On the Game-Theoretic Nature of Optimality in Active Hypothesis Testing

Sushant Vijayan [1]

Abstract

Active Hypothesis Testing (AHT) is a fundamental problem across information theory, multi-armed bandits, and operations research, but a clear characterization of its optimal large deviation behaviour has long been an open challenge. We build upon the work of Komiyama et al. (NeurIPS 2022), which presents a computationally intractable and conceptually unclear formulation of the optimal exponent R_∞^{go} of the error probability. We characterize this optimal error-exponent, in the limit that the sample budget $T \to \infty$, as the value of a certain two-player zero-sum differential game. This dynamic game, played over the normalized interval $[0, 1]$, has two players, Nature and Agent. The state variable y_t at time t is the negative log-likelihood of the observed samples till then. We show that the players aim to optimize the second-smallest coordinate of the terminal state y_1. As a result, we obtain a time-dependent Hamilton-Jacobi-Isaacs (HJI) Partial Differential Equation (PDE) that is used to develop the first numerical schemes in the literature to accurately approximate R_∞^{go}. Additionally, we can derive existing and new alternate bounds for R_∞^{go} using the viscosity solution theory of HJI PDEs. Finally, we establish a remarkable time-scaling property of the HJI viscosity solution, which reduces the dynamics to a new time-invariant PDE and provides a novel geometric description of optimality in AHT in terms of approachability.

1. Introduction

We study the problem of Active Hypothesis Testing (AHT), wherein an agent is tasked with identifying the true hypothesis from a known set of m possible hypotheses after taking T adaptive samples. Specifically, the agent is presented with a K-armed bandit instance, which is guaranteed to be exactly one of m possible bandit models $\nu^1, \ldots, \nu^m$. At each step, the agent chooses an action (arm) $a \in [K]$ and observes a sample from the corresponding distribution. At the end of the budget T, the agent must output a decision $\hat{i}_T \in [m]$ estimating the true model. AHT captures the essence of controlled sample generation and acts as a fundamental model for diverse applications, such as identifying the most effective treatment in a cosmetic trial, selecting the optimal communication channel before transmission, or placing sensors to minimize signal error (Nitinawarat et al., 2013; Hong et al., 2021).

AHT is a specialized, theoretically foundational variant of the Fixed Budget Best Arm Identification (FB-BAI) problem. While FB-BAI allows for composite or continuum hypotheses, AHT restricts the hypothesis space to m distinct, simple models. Despite this simplification, the AHT problem remains challenging and is not fully understood in terms of its optimal theoretical limits (Nitinawarat et al., 2013; Qin, 2022).

We analyze AHT in the large deviation regime ($T \to \infty$), where the performance metric of interest is the minimax error exponent—the worst-case exponential rate of decay of the error probability across all m hypotheses. Maximizing this exponent guarantees robust, exponentially fast convergence to the correct hypothesis. Recently, (Komiyama et al., 2022) provided a theoretical characterization of this optimal minimax error exponent, denoted as R_∞^{go}. However, their formulation expresses R_∞^{go} as a complex optimization. This representation is computationally intractable to evaluate and conceptually opaque, offering little geometric intuition or practically computable bounds.

In this work, we establish a novel bridge between large deviation theory in AHT, continuous-time zero-sum differential games, and Partial Differential Equations (PDEs).

Contributions. We briefly summarize our main contributions below:

- **Game-Theoretic Formulation:** We recast the optimal error exponent R_∞^{go} as the value of a continuous-time, two-player zero-sum differential game played over the normalized time interval $[0, 1]$ between Nature and

[1]Department of CS & AI, Plaksha University, Mohali, India. Correspondence to: Sushant Vijayan <sushant.vijayan@plaksha.edu.in>.

ICML 2026 Workshop on Hypothesis Testing, Seoul, South Korea, 2026. Copyright 2026 by the author(s).

the Agent. The state variable y_t is the negative log-likelihood of the observed samples.

- **HJI PDE and Viscosity Bounds:** We characterize this optimal error exponent as the viscosity solution to a time-dependent Hamilton-Jacobi-Isaacs (HJI) PDE. Using the theory of viscosity solutions (e.g., the Hopf formula and comparison principles), we demonstrate how to recover existing information-theoretic bounds and derive novel upper bounds for R_∞^{go}.
- **Time-Scaling Property & Geometric Insights:** We uncover a remarkable time-scaling property of the HJI viscosity solution, $V(y,t) = (1-t)V\left(\frac{y}{1-t}, 0\right)$. This reduces the dynamics to a new time-invariant PDE and provides a geometric description of optimality in AHT based on the idea of *approachability*.
- **Numerical Approximation Schemes:** Leveraging the continuous-time PDE formulation, we propose the first numerically tractable finite-difference schemes in the literature to accurately compute R_∞^{go} for small m. We utilize this scheme to explicitly compute R_∞^{go} for a 3-hypothesis problem, numerically refuting a recent conjecture by (Komiyama et al., 2022) regarding the equivalency of single-batch and infinite-batch optimal exponents.

We introduce the problem in Section 2 and review relevant prior work. The reformulation of the optimal exponent in terms of a zero-sum differential game is presented in Section 3. The connection with PDEs and various properties of the corresponding viscosity solution is given in Section 4. In Section 5, we provide the first numerical scheme to approximate the optimal exponent. Section 6 concludes with future directions for further work.

2. Problem Formulation and Preliminaries

We study the **Active Hypothesis Testing (AHT)** problem. AHT is a foundational, specialized variant of the Fixed Budget Best Arm Identification (FB-BAI) problem. While FB-BAI allows for general composite or even continuous hypothesis spaces, AHT restricts the hypothesis space to a discrete set of known, simple models.

Suppose an agent is presented with an unknown K-armed bandit instance, which is given to be one of m possible bandit models from a known set $\mathcal{V} = \{\nu^1, \dots, \nu^m\}$. The goal of the agent is to determine the true underlying model $i^* \in [m]$. We denote the arm distribution of arm $a \in [K]$ under hypothesis $i \in [m]$ as ν_a^i. Furthermore, we make the standard assumption that all arm distributions have full support over a common, finite set $\mathcal{X}$. Let $D(p||q)$ denote the Kullback-Leibler (KL) divergence between two distributions p and q.

Sampling and Decision Rules. Let T denote the total sample budget. An active hypothesis testing algorithm $\mathbb{A}$ consists of a sampling rule $\{a_t\}_{t=1}^T$, where each action $a_t \in [K]$ is adapted to the historical filtration of actions and observations up to time $t-1$, and a final decision rule $\hat{i}_T \in [m]$ adapted to the filtration at time T.

The error probability of an algorithm $\mathbb{A}$ on the true instance ν^i is defined as $P_e(\nu^i, \mathbb{A}) = \mathbb{P}_{\nu^i,\mathbb{A}}(\hat{i}_T \neq i)$, where the randomness arises from both the samples generated by ν^i and any internal randomization of $\mathbb{A}$. The performance of the algorithm is measured by its error exponent:

$$e(\nu^i, \mathbb{A}) \triangleq \liminf_{T\to\infty} -\frac{1}{T}\log\left(P_e(\nu^i, \mathbb{A})\right). \tag{1}$$

The minimal exponent of $\mathbb{A}$ over the hypothesis set $\mathcal{V}$ is defined as $e_{\min}(\mathcal{V}, \mathbb{A}) \triangleq \min_{i\in[m]} e(\nu^i, \mathbb{A})$. Consequently, an algorithm $\mathbb{A}^*$ is said to be minimax optimal if it maximizes this worst-case exponent: $e_{\min}(\mathcal{V}, \mathbb{A}^*) = \max_{\mathbb{A}} e_{\min}(\mathcal{V}, \mathbb{A})$. Our goal is to characterize this minimax exponent asymptotically and design computationally tractable approximation schemes for it.

2.1. Prior Bounds and the Minimax Exponent

For the binary hypothesis case ($m = 2$), (Hayashi, 2009) completely resolved the optimal exponent, showing that $e_{\min}(\mathcal{V}, \mathbb{A}^*) = \max_{a\in[K]} C(\nu_a^1, \nu_a^2)$, where $C(p,q)$ is the Chernoff Information. This implies that a static strategy (pulling the single best arm) is strictly optimal for $m = 2$.

For the general $m \geq 2$ setting, (Nitinawarat et al., 2013) derived the upper bound $R^{ub} \triangleq \min_{i\neq j} \max_{a\in[K]} C(\nu_a^i, \nu_a^j)$ and characterized the best possible exponent achievable by strictly non-adaptive (static) strategies, denoted R_{static}. Crucially, they showed that for $m > 2$, adaptive sampling yields strictly larger exponents than static sampling.

Recently, (Komiyama et al., 2022) provided a tighter, batch-based formulation of the optimal adaptive exponent. Let $B \in \mathbb{N}$ denote a number of batches, with each batch containing T/B samples. Let $Q_l = (Q_{l,1}, \dots, Q_{l,K})$ be the K-tuple of empirical distributions observed for the arms during the l-th batch, and let the history of these distributions be denoted as $Q_{1:l} = (Q_1, \dots, Q_l)$. Consider a sample allocation rule $r_l(Q_{1:l}) \in \Delta_K$ for the l-th batch, and a causal final decision rule $J(Q_{1:B}) \in [m]$.

