \right\} \end{aligned} \quad (1.2)$$

and its discrete approximation given a small δ

$$\begin{aligned} \varphi(x, t) &\approx \max_{v \in \mathbb{R}^d} \left\{ \langle x_N, v_N \rangle - g^*(p_1(x, v)) - \delta \sum_{n=1}^N H(x_n(x, v), p_n(x, v), t_n) + \delta \sum_{n=1}^{N-1} \langle x_n(x, v), \partial_x H(x_n(x, v), p_n(x, v), t_n) \rangle : \right. \\ &\quad \left. \begin{aligned} x_n(x, v) - x_{n-1}(x, v) &= \delta \partial_p H(x_n(x, v), p_n(x, v), t_n), \\ p_n(x, v) - p_{n+1}(x, v) &= \delta \partial_x H(x_n(x, v), p_n(x, v), t_n), \\ x_N &= x, p_N = v \end{aligned} \right\} \end{aligned}$$

and in both cases the argument v attaining the minimum or the maximum in the respective formula is $\partial_x \varphi(x, t)$ when φ is smooth at (x, t) . These are the key conjectures from a practical point of view. When $H(x, p, t)$ is non-smooth (e.g. in the case when $H(x, p, t)$ is homogeneous of degree 1 w.r.t. p), more general forms of the respective conjectures are also available. We must emphasize that these modifications of the formulas are **necessary** when H is non-smooth or otherwise some part of sub-gradient flow will be missed and the formula will be incorrect. We notice that, in particular, our conjectured (Hopf-type) maximization principle is a generalization of the well-known Hopf formula in [18,25,50].

Our approach of using the minimization or maximization principle goes between the **indirect method** (Pontryagin's maximum principle) and the **direct method** (direct optimization over the spaces of curves) and is optimal for computational efficiency. This way we are able to minimize the number of variables to be optimized; since a discretization of curve needs a lot of variables, meanwhile, we keep a variable and the functional to maintain a descent algorithm that guarantee convergence to a local minimum. The correctness of the computed limit (i.e. if it is a global minimum) can be checked by the condition $p(0) \in \partial g(\gamma(0))$.

The rest of our paper is organized as follows: in subsection 2.1 we introduce the general class of HJ-PDE that we are interested in, and then we briefly explain the connection between HJ-PDE and differential games subsection 2.2. Then in section 3, first in subsection 3.2, we showed the formula are true in a restricted set of assumptions; and then we discuss the major formulae that we use in our work: the conjectured minimization principle (Lax formulation) in subsection 3.3 and the conjectured maximization principle (Hopf formulation) in subsection 3.4. We then go on to explain briefly our numerical techniques in section 4 and the certificate of correctness. Numerical examples are given in section 5.

We would like to bring the readers to notice that after a previous version of our paper in arXiv is out, [62] showed that our conjectures hold in a more general setting than we did. However for the sake of completeness, we keep the proof with restricted assumptions, and we refer our readers to the more general case shown in [62].

2. Review of Hamilton-Jacobi equations and differential games

2.1. Hamilton-Jacobi equations

In this work, we are concerned with an approximation scheme for solving the following HJ PDE:

$$\frac{\partial}{\partial t} \varphi(x, t) + H(x, \nabla_x \varphi(x, t), t) = 0 \quad \text{in } \mathbb{R}^d \times (0, \infty), \quad (2.1)$$

where $H : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$ is a continuous Hamiltonian function bounded from below by an affine function, $\frac{\partial}{\partial t} \varphi$ and $\nabla_x \varphi$ respectively denote the partial derivatives with respect to t and the gradient vector with respect to x of the function $\varphi : \mathbb{R}^d \times (0, \infty) \rightarrow \mathbb{R}$. We are also given the initial data

$$\varphi(x, 0) = g(x) \quad \text{in } \mathbb{R}^d. \quad (2.2)$$

We aim to compute the viscosity solution [12,13] at a given point $x \in \mathbb{R}^d$ and time $t \in (0, \infty)$.

Viscosity solution to (2.1)–(2.2) is explicitly given by the Hopf-Lax formulae when H is x independent, since the integral curves of the Hamiltonian vector field (i.e., bi-characteristics in the phase space) are straight lines when projected to x -space. In general, when H is x -dependent, we have shown in section 3.2 that, under restricted assumptions, a minimization or maximization principle approximates the solution to the HJ PDE. They are formulated using a KKT condition which gives bi-characteristics in the phase (i.e. follow the flow along the Hamiltonian vector field generated by H). We then postulate in section 3 that the minimization and maximization principles may hold true for more general cases. In [62] the authors verified our main conjectures under a more relaxed set assumptions using the notions of minimax, in both the Hamilton-Jacobi-Bellman situation and in some special cases of Hamilton-Jacobi-Issac situation (i.e. when the Hamiltonian is non-convex). This is a first and monumental work which suggests the validity of our conjecture in a more general situation.

In fact, another closely-related theory, namely the minimax viscosity solution, is introduced in [32,54,55], and connections are made to control problems and differential games can be found in [30,31,53,56]. The theory goes in parallel with the definition and theory of viscosity solution [20] coming from differential game, and in the case with convex Hamiltonian, they always coincide. A general notion of minimax viscosity solution (with non-convex but smooth Hamiltonian) helps to recast the solution using a formulation that involves patching the graph of multi-valued geometric solutions in a correct manner, e.g. in [6,5]. In the convex case, under further assumption, the solution can be formulated as a mini-max saddle point problem of a functional over the spaces of curves [5]. With convex Hamiltonian, the minimax solution is actually also known as the principle of least action, e.g. [5,48]. A slightly different but very similar direction, which uses Wentzell-Freidlin theory and construct minimum action methods, are also developed, as in e.g. [17,22,23,60]. However, it is known that the general formulation of minimax solution coincide with the viscosity solution only if the dynamically programming principle (i.e. the semigroup property) is satisfied for the minimax solution [4,61]. In fact, in the work [61], the author shows further that the viscosity solution can be constructed by a limiting sequence of minimax solutions, but such a limiting process may defy the purpose of formulating the solution of the viscosity solution as a calculable process in the form of an optimization. Besides, in full generality e.g. in [6,5] the generating function is assumed to be quadratic at infinity and they are described using deformation retract and relative cohomology (in the spirit of Morse theory), and this may bring difficulty for computational purpose. For more exposition of the minimax solution and development of this relationship with control problem, we refer the readers to [62].

On the other hand, there is a well-known notion of Pontryagin's maximum principle [45] giving us an optimality condition for the control. In particular, work e.g. in [10] links the maximum principle and dynamic programming, which discussed in the non-smooth case when will the costate variable coincide with a partial generalized gradient of the value function (cf. [10]) for a clear definition. However the optimality condition has a final condition for the state variable and an initial condition for the co-state variable. We hope to avoid forward-backward iteration to compute for an optimal. The motivation of our conjecture over this KKT condition is to keep a Lyapunov functional such that a numerical algorithm can minimize, and therefore guarantee descent of a functional and convergence of a subsequence of an algorithm even with more pathological behaviors of H and g , e.g. non-convexity, etc.

The purpose of our conjecture is to find an appropriate generalization of Hopf formula such that it still involves a d dimensional optimization problem, thus one may apply a good optimization algorithm for computing the viscosity solution. We are not aiming at full generality; as one will need more sophisticated mathematical language while the resulting formulation may not be easy for computational purpose. On the other hand, we would also prefer to obtain generality that is sensible and not restricted to a restricted class of Hamiltonian $H(x, p, t)$ (convex w.r.t. p , concave w.r.t. x and satisfying a finite concavity-convexity assumption, cf. Assumptions (A) and Theorem 2.3 in [48]) (see also Theorem 4.8 in [49]). Since it is unclear how far such conditions and notions can be generalized as well as remained computable, we only show in a restricted case that our formulas hold in subsection 3.2. We leave a more general version as a conjecture.

2.2. Differential games and its connection with non-convex Hamilton-Jacobi equations

In this subsection, we give a brief introduction to the specific optimal control and differential game problems we are considering, and a brief explanation as to how we recast them as problems of solving HJ PDE's. We follow discussions in [7,9,15,20], see also [18], about optimal control and also for differential games, and their links with HJ PDE.

To start with, we first consider two convex compact sets C and D , in which control parameters lie. Then let us denote $\mathcal{A} = \{a : [t, T] \rightarrow C : a \text{ is measurable}\}$, which is referred to as the admissible set of Player I; and $\mathcal{B} = \{b : [t, T] \rightarrow D : b \text{ is measurable}\}$, which is referred to as the admissible set of Player II. We call the measurable functions $a : [t, T] \rightarrow C$ in the set $\mathcal{A}$ and the function $b : [t, T] \rightarrow D$ in the set $\mathcal{B}$ as *controls* performed by Players I and II respectively.

We start with a system of differential equations given as follows. Fix $0 \leq t < T$, $x \in \mathbb{R}^d$. We consider

$$\begin{cases} \frac{dx}{ds}(s) = f(s, x(s), a(s), b(s)) & t \leq s \leq T, \\ x(t) = x. \end{cases} \quad (2.3)$$

We assume that the function

$$f : [0, T] \times \mathbb{R}^d \times A \times B \rightarrow \mathbb{R}^m$$

is uniformly continuous and

$$\begin{cases} |f(t, x, a, b)| \leq C_1 \\ |f(t, x, a, b) - f(t, y, a, b)| \leq C_1 |x - y|, \end{cases}$$

for some constant C_1 and for all $0 \leq t \leq T$, $x, y \in \mathbb{R}^m$, $a \in A$, $b \in B$. The unique solution to (2.3) is called the response of the controls $a(\cdot), b(\cdot)$. Then we introduce the *payoff* functional for a given pair of (x, t) :

$$P(a, b) := P_{t,x}(a(\cdot), b(\cdot)) := \int_t^T h(s, x(s), a(s), b(s)) ds + g(x(T)),$$

where $g : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfies

$$\begin{cases} |g(x)| \leq C_2 \\ |g(x) - g(y)| \leq C_2 |x - y|, \end{cases}$$

and h satisfies

$$\begin{cases} |h(t, x, a, b)| \leq C_3 \\ |h(t, x, a, b) - h(t, y, a, b)| \leq C_3 |x - y|, \end{cases}$$

for some constants C_2, C_3 and all $0 \leq t \leq T$, $x, y \in \mathbb{R}^m$, $a \in A$, $b \in B$. In a differential game, the goal of player I is to maximize the functional P by choosing his control a whereas that of player II is to minimize P by choosing his control b .

Now we are ready to define the lower and upper values of the differential game, based on the notation introduced above. We first define the two sets containing the respective controls of players I and II:

$$M(t) := \{a : [t, T] \rightarrow A : a \text{ is measurable}\},$$

$$N(t) := \{b : [t, T] \rightarrow B : b \text{ is measurable}\}.$$

Define a strategy for player I as the map

$$\alpha : N(t) \rightarrow M(t)$$

for each $t \leq s \leq T$ and $b, \hat{b} \in B$ such that

$$b(\tau) = \hat{b}(\tau) \text{ for a.e. } t \leq \tau \leq s \Rightarrow \alpha[b](\tau) = \alpha[\hat{b}](\tau) \text{ for a.e. } t \leq \tau \leq s.$$

Therefore a strategy for player I $\alpha[b]$ is the control of player I given that of player II as b . Similarly, let us define a strategy for player II as

$$\beta : M(t) \rightarrow N(t)$$

for each $t \leq s \leq T$ and $a, \hat{a} \in A$ such that

$$a(\tau) = \hat{a}(\tau) \text{ for a.e. } t \leq \tau \leq s \Rightarrow \beta[a](\tau) = \beta[\hat{a}](\tau) \text{ for a.e. } t \leq \tau \leq s.$$

Again a strategy for player II $\beta[a]$ is the control of player II given that of player I as a .

Now let $\Gamma(t)$ denote the set of all strategies for I and $\Delta(t)$ for II beginning at time t . We are well equipped to define the upper and lower values of the differential game. The lower value $V(x, t)$ is defined as

$$\begin{aligned} V(x, t) &:= \inf_{\beta \in \Delta(t)} \sup_{a \in M(t)} P_{t,x}(a, \beta[a]) \\ &:= \inf_{\beta \in \Delta(t)} \sup_{a \in M(t)} \left\{ \int_t^T h(s, x(s), a(s), \beta[a](s)) ds + g(x(T)) \right\}, \end{aligned}$$

where $x(\cdot)$ solves (2.3) for a given pair of (x, t) . Likewise, the upper value $U(x, t)$ is defined as

$$\begin{aligned} U(x, t) &:= \sup_{\alpha \in \Gamma(t)} \inf_{b \in N(t)} P_{t,x}(\alpha[b], b) \\ &:= \sup_{\alpha \in \Gamma(t)} \inf_{b \in N(t)} \left\{ \int_t^T h(s, x(s), \alpha[b](s), b(s)) ds + g(x(T)) \right\}, \end{aligned}$$

where $x(\cdot)$ again solves (2.3) for a given pair of (x, t) .

In fact, derived from dynamic programming optimality conditions in [20], the lower and upper values V and U are the viscosity solutions of a certain possibly nonconvex HJ PDE. For the sake of exposition, we first define the following two Hamiltonians:

$$\begin{aligned}\tilde{H}^+(x, p, t) &= \max_{b \in B} \min_{a \in A} \{-\langle f(t, x, a, b), p \rangle - h(t, x, a, b)\}, \\ \tilde{H}^-(x, p, t) &= \min_{a \in A} \max_{b \in B} \{-\langle f(t, x, a, b), p \rangle - h(t, x, a, b)\}.\end{aligned}$$

A very important case of this class of Hamiltonian is when $\tilde{H}^\pm(t, x, p)$ are homogeneous of degree 1, which is highlighted in this work. In fact, in the case where

$$\begin{aligned}f(t, x, a, b) &= a - b \\ h(t, x, a, b) &= -\mathcal{I}_{c_1(x, t)A(x, t)}(a) + \mathcal{I}_{c_2(x, t)B(x, t)}(b),\end{aligned}$$

where $\mathcal{I}_\Omega$ is the indicator functions of the sets Ω , i.e.

$$\mathcal{I}_\Omega(x) = \begin{cases} 0 & \text{if } x \in \Omega \\ \infty & \text{if } x \notin \Omega \end{cases}$$

and $A(x, t)$ and $B(x, t)$ are balanced, then

$$\begin{aligned}\tilde{H}^+(x, p, t) &= \max_{b \in B} \min_{a \in A} \{\langle b - a, p \rangle + \mathcal{I}_{c_1(x, t)A(x, t)}(a) - \mathcal{I}_{c_2(x, t)B(x, t)}(b)\} \\ &= \max_{b \in c_2(x, t)B(x, t)} \min_{a \in c_1(x, t)A(x, t)} \{\langle b - a, p \rangle\} \\ &= \max_{[c_2(x, t)]^{-1}b \in B(x, t)} \min_{[c_1(x, t)]^{-1}a \in A(x, t)} \{\langle b - a, p \rangle\} \\ &= \max_{b \in B(x, t)} \min_{a \in A(x, t)} \{\langle c_2(x, t)b - c_1(x, t)a, p \rangle\}, \\ \tilde{H}^-(x, p, t) &= \min_{a \in A} \max_{b \in B} \{\langle b - a, p \rangle + \mathcal{I}_{c_1(x, t)A(x, t)}(a) - \mathcal{I}_{c_2(x, t)B(x, t)}(b)\} \\ &= \min_{a \in c_1(x, t)A(x, t)} \max_{b \in c_2(x, t)B(x, t)} \{\langle b - a, p \rangle\} \\ &= \min_{[c_1(x, t)]^{-1}a \in A(x, t)} \max_{[c_2(x, t)]^{-1}b \in B(x, t)} \{\langle b - a, p \rangle\} \\ &= \min_{a \in A(x, t)} \max_{b \in B(x, t)} \{\langle c_2(x, t)b - c_1(x, t)a, p \rangle\},\end{aligned}$$

thus it holds that $\tilde{H}^+$ and $\tilde{H}^-$ coincide, as well as the following relationship:

$$\begin{aligned}\tilde{H}^\pm(x, p, t) &= \max_{b \in B(x, t)} \min_{a \in A(x, t)} \{c_1(x, t)\langle a, p \rangle - c_2(x, t)\langle b, p \rangle\} \\ &= c_1(x, t) \min_{a \in A(x, t)} \{\langle a, p \rangle\} - c_2(x, t) \max_{b \in B(x, t)} \{\langle b, p \rangle\} \\ &= -c_1(x, t)\mathcal{I}_{A(x, t)}^*(p) + c_2(x, t)\mathcal{I}_{B(x, t)}^*(p).\end{aligned}$$

In this case, $H^\pm(x, p, t)$ can be written as a difference of two positively homogeneous (of degree 1) Hamiltonians $\Phi_1(x, \cdot, t)$, $\Phi_2(x, \cdot, t)$, namely,

$$\tilde{H}^\pm(x, p, t) = -c_1(x, t)\Phi_1(x, p, t) + c_2(x, t)\Phi_2(x, p, t)$$

where $\Phi_1(x, \cdot, t)$ and $\Phi_2(x, \cdot, t)$ have their respective Wulff sets as $A(x, t)$ and $B(x, t)$ (see [21,24,43] for more details of the Wulff set).

Now, for a general pair of $\tilde{H}^\pm(x, p, t)$, we have the following well-known theorem:

Theorem 2.1. [20] *The function U is the viscosity solution to the HJ PDE :*

$$\begin{cases} \frac{\partial}{\partial t} U - \tilde{H}^+(x, \nabla_x U, t) = 0 & \text{on } \mathbb{R}^d \times (-\infty, T), \\ U(x, T) = g(x) & \text{on } \mathbb{R}^d. \end{cases}$$

Similarly, the function V is the viscosity solution to the HJ PDE:

$$\begin{cases} \frac{\partial}{\partial t} V - \tilde{H}^-(x, \nabla_x V, t) = 0 & \text{on } \mathbb{R}^d \times (-\infty, T), \\ V(x, T) = g(x) & \text{on } \mathbb{R}^d. \end{cases}$$

It is worth mentioning again that, in a general setting where h is possibly nonconvex, the two Hamiltonians $\tilde{H}^+(x, p, t)$ and $\tilde{H}^-(x, p, t)$ may not coincide. But, when they do, there is the following corollary:

Corollary 2.2. [20] *If*

$$\tilde{H}^+(x, p, t) = \tilde{H}^-(x, p, t) \text{ on } [t, T] \times \mathbb{R}^d \times \mathbb{R}^d,$$

then it holds that $U = V$.

Hereafter, when $U = V$, we write $\varphi(x, t) := U(x, T - t) = V(x, T - t)$, and write $H(x, p, t) = \tilde{H}^\pm(x, p, T - t)$, then

$$\begin{cases} \frac{\partial}{\partial t} \varphi + H(x, \nabla_x \varphi, t) = 0 & \text{on } \mathbb{R}^d \times (0, \infty), \\ \varphi(x, 0) = g(x) & \text{on } \mathbb{R}^d. \end{cases}$$

Note that in general, the Hamiltonians H can be nonconvex and/or nonconcave, and this is one very important occasion in which nonconvex HJ PDE arises. We would like to mention that the convention to write the HJ-PDE as an initial value problem or terminal value problem is a matter of convention. One may write that either with the variable t or with the variable $T - t$ and switch between the two formulations. Since for many applications it is stated as an initial value problem, we would like to stick to the convention using initial value problem.

In the next section, we will discuss possible representation formulae of the HJ-PDE equation, which may help us to compute the solution quickly and in parallel. We will prove they hold for restricted assumptions, and refer to the readers to [62] for the case with less restricted assumptions.