(Komiyama et al., 2022) established the following upper bound on the minimax exponent, $e_{\min}(\mathcal{V}, \mathbb{A}^*) \leq R_B^{go}$, where:

$$R_B^{go} \triangleq \sup_{r_{1:B}, J} \inf_{\substack{Q_{1:B}; i\in[m] \\ i\neq J(Q_{1:B})}} \left[\frac{\sum_{l=1}^B \sum_{a\in[K]} r_l(Q_{1:l})(a) D(Q_{l,a}||\nu_a^i)}{B}\right]. \tag{2}$$

It is important to note the order of operations in the minmax game above: because the supremum over the allocation

rule r_l is evaluated outside the infimum over the empirical realizations $Q_{1:B}$, the allocation rule effectively acts with a "peek-ahead" advantage, allowing r_l to depend on the empirical distribution Q_l of the current batch.

(Komiyama et al., 2022) proved that the sequence $\{R_B^{go}\}_{B=1}^{\infty}$ is bounded, and that its limit defines the true optimal minimax exponent: $\lim_{B\to\infty} R_B^{go} \triangleq R_\infty^{go}$. The hierarchy of known bounds can be succinctly summarized as, $\forall B \in \mathbb{N}$:

$$R_{static} \leq e_{\min}(\mathcal{V}, \mathbb{A}^*) = R_\infty^{go} \leq R_B^{go} \leq R_1^{go} \leq R^{ub}. \quad (3)$$

While R_∞^{go} exactly characterizes the optimal minimax exponent, the definition in Equation (2) relies on a sequential, infinite-batch limit of min-max optimizations. This representation suffers from the curse of dimensionality even for moderate B, making it computationally intractable to evaluate and opaque for algorithm design. In the following sections, we demonstrate how this infinite-batch limit can be reformulated as a continuous-time differential game.

3. A Zero-Sum Differential Game for AHT

Let $\hat{i}_{ML} \in [m]$ denote the decision rule that outputs the hypothesis with the Maximum Likelihood (ML) after the sample budget is exhausted. The ML decision rule is optimal for AHT asymptotically. To see this, note that the optimal Bayes exponent under a fixed prior matches the minimax exponent $e_{\min}(\mathcal{V}, \mathbb{A}^*)$, as the effect of the prior vanishes in the limit $T \to \infty$. Since the Maximum-a-Posteriori (MAP) rule is Bayes optimal and reduces to the ML rule under a uniform prior, $\hat{i}_{ML}$ attains the optimal exponent. Hereon, we fix the decision rule to be $\hat{i}_{ML}$.

3.1. From Discrete Batches to Continuous Dynamics

We first reinterpret the upper bound R_B^{go} (Equation 2) as a discrete-time repeated game (proof in Appendix B.1):

Proposition 3.1. *For any number of batches $B \in \mathbb{N}$, we have:*

$$R_B^{go} = \underbrace{\inf_{Q_1} \sup_{w_1} \dots \inf_{Q_B} \sup_{w_B}}_{B \text{ times}} \left[\max_{j\in[m]} \min_{i\neq j} \left(\frac{1}{B} \sum_{l=1}^{B} \sum_{a\in[K]} w_l(a) D(Q_{l,a} \| \nu_a^i) \right) \right], \quad (4)$$

where $Q_l \in \Delta_{\mathcal{X}}^K$ and $w_l \in \Delta_K$.

This representation interprets R_B^{go} as the value of an adversarial, B-stage Stackelberg game. Note that to address the informational advantage present in Equation 2, the order of the limits here perfectly captures the "peek-ahead" nature: in each batch l, Nature acts as the leader by choosing the empirical distribution Q_l, and the Agent acts as the follower by choosing the allocation w_l with full knowledge of Q_l.

For a true bandit ν^i, the inner summation computes the negative log-likelihood of the samples. Under the $\hat{i}_{ML}$ rule, an error occurs if an incorrect hypothesis accumulates a higher likelihood than the true hypothesis. Thus, the terminal cost inside the square brackets represents the *second-smallest* negative log-likelihood (i.e., the likelihood of the most confusing alternative). Nature aims to minimize this gap (maximizing confusion), while the Agent maximizes it.

Since the number of samples T/B in a batch decreases with B, the clairvoyance granted to the Agent diminishes as $B \to \infty$, and R_B^{go} tends toward $e_{\min}(\mathcal{V}, \mathbb{A}^*)$. This limit transitions the discrete-time game into a continuous-time differential game.

3.2. Continuous-Time Differential Game Formulation

In the limit $B \to \infty$, we formulate a zero-sum differential game over the normalized time interval $t \in [0, 1]$. Two adversaries, Nature (the leader) and the Agent (the follower), control the evolution of a state vector $y_t \in \mathbb{R}^m$. In any infinitesimal interval $(t, t+dt]$, Nature chooses an empirical distribution $Q_t \in \Delta_{\mathcal{X}}^K$, and the Agent responds with an allocation $w_t \in \Delta_K$.

The dynamical equation governing the state y_t (tracking the Negative Log-Likelihoods (NLLs)) is given by:

$$\begin{aligned} \dot{y}_{t,i} &= \sum_{a\in[K]} w_t(a) D(Q_{t,a} \| \nu_a^i), \\ y_{0,i} &= 0, \quad \forall i \in [m]. \end{aligned} \quad (5)$$

The dynamics captures how the players choices in w_t, Q_t affect the growth of the statistical evidence in terms of the NLLs.

Both players aim to optimize the terminal cost at $t = 1$. Nature minimizes the cost, while the Agent maximizes it. The terminal cost $g(y_1)$ evaluates the performance of the ML rule:

$$g(y_1) = \max_{j\in[m]} \min_{i\neq j} y_{1,i} = y_{1,(2)}, \quad (6)$$

where $z_{(i)}$ denotes the i-th smallest order statistic of a vector z. On any given sample path, the hypothesis with second smallest NLL, under the $\hat{i}_{ML}$ decision rule, has the highest likelihood of error. This fact motivates the given terminal function $g(y)$.

It is a well-known paradigm in large deviation theory that error exponents can often be formulated as optimal control problems (Dembo, 2009). Here, the introduction of the adversarial Agent extends this to a zero-sum differential

game. We give a primer on differential games and associated viscosity solution theory in Appendix A for additional background.

4. The Hamilton-Jacobi-Isaacs (HJI) PDE and Viscosity Bounds

We now characterize the optimal exponent R_∞^{go} as the value of the differential game defined above. The *upper* value of such a game can be computed via the viscosity solution of its associated Hamilton-Jacobi-Isaacs (HJI) Partial Differential Equation (PDE). The next theorem establishes this rigorously (proof in Appendix B.2).

Theorem 4.1. *Consider the Hamilton-Jacobi-Isaacs (HJI) equation associated with the differential game in Section 3.2:*

$$\frac{\partial V}{\partial t} + \min_{Q} \max_{w \in \Delta_K} \left[\sum_{i=1}^{m} \sum_{a \in [K]} w_t(a) D(Q_{t,a} \| \nu_a^i) \frac{\partial V}{\partial y_i} \right] = 0,$$
$$V(y,1) = g(y) = y_{(2)}, \quad \forall y \in \mathbb{R}^m. \tag{7}$$

If $V(y,t)$ is the unique viscosity solution to this PDE, then:

$$R_\infty^{go} = V(0,0). \tag{8}$$

Proof Sketch. It is a known result from the theory of differential games that the value of discrete-time games with piecewise-constant controls converges to the upper value of the corresponding continuous-time game (detailed in Theorem A.11). By combining this convergence guarantee with the discrete-time Stackelberg game representation of R_B^{go} established in Proposition 3.1, it immediately follows that $\lim_{B\to\infty} R_B^{go} = R_\infty^{go} = V(0,0)$. The PDE representation of V follows from Theorem A.9. □

The Hamiltonian of this system dictates the local optimal play. Let $p = \nabla_y V$ denote the costate vector. The *upper* Hamiltonian is defined as $H^+(p) = \min_Q \max_w \sum_{i,a} w(a) D(Q_a \| \nu_a^i) p_i$, and the *lower* Hamiltonian $H^-(p)$ by the corresponding max-min. We derive a closed-form expression for this Hamiltonian and prove the game has a definitive value (proof in Appendix B.3).

Lemma 4.2. *The upper and lower Hamiltonians satisfy Isaacs' condition ($H^+ = H^-$). The Hamiltonian $H(p) \triangleq H^+(p) = \max_{a\in[K]} H_a(p)$ admits the closed-form:*

$$H_a(p) = \begin{cases} -\left(\sum_i p_i\right) \log\left(\sum_{x\in\mathcal{X}} \prod_{i=1}^m \left(\nu_a^i(x)\right)^{\frac{p_i}{\sum_i p_i}}\right) & \text{if } \sum_i p_i > 0, \\ \min_{x\in\mathcal{X}} -\sum_i p_i \log\left(\nu_a^i(x)\right) & \text{if } \sum_i p_i \le 0. \end{cases}$$

Further, assuming $\nu_a^i(x) \ge \epsilon > 0$ for all x, a, i, $H(p)$ is Lipschitz continuous with constant $\sqrt{m}\log(1/\epsilon)$ and is non-decreasing in each coordinate p_i.

The Hamiltonian H has this `Log-Sum-Exp` form since it is the Fenchel-conjugate of weighted entropic terms like KL-divergences. Since Isaacs' condition holds, the order of play between Nature and the Agent in the continuous game is irrelevant, providing an indirect proof of the achievability of R_∞^{go}.