3. Representation formulae for viscosity solution of HJ PDE

In this section, we prove in subsection 3.2 that the two formulas hold under restricted assumptions. Then we go on to making a conjecture of a (Lax-type) minimization principle for the viscosity solution to (2.1)–(2.2) when H is convex, and a (Hopf-type) maximization principle when H is non-convex but when g is convex that they shall still hold in this case other less restricted assumptions. In several examples given in the paper, when a Hopf formula is known for the solution, our conjectured representation reduces to these known formulae. We refer our readers to [62] for a proof that our conjectures shall hold for a less restricted set of assumptions.

Before we provide our formal statements of the conjectures, for the sake of exposition, let us emphasize that the formulae stated in Section 1 are the key conjectures from a practical point of view. When $H(x, p, t)$ is non-smooth, more general forms of the respective conjectures are necessary. The precise statements and proof under restricted assumptions will be given in the following subsections. Then we make our conjecture that the two formula still hold under less restricted assumptions. One point to remark is that our conjectured Hopf formula is a generalization of the well-known Hopf formula in [18,25,50].

3.1. A simplified system

In order to show the formula is true in a simplified setting, for the sake of exposition, let us consider first a **simplified** system of differential equations: Fix $0 \leq t < T$, $x \in \mathbb{R}^d$. We consider

$$\begin{cases} \frac{dx}{ds}(s) = f(s, x(s), u(s)) & t \leq s \leq T, \\ x(t) = x, \end{cases}$$

where $u(\cdot) \in \mathcal{U}(t) = \{u : [t, T] \rightarrow U : u \text{ is measurable}\}$ is again a control and U is the admissible set. We again assume that the function

$$f : [0, T] \times \mathbb{R}^d \times U \rightarrow \mathbb{R}^m$$

is uniformly continuous and

$$\begin{cases} |f(t, x, u)| \leq C_1 \\ |f(t, x, u) - f(t, y, u)| \leq C_1 |x - y|, \end{cases}$$

for some constant C_1 and for all $0 \leq t \leq T$, $x, y \in \mathbb{R}^m$, $u \in U$.

We also consider the following simplified payoff function

$$P(u) := P_{t,x}(u(\cdot)) := \int_t^T h(s, x(s), u(s)) ds + g(x(T)).$$

We consider also the value function

$$U(x, t) := \inf_{u \in \mathcal{U}(t)} P_{t,x}(u) = \inf_{u \in \mathcal{U}(t)} \left\{ \int_t^T h(s, x(s), u(s)) ds + g(x(T)) \right\},$$

and the Hamiltonian

$$\tilde{H}(x, p, t) := \max_{u \in U} \{-\langle f(t, x, u), p \rangle - h(t, x, u)\} \text{ as well as } H(x, p, t) = \tilde{H}(x, p, T - t). \quad (3.1)$$

This is actually the special case of the setting as discussed in Section 2.2 when the set $C = \{0\}$ is a singleton, after we denote $b(\cdot)$ as $u(\cdot)$ instead. In this special case $H(t, x, p)$ is always convex w.r.t. p .

One point to note is that, the argument to get either a Lax and a Hopf formula in the general case with differential games as discussed in section 2.2 shall be similar with the standard assumption on the set of strategy following the causality.

3.2. Verification of minimization/maximization principles under restricted assumptions

Before we go to our statement of conjectures, let us show that the formula in section 1 holds for some restricted assumptions for the sake of completeness. A proof under less restricted assumptions can be found in [62]. In this subsection, we suppress the time dependency of f and h (and thus H) to simplify the statements and proofs of the theorems. We note that the same statements remain true with time dependency. We keep the time dependency in the next subsection.

In what follows, we first state the set of assumptions that we may use. For notational sake, let us write

$$\mathcal{H}(x, p, u) := h(x, u) + \langle p, f(x, u) \rangle.$$

Let us consider the following list of assumptions that we will consider:

- (A1) U is a compact convex set in $\mathbb{R}^d$;
- (A2) $\mathcal{H}(x, p, u)$ is proper lower semi-continuous and quasi-convex w.r.t. u .
- (A3) $H(x, p)$ as defined in (3.1) is proper upper semi-continuous and quasi-concave w.r.t. x .
- (A4) $H(x, p)$ is equi-coercive (under parameters p) w.r.t. x in the following sense: for all $K > 0$, there exists N (independent of p)

$$|H(x, p)| \geq K.$$

whenever $\|x\| \geq N$.

- (A5) $g(x)$ is proper lower semi-continuous and convex w.r.t. x , and is coercive w.r.t. x in the following sense:

$$\|g(x)\| \rightarrow \infty \text{ as } \|x\| \rightarrow \infty$$

- (A6) $H(x, p)$ is proper concave w.r.t. x , and is $H(x, p)$ is equi-coercive (under parameters x) w.r.t. p .

- (A7) $\mathcal{H}(x, p, u)$, $H(x, p)$, $g(x)$ and $g^*(p)$ are all in C^2 in all its variables.

3.2.1. Verification of minimization principle under restricted assumptions

Before we proceed, we would also like to state the following lemma directly from definition:

Lemma 3.1. Under the assumption (A2), $H(x, p, t)$ as defined in (3.1) is convex and lower semi-continuous w.r.t. p .

Proof. Both properties follow from the fact that $\mathcal{H}(p, x, u, s)$ is linear w.r.t. p . $\square$

With these assumptions at hand, we have the following lemmas. We would like to remark that the following is a more discrete version of the minimax formula appeared in [5].

Lemma 3.2. Let $(\mathbf{X}, \mathbf{U}, \mathbf{P}) := (\{x_n\}_{n=0}^{N-1}, \{u_n^*\}_{n=0}^{N-1}, \{p_n\}_{n=0}^{N-1}) \in \mathbb{R}^{3dN}$ and

$$F_1(\mathbf{X}, \mathbf{U}, \mathbf{P}) := g(x_0) + \delta \sum_{n=0}^{N-1} h(x_n, u_n) + \sum_{n=0}^{N-1} \langle p_n, x_{n+1} - x_n \rangle + \delta \sum_{n=0}^{N-1} \langle p_n, f(x_n, u_n) \rangle$$

and

$$\tilde{F}_1(\mathbf{X}, \mathbf{P}) := g(x_0) + \sum_{n=0}^{N-1} \langle p_n, x_{n+1} - x_n \rangle - \delta \sum_{n=0}^{N-1} H(x_n, p_n, s_n).$$

If (A1) and (A2) are satisfied, then we have

$$\min_{\{u_n\}_{n=0}^{N-1} \in U^N} \inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} F_1(\mathbf{X}, \mathbf{U}, \mathbf{P}) = \inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(\mathbf{X}, \mathbf{P}). \quad (3.2)$$

Proof. Fixing $\mathbf{X}$, consider $F_1(\mathbf{X}, \cdot, \cdot) : (\mathbf{U}, \mathbf{P}) \mapsto F_1(\mathbf{X}, \mathbf{U}, \mathbf{P})$. Since F_1 is affine in P , together with (A1) and (A2), we apply a classical minimax theorem, e.g. strong Lagrange duality satisfying Slater's condition [3,47] or Sion's minimax theorem [29, 52] (an extension of Von Neumann minimax theorem[39], see also [46]), to deduce for a fixed X ,

$$\inf_{\{u_n\}_{n=0}^{N-1} \in U} \sup_{\{p_n\}_{n=0}^{N-1}} F_1(\mathbf{X}, \mathbf{U}, \mathbf{P}) = \sup_{\{p_n\}_{n=0}^{N-1}} \inf_{\{u_n\}_{n=0}^{N-1} \in U} F_1(\mathbf{X}, \mathbf{U}, \mathbf{P}).$$

Now taking the infimum at both hand sides w.r.t $\mathbf{X}$, and noticing

$$\min_{\{u_n\}_{n=0}^{N-1} \in U^N} \inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} F_1(\mathbf{X}, \mathbf{U}, \mathbf{P}) = \inf_{\{x_n\}_{n=0}^{N-1}} \min_{\{u_n\}_{n=0}^{N-1} \in U^N} \sup_{\{p_n\}_{n=0}^{N-1}} F_1(\mathbf{X}, \mathbf{U}, \mathbf{P}),$$

the result now follows from the definition of $H(t, x, p)$ in (3.1). $\square$

Notice Lemma 3.2 is a version of representation formula with a minimal amount of assumption, and this expression gives the finite dimensional analogy well-known principle of least action as in e.g. [5,48].

Lemma 3.3. If (A1), (A2), (A4), (A6), (A7) are satisfied, then we have

$$\begin{aligned} & \min_{\{u_n\}_{n=0}^{N-1} \in U^N} \inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} F_1(\mathbf{X}, \mathbf{U}, \mathbf{P}) \\ &= \inf_{v \in \mathbb{R}^d} \left\{ g(x_0) + \delta \sum_{n=0}^{N-1} \langle p_n, \frac{x_{n+1} - x_n}{\delta} \rangle - \delta \sum_{n=0}^{N-1} H(x_n, p_n) : \right. \\ & \quad \left. \begin{aligned} x_{n+1}(x, v) - x_n(x, v) &= \delta \partial_p H(x_n(x, v), p_n(x, v)), \\ p_{n-1}(x, v) - p_n(x, v) &= \delta \partial_x H(x_n(x, v), p_n(x, v)), \\ x_N &= x, p_N = v \end{aligned} \right\}. \end{aligned} \quad (3.3)$$

Proof. Let $(\tilde{\mathbf{X}}, \mathbf{P}) := (\{x_n\}_{n=1}^{N-1}, \{p_n\}_{n=0}^{N-1}) \in \mathbb{R}^{(2N-1)d}$. From Lemma 3.2, since (A1) and (A2), are satisfied, we have that (3.2) holds. Therefore it remains to show that the term $\inf_{\{x_n\}_{n=1}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(\mathbf{X}, \mathbf{P})$ equals to (3.3). Notice that

$$\inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(\mathbf{X}, \mathbf{P}) = \inf_{x_0} \inf_{\{x_n\}_{n=1}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P}).$$

We now wish to argue that under assumptions (A6), (A7) and (A8), for a fixed x_0 , either that there exists $(\tilde{\mathbf{X}}^*(x_0), \mathbf{P}^*(x_0))$ (depending on x_0) such that the mini-max problem $\inf_{\{x_n\}_{n=1}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P}) = \tilde{F}_1(x_0, \tilde{\mathbf{X}}^*(x_0), \mathbf{P}^*(x_0))$, or one has $\inf_{\{x_n\}_{n=1}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(\mathbf{X}, \mathbf{P}) = \infty$.

In fact, for a fixed set of $(x_0, \tilde{\mathbf{X}})$, by (A6), either $\sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P})$ is attained and thus, by (A7) there is $\mathbf{P}^*(x_0, \tilde{\mathbf{X}})$ smoothly depending on $(x_0, \tilde{\mathbf{X}})$ s.t.

$$x_{n+1}(x, v) - x_n(x, v) = \delta \partial_p H(x_n(x, v), p_n^*(x, v, (x_0, \tilde{\mathbf{X}})))$$

holds, or the supremum is infinity. Now let us take infimum over $\tilde{\mathbf{X}}$. For x_0 such that for all $\tilde{\mathbf{X}}$ the value $\sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P}) = \infty$, we have that $\inf_{\{x_n\}_{n=1}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P}) = \infty$. Otherwise, for x_0 such that there exists $\tilde{\mathbf{X}}$ with $\sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P}) \neq \infty$, by (A4) and (A7), we again either have $\tilde{\mathbf{X}}^*$ satisfies

$$\partial_{\tilde{\mathbf{X}}} \left(\sum_{n=0}^{N-1} \langle p_n^*(x, v, (x_0, \tilde{\mathbf{X}}^*)), x_{n+1}^*(x, v, x_0) - x_n^*(x, v, x_0) \rangle - \delta \sum_{n=0}^{N-1} H(x_n^*(x, v, x_0), p_n^*(x, v, (x_0, \tilde{\mathbf{X}}^*))) \right) = 0,$$

which gives, fixing x_0 , the following for all $n = 1, \dots, N-1$, using the smooth dependence of $(x_0, \tilde{\mathbf{X}}) \mapsto \mathbf{P}^*(x_0, \tilde{\mathbf{X}})$ and a chain rule, that

$$\begin{aligned} & p_{n-1}^*(x, v, (x_0, \tilde{\mathbf{X}}^*)) - p_n^*(x, v, (x_0, \tilde{\mathbf{X}}^*)) - \delta \partial_x H(x_n^*(x, v, x_0), p_n^*(x, v, (x_0, \tilde{\mathbf{X}}^*))) \\ &= - \left(\sum_{l=0}^{N-1} \left(x_{l+1}^*(x, v, x_0) - x_l^*(x, v, x_0) - \delta \partial_p H(x_l^*(x, v, x_0), p_l^*(x, v, (x_0, \tilde{\mathbf{X}}^*))) \right) \partial_{\tilde{\mathbf{X}}_n} p_l^*(x, v, (x_0, \tilde{\mathbf{X}}^*)) \right), \\ &= - \left(\sum_{l=0}^{N-1} \left(x_{l+1}(x, v) - x_l(x, v) - \delta \partial_p H(x_l(x, v), p_l^*(x, v, (x_0, \tilde{\mathbf{X}})), s_l) \right) \right) \Big|_{\mathbf{X}=\tilde{\mathbf{X}}^*} \partial_{\tilde{\mathbf{X}}_n} p_l^*(x, v, (x_0, \tilde{\mathbf{X}}^*)), \\ &= 0 \end{aligned}$$

where the last equality comes from the optimality condition defining $(x_0, \tilde{\mathbf{X}}) \mapsto \mathbf{P}^*(x_0, \tilde{\mathbf{X}})$; or that $\inf_{\{x_n\}_{n=1}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P}) = -\infty$. However the case that the infimum get to $-\infty$ will arrive at absurdity since we have tracing back Lemma 3.2

$$-\infty < \inf_{\{u_0\}_{n=1}^{N-1}} \inf_{\{x_n\}_{n=1}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P}) = \inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(\mathbf{X}, \mathbf{P}) = -\infty$$

which arrives at contradiction.

Concluding the above argument, for each x_0 , either we have $\inf_{\{x_n\}_{n=1}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P}) = \infty$ or $\inf_{\{x_n\}_{n=1}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P}) = \tilde{F}_1(x_0, \tilde{\mathbf{X}}^*(x_0), \mathbf{P}^*(x_0))$ for some $(\mathbf{X}^*(x_0), \mathbf{P}^*(x_0)) = (\tilde{\mathbf{X}}^*(x_0), \mathbf{P}^*(x_0, \tilde{\mathbf{X}}^*))$ s.t.

$$\begin{cases} x_{n+1}(x, v) - x_n(x, v) = \delta \partial_p H(x_n(x, v), p_n(x, v)), \\ p_{n-1}(x, v) - p_n(x, v) = \delta \partial_x H(x_n(x, v), p_n(x, v)), \\ x_N = x. \end{cases} \quad (3.4)$$

Notice the condition as the initial condition $\partial_x g(x_0) = p_0$ do not appear because we fixed a value x_0 . Now for any choice of $(\mathbf{X}(x_0), \mathbf{P}(x_0))$ satisfying (3.4) will give the same value $\tilde{F}_1(x_0, \tilde{\mathbf{X}}(x_0), \mathbf{P}(x_0))$ by concavity of $\tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P})$ w.r.t. $\mathbf{P}$. Therefore,

$$\inf_{x_0} \inf_{\{x_n\}_{n=1}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P}) = \min \left\{ \infty, \inf_{x_0 \in \{x_0 : \exists (\mathbf{X}^*(x_0), \mathbf{P}^*(x_0)) \text{ satisfying (3.4)}\}} \inf_{\mathbf{P}^N} \tilde{F}_1(x_0, \tilde{\mathbf{X}}^*(x_0), \mathbf{P}^*(x_0)) \right\}.$$

The conclusion of the lemma follows from the surjection between $v \in \mathbb{R}^d$ and $x_0 \in \{x_0 : \exists (\mathbf{X}^*(x_0), \mathbf{P}^*(x_0)) \text{ satisfying (3.4)}\}$ via the correspondence $p_N^*(x_0) = v$. $\square$

We would like to have the following two remarks: (1) in fact if $\mathcal{H}$ is non-smooth w.r.t. u , even $H(x, p, s) = \max_{u \in U} \{-\mathcal{H}(x, p, u)\}$, in general we do not have $\partial_p H(x, p, s) = -\partial_p \mathcal{H}(x, p, u^*)$ where u^* is such that $\mathcal{H}(x, p, u^*) = \max_{u \in U} \{-\mathcal{H}(x, p, u)\}$. Therefore in general the KKT condition (which is also referred to as Pontryagin's maximum principle [45] in the continuous case) would be stated using $\mathcal{H}$. However under appropriate regularity assumption (e.g. (A2), (A6) and (A8)), we have the two partial derivatives coinciding, i.e. $\partial_p H(x, p) = -\partial_p \mathcal{H}(x, p, u^*)$. (2) actually our assumptions (A1), (A2), (A4), (A6), (A7) already guarantee the existence of the adjoint variables, and even smooth dependence, that there exists $\mathbf{P}^*(x_0, \tilde{\mathbf{X}})$ smoothly depending on $(x_0, \tilde{\mathbf{X}})$ in the middle of the proof. These assumptions are clearly stronger than what we need, but we assume here for the simplicity of our argument.

A clear connection made between the maximum principle and dynamic programming is discussed in literature, e.g. in [10], which discussed in the non-smooth case when will the costate variable coincide with a partial generalized gradient of the value function (cf. [10]) for a clear definition.

3.2.2. Verification of maximization principle under restricted assumptions

Using a similar but slightly different argument as in Lemma 3.2, we also obtain similarly the following lemma:

Lemma 3.4. Write $(\mathbf{X}, \mathbf{U}, \mathbf{P}) := (\{x_n\}_{n=0}^{N-1}, \{u_n^*\}_{n=0}^{N-1}, \{p_n\}_{n=1}^N) \in \mathbb{R}^{3dN}$, $(\tilde{\mathbf{X}}, \mathbf{U}, \mathbf{P}) := (\{x_n\}_{n=1}^{N-1}, \{u_n^*\}_{n=0}^{N-1}, \{p_n\}_{n=1}^N) \in \mathbb{R}^{(3N-1)d}$,

$$F_2(\mathbf{X}, \mathbf{U}, \mathbf{P}) := g(x_0) + \delta \sum_{n=0}^{N-1} h(x_{n+1}, u_{n+1}) + \sum_{n=0}^{N-1} \langle p_{n+1}, x_{n+1} - x_n \rangle + \delta \sum_{n=0}^{N-1} \langle p_{n+1}, f(x_{n+1}, u_{n+1}) \rangle,$$

and

$$\tilde{F}_2(\tilde{\mathbf{X}}, \mathbf{P}) := \langle p_N, x \rangle - g^*(p_1) - \delta \sum_{n=1}^N H(x_n, p_n) + \sum_{n=1}^{N-1} \langle p_n - p_{n+1}, x_n \rangle.$$

If (A1), (A2), (A3), (A4) and (A5) are satisfied, then we have then we have

$$\min_{\{u_n\}_{n=0}^{N-1} \in U^N} \inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=1}^N} F_2(\mathbf{X}, \mathbf{U}, \mathbf{P}) = \sup_{\{p_n\}_{n=1}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, \mathbf{P}). \quad (3.5)$$

Proof. The first part of the argument goes similar as in the proof of the previous lemma. In fact, fixing $\mathbf{X}$, we see that $F_2(\mathbf{X}, \cdot, \cdot) : (\mathbf{U}, \mathbf{P}) \mapsto F_2(\mathbf{X}, \mathbf{U}, \mathbf{P})$ is affine w.r.t. P . Together with (A1), (A2) and Definition (3.1) of $H(t, x, p)$ and summation by part, we again apply a classical minimax theorem, e.g. strong Lagrange duality [3,47], to get

$$\inf_{\{u_n\}_{n=0}^{N-1} \in U} \sup_{\{p_n\}_{n=1}^N} F_2(\mathbf{X}, \mathbf{U}, \mathbf{P}) = \sup_{\{p_n\}_{n=1}^N} \inf_{\{u_n\}_{n=0}^{N-1} \in U} F_2(\mathbf{X}, \mathbf{U}, \mathbf{P}) = \sup_{\{p_n\}_{n=1}^N} \hat{F}_2(\mathbf{X}, \mathbf{P}),$$

where

$$\hat{F}_2(\mathbf{X}, \mathbf{P}) := g(x_0) - \langle p_1, x_0 \rangle - \delta \sum_{n=0}^{N-1} H(x_{n+1}, p_{n+1}) + \sum_{n=1}^{N-1} \langle p_n - p_{n+1}, x_n \rangle + \langle p_N, x \rangle.$$

Hence we get

$$\min_{\{u_n\}_{n=0}^{N-1} \in U^N} \inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=1}^N} F_2(\mathbf{X}, \mathbf{U}, \mathbf{P}) = \inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=1}^N} \hat{F}_2(\mathbf{X}, \mathbf{P}).$$

Now since $H(s, x, p)$ satisfies (A3), we have $\hat{F}_2(\mathbf{X}, \mathbf{P})$ is lower-semicontinuous and quasi-convex w.r.t. X . Since (A4) and (A5) are satisfied, there exists a compact (and convex, w.l.o.g.) $C^N \subset \mathbb{R}^{dN}$ such that

$$\inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=1}^N} \hat{F}_2(\mathbf{X}, \mathbf{P}) = \inf_{\{x_n\}_{n=0}^{N-1} \in C^N} \sup_{\{p_n\}_{n=1}^N} \hat{F}_2(\mathbf{X}, \mathbf{P})$$

From Lemma 3.1 $\hat{F}_2(X, U, P)$ is upper-semicontinuous and quasi-concave w.r.t. P . Therefore we may apply Sion's minimax theorem [29,52] to get

$$\inf_{\{x_n\}_{n=0}^{N-1} \in C^N} \sup_{\{p_n\}_{n=1}^N} \hat{F}_2(\mathbf{X}, \mathbf{P}) = \sup_{\{p_n\}_{n=1}^N} \inf_{\{x_n\}_{n=0}^{N-1} \in C^N} \hat{F}_2(\mathbf{X}, \mathbf{P})$$

and in fact $\inf_{\{x_n\}_{n=0}^{N-1} \in C^N}$ can be replaced by $\min_{\{x_n\}_{n=0}^{N-1} \in C^N}$. Now considering (A4) and (A5) again, we obtain

$$\sup_{\{p_n\}_{n=1}^N} \inf_{\{x_n\}_{n=0}^{N-1} \in C^N} \hat{F}_2(\mathbf{X}, \mathbf{P}) = \sup_{\{p_n\}_{n=1}^N} \inf_{\{x_n\}_{n=0}^{N-1}} \hat{F}_2(\mathbf{X}, \mathbf{P}).$$

Our conclusion now comes from the fact that

$$\inf_{x_0} \hat{F}_2(\mathbf{X}, \mathbf{P}) = \tilde{F}_2(\tilde{\mathbf{X}}, \mathbf{P})$$

by the definition of Fenchel-Legendre transform [35]. $\square$

Lemma 3.5. If (A1), (A2), (A4), (A5), (A6), (A7) are satisfied, then we have

$$\begin{aligned} & \min_{\{u_n\}_{n=0}^{N-1} \in U^N} \inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=1}^N} F_2(\mathbf{X}, \mathbf{U}, \mathbf{P}) \\ &= \sup_{v \in \mathbb{R}^d} \left\{ \langle p_N, x \rangle - g^*(p_1) - \delta \sum_{n=1}^N H(x_n, p_n) + \delta \sum_{n=1}^{N-1} \left\langle \frac{p_n - p_{n+1}}{\delta}, x_n \right\rangle : \right. \\ & \quad \left. \begin{aligned} x_n(x, v) - x_{n-1}(x, v) &= \delta \partial_p H(x_n(x, v), p_n(x, v)), \\ p_n(x, v) - p_{n+1}(x, v) &= \delta \partial_x H(x_n(x, v), p_n(x, v)), \\ x_N &= x, p_N = v \end{aligned} \right\}. \end{aligned} \quad (3.6)$$

Proof. Let $(\tilde{\mathbf{X}}, \tilde{\mathbf{P}}) := (\{x_n\}_{n=1}^{N-1}, \{p_n\}_{n=2}^N) \in \mathbb{R}^{(2N-2)d}$. Again since (A6) and (A7) imply (A3), from Lemma 3.4, we have that (3.5) holds. Therefore it remains to show that the term $\sup_{\{p_n\}_{n=2}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, \mathbf{P})$ equals to (3.6). Notice again that

$$\sup_{\{p_n\}_{n=2}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, \mathbf{P}) = \sup_{p_1} \sup_{\{p_n\}_{n=2}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, p_1, \tilde{\mathbf{P}}).$$

Now as in Lemma 3.4, we now wish to argue that under assumptions (A6), (A7) and (A8), for a fixed p_1 , either that there exists $(\tilde{\mathbf{X}}^*(p_1), \tilde{\mathbf{P}}^*(p_1))$ (depending on p_1) such that the max-min problem $\sup_{\{p_n\}_{n=2}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, p_1, \tilde{\mathbf{P}}) = \tilde{F}_2(\tilde{\mathbf{X}}^*(p_1), p_1, \tilde{\mathbf{P}}^*(p_1))$, or one has $\sup_{\{p_n\}_{n=2}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, p_1, \tilde{\mathbf{P}}) = -\infty$.

Now, for a fixed set of $(p_1, \tilde{\mathbf{P}})$, by (A4) and (A5), either $\inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, p_1, \tilde{\mathbf{P}})$ is attained and thus, by (A7) there is $\tilde{\mathbf{X}}^*(p_1, \tilde{\mathbf{P}})$ smoothly depending on $(p_1, \tilde{\mathbf{P}})$ s.t.

$$p_n(x, v) - p_{n+1}(x, v) = \delta \partial_x H(x_n^*(x, v, (p_1, \tilde{\mathbf{P}})), p_n(x, v)),$$

holds, or the infimum is minus infinity. Now let us take supremum over $\tilde{\mathbf{P}}$. For p_1 such that for all $\tilde{\mathbf{P}}$ the value $\inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, p_n, \tilde{\mathbf{P}}) = -\infty$, we have that $\sup_{\{p_n\}_{n=2}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, p_n, \tilde{\mathbf{P}}) = -\infty$. Otherwise, for p_1 such that there exists $\tilde{\mathbf{P}}$ with $\inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, p_1, \tilde{\mathbf{P}}) \neq -\infty$, by (A6) and (A7), we again either have $\tilde{\mathbf{P}}^*$ satisfies

$$\begin{aligned} \partial_{\tilde{\mathbf{P}}} \left(\langle p_N^*(x, v, p_1), x \rangle + \sum_{n=1}^{N-1} \langle p_n^*(x, v, p_1) - p_{n+1}^*(x, v, p_1), x_n^*(x, v, (p_1, \tilde{\mathbf{P}}^*)) \rangle \right. \\ \left. - \delta \sum_{n=1}^N H(x_n^*(x, v, (p_1, \tilde{\mathbf{P}}^*)), p_n^*(x, v, p_1)) \right) = 0, \end{aligned}$$

where we notice the last term is independent of $\tilde{\mathbf{P}}$, and thus gives, fixing p_1 , the following for all $n = 2, \dots, N$, using the smooth dependence of $(p_1, \tilde{\mathbf{P}}) \mapsto \tilde{\mathbf{X}}^*(p_1, \tilde{\mathbf{P}})$ and a chain rule, that

$$\begin{aligned} & x_n^*(x, v, (p_1, \tilde{\mathbf{P}}^*)) - x_{n-1}^*(x, v, (p_1, \tilde{\mathbf{P}}^*)) - \delta \partial_p H(x_n^*(x, v, (p_1, \tilde{\mathbf{P}}^*)), p_n^*(x, v, p_1)) \\ &= - \left(0 + \sum_{l=0}^{N-1} \left(p_l^*(x, v, p_1) - p_{l+1}^*(x, v, p_1) - \delta \partial_x H(x_l^*(x, v, (p_1, \tilde{\mathbf{P}}^*)), p_l^*(x, v, p_n)) \right) \partial_{\tilde{\mathbf{P}}_n} x_l^*(x, v, (p_1, \tilde{\mathbf{P}}^*)) \right), \\ &= - \left(\sum_{l=0}^{N-1} \left(p_l(x, v) - p_{l+1}(x, v) - \delta \partial_x H(x_l^*(x, v, (p_1, \tilde{\mathbf{P}})), p_l(x, v)) \right) \right) \Big|_{\tilde{\mathbf{P}}=\tilde{\mathbf{P}}^*} \partial_{\tilde{\mathbf{P}}_n} x_l^*(x, v, (p_1, \tilde{\mathbf{P}}^*)), \\ &= 0 \end{aligned}$$

where the second equality used the fact that $\partial_{\tilde{\mathbf{P}}_n} x_N^*(x, v, (p_1, \tilde{\mathbf{P}}^*)) = \partial_{\tilde{\mathbf{P}}_n} x = 0$ the last equality comes from the optimality condition defining $(p_1, \tilde{\mathbf{P}}) \mapsto \tilde{\mathbf{X}}^*(p_1, \tilde{\mathbf{P}})$; or that $\sup_{\{p_n\}_{n=2}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, p_1, \tilde{\mathbf{P}}) = \infty$. However again case that the supremum get to ∞ will arrive at absurdity since we have tracing back Lemma 3.2

$$\infty > \inf_{\{u_0\}_{n=1}^{N-1}} \inf_{\{x_n\}_{n=1}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P}) = \sup_{\{p_n\}_{n=1}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, p_n, \tilde{\mathbf{P}}) = \infty$$

which arrives at contradiction.

Concluding the above argument, for each p_1 , either we have $\sup_{\{p_n\}_{n=2}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, p_1, \tilde{\mathbf{P}}) = -\infty$ or $\sup_{\{p_n\}_{n=2}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, p_1, \tilde{\mathbf{P}}) = \tilde{F}_2(\tilde{\mathbf{X}}^*(p_1), p_1, \tilde{\mathbf{P}}^*(p_1))$ for some $(\mathbf{X}^*(p_1), \mathbf{P}^*(p_1)) = (\tilde{\mathbf{X}}^*(p_1), \mathbf{P}^*(p_1, \tilde{\mathbf{X}}^*))$ s.t.

$$\begin{cases} x_n(x, v) - x_{n-1}(x, v) = \delta \partial_p H(x_n(x, v), p_n(x, v)), \\ p_n(x, v) - p_{n+1}(x, v) = \delta \partial_x H(x_n(x, v), p_n(x, v)), \\ x_N = x. \end{cases} \quad (3.7)$$

Notice the condition as the initial condition $\partial_x g^*(p_1) = x_1$ do not appear because we fixed a value p_1 . Now for any choice of $(\mathbf{X}(x_0), \mathbf{P}(x_0))$ satisfying (3.4) will give the same value $\tilde{F}_1(x_0, \tilde{\mathbf{X}}(x_0), \mathbf{P}(x_0))$ by concavity of $\tilde{F}_1(x_0, \tilde{\mathbf{X}}, \mathbf{P})$ w.r.t. $\mathbf{P}$. Therefore,

$$\sup_{p_1} \sup_{\{p_n\}_{n=2}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, p_1, \tilde{\mathbf{P}}) = \max \left\{ -\infty, \sup_{p_1 \in \{p_1 : \exists (\mathbf{X}^*(p_1), \mathbf{P}^*(p_1)) \text{ satisfying (3.7)}\}} \sup_{v_N} \tilde{F}_2(\tilde{\mathbf{X}}^*, p_1, \tilde{\mathbf{P}}^*) \right\}.$$

The conclusion of the lemma follows from the surjection between $v \in \mathbb{R}^d$ and $p_1 \in \{p_1 : \exists(\mathbf{X}^*(p_1), \mathbf{P}^*(p_1)) \text{ satisfying (3.7)}\}$ via the correspondence $p_N^*(p_1) = v$. $\square$

Again, assumptions (A1), (A2), (A4), (A5), (A6), (A7) already guarantee the existence and smooth dependence of $(p_1, \tilde{\mathbf{P}}) \mapsto \tilde{\mathbf{X}}^*(p_1, \tilde{\mathbf{P}})$. These assumptions are stronger than what we need, but we assume them in this work for the simplicity of our argument.

Before we proceed, we again notice that all the statements above would with h and f depending on time s . From now on, we do not suppress the time dependence in the notation to indicate sometimes the necessity of a change of variable $s \mapsto T - s$. This, although mathematically trivial, may bring computational error if one is not fully aware of it, and is hence kept from now on.

3.3. Generalized Lax minimization principle for Convex/Concave Hamiltonian

We first describe how we obtain a Lax formula under restricted assumptions and applying Lemma 3.2 and Lemma 3.4:

$$\varphi(x, t) := U(x, T - t) := \inf_{u \in \mathcal{U}(T-t)} \left\{ \int_{T-t}^T h(s, x(s), u(s)) ds + g(x(T)) \right\}.$$

We derive our formula as follows. Following [40], we shall first consider the following discretization (approximation) for a given δ such that $\delta N = t$, by denoting $s_n = \delta n$ and $x_N = x$ (and flipping the sign),

$$\varphi(x, t) \approx \varphi_N^1(x, t)$$

where

$$\begin{aligned} & \varphi_N^1(x, t) \\ &:= \min_{\{u_n\}_{n=0}^{N-1} \in U} \left\{ g(x_0) + \delta \sum_{n=0}^{N-1} h(T - s_n, x_n, u_n) : x_{n+1} - x_n = -\delta f(T - s_n, x_n, u_n) \text{ for } n = 0, \dots, N-1, x_N = x \right\}, \\ &= \min_{\{u_n\}_{n=0}^{N-1} \in U^N} \inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} F_1(\mathbf{X}, \mathbf{U}, \mathbf{P}), \end{aligned}$$

where F_1 is given as in Lemma 3.2, and the second equality comes from reformulating the problem with Lagrange multiplier. Therefore, applying Lemma 3.2 and Lemma 3.4, we get the following:

Theorem 3.6. *If (A1), (A2) are satisfied, then*

$$\varphi_N^1(x, t) = \inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=0}^{N-1}} \tilde{F}_1(\mathbf{X}, \mathbf{P}).$$

If (A1), (A2), (A4), (A6), (A7) are satisfied, then we have

$$\begin{aligned} \varphi_N^1(x, t) = \inf_{v \in \mathbb{R}^d} & \left\{ g(x_0) + \delta \sum_{n=0}^{N-1} \left\langle p_n, \frac{x_{n+1} - x_n}{\delta} \right\rangle - \delta \sum_{n=0}^{N-1} H(x_n, p_n, s_n) : \right. \\ & \left. \begin{aligned} x_{n+1}(x, v) - x_n(x, v) &= \delta \partial_p H(x_n(x, v), p_n(x, v), s_n), \\ p_{n-1}(x, v) - p_n(x, v) &= \delta \partial_x H(x_n(x, v), p_n(x, v), s_n), \\ x_N &= x, p_N = v \end{aligned} \right\}. \end{aligned} \quad (3.8)$$

We would like to remark that (A5) is not needed for the validity of the above formulae.

However, we notice that the resulting formula that we conjectured seems to be correct beyond these assumptions, as the numerical results show (especially when we take the minimum over all the paths satisfying the KKT conditions). We hope to get rigorous criteria for this formula to hold in the future. In fact, passing to the limit, in the special case when $H(x, p, t)$ is smooth also w.r.t. p , we conjecture the following Lax formula.

Conjecture 3.7. *Assume $H(x, p, t) \in C^2$ and is convex w.r.t. p , and (A5) is satisfied, then there exists t_0 such that the viscosity solution to (2.1)–(2.2) can be represented as (1.1) for $t \leq t_0$. Moreover, if $\phi(x, t)$ is differentiable w.r.t. x at a neighborhood of (x, t) and the infimum is attained by $\tilde{v}$, then we have $\partial_x \varphi(x, t) = \tilde{v}$.*

We refer to it as the **minimization principle**, or the generalized Lax formula. This conjecture was validated under less restricted assumption in [62] after a previous version of our paper in arXiv in [8] was launched. [62] is a very important and monumental work which suggest the validity of our conjecture in a more general situation.

Remark 1. When $H(x, p, t)$ is not differentiable at some given point p , then we believe that in formula (1.1), the Mordukhovich subdifferential, $\partial_x^- H$, as defined in [36–38], should be used instead of $\partial_x H$. In that case the constraint becomes the inclusion $\dot{\gamma}(x, v, s) \in \partial_p^- H(\gamma(x, v, s), p(x, v, s), s)$, and infimum is taken over also all the curves $(\gamma, p) \in C^\infty$ satisfying the inclusion.

In below there are several examples for the conjecture, that we only provide a brief account.

Example 1. For Hamiltonian $H(x, p, t)$ which is convex w.r.t. p , concave w.r.t. x and satisfying Assumptions (A) in [48], the following minimization principle holds for the viscosity solution to (2.1)–(2.2) (see Theorem 2.3 in [48] and Theorem 4.8 in [49]):

$$\varphi(x, t) = \inf_{\gamma \in C^\infty, \gamma(t)=x} \{g(\gamma(0)) + \int_0^t L(\gamma(s), \dot{\gamma}(s), s) ds\}$$

where the Lagrangian L is defined as

$$L(x, q, s) = \sup_p \{ \langle p, q \rangle - H(x, p, t) \}.$$

This example may not satisfy either the assumptions in Lemma 3.2 and or Lemma 3.4, since U may not be compact.

Example 2. When $H(x, p, t)$ is a convex homogeneous degree-1 functional w.r.t. p of the following special form

$$H(x, p, t) = c(x)\Phi(p),$$

where $c \in C^\infty$ with $C_0 \geq c(x) \geq c_0$ for some $C_0, c_0 > 0$, and Φ is homogeneous of degree 1 functional. We recall the definition of the Wulff set W [43] of Φ defined as the set W such that the following equality holds:

$$\Phi(p) = \max_{x \in W} \langle p, x \rangle = \mathcal{I}_W^*(p)$$

with $*$ denoting the Fenchel-Legendre transform. We furthermore assume closed, strictly convex, balanced (i.e. $-W = W$), absorbing (i.e. for all $y \in \mathbb{R}^d$, $y \in \tau W$ for some $\tau > 0$) with smooth boundary $\partial W \in C^\infty$ [21,24]. Then it is ready to check that the subdifferential of Φ is given as follows:

$$\partial_p^- \Phi(p) = \begin{cases} \partial_p \Phi(p) \in \partial W & \text{if } p \neq 0 \\ W & \text{if } p = 0. \end{cases}$$

For convenience sake, let us also define, for a closed, strictly convex, balanced, absorbing set W with smooth boundary $\partial W \in C^\infty$, the Minkowski functional of W [21,24], as follows

$$\rho_W(y) := \inf\{\tau > 0 : y \in \tau W\}.$$

With this, we are ready to define a metric on $\mathbb{R}^d$ as follows:

$$\tilde{d}(x, y) := \inf_{t>0} \{t : \gamma \in C^\infty, \gamma(0) := x, \gamma(t) = y, \dot{\gamma}(s) \in c(\gamma(s))W\}.$$

It is ready to check that $\tilde{d}$ defines a metric and thus $(\mathbb{R}, \tilde{d})$ forms a metric space.