Alternate Bounds via Viscosity Theory. Formulating R_∞^{go} as a PDE viscosity solution unlocks powerful tools from optimal control theory to derive analytical bounds. For example, by using simple convex upper bounds on the terminal function g, we can apply the Hopf formula and comparison principles for viscosity solutions (details in Appendix A.2) recovers the exponent bounds of (Nitinawarat et al., 2013) and (Hayashi, 2009) as strictly special cases of our framework (detailed in Appendix B.4). Using this same approach, we derive a new tighter bounds than Hayashi (2009); Nitinawarat et al. (2013) by simply choosing tighter convex upper bounds on g:

Proposition 4.3. *Let S denote the set of all pairs (i,j) with $i<j$. For each pair, let $\beta_{i,j} \in \Delta_m$ be a distribution supported strictly on $\{i,j\}$. For any $\lambda \in \Delta_S$, define the mixture $P(\lambda,\beta) \triangleq \sum_{(i,j)\in S} \lambda_{i,j}\beta_{i,j} \in \Delta_m$. Then, the minimax exponent is bounded by $R_\infty^{go} \le R_{Hopf}$, where*

$$R_{Hopf} \triangleq \inf_{\lambda\in\Delta_S} \sup_{\beta} \max_{a\in[K]} \left[-\log\left(\sum_{x\in\mathcal{X}} \prod_{k=1}^m \left(\nu_a^k(x)\right)^{P(\lambda,\beta)} \right) \right].$$

The next proposition establishes additional properties of the solution V for the HJI PDE (7) (proof in Appendix B.6):

Proposition 4.4. *The solution $V(y,t)$ satisfies the following three properties:*

- (Monotonicity) *If $a \succeq b$, then $V(a,t) \ge V(b,t)$, where $a \succeq b$ means a dominates b co-ordinate wise.*
- (Additive Homogenity) *For every $y \in \mathbb{R}^m$, we have $V(y + c\mathbb{1}, t) = V(y) + c$. Here $\mathbb{1}$ denotes the vector with all co-ordinates as 1.*
- (Scale Invariance) *We have that $\forall t \in [0,1], y \in \mathbb{R}^m$: $V(y,t) = (1-t)V\left(\frac{y}{1-t}, 0\right)$.*

These properties follow from the basic nature of the differential game we consider. Monotonicity follows from the fact that if the likelihood evidence at a given time is larger for all hypothesis, then it must be larger even at the terminal time. The additive homogenity is due to the intuitive fact that in hypothesis testing only the difference between log-likelihoods are important and not the absolute values

themselves. The scale-invariance is more surprising and comes from the fact that the terminal function g is both monotone and positively homogenous and hence a upscaling in space and is compensated for by a downscaling in time.

Geometric characterization by a Clairaut PDE: The scale invariance property in Proposition 4.4 allows us to derive the following alternate characterization of R^{go}_{∞} (proof in Appendix B.7):

Theorem 4.5. *Consider the following Clairaut PDE:*

$$\hat{V}(y) = \langle \nabla \hat{V}, y \rangle + H(\nabla \hat{V}), \qquad \lim_{||y|| \to \infty} \frac{\hat{V}(y)}{||y||} = g(\hat{y}), \tag{9}$$

where $y = ||y||\hat{y}$ with the unit vector $\hat{y}$ held fixed while taking the limit. Here, H is the same as the Hamiltonian in (7). Then the unique viscosity solution $\hat{V}$ to this time-invariant PDE satisfies $R^{go}_{\infty} = \hat{V}(0)$.

The importance of this result is that the characterization of R^{go}_{∞} is now independent of any temporal co-ordinate. It is natural to expect a time dependence for a dynamic optimization like AHT. But remarkably this Clairaut type PDE is time invariant. Moreover, it is easy to show from Proposition 4.4, that the solution $\hat{V}$ to Clairaut PDE satisfies both the monotonicity and additive homogenity conditions. The Clairaut type PDE geometrically embodies *flow-invariant* level sets to the time-averaged dynamics of (5) as discussed next.

Definition 4.6. A set S is **flow-invariant** if there exists a strategy for the Agent, to always ensure that if a trajectory starts in S at $t = 0$, it terminates in S at $t = 1$, regardless of the play by Nature.

Given a set S, one natural condition (called as *Nagumo's* condition (Blanchini, 1999)) to ensure flow-invariance is that at any point in the boundary ∂S, there exists atleast some $w \in \Delta_K$, such that

$$\sum_i \sum_a p_i w_a D(Q_a || \nu^i_a) \leq 0, \quad \text{for every } Q \in \Delta^K_{\mathcal{X}}.$$

Here p, is outwards *proximal* normal [1] at the boundary point. The intuition is that as the flow is about leave the boundary of S, the Agent can always play w satisfying Nagumo's condition, so that the flow moves back into the set S. The above condition can be conveniently re-phrased as $H(-p) \geq 0$, with H the Hamiltonian of (7).

A closely related concept is that of **Approachability**. We define a set S to be *approachable* if the dynamics of the time-averaged vector $\bar{y}_t := \frac{y_t}{t}$ is flow-invariant. This idea of approachability was introduced separately in the field of online learning by the seminal work (Blackwell, 1956). A simple change-of-formula calculation yields:

$$\frac{d\bar{y}_t}{dt} = \frac{(dy_t/dt - \bar{y}_t)}{t}.$$

Applying Nagumo's condition to this time-averaged dynamics yields the criteria $H(-p) + \langle -p, -\bar{y} \rangle \geq 0$. Now, if we assume $\hat{V}$ was differentiable, then the normal p for the sublevel set $S_0 := \{\bar{y} \mid \hat{V}(-\bar{y}) \leq 0\}$ at a boundary point $\bar{y}$ would be $-\nabla \hat{V}(-\bar{y})$. From the Clairaut PDE we then have that:

$$\hat{V}(-\bar{y}) = \langle -p, -\bar{y} \rangle + H(-p).$$

Since $\bar{y}$ is a boundary point, it is the case that $\hat{V}(-\bar{y}) = 0$. Hence, S_0 satisfies Nagumo's condition and is an approachable set. The above argument assumed $\hat{V}$ was differentiable, but a viscosity solution can be interpreted in terms of the proximal normals of sublevel sets as shown in Chapter 4 of (Clarke et al., 1998). This enables us to make the above argument rigorous. We record the result as (proof in Appendix B.8):

Theorem 4.7. *The sublevel set, S_0, is approachable to the time-averaged dynamics of (5).*

Remark 4.8. In fact, it can be further showed that amongst all similarly defined sublevel sets S_c, for some $c \in \mathbb{R}$, S_0 is the largest such set that is approachable.

We note that level set $l_0 := \{y \mid \hat{V}(-y) = 0\}$, is in fact the only important geometric object in the Clairaut PDE (9). This is so because any other level set maybe obtained by translating an element in l_0 by $-c\mathbb{1}$ to obtain an element in l_c by using the additive homogenity of $\hat{V}$.

This geometric characterization is likely to have algorithmic consequences for AHT. If a set is known to be approachable, then many prior results from the online learning literature like (Blackwell, 1956; Abernethy et al., 2011; Shimkin, 2016) can be leveraged to design algorithms for the AHT setting. One, structural issue though is that this prior work assumes the approachable set is convex and it is the case that S_0 is non-convex.

5. Explicit Upwind Scheme to approximate R^{go}_{∞}.

It was intractable to approximate R^{go}_{∞} using the representation in Komiyama et al. (2022) even when the number of hypothesis m was small. We utilize the PDE formulation (7) along with the following simple explicit upwind finite difference scheme to approximate R^{go}_{∞}:

$$\textit{FD scheme:} \qquad V^n_i = V^{n+1}_i + \Delta t H\left(\left(\frac{\Delta_j V^{n+1}_i}{\Delta h}\right)_{j \in [m]}\right), \tag{10}$$

[1] See (Clarke et al., 1998) for a formal definition

where V_i^n refers to the approximate value function at time $n\Delta t$ at grid point i. Δt and Δh are time and space discretizations respectively. Each grid point i corresponds to spatial point $(i_1\Delta h, i_2\Delta h, \ldots, i_m\Delta h)$. H is the Hamiltonian of the PDE (7). Further

$$\Delta_j V_i^n = V^n_{(i_1,i_2,\ldots,i_j+1,\ldots,i_m)} - V_i^n,$$

is the difference of approximated value functions at the forward point $(i_1, \ldots, i_j + 1, \ldots, i_m)$ from $i = (i_1, \ldots, i_j, \ldots, i_m)$.

To utilize this scheme in a practical computation we must first restrict the domain to compact set. Since the *speed of information propagation*[2] $v = \sqrt{m}\log(1/\epsilon)$, of PDE (7) is finite, only points reachable within time t by speeds less than v from y, will affect the value $V(y,t)$. The next Lemma formalizes this (proof in Appendix C.1):

Lemma 5.1. *In the PDE (7), suppose we replace the terminal condition g with a new function $g^{'}$ such that*

$$g(y) = g^{'}(y), \quad \forall y \in \left[0, \log\left(\frac{1}{\epsilon}\right)\right]^m.$$

Further, assume that $g^{'}$ is bounded, lipschitz and has compact support. This new terminal value problem is guaranteed to have a unique viscosity solution, denoted as $V^{'}$. Then, we have that:

$$V(y,t) = V^{'}(y,t),$$

whenever $0 \le y_i \le \log\left(\frac{1}{\epsilon}\right) t$, $\forall i \in [m]$ and $t \in [0,1]$.