Lemma 3.8. Assume the metric space $(\mathbb{R}, \tilde{d})$ is complete. If we define

$$C(x, t) := \bigcup_{v \in \mathbb{R}^d} \left\{ \begin{array}{l} (\gamma, p) \in C^\infty \\ \dot{\gamma}(x, v, s) = \partial_p^- H(\gamma(x, v, s), p(x, v, s)), \\ \dot{p}(x, v, s) = -\partial_x H(\gamma(x, v, s), p(x, v, s)), \\ \gamma(x, v, t) = x, p(x, v, t) = v \end{array} \right\}$$

and

$$B(x, t) := \bigcup_{0 \leq r \leq t} \bigcup_{v \in \mathbb{R}^d \setminus \{0\}} \left\{ \begin{array}{l} \dot{\gamma}(x, v, s) = \partial_p H(\gamma(x, v, s), p(x, v, s)), \\ \gamma(x, v, 0) : \dot{p}(x, v, s) = -\partial_x H(\gamma(x, v, s), p(x, v, s)), \\ \gamma(x, v, r) = x, p(x, v, r) = v \end{array} \right\}.$$

Then for all $(x, t) \in \mathbb{R}^{d+1} \times (0, T)$, we have $B(x, t)$ is well-defined and

$$C(x, t) = B(x, t)$$

Proof. Before we get to the proof of well-definedness of $B(x, t)$ and the equivalence of $B(x, t)$ and $C(x, t)$, let us first notice that for any curve $(\gamma, p) \in C^\infty$ satisfying the following inclusion

$$\dot{\gamma}(x, v, s) \in c(\gamma(x, v, s)) \partial_p^- \Phi(p(x, v, s)), \quad \dot{p}(x, v, s) = -\partial_x c(\gamma(x, v, s)) \Phi(p(x, v, s)) \quad (3.9)$$

We have that

$$\dot{\Phi}(p(x, v, s)) = -\langle l(x, v, s), \partial_x c(\gamma(x, v, s)) \rangle \Phi(p(x, v, s))$$

where $l(x, v, s) \in \partial_p^- \Phi(p(x, v, s))$, i.e. we get that

$$\Phi(p(x, v, s)) = \Phi(v) \exp(\langle l(x, v, s), \partial_x c(\gamma(x, v, s)) \rangle s) > 0 \text{ for all } s \Leftrightarrow \Phi(v) > 0$$

From the fact that W is absorbing, we conclude that

$$p(x, v, s) \neq 0 \text{ for all } s \Leftrightarrow v \neq 0.$$

With the above observation, now we would get to the proof of our lemma:

1. $B(x, t)$ is well-defined:

Since for all $(\gamma, p) \in C^\infty$ satisfying (3.9) with $\gamma(x, v, r) = x, p(x, v, r) = v$, if $v \neq 0$, by the above statement, we have that $p(x, v, s) \neq 0$ for all $0 \leq s \leq r$, and therefore $\Phi(p(x, v, s))$ is smooth. Therefore $\partial_p \Phi(p(x, v, s))$ is well-defined for all s . By uniqueness and existence of ODE system, we have well-definedness of $B(x, t)$ for all $(x, t) \in \mathbb{R}^{d+1} \times (0, T)$.

2. $B(x, t) \subset C(x, t)$:

For $y \in B(x, t)$, there exists a pair of curves $(\gamma, p) \in C^\infty$ and $v \neq 0, 0 \leq r \leq t$ such that

$$\begin{aligned} \dot{\gamma}(x, v, s) &= \partial_p H(\gamma(x, v, s), p(x, v, s)), \\ \dot{p}(x, v, s) &= -\partial_x H(\gamma(x, v, s), p(x, v, s)), \\ \gamma(x, v, r) &= x, p(x, v, r) = v \end{aligned}$$

and $\gamma(0) = y$. Then $\dot{\gamma}(x, v, s) = \partial_p H(\gamma(x, v, s), p(x, v, s)) \in \partial W \in W$. Now let us define $\tilde{\gamma}$ as a rescaling by $\tilde{\gamma} := \gamma(x, v, sr/t)$ where $r/t \leq 1$, then we notice that $\tilde{\gamma} \in C^\infty, \dot{\tilde{\gamma}} \in r/t \partial W \in W$ and $\tilde{\gamma}(t) = x, \tilde{\gamma}(0) = y$. Now choose $\tilde{p}(s) = 0$ for all $0 \leq s \leq t$. Thus we have $(\tilde{\gamma}, \tilde{p}) \in C^\infty$ satisfying the differential inclusion defining $C(x, t)$ and $\tilde{\gamma}(t) = x, \tilde{p}(t) = 0$ initial value. Thus $y = \tilde{\gamma}(x, v, t) \in C(x, t)$.

3. $C(x, t) \subset B(x, t)$:

For all $y \in C(x, t)$, there exists $(\gamma, p) \in C^\infty$ such that

$$\dot{\gamma}(x, v, s) = \partial_p H(\gamma(x, v, s), p(x, v, s)), \quad \dot{p}(x, v, s) = -\partial_x H(\gamma(x, v, s), p(x, v, s))$$

and

$$\gamma(x, v, 0) = y, \gamma(x, v, t) = x, p(x, v, t) = v.$$

Since the metric space $(\mathbb{R}, \tilde{d})$ is complete (and locally compact), by the Hopf-Rinow-Cohn-Vossen theorem [11,26], there exists a (globally) length minimizing geodesic $\tilde{\gamma} \in C^\infty$ such that $\tilde{\gamma}(0) = y, \tilde{\gamma}(r) = x$, and $\tilde{d}(x, y) = r$ for some r . Notice that

$$r = \tilde{d}(x, y) \leq t$$

by definition since $\gamma(x, v, \cdot)$ satisfies the conditions $\gamma \in C^\infty, \gamma(0) := x, \gamma(t) = y, \dot{\gamma}(s) \in c(\gamma(s))W$. Consider the vector $l(s) := \dot{\tilde{\gamma}}(s)/c(\tilde{\gamma}(s))$. Now rewrite

$$\tilde{d}(x, y) = \inf_{\gamma \in C^\infty, t > 0, \gamma(0) := x, \gamma(t) = y} \int_0^t \left(1 + \mathcal{I}_W \left(\frac{\dot{\gamma}(s)}{c(\gamma(s))} \right) \right) ds.$$

From the fact that for any given $\gamma \in C^\infty$,

$$\frac{\dot{\gamma}(s) + \dot{h}(s)}{c(\gamma(s) + h(s))} = \frac{\dot{\gamma}(s)}{c(\gamma(s))} + \frac{\dot{h}(s)}{c(\gamma(s))} - \frac{\dot{\gamma}(s) \langle \nabla_x c(\gamma(s)), h(s) \rangle}{[c(\gamma(s))]^2} + O(|h(s)|^2 + |\dot{h}(s)|^2)$$

for all $h \in C_c^\infty$, we obtain directly from variational calculus that the following optimality condition that $\bar{\gamma}$ necessarily satisfies, for all perturbation $h \in C_c^\infty$:

$$\int_0^r \langle h(s), (\dot{\bar{p}}(s) + \nabla_x c(\bar{\gamma}(s)) \langle \bar{p}(s), l(s) \rangle) \rangle ds = 0 \quad \text{for some } \bar{p}(s) \in \partial^-[\mathcal{I}_W](l(s)).$$

By the definition of Legendre transformation, the above condition is equivalent to the existence of $\bar{p}$ such that

$$\dot{\bar{p}}(s) = -\nabla_x c(\bar{\gamma}(s)) \Phi(\bar{p}(s)), \quad l(s) \in \partial_p^- \Phi(\bar{p}(s)). \quad (3.10)$$

Now we would like to show that we may furthermore choose the dual vector $\bar{p}(s) \neq 0$ for some s (i.e. the strict complementary condition is satisfied and $\bar{p} \neq 0$ as a function). In order to do so, we would like to argue that if we enlarge our constraint set from W to $W^\epsilon := (1 + \epsilon)W$ for any $\epsilon > 0$, we still have a global optimizer but the optimizer differs from the original case when $\epsilon = 0$. In fact, for all $\epsilon > 0$, consider the problem

$$\tilde{d}^\epsilon(x, y) := \inf_{t > 0} \{t : \gamma \in C^\infty, \gamma(0) := x, \gamma(t) = y, \dot{\gamma}(s) \in c(\gamma(s))W^\epsilon\},$$

then by reparametrization of each curve $\gamma \in C^\infty$ by $a : s \mapsto s/(1 + \epsilon)$, we get that

$$\tilde{d}^\epsilon(x, y) = (1 + \epsilon)^{-1} \tilde{d}(x, y)$$

and the new metric space $(\mathbb{R}^d, \tilde{d}^\epsilon)$ is also complete, and by Hopf-Rinow-Cohn-Vossen theorem [11,26], we have a unique minimizing geodesic $\tilde{\gamma}^\epsilon$. Since the new optimization problem to obtain $\tilde{d}^\epsilon(x, y)$ comes from a rescaling of the original problem $\tilde{d}(x, y)$, we have $\tilde{\gamma}^\epsilon = \tilde{\gamma} \circ a$. Therefore the optimizer $\tilde{\gamma}^\epsilon \neq \tilde{\gamma}$ unless $\epsilon = 0$. Hence there is a dual vector $\bar{p}$ satisfying (3.10) that satisfies a strict complementary condition, i.e. $\bar{p}(s) \neq 0$ for some s . Now since $(\tilde{\gamma}, \bar{p})$ satisfies (3.9) and $\bar{p}(s) \neq 0$ for some s , one concludes that $\bar{p}(s) \neq 0$ for all s , and therefore $\partial_p^- \Phi(\bar{p}(s)) = \partial_p \Phi(\bar{p}(s))$. Hence there exists $(\tilde{\gamma}, \bar{p}) \in C^\infty$ that satisfies the bi-characteristic equation and $\tilde{\gamma}(0) = y$, $\tilde{\gamma}(r) = x$ and $0 \leq r \leq t$. Hence $y \in B(x, t)$. $\square$

Corollary 3.9. Assume Conjecture 3.7, i.e. (1.1), and Remark 1 are true. Assume the metric space $(\mathbb{R}, \tilde{d})$ is complete. Then

$$\varphi(x, t) = \inf_{v \in \mathbb{R}^d \setminus \{0\}} \min_{0 \leq r \leq t} \left\{ g(\gamma(x, v, r)) : \begin{aligned} \dot{\gamma}(x, v, s) &= -c(\gamma(x, v, s)) \partial_p \Phi(p(x, v, s)), \\ \dot{p}(x, v, s) &= \partial_x c(\gamma(x, v, s)) \Phi(p(x, v, s)), \\ \gamma(x, v, 0) &= x, p(x, v, 0) = v \end{aligned} \right\}. \quad (3.11)$$

Proof. Apply Lemma 3.8 and a change of variable from v to $-v$ and from s to $t - s$. $\square$

We wish to remind that if $\Phi(p) = |p|_2$, we have $\partial_p \Phi(p(v, s)) = p(v, s)/|p(v, s)|_2$. However, notice this example satisfies the assumption in Lemma 3.2, but does not satisfy the assumptions in Lemma 3.4. In fact the formula (3.11) is the Huygens principle in disguise.

3.4. Generalized Hopf maximization with Hamiltonian H that are possibly neither convex nor concave, but the initial data g is convex

When the Hamiltonian H is neither convex nor concave, but the initial data g is convex, the aforementioned conjectured minimization principle does not seem to hold any longer. In light of the fact that in some special cases a Hopf formula holds, we conjecture that a generalized Hopf-type maximization principle shall hold for a wide class of problem.

We now describe how we get to a Hopf formula under restricted assumption by Lemma 3.4 and Lemma 3.5:

$$\varphi(x, t) := U(x, T - t) := \inf_{u \in \mathcal{U}(T-t)} \left\{ \int_{T-t}^T h(s, x(s), u(s)) ds + g(x(T)) \right\}.$$

We again derive our formula formally as follows. Again, following [40], we shall first consider the following discretization (approximation) for a given δ such that $\delta N = t$, by denoting $s_n = \delta n$ and $x_N = x$

$$\varphi(x, t) \approx \varphi_N^2(x, t)$$

where

$$\begin{aligned} & \varphi_N^2(x, t) \\ &:= \min_{\{u_n\}_{n=1}^N \in U} \left\{ g(x_0) + \delta \sum_{n=1}^N h(T - s_n, x_n, u_n) : x_{n+1} - x_n = -\delta f(T - s_{n+1}, x_{n+1}, u_{n+1}) \text{ for } n = 0, \dots, N-1, x_N = x \right\}, \\ &= \inf_{\{x_n\}_{n=0}^{N-1}} \sup_{\{p_n\}_{n=1}^N} F_2(\mathbf{X}, \mathbf{U}, \mathbf{P}), \end{aligned}$$

where F_2 is defined as in Lemma 3.4 and the second equality comes from reformulating the problem with Lagrange multiplier.

Now again if we apply Lemma 3.4 and Lemma 3.5, we get the following:

Theorem 3.10. *If (A1), (A2), (A3), (A4) and (A5) are satisfied, then we have then we have*

$$\varphi_N^2(x, t) = \sup_{\{p_n\}_{n=1}^N} \inf_{\{x_n\}_{n=1}^{N-1}} \tilde{F}_2(\tilde{\mathbf{X}}, \mathbf{P}). \quad (3.12)$$

If (A1), (A2), (A4), (A5), (A6), (A7) are satisfied, then we have

$$\begin{aligned} \varphi_N^2(x, t) = \sup_{v \in \mathbb{R}^d} & \left\{ \langle p_N, x \rangle - g^*(p_1) - \delta \sum_{n=1}^N H(x_n, p_n, s_n) + \delta \sum_{n=1}^{N-1} \left\langle \frac{p_n - p_{n+1}}{\delta}, x_n \right\rangle : \right. \\ & \left. \begin{aligned} x_n(x, v) - x_{n-1}(x, v) &= \delta \partial_p H(x_n(x, v), p_n(x, v), t_n), \\ p_n(x, v) - p_{n+1}(x, v) &= \delta \partial_x H(x_n(x, v), p_n(x, v), t_n), \\ x_N &= x, p_N = v \end{aligned} \right\} \end{aligned}$$

Again, nonetheless, we notice that the resulting formula that we conjectured seems to be correct beyond these assumptions, as the numerical results show (again especially when we take the maximum over all the paths satisfying the KKT conditions). We hope to get some rigorous criteria for this formula to hold in the future.

In order to state our conjecture, in view of [50], let us define the following before we proceed.

Definition 3.11. Given a set $S \in \mathbb{R}^d$. A function $G : \text{co}(S) \rightarrow \mathbb{R}$ is pseudoconvex on S , where $\text{co}(S)$ is the convex hull of S , if for all $\{s_i\}_{i=1}^{d+1} \in S$ and $\{\alpha_i\}_{i=1}^{d+1}$ such that $\alpha_i \geq 0$ and $\sum_{i=1}^{d+1} \alpha_i = 1$, we have

$$G\left(\sum_{i=1}^{d+1} \alpha_i s_i\right) \leq \sum_{i=1}^{d+1} \alpha_i G(s_i).$$

Now, passing N to the limit, we arrive at the following conjectured Hopf formula:

Conjecture 3.12. *Assume $H(x, p, t) \in C^2$ and $g(p) \in C^2$ is convex w.r.t. p that satisfies (A5). Consider the set*

$$\begin{aligned} & S_0(x, t) \\ &:= \operatorname{argmax}_{v \in \mathbb{R}^d} \left\{ \langle x, v \rangle - g^*(p(x, v, 0)) - \int_0^t \left\{ H(\gamma(x, v, s), p(x, v, s), s) - \langle \partial_x H(\gamma(x, v, s), p(x, v, s), s), \gamma(x, v, s) \rangle \right\} ds : \right. \\ & \quad \left. \begin{aligned} \dot{\gamma}(x, v, s) &= \partial_p H(\gamma(x, v, s), p(x, v, s), s), \\ \dot{p}(x, v, s) &= -\partial_x H(\gamma(x, v, s), p(x, v, s), s), \\ \gamma(x, v, t) &= x, p(x, v, t) = v \end{aligned} \right\}. \end{aligned}$$

Assume further that for all (x, t) , we have that the function

$$\begin{aligned} G : \text{co}(S_0(x, t)) &\rightarrow \mathbb{R} \\ v &\mapsto H(x, v, s) - \langle \partial_x H(x, v, t), x \rangle \end{aligned}$$

is pseudoconvex on $S_0(x, t)$. Then there exists t_0 such that the viscosity solution to (2.1)–(2.2) can be represented as (1.2) for $t < t_0$. Moreover, if $\phi(x, t)$ is differentiable w.r.t. x at a neighborhood of (x, t) and the infimum is attained by $\tilde{v}$, then we have $\partial_x \phi(x, t) = \tilde{v}$.

We may refer to it as the **maximization principle**, or the generalized Hopf formula. One point to remark is that the conjectured Hopf formula is a generalization of the well-known Hopf formula in [18,25,50]. A proof that our formula holds under less restricted assumption that shown above is also done in [62], after a previous version of our paper [8] was

arXiv-ed. A proof that our formula holds under less restricted assumption that shown above is also done in [62], after a previous version of our paper [8] was arXiv-ed. We suspect the weakest assumption of our formula to hold is the above assumption of pseudoconvexity, similar to one stated in [50].

Remark 2. When $H(x, p, t)$ is not differentiable at some given point p , then we believe again that in formula (1.2), the Mordukhovich subdifferential, $\partial_x^- H$, as defined in [36–38], should be used instead of $\partial_x H$. In that case the constraint becomes the inclusion $\dot{\gamma}(x, v, s) \in \partial_p^- H(\gamma(x, v, s), p(x, v, s), s)$, and infimum is taken over also all the curves $(\gamma, p) \in C^\infty$ satisfying the inclusion before supremum over v is taken.

Remark 3. We expect a candidate of less restricted assumption than [62] as the above convexity assumption in the variable p in view of the theorem in [50]. A rigorous approach toward this formula might be following an approach of [50] to show the postulated formula is a minimax viscosity solution following the notations in [50] and references therein.

Below we present several examples where this conjecture is valid. We notice that **none** of the followings satisfies the assumptions of Lemma 3.4 and Lemma 3.5 given in the proof of our formula, since by Lemma 3.1, the Hamiltonian $H(x, p, t)$ given as in (3.1) under our assumptions are automatically convex.

Example 1. When $H(x, p, t) = H(p, t)$, we have $\dot{p}(v, s) = \partial_x H(\gamma(v, s), p(v, s), s) = 0$. Therefore assuming Conjecture 3.12, i.e. (1.2), and Remark 2, the conjectured formula gives

$$\begin{aligned} \varphi(x, t) &= - \inf_{v \in \mathbb{R}^d} \left\{ g^*(p(x, v, 0)) + \int_0^t \left\{ H(p(x, v, s), s) - \langle 0, \gamma(x, v, s) \rangle \right\} ds - \langle x, v \rangle : \begin{array}{l} (\gamma, p) \in C^\infty \\ \dot{\gamma}(x, v, s) \in \partial_p^- H(p(x, v, s), s), \\ \dot{p}(x, v, s) = 0, \\ \gamma(x, v, t) = x, \\ p(x, v, t) = v \end{array} \right\} \\ &= - \inf_{v \in \mathbb{R}^d} \left\{ g^*(v) + \int_0^t H(v, s) ds - \langle x, v \rangle \right\} \end{aligned}$$

which gets us back to the Hopf formula [50]. Note that in this example assumption in Lemma 3.4 or in Lemma 3.5 may not be satisfied since U may not be compact.