Thus when computing, $V(0,0) = R_\infty^{go}$ (from Theorem 7), we replace g with g' such that $g = g'$ in an appropriate compact domain. One such g' is constructed next (proof in Appendix C.2):

Lemma 5.2. *Let $C = [0,v]^m$, with its l_∞ expansion by a radius $v/2$ denoted as $C_{v/2} = [-v/2, 3v/2]^m$. Define:*

$$g^{'}(y) = \begin{cases} g\left(\prod_C^{(2)}(y)\right)\left(1 - \frac{2d_\infty(y,C)}{v}\right), & y \in C_{v/2} \\ 0 & y \notin C_{v/2}, \end{cases} \tag{11}$$

where $\prod_C^{(2)}(y)$ and $d_\infty(y,C)$ denote, respectively, the l_2 projection and l_∞ distance of y from C. Then, $g^{'}(y)$ is a function with requisite properties needed in Lemma 5.1 and $V^{'}(y,t) = 0$[3] whenever $y \notin [-3v/2, 3v/2]^m$, $\forall t \in [0,1]$.

In the limit of $\Delta h, \Delta t \to 0$, Crandal & Lions (1984) showed that explicit schemes satisfying certain *consistency* and *montonicity* conditions are convergent to the viscosity solution.

[2]Lipschitz constant of H is called the speed of propagation in PDE literature, see (Tran, 2021).

[3]recall the definition of $V^{'}$ from lemma 5.1.

Definition 5.3. Any finite-difference scheme of the form:

$$V_i^n = V_i^{n+1} + \Delta t \times g\left(\frac{\Delta_j V_{i-k}^{n+1}}{\Delta h}_{j\in[m]}, \ldots, \frac{\Delta_j V_{i+l}^{n+1}}{\Delta h}_{j\in[m]}\right), \tag{12}$$

is called **consistent** with the PDE in (7), if $g(p,\ldots,p) = H(p)$, for any given vector $p \in \mathbb{R}^m$.

For the scheme (10), consistency with (7) follows since we use only one forward difference $\frac{\Delta_j V_i^n}{\Delta h}$ (here $k = l = 0$ in the consistency definition) per direction. Next, we define the monotonicity condition for a finite-difference scheme:

Definition 5.4. A scheme of the form in equation (12) is called **monotone** for the terminal value problem (7), if the RHS in (12) is non-decreasing V_r^{n+1} for every r, provided $\left|\left|\frac{\Delta_j V_r^{n+1}}{\Delta h}_{j\in[m]}\right|\right| \le L$, where L is the lipschitz constant of the terminal function g.

The monotonicity in V_i^n is satisfied if we set $\Delta h \ge m\Delta t \log\left(\frac{1}{\epsilon}\right)$ (see Lemma C.2). This is a Courant-Friedrichs-Lewy (CFL) stability condition known to be necessary for explicit schemes to converge (see (Courant et al., 1967)). Then, we have the following convergence result for the scheme (10) follow from Theorem 2 of Crandal & Lions (1984):

Theorem 5.5. *The explicit upwind scheme (10), with $\Delta h \ge m\Delta t \log\left(\frac{1}{\epsilon}\right)$, has the following error estimate:*

$$|V_i^n - V((i_1\Delta h, i_2\Delta h, \ldots, i_m\Delta h), n\Delta t)| \le c\sqrt{\Delta t}, \tag{13}$$

where c is a constant dependent only on $\sup_x |g^{'}(x)|$ and the Lipschitz constants of g' and H.

This guarantees atmost an $O(\sqrt{\Delta t})$ additive error when approximating R_∞^{go} using (5), for a chosen Δt.

Computational effort for approximating R_∞^{go}. The state space of the differential game is m-dimensional, requiring a spatial grid of size $\mathcal{O}(1/(\Delta t)^m)$. At each grid point, evaluating the Hamiltonian (Eq. (4.2)) requires $\mathcal{O}(K|\mathcal{X}|)$ operations. Updating the value function for a single time step via Eq. (10) thus requires $\mathcal{O}\left(K|\mathcal{X}|\frac{1}{(\Delta t)^m}\right)$ operations. Since the scheme integrates over $1/\Delta t$ time steps, the total computational cost is $\mathcal{O}\left(\frac{K|\mathcal{X}|}{(\Delta t)^{m+1}}\right)$. By setting the time step proportional to the batch size fraction ($\Delta t = 1/B$), the overall complexity becomes $\mathcal{O}(K|\mathcal{X}|B^{m+1})$.

Instead, suppose we are interested in evaluating the repeated Stackelberg game in Eq. (4) for a sufficiently large B. This can be expressed equivalently as:

$$R_B^{go} = \underbrace{\inf_{Q_1}\sup_{w_1}\ldots\inf_{Q_B}\sup_{w_B}}_{B \text{ times}} g(w^B, Q^B),$$

where g is the terminal value function defined in Eq. (6), and w^B, Q^B denote the sequence of controls applied by the Agent and Nature, respectively, throughout the B-stage game. To understand the computational burden in this approach, suppose we discretize the control spaces Δ_K and $(\Delta_{\mathcal{X}})^K$. At the very last stage, the Agent always chooses w_B to maximize the final cost $g(w^B, Q^B)$. Solving this via grid search requires $\mathcal{O}(me^K)$ operations, as the grid size scales exponentially with the dimension of the Agent's control set. Anticipating this optimal response, Nature must choose Q_B to minimize the terminal cost, which requires evaluating the Agent's response for each discretized Q_B. This drives the complexity of the final stage to $\mathcal{O}(me^{K|\mathcal{X}|+K})$. Propagating this backward induction over all B stages results in a staggering overall computational complexity of $\mathcal{O}(me^{BK|\mathcal{X}|})$. Because there are no known efficient methods for solving such repeated Stackelberg games, this exponential blow-up is unavoidable. Consequently, even for small values of m, K, and $|\mathcal{X}|$, evaluating the game for a moderate number of batches (e.g., $B \sim 10$) is computationally impossible.

For small m, this polynomial dependence on B for the explcit scheme (5) is vastly more efficient than the exponential intractability of the repeated Stackelberg game formulation. Setting $B = O(1/\epsilon^2)$, we see that the explicit upwind scheme (10) needs computational effort of $O\left(\frac{K}{\epsilon^{2(m+1)}}\right)$ to approximate R_∞^{go} to within ϵ additive accuracy while the repeated min-max representation (3.1) in equation (2) would require an effort of $O(me^{K/\epsilon^2})$. Note that the scheme (10) still suffers an exponential dependence on m. Thus, although the FD scheme is suitable only for small $m(\leq 4)$, it is efficient compared to (3.1), which is intractable even for such small m.

The geometric characterization of R_∞^{go} in Theorem 4.7, will enable much more efficient usage of finite-difference schemes. Firstly, the time discretization is no longer required while additive homogenity enables a further one-dimensional reduction in the state space to $\mathbb{R}^{m-1}$. We leave the task of developing more efficient schemes based on the geometric ideas of Theorem 4.7 for further work.

Numerical disproof of a conjecture. (Komiyama et al., 2022) had conjectured if R_1^{go} was achievable generally (see footnote 10 in page 4 of (Komiyama et al., 2022).). They had been unable to calculate R_∞^{go}, but using the scheme (10), we provide a simple example in Table 1 where for $m = K = 3$ where $R_\infty^{go} < R_1^{go}$ and thereby numerically disproving this conjecture.

The simple bandit class ξ shown in Table 1 contains three Bernoulli bandits hypotheses - $\nu^{(1)}, \nu^{(2)}, \nu^{(3)}$. Each of these bernoulli bandit have three arms, whose parameters are as shown in Table 1. In Table 1, the computed value of R_∞^{go} is $0.1234 \pm 2 \times 10^{-4}$, while R_1^{go} is $0.13205 \pm 1 \times 10^{-5}$, thereby falsifying the conjecture. The confidence intervals for R_∞^{go} are derived from Theorem 5.5, while those for R_1^{go} are derived by estimates outputted by standard optimizers [4] we employ in calculating it. Given our repeated Stackelberg game formulation of R_B^{go} (4) and our interpretation of R_∞^{go} as the value of a differential game (7) this seems natural. Indeed, most simple examples have the property $R_\infty^{go} < R_1^{go}$.

Quantity	Value
$\nu^{(1)}$	$(0.6, 0.3, 0.23)$
$\nu^{(2)}$	$(0.35, 0.8, 0.35)$
$\nu^{(3)}$	$(0.2, 0.3, 0.75)$
R_∞^{go}	$0.1234 \pm 2 \times 10^{-4}$
R_1^{go}	$0.13205 \pm 1 \times 10^{-5}$
R_{static}	$0.07814 \pm 3 \times 10^{-5}$

Table 1. A simple bandit class $\xi = \{\nu^{(1)}, \nu^{(2)}, \nu^{(3)}\}$, with three $(m = 3)$ three-armed bernoulli bandits $(K = 3)$.

6. Conclusion

In this work, we have established a novel bridge between large deviation theory in Active Hypothesis Testing (AHT) and the mathematical framework of two-player zero-sum differential games. By recasting the optimal minimax error exponent R_∞^{go} as the value of a continuous-time game, we move beyond the computationally intractable discrete-time limits proposed in prior literature.

Our formulation as a Hamilton-Jacobi-Isaacs (HJI) Partial Differential Equation provides several key advantages. Analytically, the viscosity solution theory allows for a unified derivation of existing information-theoretic bounds and the development of new analytical constraints via the Hopf representation. Geometrically, the discovered time-scaling property reduces the dynamic problem to a time-invariant Clairaut PDE, revealing a deep connection between optimal testing strategies and Blackwell approachability. Numerically, we have provided the first stable finite-difference scheme capable of approximating R_∞^{go} for small hypothesis sets. Crucially, our numerical results provide a definitive counterexample to the conjecture that single-batch and infinite-batch exponents are equivalent, demonstrating that the value of adaptivity in AHT is strictly more nuanced than previously hypothesized.

Several promising directions for future research remain. First, while this work characterizes the optimal exponent, a vital next step is the synthesis of closed-loop sampling policies directly from the HJI viscosity solution. Second, the

[4] We use scipy.optimize modules differential_evolution and shgo.

extension of this game-theoretic framework to the more general case of Fixed-Budget Best Arm Identification (FB-BAI) with composite hypotheses remains an open challenge. Finally, investigating the implications of the Clairaut PDE characterization for finite-time, non-asymptotic regimes may provide tighter guarantees for approximating R_∞^{go} and designing algorithms for AHT.

References

Abernethy, J., Bartlett, P. L., and Hazan, E. Blackwell approachability and no-regret learning are equivalent. In *Proceedings of the 24th Annual Conference on Learning Theory*, pp. 27–46. JMLR Workshop and Conference Proceedings, 2011.

Bardi, M. and Evans, L. C. On hopf's formulas for solutions of hamilton-jacobi equations. *Nonlinear Analysis: Theory, Methods & Applications*, 8(11):1373–1381, 1984.

Blackwell, D. An analog of the minimax theorem for vector payoffs. 1956.

Blanchini, F. Set invariance in control. *Automatica*, 35(11): 1747–1767, 1999.

Clarke, F. H., Ledyaev, Y. S., Stern, R. J., and Wolenski, R. *Nonsmooth analysis and control theory*. Springer, 1998.

Courant, R., Friedrichs, K., and Lewy, H. On the partial difference equations of mathematical physics. *IBM journal of Research and Development*, 11(2):215–234, 1967.

Crandal, M. and Lions, P. Two approximations of solutions of hamilton–jacobi equations. *Math. Comp*, 43(167): 1–19, 1984.

Dembo, A. *Large deviations techniques and applications*. Springer, 2009.

Elliott, R. J. and Kalton, N. J. *The existence of value in differential games*, volume 126. American Mathematical Soc., 1972.

Glimm, J., Kranzer, H., Tan, D., and Tangerman, F. Wave fronts for hamilton-jacobi equations: The general theory for riemann solutions in. *Communications in mathematical physics*, 187(3):647–677, 1997.

Hayashi, M. Discrimination of two channels by adaptive methods and its application to quantum system. *IEEE Transactions on Information Theory*, 55(8):3807–3820, 2009.

Hong, L. J., Fan, W., and Luo, J. Review on ranking and selection: A new perspective. *Frontiers of Engineering Management*, 8(3):321–343, 2021.

Komiyama, J., Tsuchiya, T., and Honda, J. Minimax optimal algorithms for fixed-budget best arm identification. *Advances in Neural Information Processing Systems*, 35: 10393–10404, 2022.

Nitinawarat, S., Atia, G. K., and Veeravalli, V. V. Controlled sensing for multihypothesis testing. *IEEE Transactions on automatic control*, 58(10):2451–2464, 2013.

Qin, C. Open problem: Optimal best arm identification with fixed-budget. In Loh, P.-L. and Raginsky, M. (eds.), *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pp. 5650–5654. PMLR, 02–05 Jul 2022. URL https://proceedings.mlr.press/v178/open-problem-qin22a.html.

Shimkin, N. An online convex optimization approach to blackwell's approachability. *Journal of Machine Learning Research*, 17(129):1–23, 2016.

Souganidis, P. E. Two-player, zero-sum differential games and viscosity solutions. In *Stochastic and differential games: Theory and numerical methods*, pp. 69–104. Springer, 1999.

Tran, H. V. *Hamilton–Jacobi equations: theory and applications*, volume 213. American Mathematical Soc., 2021.

A. Primer on zero sum differential games and viscosity solutions.

A differential game is a continuous time game played over the interval $[0,1]$ by N agents. Each agent $l \in [N]$ can use a control $u_{t,l}$, $t \in [0,1]$ to control a state $y_t \in \mathbb{R}^m$. Each agent aims to minimize their own functional $g_l(.)$ of the terminal state y_1. The dynamical equation governing the evolution of state y_t is given by:

$$\dot{y}_{i,t} = f_i(u_t), \quad \forall i \in [m] \tag{14}$$

where $u_t = (u_{t,l})_{l \in [N]}$. This system of differential equations is solved with the initial condition $y_0 = y$ for a fixed set of controls.

Hereafter, we specialize to the two player (N=2) zero-sum case ($g := g_1(.) = -g_2(.)$). We follow the setup given in Souganidis (1999) but specialized to our specific case. We will assume that both players can play controls w and Q from the compact sets $\mathcal{W}$ and $\mathcal{Q}$. The maximizing player will be called *agent* and the minimizing player called *nature*. We assume the following:

1. The function $f(.)$ is uniformly continuous over the set $\mathcal{W} \times \mathcal{Q}$.
2. The function $g(.)$ is Lipschitz.

One of the main aims of this section is to relate the above differential game with the following Hamilton Jacobi Isaacs' PDE:

$$\begin{aligned} &\frac{\partial V}{\partial t} + H\left(\nabla V\right) = 0 \\ &V(y,1) = g(y), \quad \forall y \in R^m, \end{aligned} \tag{15}$$

where the function $H(.)$ is called as the *Hamiltonian* and can take the form of the *upper Hamiltonian*:

$$H^+(p) = \min_{Q \in \mathcal{Q}} \max_{w \in \mathcal{W}} \left\{ \sum_{i=1}^{m} f_i(w,Q) p_i \right\}$$

or the *lower Hamiltonian*

$$H^-(p) = \max_{w \in \mathcal{W}} \min_{Q \in \mathcal{Q}} \left\{ \sum_{i=1}^{m} f_i(w,Q) p_i \right\}.$$

For any $p \in \mathbb{R}^m$, we have that : $H^-(p) \leq H^+(p)$. If the equality holds for all p, that is, $H^+(p) = H^-(p)$, then we say the *Isaacs' condition* is satisfied.

A.1. A dynamic programming representation of the value functions of a differential game.

In this section we will give a Dynamic Programming (DP) representation of $V(y,t)$ in terms of non-anticipating strategies (defined below) first defined in (Elliott & Kalton, 1972). First let us define set of measurable controls:

Definition A.1. A measurable control for the agent from time t is a measurable mapping :

$$w_{[t,1]} : [t,1] \longrightarrow \mathcal{W}$$

for each $t \in [0,1)$. Similarly a measurable control for nature from time t is defined as the measurable mapping:

$$Q_{[t,1]} : [t,1] \longrightarrow \mathcal{Q}.$$

Denote the set of such measurable controls as $\mathcal{W}(t)$ and $\mathcal{Q}(t)$ respectively for agent and nature respectively. Next, we define non-anticipative strategies for both nature and agent:

Definition A.2 (Non-anticipative strategy for agent)**.** A non-anticipative strategy for the agent beginning from time t is given by the mapping:

$$\alpha : \mathcal{Q}(t) \longrightarrow \mathcal{W}(t) \tag{16}$$

such that for each $s \in [t,1]$ and Q^0,$Q^1 \in \mathcal{Q}(t)$, if $Q^0 = Q^1$ almost everywhere in $[t,s]$ then we have that $\alpha(w^0) = \alpha(w^1)$ almost everywhere in $[t,s]$.

Denote the set of all non-anticipative strategies for the agent beginning from t as $\mathcal{A}(t)$. Similarly, we can defined non-anticipative strategies for nature:

Definition A.3 (Non-anticipative strategy for nature)**.** A non-anticipative strategy for nature beginning from time t is given by the mapping:

$$\gamma : \mathcal{W}(t) \longrightarrow \mathcal{Q}(t) \tag{17}$$

such that for each $s \in [t, 1]$ and w^0,$w^1 \in \mathcal{W}(t)$, if $w^0 = w^1$ almost everywhere in $[t, s]$ then we have that $\gamma(w^0) = \gamma(w^1)$ almost everywhere in $[t, s]$.