Example 2. When $H(x, p, t)$ is a **non-convex** homogeneous degree-1 functional w.r.t. p in the following form:

$$H(x, p, t) = c_1(x, t)\Phi_1(p) - c_2(x, t)\Phi_2(p)$$

where Φ_1 and Φ_2 are with their Wulff sets W_1, W_2 [43] as strictly convex set with smooth boundary $\partial W_i \in C^\infty, i = 1, 2$, then by definition of Mordukhovich subdifferential, we have

$$\partial_p^- H(x, p, t) = \begin{cases} c_1(x, t)\partial_p \Phi_1(p) - c_2(x, t)\partial_p \Phi_2(p) & \text{if } p \neq 0 \\ \emptyset & \text{if } p = 0. \end{cases}$$

Therefore assuming Conjecture 3.12, i.e. Equation (1.2) and Remark 2, we have:

$$\begin{aligned} \varphi(x, t) &= - \inf_{v \in \mathbb{R}^d \setminus \{0\}} \left\{ g^*(p(x, v, 0)) + \int_0^t \left\{ \left(c_1(\gamma(x, v, s), s) - \langle \partial_x c_1(\gamma(x, v, s), s), \gamma(x, v, s) \rangle \right) \Phi_1(p(x, v, s)) \right. \right. \\ &\quad \left. \left. - \left(c_2(\gamma(x, v, s), s) - \langle \partial_x c_2(\gamma(x, v, s), s), \gamma(x, v, s) \rangle \right) \Phi_2(p(x, v, s)) \right\} ds - \langle x, v \rangle : \right. \\ &\quad \left. \begin{array}{l} \dot{\gamma}(x, v, s) = c_1(\gamma(x, v, s), s)\partial_p \Phi_1(p(x, v, s)) - c_2(\gamma(x, v, s), s)\partial_p \Phi_2(p(x, v, s)) \\ \dot{p}(x, v, s) = -\partial_x c_1(\gamma(x, v, s), s)\Phi_1(p(x, v, s)) - \partial_x c_2(\gamma(x, v, s), s)\Phi_2(p(x, v, s)), \\ \gamma(x, v, t) = x, p(x, v, t) = v \end{array} \right\}. \end{aligned} \quad (3.13)$$

When $H(x, p, t) = -c(x)|p|_2$, we have, from (3.13) and integration by parts, that

$$\varphi(x, t) = \sup_{v \in \mathbb{R}^d \setminus \{0\}} \left\{ g(\gamma(x, v, t)) : \begin{array}{l} \dot{\gamma}(x, v, s) = c_2(\gamma(x, v, s))p(x, v, s)/|p(x, v, s)|_2 \\ \dot{p}(x, v, s) = -\partial_x c_2(\gamma(x, v, s))|p(x, v, s)|_2, \\ \gamma(x, v, 0) = x, p(x, v, 0) = v \end{array} \right\}.$$

This is again Huygens principles in disguise. In these examples, assumption in Lemma 3.4 or in Lemma 3.5 are not satisfied.

When $H(x, p, t)$ is, on the other hand, a **convex** homogeneous degree-one functional w.r.t. p , i.e. when $c_2(x, t) = 0$, and hence $H(x, p, t) = c(x, t)\Phi(p)$ for some $c > 0$ and Φ , then assuming Conjecture 3.12, i.e. Equation (1.2) and Remark 2, we obtain

$$\begin{aligned} \varphi(x, t) &= - \inf_{v \in \mathbb{R}^d \setminus \{0\}} \left\{ g^*(p(x, v, 0)) + \int_0^t \left\{ \left(c(\gamma(x, v, s), s) - \langle \partial_x c(\gamma(x, v, s), s), \gamma(x, v, s) \rangle \right) \Phi(p(v, s)) \right\} ds - \langle x, v \rangle : \right. \\ &\quad \left. \begin{aligned} &(\gamma, p) \in C^\infty \\ &\dot{\gamma}(x, v, s) \in c(\gamma(x, v, s), s) \partial_p^- \Phi(p(x, v, s)) \\ &\dot{p}(x, v, s) = -\partial_x c(\gamma(x, v, s), s) \Phi(p(x, v, s)), \\ &\gamma(x, v, t) = x, p(x, v, t) = v \end{aligned} \right\}. \end{aligned}$$

Note that in this example assumptions in Lemma 3.4 are satisfied but that of Lemma 3.5 are not satisfied.

Example 3. When the HJ PDE comes from an ODE system in a differential game, we have the following finite horizon problem with initial state $x \in \mathbb{R}^d$. We consider the (Lipschitz) solution $x : [t, T] \rightarrow \mathbb{R}^d$ of the following linear dynamic system with initial condition x at time t :

$$\begin{cases} \frac{dx}{ds}(s) = Mx(s) + N_C(s)a(s) + N_D(s)b(s) & \text{in } (t, T) \\ x(t) = x \end{cases}$$

where M is a given $d \times d$ matrix independent of time, and $\{N_C(s)\}_{t < s < T}$, $\{N_D(s)\}_{t < s < T}$ are two families of $d \times d$ matrices with real entries. Using the notation in section 2.2, if we let

$$\begin{aligned} f(t, x, a, b) &= Mx(s) + N_C(s)a + N_D(s)b \\ h(t, x, a, b) &= -\mathcal{I}_{C(t)}(p) + \mathcal{I}_{D(t)}(p), \end{aligned}$$

for some family of convex sets $\{C(s)\}_{t < s < T} \subset C$, $\{D(s)\}_{t < s < T} \subset D$, our Hamiltonian read:

$$\tilde{H}^\pm(x, p, t) = \max_{b \in D(t)} \min_{a \in C(t)} \{-\langle Mx + N_C(t)a + N_D(t)b, p \rangle\} = -\langle Mx, p \rangle + \Phi_{C(t)}(-N_C^*(t)p) - \Phi_{D(t)}(-N_D^*(t)p),$$

where

$$\Phi_W(p) := \max_{x \in W} \langle x, p \rangle.$$

After a change of variable to get from a finite time PDE problem $(-\infty, T)$ to an initial time PDE problem $(0, \infty)$, we have the upper/lower values $\varphi(x, t) := U(x, T - t) = V(x, T - t)$ satisfy:

$$\begin{cases} \frac{\partial}{\partial t} \varphi + H(x, \nabla_x \varphi, t) = 0 & \text{on } \mathbb{R}^d \times (0, \infty), \\ \varphi(x, 0) = g(x) & \text{on } \mathbb{R}^d \end{cases}$$

where the Hamiltonian H is now

$$H(x, p, t) = \tilde{H}^\pm(x, p, T - t) = -\langle Mx, p \rangle + \Phi_{C(T-t)}(-N_C^*(T-t)p) - \Phi_{D(T-t)}(-N_D^*(T-t)p).$$

Now, notice that H is smooth w.r.t. x , assuming Conjecture 3.12, i.e. Equation (1.2) and Remark 2, together with the fact that $\partial_x H(x, p, t) = -M^*p$ and a change of variable, we get the same formula as in (2.5) in [9]. In fact, the generalized Hopf formula gives

$$\begin{aligned} \varphi(x, t) &= - \min_{p \in \mathbb{R}^d} \left\{ g^*(e^{-M^*T}p) + \int_0^t \left\{ \Phi_{C(T-s)} \left([-e^{-M(T-s)} N_C(T-s)]^* p \right) - \Phi_{D(T-s)} \left([-e^{-M(T-s)} N_D(T-s)]^* p \right) \right\} ds \right. \\ &\quad \left. - \langle e^{-M(T-t)}x, p \rangle \right\}. \end{aligned}$$

If we write $z = e^{-M(T-t)}x$ and write $J(z) = g(e^{MT}z)$, then this change of coordinate in the state variable by e^{MT} gives rise to a symplectic change of coordinate in the phase variables by $\text{diag}(e^{MT}, e^{-M^*T})$, hence

$$J^*(p) = g^*(e^{-M^*T}p),$$

and therefore we get to

$$\begin{aligned} & \varphi(x(z), t) \\ &= -\min_{p \in \mathbb{R}^d} \left\{ J^*(p) + \int_0^t \left\{ \Phi_{C(T-s)} \left([-e^{-M(T-s)} N_C(T-s)]^* p \right) - \Phi_{D(T-s)} \left([-e^{-M(T-s)} N_D(T-s)]^* p \right) \right\} ds - \langle z, p \rangle \right\} \end{aligned}$$

which is the same formula as in (2.5) in [9]. Note that in this example, both the assumptions in Lemma 3.4 and Lemma 3.5 are satisfied.

4. Numerical methods

4.1. Optimization methods: coordinate descent

In order for computation of optimization (in either the Lax formulation or the Hopf formulation) to be efficient, we have recast the initial value HJ PDE problem to minimization problem in d dimensions, where the curves γ and p inside the function evaluation (given a vector v) are defined explicitly as the solutions to the bi-characteristic equation. We suggest to solve the ODE's numerically, given a pair of (x, v) , up to time t using any ODE solver. This way the minimization/maximization problem reduces to optimization of a finite-dimensional problem (as a function of v). Similar to [9], we suggest to apply coordinate descent to the following functionals with argument v for a given pair (x, t) :

$$\mathcal{F}_{x,t}^1(v) := g(\gamma(x, v, t, 0)) + \int_0^t \left\{ \langle p(x, v, t, s), \partial_p H(\gamma(x, v, t, s), p(x, v, t, s), s) \rangle - H(\gamma(x, v, t, s), p(x, v, t, s), s) \right\} ds$$

or (noticed we omitted r in this functional, but r is actually the final time of the ODE (γ, p) stated below)

$$\mathcal{F}_{x,t}^2(v) := \min_{0 \leq r \leq t} \{g(\gamma(x, v, r, 0))\}$$

or

$$\begin{aligned} \mathcal{G}_{x,t}(v) &:= g^*(p(x, v, t, 0)) + \int_0^t \left\{ H(\gamma(x, v, t, s), p(x, v, t, s), s) - \langle \partial_x H(\gamma(x, v, t, s), p(x, v, t, s), s), \gamma(x, v, t, s) \rangle \right\} ds \\ &\quad - \langle x, v \rangle \end{aligned}$$

where in either case, the pair $\gamma(x, v, t, s), p(x, v, t, s)$ solves the following final value problem for the given pair of x, t (with the dependence of the curves w.r.t. x, t clearly indicated in the notion):

$$\begin{cases} \dot{\gamma}(x, v, t, s) = \partial_p H(\gamma(x, v, t, s), p(x, v, t, s), s), \\ \dot{p}(x, v, t, s) = -\partial_x H(\gamma(x, v, t, s), p(x, v, t, s), s), \\ \gamma(x, v, t, t) = x, \\ p(x, v, t, t) = v. \end{cases}$$

Now, to minimize $\mathcal{F}_{x,t}^1(\cdot)$, $\mathcal{F}_{x,t}^2(\cdot)$ or $\mathcal{G}_{x,t}(\cdot)$, we utilize a cyclic coordinate descent algorithm. We illustrate our algorithm with the functional $\mathcal{G}_{x,t}(\cdot)$:

Algorithm 1. Take an initial guess of the Lipschitz constant L , and set $count := 0$. Initialize $j_1 := 1$ and a parameter $\alpha := 1/L$. For $k = 1, \dots, M$, do:

1:

$$\begin{cases} v_i^{k+1} = v_i^k - \alpha \partial_i \mathcal{G}_{x,t}(v^k) & \text{if } i = j_k, \\ v_i^{k+1} = v_i^k & \text{otherwise.} \end{cases}$$

2:

$$j_{k+1} := j_k + 1.$$

If $j_{k+1} = d + 1$, then reset $j_{k+1} = 1$.

3: If $|v_i^{k+1} - v_i^k| > \varepsilon$, then set $count := 0$. If $k = M$, then reset $k := 0$ and set $\alpha := \alpha/2$ (i.e. let $L := 2L$).

4: If $|v^{k+1} - v^k| < \varepsilon$, set $\text{count} := \text{count} + 1$.
 5: If $\text{count} = d$, stop.

Return $v_{\text{final}} = v^{k+1}$.

We minimize $\mathcal{F}_{x,t}^1(\cdot)$, $\mathcal{F}_{x,t}^2(\cdot)$ in a similar fashion. In this algorithm, we will need to discuss how to evaluate the functional values and also their numerical derivatives. This will be discussed in the next subsection.

4.2. Evaluation of functional and its derivatives: ODE solver, numerical differentiation and integration

In this subsection, we discuss several numerical approximation used in our numerical experiments. First, in our step, we need to devise by ODE solvers. We suggest to use the standard forward Euler solver for a given stepsize Δs . Of course the performance can be improved by using more advanced solvers, such as the pseudo-spectral method, etc. We also need to deal with approximating integrals. Similar to [9], we suggest to evaluate either the derivatives of $\mathcal{F}_{x,t}^1$ or $\mathcal{G}_{x,t}$ by numerical quadrature rules for integral computations. As for $\mathcal{F}_{x,t}^2$, we compute the maximum by directly choosing the maximum around the computed grid from the ODE solver in the case when $H(x, p, t) = H(x, p)$, since the ODE will be the same in this case and the ODE solver will get to the function values of $\gamma(v, s)$ easily.

$$\mathcal{F}_{x,t}^1(v) \approx \min_{r_i=i\Delta s, i=0, \dots, t/\Delta s} \{g(\gamma(x, v, r_i, 0))\} \quad (4.1)$$

with a same choice of Δs as the ODE solver. We also suggest, as in [9], a numerical differentiation rule for derivative computations. We integrate using a standard rectangular quadrature rule (we use $\mathcal{F}_{x,t}^1$ to illustrate):

$$\begin{aligned} \mathcal{F}_{x,t}^1(v) &\approx g(\gamma(x, v, t, 0)) \\ &+ \sum_i \{ \langle p(x, v, t, i\Delta s), \partial_p H(\gamma(x, v, t, i\Delta s), p(x, v, t, i\Delta s), i\Delta s) \rangle - H(\gamma(x, v, t, i\Delta s), p(x, v, i\Delta s), t, i\Delta s) \} \Delta s \end{aligned} \quad (4.2)$$

again with the same choice of Δs as in the ODE solver. We suggest approximating the partial derivative $\partial_i \mathcal{F}_{x,t}(p)$ (which means the differentiation of the function w.r.t. the direction e_i) by a finite difference:

$$\partial_i \mathcal{F}_{x,t}(v) \approx \frac{\mathcal{F}_{x,t}^1(v + \sigma e_i) - \mathcal{F}_{x,t}^1(v)}{\sigma} \quad (4.3)$$

with a given choice of σ . By using numerical differentiation, we have the advantage of not necessarily handling tedious analytic computations of the derivative of Hamiltonian which might be singular at times. Also, we only have two evaluations of the function value per iteration. By performing numerical approximations, either ODE solvers, differentiation or integration, we are bound to introduce numerical errors. These errors introduced by numerical approximation can be effectively controlled by choosing appropriately small sizes of Δs and σ .

4.3. Certificate of correctness

The method we compute a sequence is guaranteed to converge to a local minimum under an assumption of lower semi-continuity and boundedness of the functional. However such a descent type algorithm cannot guarantee convergence to global optimal. However, we can check the correctness of the vector v that is computed $p(0) \in \partial g(\gamma(0))$. With this certificate, in case a local optimal does not satisfy the assumption, we can discard the value thus computed and restart the algorithm with another initial guess.

5. Numerical results

In this section, we provide numerical experiments which compute viscosity solutions to HJ PDE with a time-dependent Hamiltonian arising from control system. For a given set of points (t, z) , we use **Algorithm 1** to compute (2.1)–(2.2). We set $M = 500$ and have a different initial guess of the Lipschitz constant L in each example. We evaluate (x, t) in a given set of grid points over some 2 dimensional cross-sections of the form $[-3, 3]^2 \times \{0\}^{d-2}$. We choose our error tolerance in the coordinate descent iteration as $\varepsilon = 0.5 \times 10^{-7}$, which acts as our stopping criterion. The step-size in the numerical quadrature rule in (4.1), (4.2) as well as the forward Euler ODE solver is set to be different in each example, and they are all denoted as Δs . The step-size for numerical differentiation in (4.3) denoted by σ , and the Lipschitz constants L in **Algorithm 1** are also chosen differently in each example. In all our examples, we set random initial starting points uniformly

distributed in $[-2, 2]^d$. We always consider the initial value to be a function with zero level set as an ellipse enclosed by the equation $\langle x, Ax \rangle = 1$ where $A^{-1} = \text{diag}(1, 25/4, 1/4, 1/4, \dots, 1/4)$, i.e. our initial condition for the HJ PDE is

$$g(x) = \frac{1}{2}(\langle x, Ax \rangle - 1)$$

with the aforementioned A . For a convex Hamiltonian, we make one initial guess, and for a non-convex Hamiltonian, we perform at most 20 independent trials of random initial guesses to get rid of possible local minima or in places when the derivative of the viscosity solution does not exist. We present our run-time in the format of $a \times 10^{-c} s \times k$ where k is the number of initial guesses made in the respective example. When $d = 2$, clear comparison is performed with our solution to the solution computed by a first order Lax-Friedrichs monotone scheme [41] with $\Delta t = 0.001$ and $\Delta x = 0.005$. We perform such comparison because we do not have an explicit solution for neither one of our examples, and therefore Lax-Friedrichs scheme (which has theoretical convergence guarantee) is used to compare with the solution we computed using the minimization/maximization principles. Our algorithm is implemented in C++ on an 1.7 GHz Intel Core i7-4650U CPU. In what follows, we present some examples.

Example 1. We solve for the state-dependent Hamiltonian of the linear form

$$H(x, p, t) = -0.2c(x) - \nabla c(x) \cdot p,$$

where

$$c(x) = 1 + 3 \exp(-4|x - (1, 1, 0, 0, \dots, 0)|_2^2).$$

The minimization principle (1.1) is used to compute the solution φ . The temporal stepsize is chosen as $\Delta s = 0.02$. In this example, since the Hamiltonian is linear, there exists only a **unique** v such that the constraints of (γ, p) in the bi-characteristics are satisfied, and therefore there is nothing to optimize, thus no choices of σ and L are necessary. Fig. 1 (left) gives the solutions when $d = 2$, $T = 0.12$. The runtime using C++ is $9.929 \times 10^{-7} s \times 1$ per point. Fig. 1 (right) is the solution computed by the Lax-Friedrichs scheme, which illustrates that the two solutions coincide. Fig. 2 gives the solutions when $d = 1024$, $T = 0.5$. The runtime using C++ is $4.342 \times 10^{-2} s \times 1$ per point. The computation time is reasonably fast for a 1024 dimensional problem (although in this toy example, there is no optimization to solve). This example satisfies the assumptions in Lemma 3.2, but does not satisfy the assumptions in Lemma 3.3, Lemma 3.4 and in Lemma 3.5.

Example 2. Next we solve for the state-dependent Hamiltonian, the well-known harmonic oscillator:

$$H^\pm(x, p, t) = \pm \frac{1}{2}(|p|_2^2 + |x|_2^2).$$

The maximization principle (1.2) is used to compute the solution φ for both of the $\pm$ cases. The temporal stepsize is chosen as $\Delta s = 0.02$. The other constants are chosen as follows: stepsize $\sigma = 0.001$ and $L = 3$. Fig. 3 (left) gives the solutions when $d = 2$ and sign of H is negative, i.e. H^- , $T = 0.5$. The runtime using C++ is $2.366 \times 10^{-5} s \times 5$ per point. Fig. 3 (right) is the solution computed by the Lax-Friedrichs scheme for comparison. Fig. 4 (left) gives the solutions when $d = 2$ and sign of H is positive, i.e. H^+ , $T = 0.5$. The runtime using C++ is $2.605 \times 10^{-5} s \times 5$ per point. Fig. 4 (right) is the solution computed by the Lax-Friedrichs scheme for comparison. Fig. 6 gives the solutions when $d = 7$ and sign of H is negative, i.e. H^- , $T = 0.2$. The runtime using C++ is $3.717 \times 10^{-4} s \times 5$ per point. The computation time is minimal for a 7 dimensional problem.

In order to test for an example with convex Hamiltonian but a non-convex initial condition, we take another initial function $g(x)$ as the following Rosenbrock function:

$$g(x) = 0.4 \times 10^{-3} \times \left(-100 + (1 - x_1)^2 + 100(1 + x_2 - x_1^2)^2 \right). \quad (5.1)$$

The Hamiltonian is chosen as H^+ which is convex, and $T = 0.5$. We choose $d = 2$ for comparison. Fig. 5 (left) gives the solutions using the minimization principle. The runtime using C++ is 3.847×10^{-5} per point. Fig. 5 (right) is the solution computed by the Lax-Friedrichs scheme for comparison. To remark, both the convex case H^+ and nonconvex case H^- in this example does not satisfy the assumptions of any of the lemmas proved in this work since U may not be compact.