Denote the set of such non-anticipative strategies of nature starting from t as $\mathcal{N}(t)$. We define next thew upper and lower values of the differential game as:

Definition A.4 (Upper and Lower values of the differential game)**.** Suppose the state is y at time $t \in [0, 1)$, then we define the **upper** value of the game from (y, t) as:

$$V^+(y, t) = \sup_{w \in \mathcal{A}(t)} \inf_{Q_{[t,1]} \in \mathcal{Q}(t)} g(y_1) \tag{18}$$

and the **lower** value of the game is defined as:

$$V^-(y, t) = \inf_{Q \in \mathcal{N}(t)} \sup_{w_{[t,1]} \in \mathcal{W}(t)} g(y_1). \tag{19}$$

The upper value represents a situation where the agent can respond to the strategy of nature and hence has the advantage. A similar conclusion can be drawn for the lower value. Next, we state the following result which shows the DP principle for upper and lower values:

Theorem A.5 (Theorem 2.2 in Souganidis (1999))**.** *For every $(y, t) \in \mathbb{R}^m \times [0, 1]$ and $\epsilon > 0$ such that $t + \epsilon \in [0, 1]$ we have:*

$$V^+(y, t) = \sup_{w \in \mathcal{A}(t)} \inf_{Q_{[t,1]} \in \mathcal{Q}(t)} V^+(y_{t+\epsilon}, t + \epsilon) \tag{20}$$

and

$$V^-(y, t) = \inf_{Q \in \mathcal{N}(t)} \sup_{w_{[t,1]} \in \mathcal{W}(t)} V^-(y_{t+\epsilon}, t + \epsilon) \tag{21}$$

where $y_{t+\epsilon}$ is the solution of (14) under the appropriate controls $w_{[t,1]}$ and $Q_{[t,1]}$.

A.2. Viscosity solution to Hamilton-Jacobi-Isaacs PDE

Many PDEs even with smooth Hamiltonians and terminal boundary conditions can develop discontinuities in the gradient, while many Lipschitz functions can satisfy the PDE at points of differentiability but are otheriwse essentially non-unique. We expect the value functions in optimal control and differential control to be the unique solution that satisfies the PDE at points of differentiability. A *viscosity* solution to a PDE is a generalized solution that has both uniqueness and stability properties necessary for them to be useful in differential games.

Definition A.6 (Viscosity solution)**.** For the PDE :

$$\begin{aligned} &\frac{\partial V}{\partial t} + H(\nabla V) = 0, \\ &V(y, 1) = g(y) \end{aligned} \tag{22}$$

a bounded uniformly continuous function V is a **viscosity subsolution** (resp. **supersolution**) if for each smooth function $\phi \in \mathcal{C}^\infty(\mathbb{R}^m \times (0, \infty))$, if $V - \phi$ has a local maximum at a point (y_0, t_0), then

$$\frac{\partial \phi(y_0, t_0)}{\partial t} + H(\nabla \phi(y_0, t_0)) \leq 0.$$

(resp. if for each smooth function $\phi \in \mathcal{C}^\infty(\mathbb{R}^m \times (0, \infty))$, if $V - \phi$ has a local minimum at a point (y_0, t_0), then

$$\frac{\partial \phi(y_0, t_0)}{\partial t} + H(\nabla \phi(y_0, t_0)) \geq 0.)$$

A **viscosity solution** u to the above PDE (22) is both a viscosity subsolution and a supersolution along with satisfying the terminal condition $V(y, 1) = g(y)$.

The following theorem guarantee the uniqueness and stability properties of the viscosity solution:

Theorem A.7 (Theorem 3.3 Souganidis (1999))**.** *Assuming the Hamiltonian H and terminal condition g in (22) are Lipschitz, we have that:*

1. *There exists a unique viscosity solution satisfying (22).*
2. *Let u_1 and u_2 be two viscosity solution for two different terminal conditions $g_1 \leq g_2$ then we have that $u_1(.,t) \leq u_2(.,t)$.*

In certain special cases one can provide an explicit representation for the Hamilton-Jacobi PDEs:

Theorem A.8 (Theorem 2.1 and 3.1 Bardi & Evans (1984))**.** *Consider the PDE problem with initial condition:*

$$\frac{\partial V}{\partial t} + H(\nabla V) = 0,$$
$$V(y,0) = g(y)$$

then

- ***Hopf-Lax formula*** *If the Hamiltonian H is* convex *and g is uniformly Lipschitz, then the viscosity solution for the above problem is given by:*
$$V(y,t) = \inf_z \sup_p \{g(z) + \langle p, y-z\rangle - tH(p)\} \tag{23}$$
- ***Hopf formula*** *If the Hamiltonian H is continuous and g is uniformly Lipschitz and* convex*, then the viscosity solution for the above problem is given by:*
$$V(y,t) = \sup_p \inf_z \{g(z) + \langle p, y-z\rangle - tH(p)\} \tag{24}$$

A.3. Value functions as viscosity solutions of Hamilton-Jacobi-Isaacs' PDE

We have the following theorem relating the upper and lower values of the differential game:

Theorem A.9 (Theorem 4.1 Souganidis (1999))**.**

1. *The upper value V^+ of the differential equation defined above is the unique viscosity solution of the following PDE:*
$$\frac{\partial V^+}{\partial t} + H^+(\nabla V^+) = 0,$$
$$V^+(y,1) = g(y)$$
where H^+ is the upper Hamiltonian.
2. *The lower value V^- of the differential equation defined above is the unique viscosity solution of the following PDE:*
$$\frac{\partial V^-}{\partial t} + H^-(\nabla V^-) = 0,$$
$$V^-(y,1) = g(y)$$
where H^- is the lower Hamiltonian.

If in addition the Isaacs' condition holds we have:

Lemma A.10 (Corollary 4.2 Souganidis (1999))**.** *We have under the Isaacs condition $H^+(p) = H^-(p)$ that:*

$$V(y,t) := V^+(y,t) = V^-(y,t)$$

along with the following representations

$$\begin{aligned} V(y,t) &= \sup_{w\in\mathcal{A}(t)} \inf_{Q_{[t,1]}\in\mathcal{Q}(t)} g(y_1) \\ &= \inf_{Q\in\mathcal{N}(t)} \sup_{w_{[t,1]}\in\mathcal{W}(t)} g(y_1). \end{aligned} \tag{25}$$

and the following DP recursion satisfied for any $s > t$:

$$\begin{aligned} V(y,t) &= \sup_{w\in\mathcal{A}(t)} \inf_{Q_{[t,1]}\in\mathcal{Q}(t)} V(y_s,s) \\ &= \inf_{Q\in\mathcal{N}(t)} \sup_{w_{[t,1]}\in\mathcal{W}(t)} V(y_s,s). \end{aligned} \tag{26}$$

We next provide a result that connects the discrete time repeated games (like the one considered in Proposition 3) with differential games in continuous time. Consider any partition $P = \{0 = t_0 < t_1 < \cdots < t_{N(P)} = 1\}$ of $[0,1]$. We define the following piece-wise constant in time functions:

$$\begin{aligned} &W_P^+(y,T) = g(y) \\ &W_P^+(y,t) = \inf_{Q\in\mathcal{Q}} \sup_{w\in\mathcal{W}} \{W_P^+(y+(t_{j+1}-t)f(w,Q),t_{j+1})\} \quad \text{if } t\in[t_j,t_{j+1}) \text{ for } j=0,1,\ldots,t_{N(P)}-1, \end{aligned}$$

and

$$\begin{aligned} &W_P^-(y,T) = g(y) \\ &W_P^-(y,t) = \sup_{w\in\mathcal{W}} \inf_{Q\in\mathcal{Q}} \{W_P^-(y+(t_{j+1}-t)f(w,Q),t_{j+1})\} \quad \text{if } t\in[t_j,t_{j+1}) \text{ for } j=0,1,\ldots,t_{N(P)}-1. \end{aligned}$$

Then we have that:

Theorem A.11 (Theorem 4.4 Souganidis (1999))**.** *Under the assumptions of this section we have that $W_P^\pm \to V^\pm$ locally uniformly in $\mathbb{R}^m \times [0,1]$ as $|P| \to 0$, where $V^\pm$, as before, are the unique viscosity solutions of the PDEs:*

$$\begin{aligned} \frac{\partial V^\pm}{\partial t} + H^\pm(\nabla V^\pm) = 0, \\ V^\pm(y,1) = g(y) \end{aligned}$$

and as before under the Isaacs' condition we have that $V^+ = V^-$.

B. Supplemental material related to and proofs for Section 3.

Claim B.1. *Given N arbitrary sets A_i, $i \in [N]$ and any collection of functions $f_i(.)$ which takes elements from A_i as input, we have:*

$$\max_j \min_{j\neq i} \inf_{a\in A_i} f_i(a) = \min_{i\neq j} \max\left\{\inf_{a\in A_i} f_i(a), \inf_{a\in A_j} f_j(a)\right\}.$$

Proof. Let us define $h_i := \inf_{a\in A_i} f_i(a)$. Then we just need to show that :

$$\max_j \min_{j\neq i} h_i = \min_{i\neq j} \max\left\{h_i, h_j\right\}.$$

Let us reorder the h_i's such that $h_{(1)} \leq h_{(2)} \leq \ldots, h_{(N)}$. Now let us interpret the LHS of the above equation. For any fixed $j \neq (1)$, we observe that $\min_{j\neq i} h_i = h_{(1)}$. However when $j = (1)$ we then have that $\min_{j\neq i} h_i = h_{(2)}$, Since $h(2) \geq h(1)$, we conclude that the LHS equals $h_{(2)}$. On the RHS it is clear the ordered pair $((1),(2))$ attains the minima of $max(h_{(1)}, h_{(2)}) = h_{(2)}$. Thus we have shown that both RHS and LHS are equal to $h_{(2)}$.

□

B.1. Proof of Proposition 3.1.