Example 3. We solve for the state-dependent Hamiltonian of the form

$$H^\pm(x, p, t) = \pm c(x)|p|_2,$$

where

$$c(x) = 1 + 3 \exp(-4|x - (1, 1, 0, 0, \dots, 0)|_2^2).$$

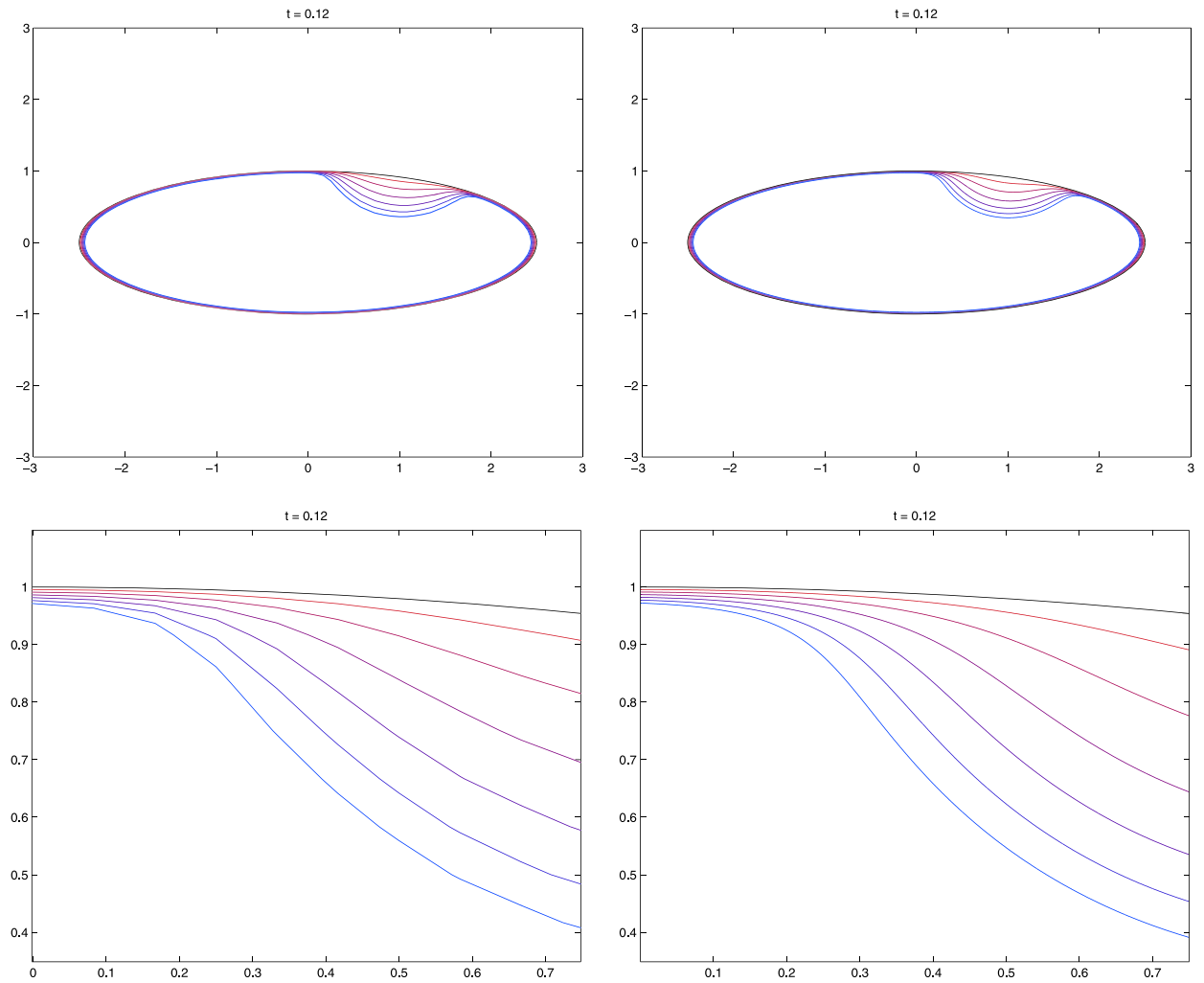


Fig. 1. Zero level sets of the solution $\phi(\cdot, t)$ for $t = 0.02, 0.04, \dots, 0.12$ in **Example 1** with $d = 2$; left: minimization/maximization principle, right: Lax-Friedrichs.

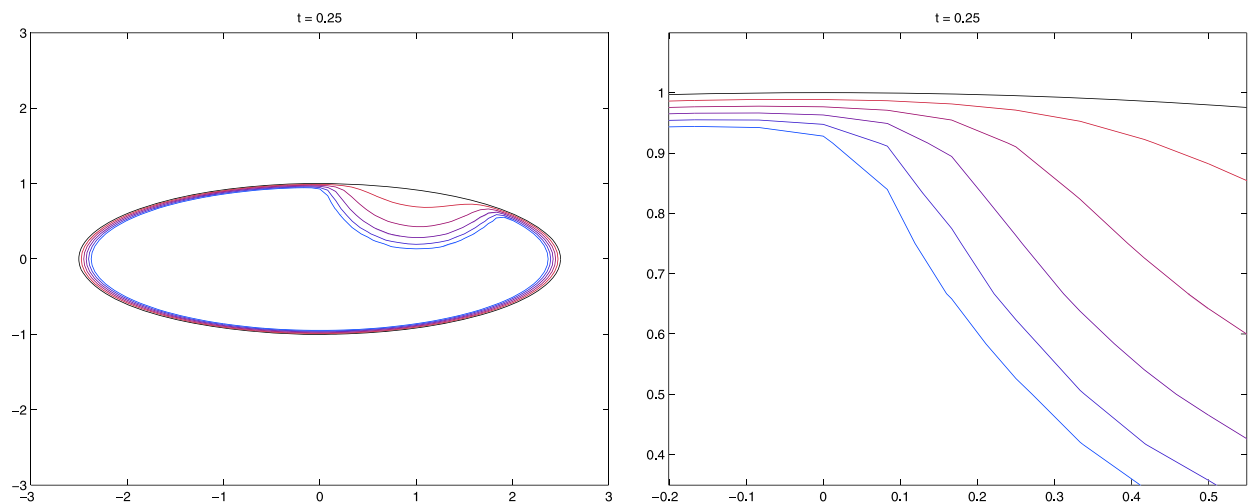


Fig. 2. Zero level sets of the solution $\phi(\cdot, t)$ for $t = 0.05, 0.1, \dots, 0.25$ by minimization/maximization principle in **Example 1** with $d = 1024$; left: full-size; right: close-up.

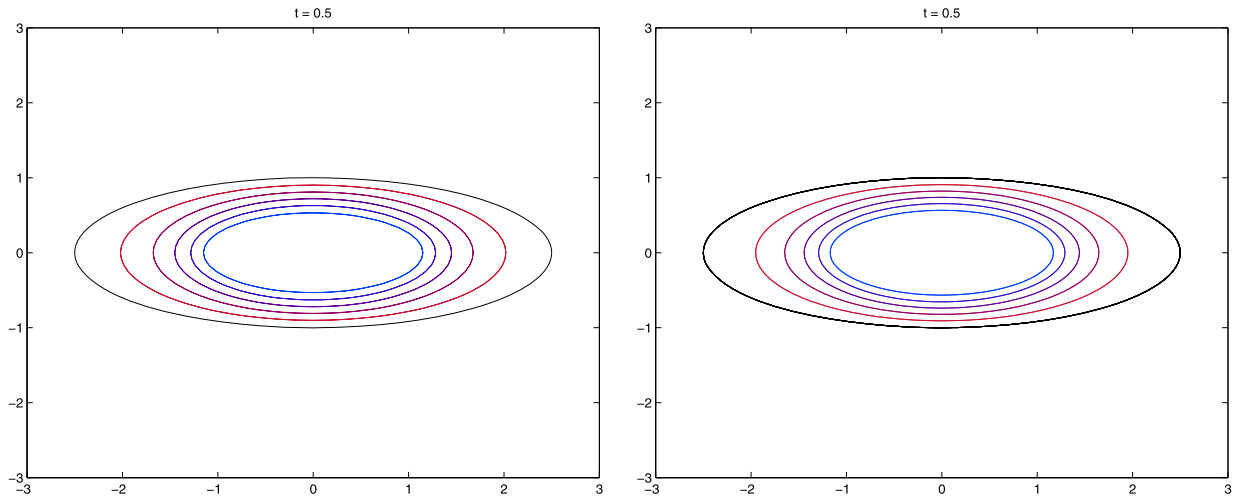


Fig. 3. Zero level sets of the solution $\phi(\cdot, t)$ for $t = 0.1, 0.2, \dots, 0.5$ by minimization/maximization principle in **Example 2** with H^- and $d = 2$; top: full-size; bottom: close-up.

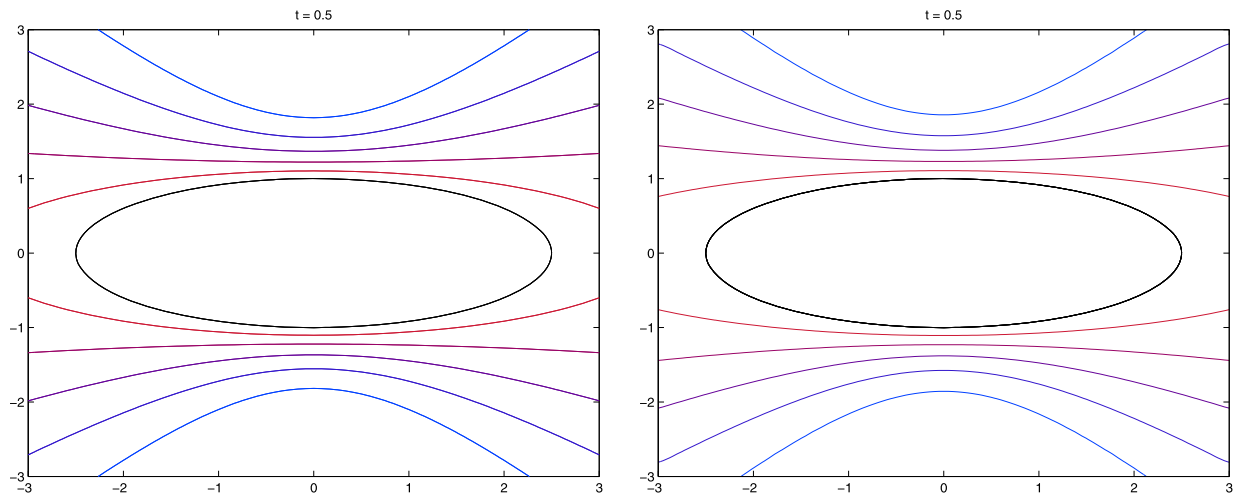


Fig. 4. Zero level sets of the solution $\phi(\cdot, t)$ for $t = 0.1, 0.2, \dots, 0.5$ by minimization/maximization principle in **Example 2** with H^+ and $d = 2$; top: full-size; bottom: close-up.

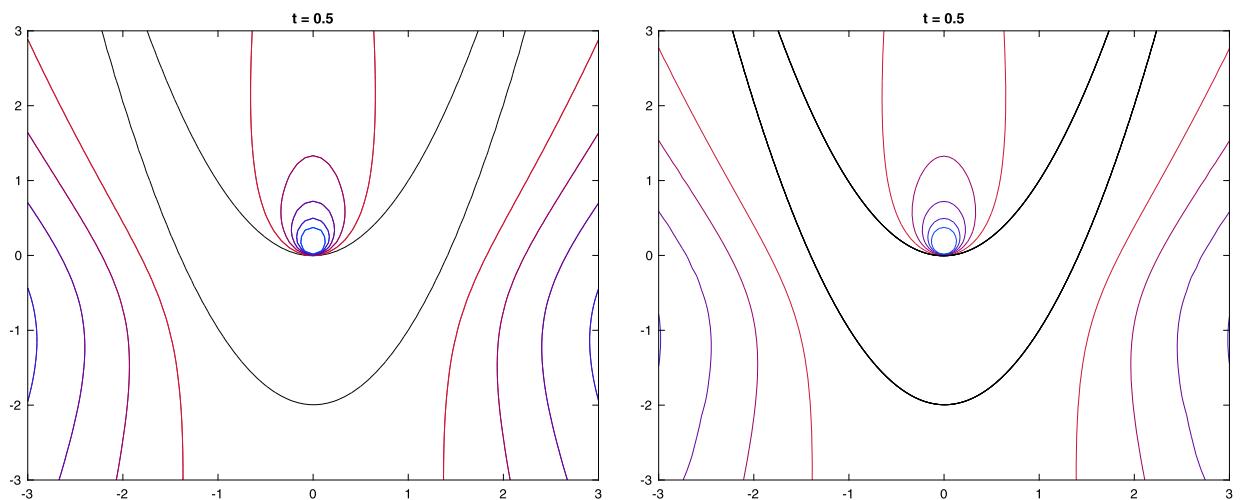


Fig. 5. Zero level sets of the solution $\phi(\cdot, t)$ for $t = 0.1, 0.2, \dots, 0.5$ by minimization/maximization principle in **Example 2** with H^+ , $d = 2$ and non-convex initial data (5.1); top: full-size; bottom: close-up.

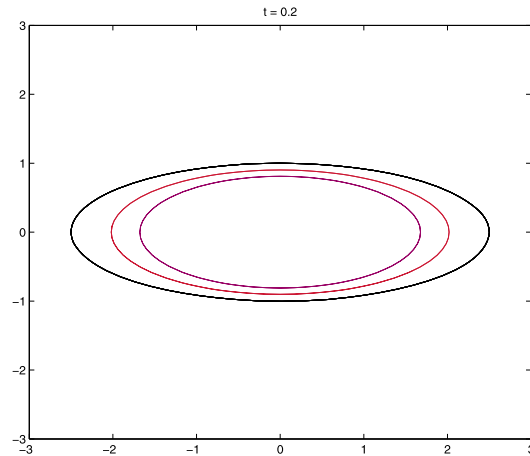


Fig. 6. Zero level sets of the solution $\phi(\cdot, t)$ for $t = 0.1, 0.2$ by minimization/maximization principle in **Example 2** with H^- and $d = 7$.

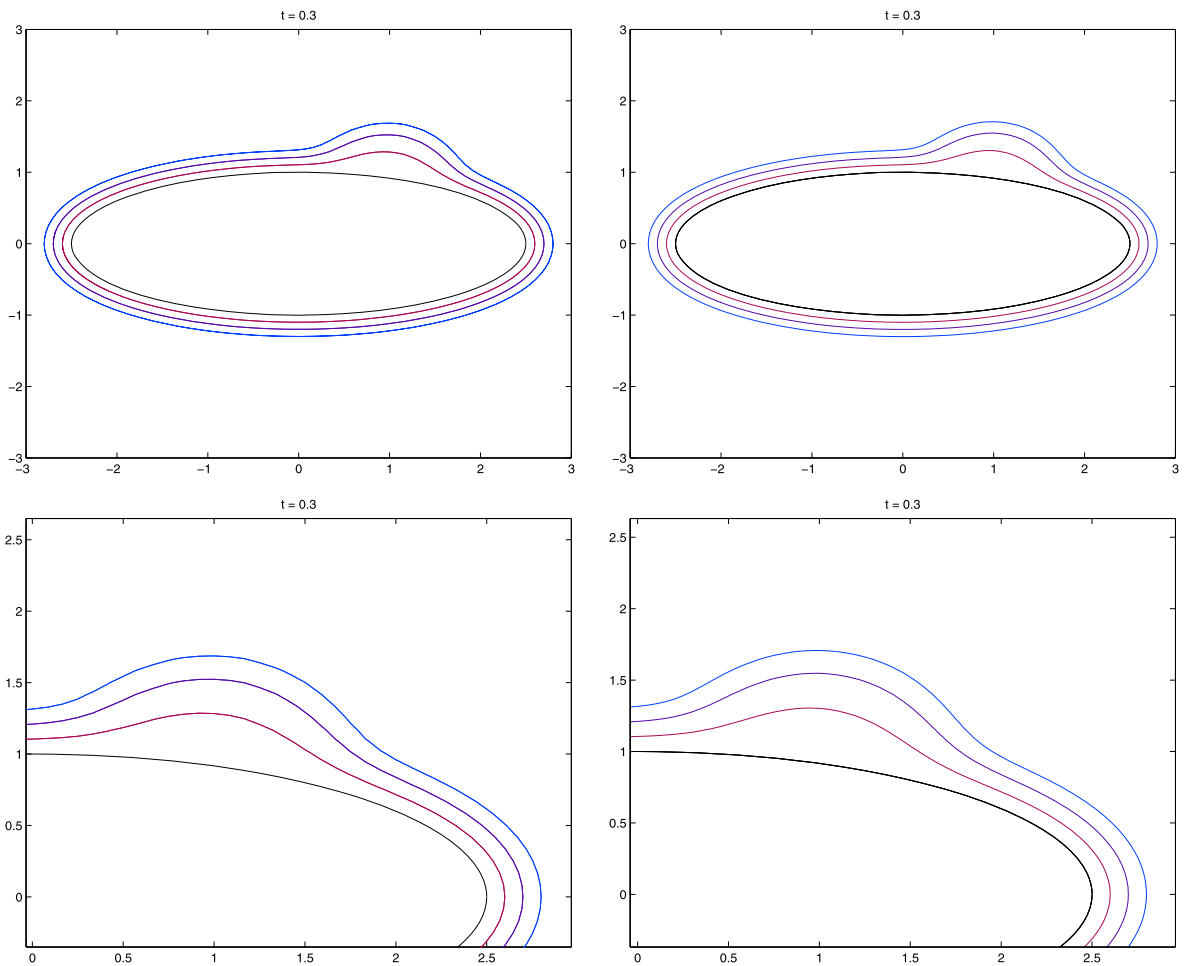


Fig. 7. Zero level sets of the solution $\phi(\cdot, t)$ for $t = 0.1, 0.2, 0.3$ in **Example 3** with H^+ and $d = 2$; left: minimization/maximization principle, right: Lax-Friedrichs.