Proof. From equation (2) we have that:

$$R_B^{go} = \sup_{r^B(.),J(.)} \inf_{Q^B} \inf_{\substack{\nu\in\xi \\ J(Q^B)\neq i^*(\nu)}} \frac{1}{B}\sum_{l=1}^{B}\sum_{a\in[K]} r_l(Q^l)(a) D(Q_{l,a}||\nu_a).$$

As $r^B(.)$ and $J(.)$ are functions taking Q^B as argument we can just optimize the function for each fixed Q^B. So assuming Q^B is fixed we want to optimize

$$\inf_{\substack{\nu\in\xi \\ J(Q^B)\neq i^*(\nu)}} \frac{1}{B}\sum_{l=1}^{B}\sum_{a\in[K]} r_l(Q^l)(a)D(Q_{l,a}||\nu_a),$$

while still respecting the sequential nature of the functions $r_1, r_2, \ldots, r_B, J$. This is then just a simple backward induction process. So, we start with assuming all $r_1, \ldots, r_B$ have been chosen and optimize only $J(.)$ given Q^B. Clearly the optimal choice here is :

$$J^*(Q^B) = \underset{j}{argmax} \min_{j\neq i} \inf_{\nu\in\xi_i} \frac{1}{B}\sum_{l=1}^{B}\sum_{a\in[K]} r_l(Q^l)(a)D(Q_{l,a}||\nu_a).$$

Now after this optimization over $J(.)$ and fixing the choices $r_1, \ldots, r_{B-1}$, we optimize over r_B. This entails just taking a supremum over Δ_K:

$$r_B^* = \underset{w_B\in\Delta_K}{argmax} \max_j \min_{j\neq i} \inf_{\nu\in\xi_i} \frac{1}{B}\sum_{l=1}^{B-1}\sum_{a\in[K]} r_l(Q^l)(a)D(Q_{l,a}||\nu_a) + \frac{1}{B}\sum_{a\in[K]} w_B(Q^B)(a)D(Q_{B,a}||\nu_a).$$

Note that the argmax here is nonempty because Δ_K is compact and the optimizing objective is continuous in w_B, since ξ_i are also compact. Now we observe that $r_1, \ldots, r-B-1$ need only Q^{B-1} as input. The R_B^{go} definition has an infimization over Q^B. This means we must now infimize over Q_B first and then again optimize over r^{B-1} assuming Q^{B-1} and $r_1, \ldots, r_{B-2}$ are held fixed. In this way we also respect the sequential nature inherent in Q^B. Hence we have:

$$r_{B-1}^* = \underset{w_{B-1}}{argmax} \inf_{Q_B} \max_{w_B} \max_j \min_{j\neq i} \inf_{\nu\in\xi_i} \Bigg\{ \frac{1}{B}\sum_{l=1}^{B-2}\sum_{a\in[K]} r_l(Q^l)(a)D(Q_{l,a}||\nu_a) + \frac{1}{B}\sum_{l=B-1}^{B}\sum_{a\in[K]} w_l(Q^l)(a)D(Q_{l,a}||\nu_a)\Bigg\}.$$

The nonemptiness of the argmax follows from the compactness of Δ_K, ξ_i and the fact that Q_B belongs to compact set by the finite support assumption. Now clearly the procedure to define the optimal r_{B-2}^* to r_1^* is similar and maybe similarly carried to get the equivalent representation of R_B^{go}. This completes the proof. □

B.2. Proof of Theorem 4.1.

Proof. We utilize the results and theory developed in Appendix A. For the specific differential game considered in section 3.2, we have that $\mathcal{W} = \Delta_K$ and $\mathcal{Q} = \Delta_{\mathcal{X}}^K$. The dynamical equations are given by:

$$f_i(w, Q) = \sum_{a\in[K]} w(a)D(Q_{ta}||\nu_a^i).$$

Since by assumption ν_a^i has full support over the finite support $\mathcal{X}$ for all $a \in [K], i \in [m]$ we have that $D(Q||\nu_a^i)$ is a continuous function. Consequently $f(w, Q)$ in our differential game is a jointly continuous function over the compact set $\mathcal{W} \times \mathcal{Q}$ and hence, also uniformly continuous.

Next, we verify that the terminal function $g(y) = y_{(2)}$ is Lipschitz continuous. We fist note that the co-ordinate function y_i are trivially Lipschitz. As minimum of finite number of Lipschitz functions is Lipschitz we have that $\min_{j\neq i} y_i$ is Lipschitz. Similarly, as maximum of finite number of Lipschitz functions is also Lipschitz, we have that $y_{(2)} = \max_j \min_{j\neq i} y_i$ is also Lipschitz and hence $g(y)$ is Lipschitz.

Since, we have verified the requisite assumptions, we can invoke Theorem A.9 to conclude the upper value functional $V(y, t)$ of the differential game in section 3.2 is the viscosity solution to Hamilton-Jacobi PDE given by:

$$\frac{\partial V}{\partial t} + \min_Q \max_{w\in\Delta_K} \Bigg\{ \sum_{i=1}^{m}\sum_{a\in[K]} w_t(a)D(Q_{t,a}||\nu_a^i)\frac{\partial V}{\partial y_i}\Bigg\} = 0$$

$$V(y, 1) = g(y) = y_{(2)}, \quad \forall y \in R^m.$$

Additionally we use Theorem A.11 with the partition $P_B = \{0 = t_0 < t_1 < \cdots < t_B = 1\}$ with each $t_i = i/B$ to get that:

$$W^+_{P_B}(0,0) = R^{go}_B$$

and $\lim_{B\to\infty} R^{go}_B = \lim_{B\to\infty} W^+_{P_B}(0,0) = V(0,0)$. Now we have by Theorem 7 in (Komiyama et al., 2022) that:

$$R^{go}_\infty = \lim_{B\to\infty} R^{go}_B.$$

and hence conclude that $R^{go}_\infty = V(0,0)$. □

B.3. Proof of Lemma 4.2.

Proof. The upper Hamiltonian is by definition:

$$H^+(p) = \min_Q \max_{w\in\Delta_K} \left\{ \sum_{i=1}^m \sum_{a\in[K]} w(a) D(Q_a||\nu^i_a) p_i \right\}.$$

Re-writing it and expanding Q for clarity we have:

$$H^+(p) = \min_{Q=(Q_a)_{a\in[K]}} \max_{w\in\Delta_K} \left\{ \sum_{a\in[K]} w(a) \sum_{i=1}^m D(Q_a||\nu^i_a) p_i \right\}.$$

Notice nature has to respond with a Q_a for each arm a in response to the agents play w. Further the objective as rewritten is linear in w. Suppose nature chooses $Q^*_a \in \underset{Q_a}{argmin} \sum_{i=1}^m D(Q_a||\nu^i_a) p_i$ for each arm a. Then we have by definition:

$$H^+(p) \le \max_{w\in\Delta_K} \left\{ \sum_{a\in[K]} w(a) \sum_{i=1}^m D(Q^*_a||\nu^i_a) p_i \right\}.$$

But for this choice Q^*_a we also have for every Q_a, $a \in [K]$ that:

$$\sum_{i=1}^m D(Q_a||\nu^i_a) p_i \ge \sum_{i=1}^m D(Q^*_a||\nu^i_a) p_i = \inf_{Q_a} \sum_{i=1}^m D(Q_a||\nu^i_a) p_i$$

and hence for every $w \in \Delta_K$:

$$\sum_a w(a) \sum_{i=1}^m D(Q_a||\nu^i_a) p_i \ge \sum_a w(a) \inf_{Q_a} \sum_{i=1}^m D(Q_a||\nu^i_a) p_i$$

and hence:

$$\sup_{w\in\Delta_K} \sum_a w(a) \sum_{i=1}^m D(Q_a||\nu^i_a) p_i \ge \sup_{w\in\Delta_K} \sum_a w(a) \inf_{Q_a} \sum_{i=1}^m D(Q_a||\nu^i_a) p_i.$$

As this is true for each Q_a, we can take an infimum over the LHS to get:

$$\begin{aligned} \inf_Q \sup_{w\in\Delta_K} \sum_a w(a) \sum_{i=1}^m D(Q_a||\nu^i_a) p_i &\ge \sup_{w\in\Delta_K} \sum_a w(a) \inf_{Q_a} \sum_{i=1}^m D(Q_a||\nu^i_a) p_i \\ &= \max_{w\in\Delta_K} \left\{ \sum_{a\in[K]} w(a) \sum_{i=1}^m D(Q^*_a||\nu^i_a) p_i \right\} \end{aligned}$$

Thus we have that:

$$H^+(p) = \max_{w\in\Delta_K} \left\{ \sum_{a\in[K]} w(a) \sum_{i=1}^m D(Q^*_a||\nu^i_a) p_i \right\}.$$

But we note that:

$$\begin{aligned}\max_{w\in\Delta_K}\left\{\sum_{a\in[K]} w(a)\sum_{i=1}^m D(Q_a^*||\nu_a^i)p_i\right\} &= \max_{w\in\Delta_K}\left\{\sum_{a\in[K]} w(a)\inf_{Q_a}\sum_{i=1}^m D(Q_a||\nu_a^i)p_i\right\}\\ &= \max_{w\in\Delta_K}\inf_{Q}\left\{\sum_{a\in[K]} w(a)\sum_{i=1}^m D(Q_a||\nu_a^i)p_i\right\}\\ &= H^-(p).\end{aligned}$$