The minimization principle (1.1) is used to compute the solution φ for both of the $\pm$ cases. The temporal stepsize is chosen as $\Delta s = 0.02$. The other constants are chosen as follows: stepsize $\sigma = 0.001$ and $L = 0.02$. Fig. 7 gives the solutions when $d = 2$ and sign of H is positive, i.e. H^+ , $T = 0.3$. The runtime using C++ is $6.865 \times 10^{-4} \text{ s} \times 5$ per point. For comparison, Fig. 7 (right) is the solution computed by Lax-Friedrichs scheme. Fig. 8 gives the solutions when $d = 2$ and sign of H is negative, i.e. H^- , $T = 0.5$. The runtime using C++ is $1.417 \times 10^{-3} \text{ s} \times 5$ per point. Fig. 8 (right) is the solution computed

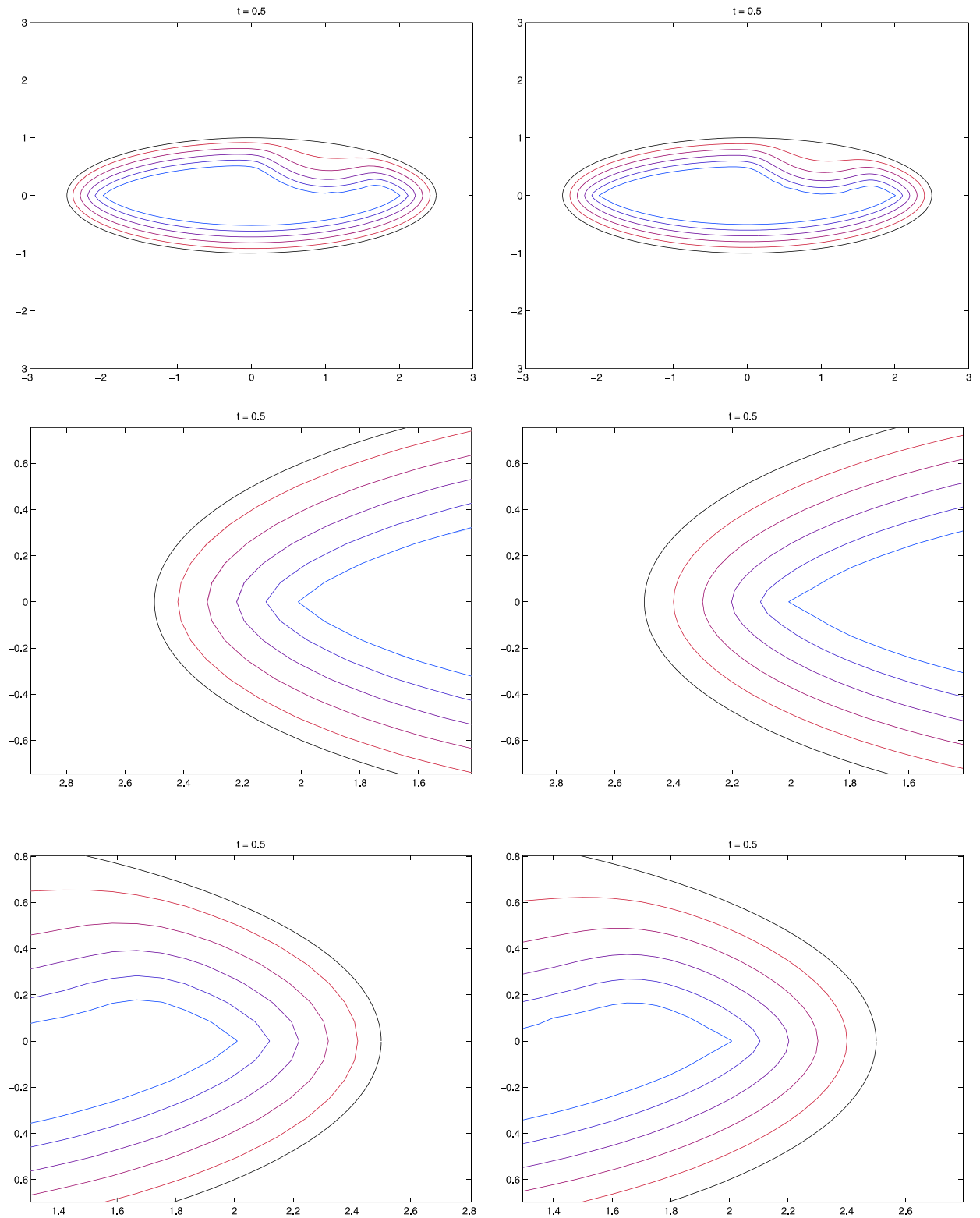


Fig. 8. Zero level sets of the solution $\phi(\cdot, t)$ for $t = 0.1, 0.2, \dots, 0.5$ in **Example 3** with H^- and $d = 2$; left: minimization/maximization principle, right: Lax-Friedrichs.

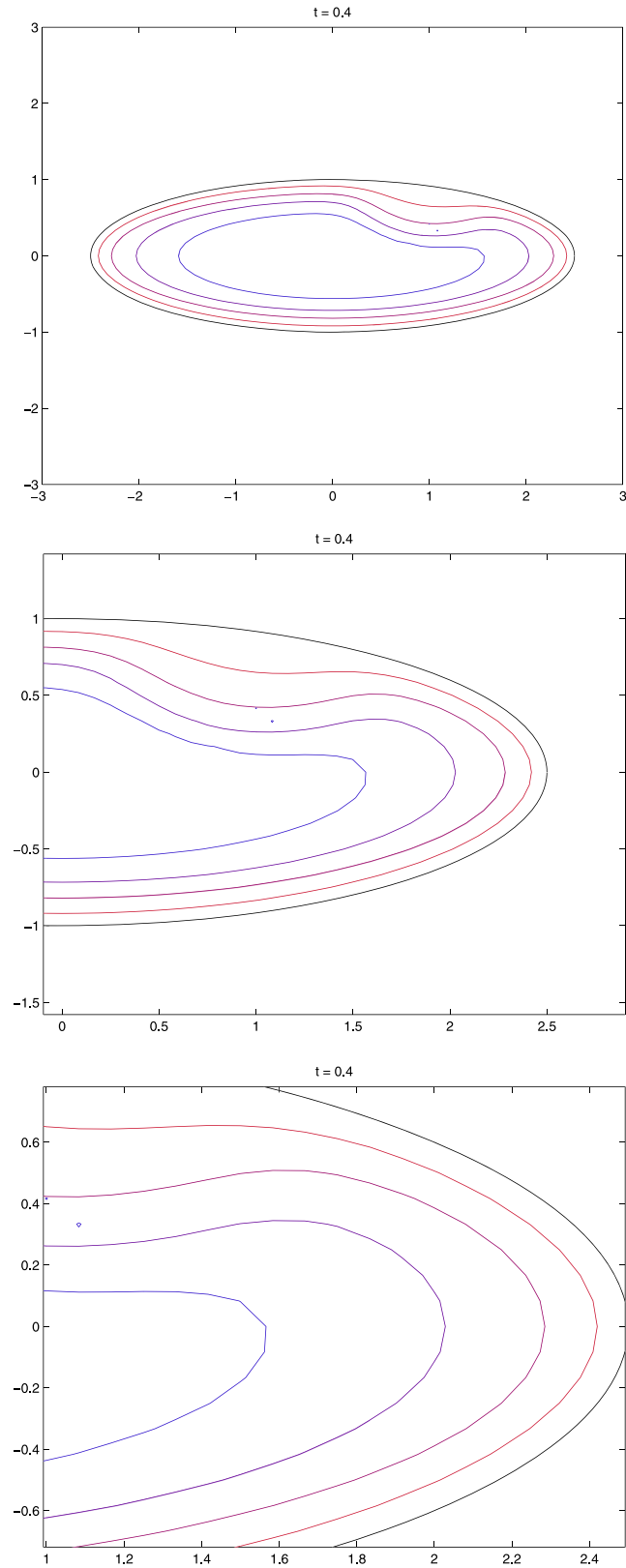


Fig. 9. Zero level sets of the solution $\phi(\cdot, t)$ for $t = 0.1, 0.2, \dots, 0.4$ by minimization/maximization principle in **Example 3** with H^- and $d = 10$; top: full-size; middle and bottom: close-up.

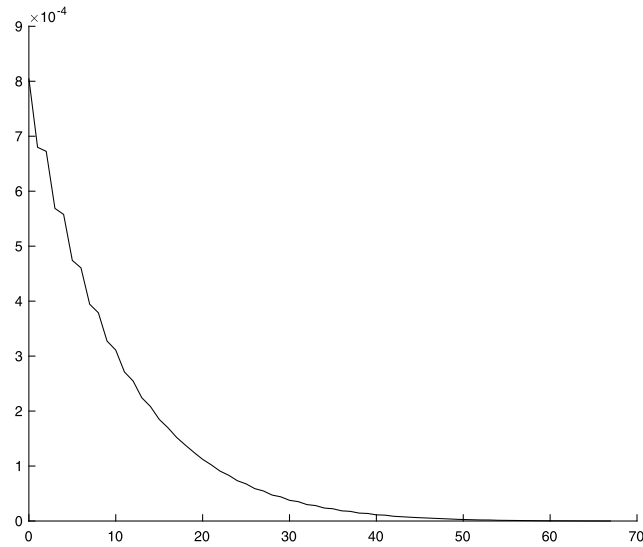


Fig. 10. Error with respect to the number of iterations with H^+ in Example 3 at the point $x = (-0.93, -0.35)$, $t = 0.3$ when $\Delta s = 0.01$ and $\sigma = 0.005$.

Table 1

Convergence table of the value with respect to $\sigma = 0.02, 0.03, \dots, 0.06$ and $\Delta s = 0.005$ with H^+ in Example 3 at the point $x = (-0.93, -0.35)$, $t = 0.3$.

σ	Error
0.06	3.539×10^{-4}
0.05	2.903×10^{-4}
0.04	4.468×10^{-4}
0.03	6.185×10^{-4}
0.02	1.818×10^{-4}

Table 2

Convergence table of the value with respect to $\sigma = 0.01$ and $\Delta s = 0.01, 0.015, \dots, 0.03$ with H^+ in Example 3 at the point $x = (-0.93, -0.35)$, $t = 0.3$.

Δs	Error
0.03	9.466×10^{-3}
0.025	7.804×10^{-3}
0.02	6.083×10^{-3}
0.015	4.314×10^{-3}
0.01	1.024×10^{-3}

by Lax-Friedrichs scheme for comparison. The angles where a discontinuity of derivative is present are computed more accurately using our maximization principle. Fig. 9 gives the solutions when $d = 10$ and sign of H is negative, i.e. H^- , $T = 0.4$. The runtime using C++ is $2.470 \times 10^{-2} s \times 5$ per point. The computation time is still excellent for a 10 dimensional problem. To remark, again the convex case H^+ in this example satisfies the assumptions in Lemma 3.2, but does not satisfy the assumptions in Lemma 3.3, Lemma 3.4 and in Lemma 3.5. Whereas for the nonconvex case H^- in this example does not satisfy the assumptions of any of the lemmas proved in this work.

In order to compare convergence of the method with respect to different discretization parameters, we compute the solutions with different Δs and σ to compare our result. We choose H^+ and compare our solutions computed at the point $x = (-0.93, -0.35)$, $t = 0.3$ with that of using the finest set of parameter $\Delta s = 0.005$ and $\sigma = 0.01$. We first fix $\Delta s = 0.005$ and compute the values of the function $\phi(x, t)$ using parameter different $\sigma = 0.02, 0.03, \dots, 0.06$. The convergence table of the value with respect to the discretization parameter is shown Table 1. In the comparison of this discretization parameter σ , the error does not show a clear trend of diminishing and shows a bit of oscillatory behaviour, although it is overall small as Δs goes to 0. We then fix $\sigma = 0.01$ and compute the values of the function $\phi(x, t)$ using parameter different $\Delta s = 0.01, 0.015, \dots, 0.03$. The convergence table of the value with respect to the discretization parameter is shown Table 2. In the comparison of this discretization parameter Δs , the error shows a clear trend of converging as goes to 0. In order to show convergence rate, as an example, Fig. 10 shows the convergence of the algorithm $\Delta s = 0.01$ and $\sigma = 0.005$. As we may observe from the figure, the convergence is sublinear.

Example 4. To test our maximization principle for a general state-dependent non-convex Hamilton-Jacobi equation, we use a state-dependent non-convex Hamiltonian of the following form given in [19] but a different problem from [19]:

$$H(x, p, t) = -c(x)p_1 + 2|p_2| - \sqrt{|p_1|^2 + |p_2|^2} - 1,$$

where we write $p = (p_1, p_2)$ and

$$c(x) = 2 \left(1 + 3 \exp(-4|x - (1, 1)|_2^2) \right).$$

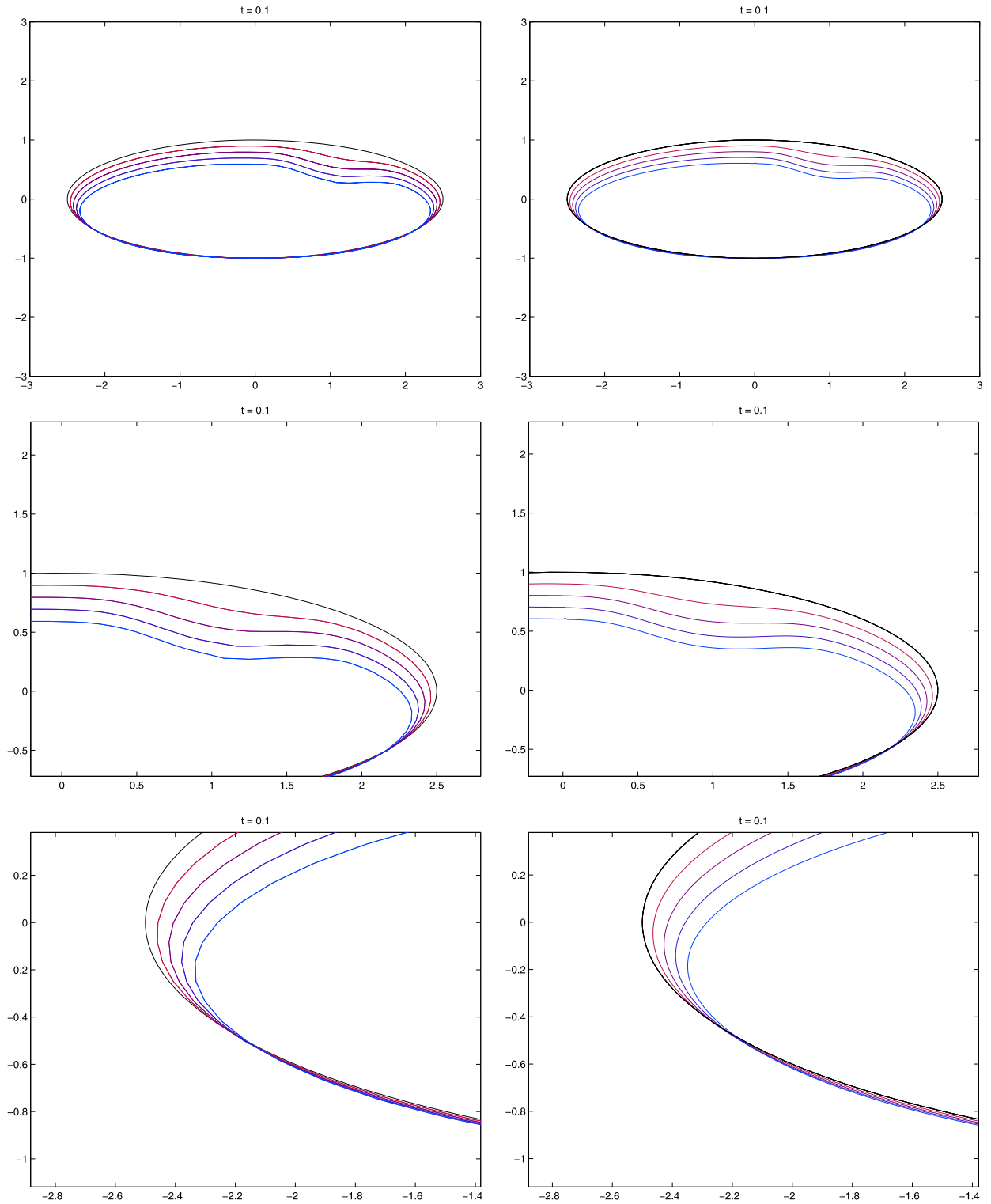


Fig. 11. Zero level sets of the solution $\phi(\cdot, t)$ for $t = 0.025, 0.05, \dots, 0.1$ in **Example 4** with $d = 2$; left: minimization/maximization principle, right: Lax-Friedrichs.

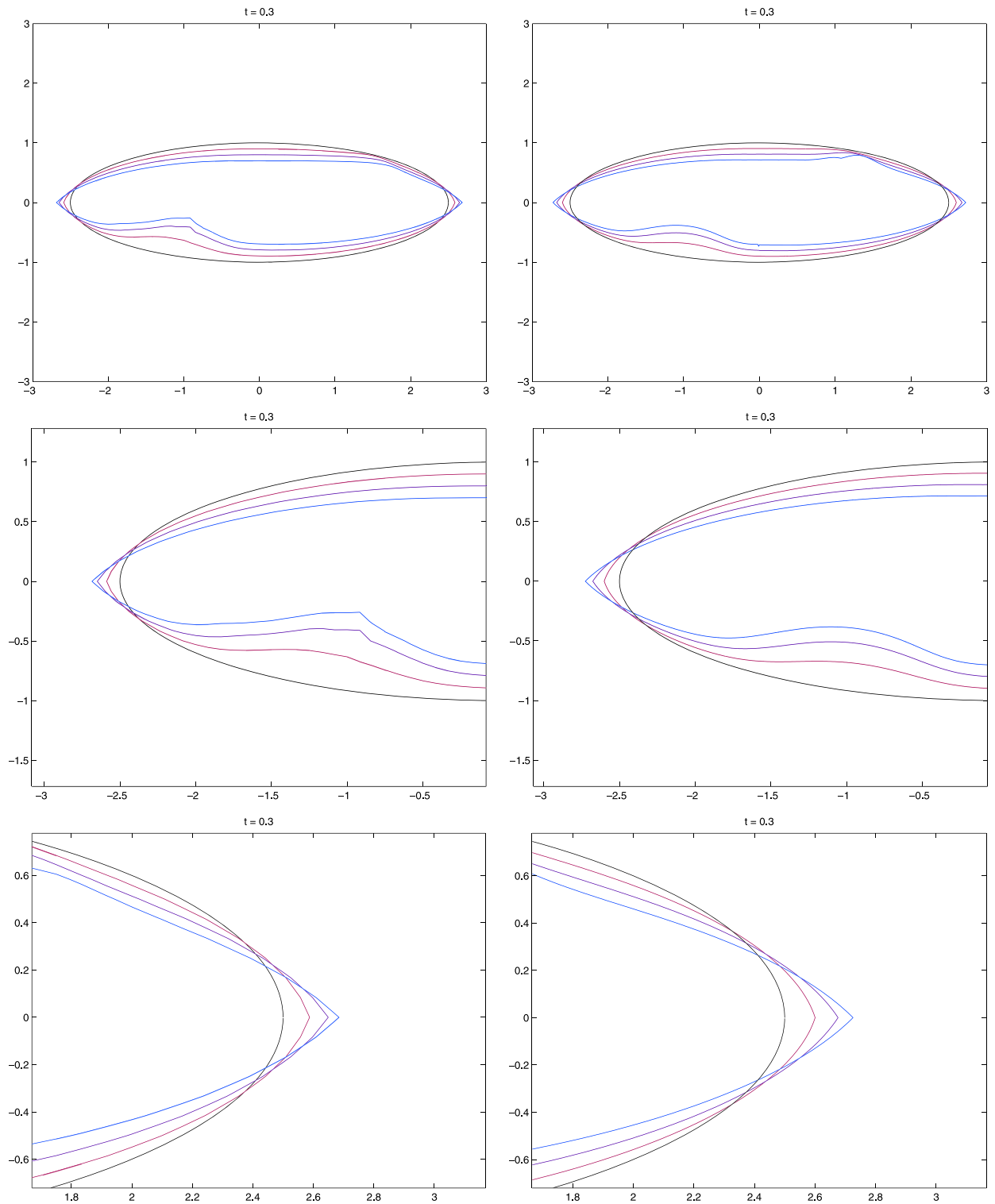


Fig. 12. Zero level sets of the solution $\phi(\cdot, t)$ for $t = 0.1, 0.2, 0.3$; left: minimization/maximization principle in **Example 5** with $k = 1$ and $d = 2$, right: Lax-Friedrichs.

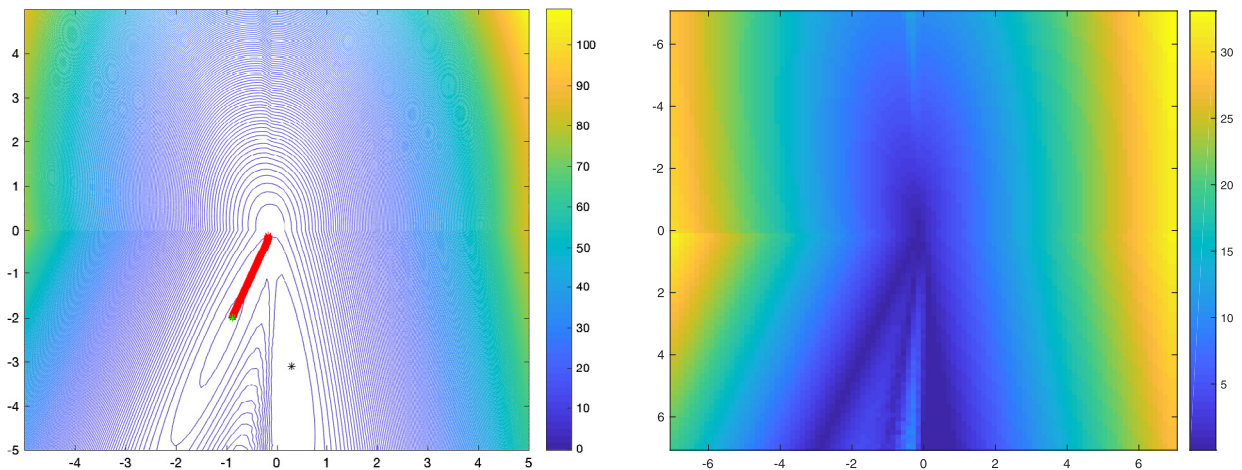


Fig. 13. Example 5 with $k = 1$ and $d = 2$. Left: functional to be minimized at $x = (-0.93, -0.35)$, $t = 0.3$ (black star: global minimum; red stars: an iterate of coordinate descent with random initial guess; green star: final iterate of the coordinate descent method.) Right: 2 norm of the gradient of the functional. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

The maximization principle (1.2) is used to compute the solution φ . The temporal stepsize is chosen as $\Delta s = 0.005$. The other constants are chosen as follows: stepsize $\sigma = 0.001$ and $L = 4$. Fig. 11 gives the solutions with $T = 0.1$. The runtime using C++ is $7.279 \times 10^{-2} s \times 5$ per point. Fig. 11 (right) is the solution computed by the Lax-Friedrichs scheme for comparison. The example is to illustrate that the maximization principle coincide with the Lax-Friedrichs solution in a quite general case. To remark, for this case, the example does not satisfy the assumptions of any of the lemmas proved in this work.

Example 5. We solve for the state-dependent non-convex Hamiltonian of the form

$$H(x, p, t) = c_1(x)|p_{1,\dots,k}|^2 - c_2(x)|p_{k+1,\dots,d}|^2,$$

where

$$c_1(x) = c(x), \quad c_2(x) = c(-x), \quad c(x) = 2 \left(1 + 3 \exp(-4|x - (1, 1)|_2^2) \right).$$

In this case the maximization principle (1.2) is used to compute the solution φ . The constants are chosen as follows: the temporal stepsize $\Delta s = 0.02$, stepsize in numerical differentiation $\sigma = 0.001$ and $L = 50$.

Fig. 12 gives the solutions when $d = 2$ and $k = 1$, $T = 0.3$. The runtime using C++ is $9.094 \times 10^{-2} s \times 20$ per point. To compare, Fig. 12 (right) is the solution computed by Lax-Friedrichs scheme for comparison. Now there is small defect in the solution computed by the maximization principle at one point of the wave-front close to $x = (-1, -0.4)$, owing to the high non-convexity and non-smoothness of the corresponding functionals around that point.

In order to pin down the exact problem causing the defect, let us fix $x = (-0.93, -0.35)$ and $t = 0.3$. Fig. 13 (left) shows the functional to be minimized. The black star is the global minimizer. It is clear that we now have a local minimizer (attractor) which has a comparable size of basin of attraction as the global minimizer. Moreover consider Fig. 13 (right), which shows the norm of the gradient for the functional to be minimized, we can clearly see that around the two basins of attractions, there is a long V-shape trench with gradient of very small magnitude. With this kind of structure, it is very likely that a gradient type method either get stuck at a local minimum (as one may see, there is almost at least 0.5 probability that one may fall into the basin of attraction of the local minimum); or it will drag very slowly along the trench and is very difficult to leave it before converging to a global minimum. In fact, this truly happen quite often. As shown in Fig. 13 (left) the red stars show the path of an iteration of coordinate descent with a random initial guess and green star is the final iterate. It is clear that the iterates drag along the trench and fall into the attraction basin of a local minimum. In fact, this structure of the functional make descent-type algorithms less easy to find the global minimum, unless better optimization methods are considered, e.g. adding momentum [44,51,57]. Fig. 14 gives the solutions when $d = 7$ and $k = 1$, $T = 0.3$. The runtime using C++ is $5.428 \times 10^{-1} s \times 20$ per point. The computation time is still acceptable for a 7 dimensional problem which is fully non-convex and state-dependent. To remark, again, this example does not satisfy the assumptions of any of the lemmas proved in this work.

Acknowledgements

We sincerely thank Dr. Gary Hewer and his colleagues (China Lake Naval Center) for providing help and guidance in practical optimal control and differential game problems, Prof. Frederic Gibou for private communication to explore the pos-

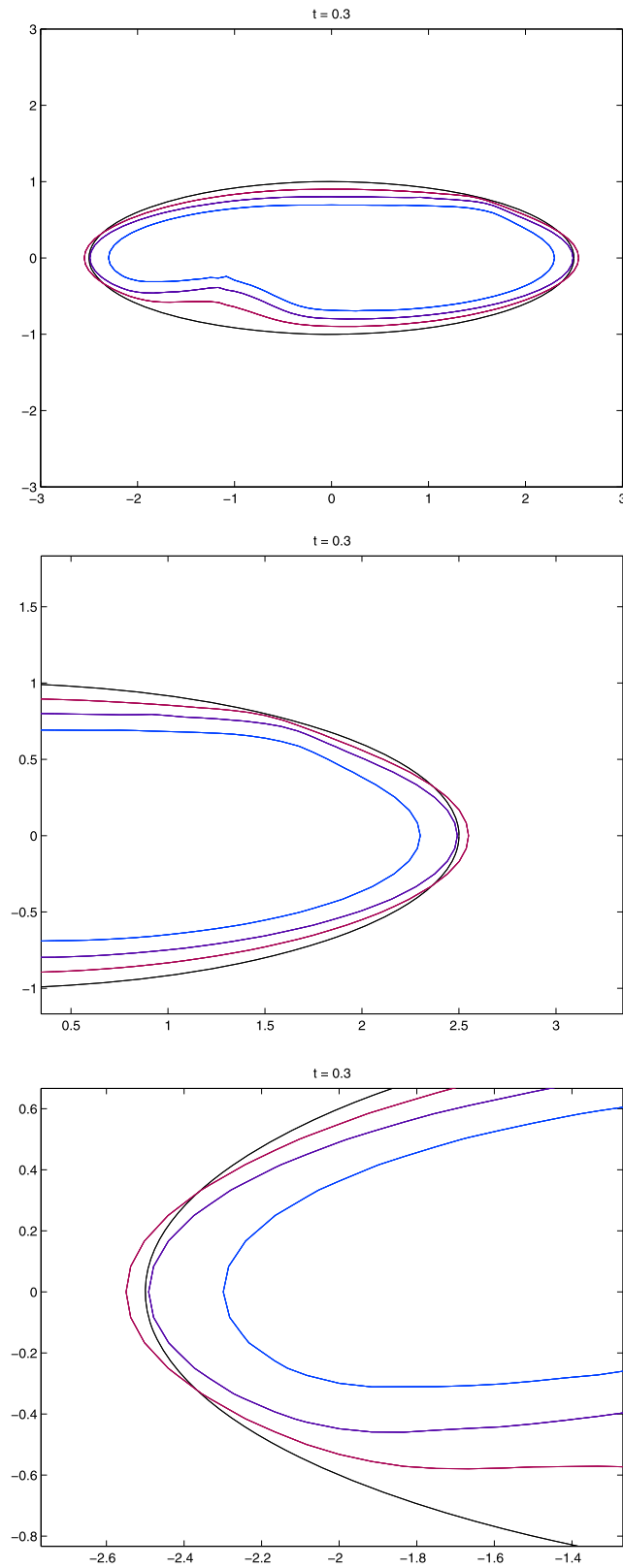


Fig. 14. Zero level sets of the solution $\phi(\cdot, t)$ for $t = 0.1, 0.2, 0.3$ by minimization/maximization principle in **Example 5** with $k = 1$ and $d = 7$; top: full-size; middle and bottom: close-up.

sibility to fully parallelize the method, and Prof. Lawrence C. Evans and Prof. Wilfrid Gangbo for exploring if the conjectures made in this paper are sound. We also thank Alex Tong Lin for many discussions of insights in our paper.

We thank sincerely also the two anonymous referees for many insightful recommendations and comments to help substantially improve our paper.

Research of the first author is supported by ONR N000141712162 and NSF DMS-1720237. Research of the second author is supported by NSF DMS-1820821. Research of the third and fourth authors are supported by ONR grants N000141410683, N000141210838, N000141712162 and DOE grant DE-SC00183838.

References

- [1] R. Bellman, *Adaptive Control Processes, a Guided Tour*, Princeton University Press, 1961.
- [2] R. Bellman, *Dynamic Programming*, Princeton University Press, 1957.
- [3] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [4] P. Bernard, The Lax-Oleinik semi-group: a Hamiltonian point of view, preprint, arXiv:1203.3569, 2012.
- [5] O. Bernardi, F. Cardin, Minimax and viscosity solutions of Hamilton-Jacobi equations in the convex case, *Commun. Pure Appl. Anal.* 5 (4) (2006) 793–812.
- [6] G. Capitanio, Generic singularities of minimax solutions to Hamilton–Jacobi equations, preprint, arXiv:math/0407067, 2004.
- [7] Y.T. Chow, J. Darbon, S. Osher, W. Yin, Algorithm for Overcoming the Curse of Dimensionality for Certain Non-convex Hamilton–Jacobi Equations, Projections and Differential Games, UCLA CAM Report 16-27, in: *Annals of Mathematical Sciences and Applications*, vol. 3, 2018, pp. 369–403.
- [8] Y.T. Chow, J. Darbon, S. Osher, W. Yin, Algorithm for overcoming the curse of dimensionality for state-dependent Hamilton–Jacobi equations, preprint, arXiv:1704.02524, 2017.
- [9] Y.T. Chow, J. Darbon, S. Osher, W. Yin, Algorithm for overcoming the curse of dimensionality for time-dependent non-convex Hamilton–Jacobi equations arising from optimal control and differential games problems, *J. Sci. Comput.* 73 (2–3) (2017) 617–643, <https://doi.org/10.1007/s10915-017-0436-5>.
- [10] F.H. Clarke, R.B. Vinter, The relationship between the maximum principle and dynamic programming, *SIAM J. Control Optim.* 25 (5) (1987) 1291–1311.
- [11] S.E. Cohn-Vossen, *Some Problems of Differential Geometry in the Large*, Fizmatgiz, Moscow, 1959.
- [12] M.G. Crandall, P.-L. Lions, Some properties of viscosity solutions of Hamilton–Jacobi equations, *Trans. Am. Math. Soc.* 282 (2) (1984) 487–502.
- [13] M.G. Crandall, P.-L. Lions, Viscosity solutions of Hamilton–Jacobi equations, *Trans. Am. Math. Soc.* 277 (1) (1983) 1–42.
- [14] J. Darbon, On convex finite-dimensional variational methods in imaging sciences, and Hamilton–Jacobi equations, *SIAM J. Imaging Sci.* 8 (4) (2015) 2268–2293.
- [15] J. Darbon, S. Osher, Algorithms for Overcoming the Curse of Dimensionality for Certain Hamilton–Jacobi Equations Arising in Control Theory and Elsewhere, preprint, UCLA CAM report 15-50, 2015.
- [16] E.W. Dijkstra, A note on two problems in connexion with graphs, *Numer. Math.* 1 (1959) 269–271.
- [17] W. E, W. Ren, E. Vanden-Eijnden, Minimum action method for the study of rare events, *Commun. Pure Appl. Math.* 57 (5) (2004) 637–656.
- [18] L.C. Evans, *Partial Differential Equations*, *Grad. Studies in Math.*, vol. 19, AMS, 2010.
- [19] L.C. Evans, Envelopes and nonconvex Hamilton–Jacobi equations, *Calc. Var. Partial Differ. Equ.* 50 (1) (2014) 257–282.
- [20] L.C. Evans, P.E. Souganidis, Differential games and representation formulas for solutions of Hamilton–Jacobi Isaacs equations, *Indiana Univ. Math. J.* 38 (1984) 773–797.
- [21] M.P. Friedlander, I. Macedo, T.K. Pong, Gauge optimization and duality, *SIAM J. Optim.* 24 (4) (2014) 1999–2022.
- [22] T. Grafke, T. Schaefer, E. Vanden-Eijnden, Long term effects of small random perturbations on dynamical systems: theoretical and computational tools, preprint, arXiv:1604.03818.
- [23] M. Heymann, E. Vanden-Eijnden, The geometric minimum action method: a least action principle on the space of curves, *Commun. Pure Appl. Math.* 61 (8) (2008) 1052–1117.
- [24] J.-B. Hiriart-Urruty, C. Lemaréchal, *Fundamentals of Convex Analysis*, Grundlehren Text Editions, Springer, 2001.
- [25] E. Hopf, Generalized solutions of nonlinear equations of the first order, *J. Math. Mech.* 14 (1965) 951–973.
- [26] H. Hopf, W. Rinow, Ueber den Begriff der vollständigen differentialgeometrischen Flächen, *Comment. Math. Helv.* 3 (1931) 209–225.
- [27] M.B. Horowitz, A. Damle, J.W. Burdick, Linear Hamilton Jacobi Bellman equations in high dimensions, arXiv:1404.1089, 2014.
- [28] W. Kang, L.C. Wilcox, Mitigating the curse of dimensionality: sparse grid characteristics method for optimal feedback control and HJB equations, preprint, arXiv:1507.04769, 2015.
- [29] H. Komiya, Elementary proof for Sion's minimax theorem, *Kodai Math. J.* 11 (1988) 5–7.
- [30] N.N. Krasovskii, A.I. Subbotin, *Game-Theoretical Control Problems*, Springer-Verlag, New York, 1988.
- [31] N.N. Krasovskii, A.I. Subbotin, *Positional Differential Games*, Nauka, Moscow, 1974.
- [32] S. Mirica, Extending Cauchy's method of characteristics for Hamilton–Jacobi equations, *Stud. Cercet. Mat.* 37 (6) (1985) 555–565.
- [33] I.M. Mitchell, A.M. Bayen, C.J. Tomlin, A time-dependent Hamilton–Jacobi formulation of reachable sets for continuous dynamic games, *Automatic Control, IEEE Trans. Autom. Control* 50 (7) (2005) 947–957.
- [34] I.M. Mitchell, C.J. Tomlin, Overapproximating reachable sets by Hamilton–Jacobi projections, *J. Sci. Comput.* 19 (1–3) (2003) 323–346.
- [35] J.J. Moreau, Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. Fr.* 93 (1965) 273–299.
- [36] B.S. Mordukhovich, Maximum principle in problems of time optimal control with nonsmooth constraints, *J. Appl. Math. Mech.* 40 (1976) 960–969.
- [37] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation, I: Basic Theory*, Grundlehren Math. Wiss., vol. 330, Springer-Verlag, Berlin, 2006.
- [38] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation, II: Applications*, Grundlehren Math. Wiss., vol. 331, Springer-Verlag, Berlin, 2006.
- [39] J. von Neumann, Zur theorie der gesellschaftsspiele, *Math. Ann.* 100 (1928) 295–320.
- [40] B. O'Donoghue, G. Stathopoulos, S. Boyd, A splitting method for optimal control, *IEEE Trans. Control Syst. Technol.* 21 (6) (2013) 2432–2442.
- [41] S. Osher, C.-W. Chu, High order essentially non-oscillatory schemes for Hamilton–Jacobi equations, *SIAM J. Numer. Anal.* 28 (4) (1991) 907–922.
- [42] S. Osher, J.A. Sethian, Fronts propagating with curvature dependent speed: algorithms based on Hamilton–Jacobi formulations, *J. Comput. Phys.* 79 (1) (1988) 12–49.
- [43] S. Osher, B. Merriman, The Wulff shape as the asymptotic limit of a growing crystalline interface, *Asian J. Math.* 1 (3) (1997) 560–571.
- [44] B.T. Polyak, Some methods of speeding up the convergence of iteration methods, *USSR Comput. Math. Math. Phys.* 4 (5) (1964) 1–17.
- [45] L.S. Pontryagin, V.G. Boltyansky, R.V. Gamkrelidze, E.F. Mishchenko, *The Mathematical Theory of Optimal Processes*, Macmillan, New York, 1964.
- [46] R.T. Rockafellar, A general correspondence between dual minimax problems and convex programs, *Pac. J. Math.* 25 (3) (1968) 597–611.
- [47] R.T. Rockafellar, *Convex Analysis*, Princeton Landmarks in Mathematics, Princeton University Press, 1997.
- [48] R. Rockafellar, P. Wolenski, Convexity in Hamilton–Jacobi theory I: Dynamics and duality, *SIAM J. Control Optim.* 39 (5) (2000) 1323–1350.

- [49] R. Rockafellar, P. Wolenski, Convexity in Hamilton-Jacobi theory II: Envelope representations, *SIAM J. Control Optim.* 39 (5) (2000) 1351–1372.
- [50] I.V. Rublev, Generalized Hopf formulas for the nonautonomous Hamilton-Jacobi equation, *Comput. Math. Model.* 11 (4) (2000) 391–400.
- [51] H. Rutishauser, Theory of gradient methods, in: *Refined Iterative Methods for Computation of the Solution and the Eigenvalues of Self-Adjoint Boundary Value Problems*, 1959, pp. 24–49.
- [52] M. Sion, On general minimax theorems, *Pac. J. Math.* 8 (1958) 171–176.
- [53] A.I. Subbotin, A.G. Chentsov, *Optimization of Guaranteed Result in Control Problems*, Nauka, Moscow, 1981.
- [54] N.N. Subbotina, Method of Cauchy characteristics and generalized solutions of Hamilton-Jacobi-Bellman equations, *Dokl. Akad. Nauk SSSR* 320 (1991) 556–561.
- [55] N.N. Subbotina, Necessary and Sufficient Optimality Conditions in Terms of Characteristics of the Hamilton-Jacobi-Bellman Equation, Report 393, Institut für Angewandte Mathematik und Statistik, Universität Würzburg, Würzburg, 1992.
- [56] N.N. Subbotina, The method of characteristics for Hamilton-Jacobi equations and applications to dynamical optimization, *J. Math. Sci.* 135 (3) (2006) 2955–3091.
- [57] I. Sutskever, J. Martens, G.E. Dahl, G.E. Hinton, On the importance of initialization and momentum in deep learning, *Int. Conf. Mach. Learn.* 28 (3) (2013) 1139–1147.
- [58] Y.H.R. Tsai, L.T. Cheng, S. Osher, H.K. Zhao, Fast sweeping algorithms for a class of Hamilton-Jacobi equations, *SIAM J. Numer. Anal.* 41 (2) (2003) 673–694.
- [59] J.N. Tsitsiklis, Efficient algorithms for globally optimal trajectories, *IEEE Trans. Autom. Control* 40 (9) (1995) 1528–1538.
- [60] E. Vanden-Eijnden, J. Weare, Rare event simulation of small noise diffusions, *Commun. Pure Appl. Math.* 65 (12) (2012) 1770–1803.
- [61] Q. Wei, Viscosity solution of the Hamilton-Jacobi equation by a limiting minimax method, *Nonlinearity* 27 (1) (2013) 17–42.
- [62] I. Yegorov, P. Dower, Perspectives on characteristics based curse-of-dimensionality-free numerical approaches for solving Hamilton-Jacobi equations, preprint, arXiv:1711.03314.