Thus we have shown that $H^+(p) = H^-(p)$ for every p. From the above we also see that:

$$\begin{aligned}H^+(p) = H^-(p) &= \max_{w\in\Delta_K}\left\{\sum_{a\in[K]} w(a)\sum_{i=1}^m D(Q_a^*||\nu_a^i)p_i\right\}\\ &= \max_{a}\left\{\sum_{i=1}^m D(Q_a^*||\nu_a^i)p_i\right\}.\end{aligned}$$

Hence to get an exact expression of the Hamiltonians we next study the following optimization problem in detail:

$$\inf_{Q_a}\ \sum_{i=1}^m D(Q_a||\nu_a^i)p_i.$$

To that end we re-write the objective as follows:

$$\begin{aligned}\sum_{i=1}^m D(Q_a||\nu_a^i)p_i &= \sum_{i=1}^m p_i\sum_{x\in\mathcal{X}} Q_a(x)\log\left(\frac{Q_a(x)}{\nu_a^i(x)}\right)\\ &= \sum_{x\in\mathcal{X}}\sum_{i=1}^m p_i Q_a(x)\log\left(\frac{Q_a(x)}{\nu_a^i(x)}\right)\\ &= \sum_{x\in\mathcal{X}} Q_a(x)\sum_{i=1}^m p_i\log\left(\frac{Q_a(x)}{\nu_a^i(x)}\right).\end{aligned} \tag{27}$$

Define $P = \sum_i p_i$. If $P \neq 0$, we can normalize eqn (27) by P to get:

$$\begin{aligned}\sum_{i=1}^m D(Q_a||\nu_a^i)p_i &= \sum_{x\in\mathcal{X}} Q_a(x)\sum_{i=1}^m p_i\log\left(\frac{Q_a(x)}{\nu_a^i(x)}\right)\\ &= P\sum_{x\in\mathcal{X}} Q_a(x)\sum_{i=1}^m \frac{p_i}{P}\log\left(\frac{Q_a(x)}{\nu_a^i(x)}\right)\\ &= P\sum_{x\in\mathcal{X}} Q_a(x)\log\left(\frac{Q_a(x)}{\prod_{i=1}^m(\nu_a^i(x))^{p_i/P}}\right).\end{aligned}$$

We then normalize inside the logarithm using the partition term $Z = \sum_{x\in\mathcal{X}}\prod_{i=1}^m(\nu_a^i(x))^{p_i/P}$ to get:

$$\begin{aligned}\sum_{i=1}^m D(Q_a||\nu_a^i)p_i &= P\sum_{x\in\mathcal{X}} Q_a(x)\log\left(\frac{Q_a(x)Z}{\prod_{i=1}^m(\nu_a^i(x))^{p_i/P}Z}\right)\\ &= P\sum_{x\in\mathcal{X}} Q_a(x)\log\left(\frac{Q_a(x)Z}{\prod_{i=1}^m(\nu_a^i(x))^{p_i/P}}\right) - P\log(Z).\end{aligned}$$

Now observing that $\tilde{Q} = \left(\frac{\prod_{i=1}^m(\nu_a^i(x))^{p_i/P}}{Z}\right)_{x\in\mathcal{X}}$ is a probability distribution over $\mathcal{X}$ we have that:

$$\sum_{i=1}^m D(Q_a||\nu_a^i)p_i = PD(Q_a||\tilde{Q}) - P\log\left(\sum_{x\in\mathcal{X}}\prod_{i=1}^m(\nu_a^i(x))^{p_i/P}\right). \tag{28}$$

Let us now study the three different cases: (i) $P > 0$, (ii) $P < 0$ and (iii) $P = 0$.

Case 1 - $P > 0$: As $P > 0$, from eqn (28) the minimization over Q_a is achieved when $Q_a^* = \tilde{Q}$. Thus whenever $P > 0$ we have that:

$$\inf_{Q_a} \sum_i p_i D(Q_a||\nu_a^i) = -P \log \left(\sum_{x \in \mathcal{X}} \prod_{i=1}^m (\nu_a^i(x))^{p_i/P} \right).$$

Case 2 - $P < 0$: As $P < 0$, from eqn (28) we have that minimization of the objective obtained by maximizing $D(Q_a||\tilde{Q})$. As KL divergence is a strictly convex function the maxima is attained at δ distributions over $\mathcal{X}$. Hence we have:

$$\begin{aligned}
\inf_{Q_a} \sum_i p_i D(Q_a||\nu_a^i) &= P \max_{Q_a} \{D(Q_a||\tilde{Q})\} - P \log \left(\sum_{x \in \mathcal{X}} \prod_{i=1}^m (\nu_a^i(x))^{p_i/P} \right) \\
&= P \max_{x \in \mathcal{X}} \log \left(\frac{Z}{\prod_{i=1}^m (\nu_a^i(x))^{p_i/P}} \right) - P \log(Z) \\
&= P \log(Z) + P \max_{x \in \mathcal{X}} - \log \left(\prod_{i=1}^m (\nu_a^i(x))^{p_i/P} \right) - P \log(Z) \\
&= \min_{x \in \mathcal{X}} - P \log \left(\prod_{i=1}^m (\nu_a^i(x))^{p_i/P} \right) \\
&= \min_{x \in \mathcal{X}} - \sum_{i=1}^m p_i \log \left(\nu_a^i(x) \right).
\end{aligned}$$

Case 3 - $P = 0$: As $P = 0$, the eqn. (28) is not valid. Instead, we will work with eqn. (27). Thus, we have:

$$\begin{aligned}
\sum_{i=1}^m D(Q_a||\nu_a^i) p_i &= \sum_{x \in \mathcal{X}} Q_a(x) \sum_{i=1}^m p_i \log \left(\frac{Q_a(x)}{\nu_a^i(x)} \right) \\
&= \sum_{x \in \mathcal{X}} Q_a(x) \log \left(\frac{Q_a(x)^P}{\prod_{i=1}^m (\nu_a^i(x))^{p_i}} \right) \\
&= \sum_{x \in \mathcal{X}} Q_a(x) \log \left(\frac{1}{\prod_{i=1}^m (\nu_a^i(x))^{p_i}} \right),
\end{aligned}$$

where last equality is because $P = 0$. The final expression is also linear in Q_a and hence we have that:

$$\begin{aligned}
\inf_{Q_a} \sum_i p_i D(Q_a||\nu_a^i) &= \inf_{Q_a} \sum_{x \in \mathcal{X}} Q_a(x) \log \left(\frac{1}{\prod_{i=1}^m (\nu_a^i(x))^{p_i}} \right) \\
&= \min_{x \in \mathcal{X}} \log \left(\frac{1}{\prod_{i=1}^m (\nu_a^i(x))^{p_i}} \right) \\
&= \min_{x \in \mathcal{X}} - \sum_{i=1}^m p_i \log(\nu_a^i(x)).
\end{aligned}$$

Next, we show the Lipschitzness of the Hamiltonian function $H(p)$. We observe that for any fixed a, the function

$$H_a(p) = \inf_{Q_a} \sum_i p_i D(Q_a||\nu_a^i)$$

is a concave function in p over $\mathbb{R}^m$ and hence continuous over $\mathbb{R}^m$. Further, we have that:

$$H(p) = \max_a H_a(p),$$

is maximum of finite number of continuous concave functions and hence is also continuous over $\mathbb{R}^m$. We have shown that :

$$H_a(p) = \begin{cases} -(\sum_i p_i) \log \left(\sum_{x \in \mathcal{X}} \prod_{i=1}^m (\nu_a^i(x))^{\frac{p_i}{\sum_i p_i}} \right), & \text{if } \sum_i p_i > 0 \\ \min_{x \in \mathcal{X}} - \sum_{i=1}^m p_i \log(\nu_a^i(x)), & \text{if } \sum_i p_i \leq 0. \end{cases}$$

In the case $\sum_i p_i \leq 0$, we observe that $H_a(p)$ is the minimum of finite number (since $|\mathcal{X}| < \infty$) of linear functions in p and hence is going to be Lipschitz in p over the domain $\mathcal{D}_{\leq 0} = \{p : \sum_i p_i \leq 0\}$. It is straightforward to calculate the Lipschitz constant by invoking Lemma B.2. Since $|\log(\nu_a^i(x))| \leq \log(\frac{1}{\epsilon})$ we have that the Lipschitz constant in the domain $\mathcal{D}_{\leq 0}$ is $\sqrt{m}\log(\frac{1}{\epsilon})$.

In the case $\sum_i p_i > 0$, we note that $H_a(p)$ is differentiable. It is useful to define $s_p(x) = \prod_{i=1}^m (\nu_a^i(x))^{\frac{p_i}{\sum_i p_i}}$ and $s_p = \sum_{x\in\mathcal{X}} s_p(x)$. So we differentiate wrt to p_j:

$$
\begin{aligned}
\frac{\partial H_a(p)}{\partial p_j} &= -\log(s_p) - \frac{(\sum_i p_i)}{s_p}\left(\sum_{i=1}^m \sum_{x\in\mathcal{X}} \left(\prod_{i=1}^m (\nu_a^i(x))^{\frac{p_i}{\sum_i p_i}}\right)\log(\nu_a^i(x))\frac{\partial\left(\frac{p_i}{\sum_i p_i}\right)}{\partial p_j}\right) \\
&= -\log(s_p) - \frac{(\sum_i p_i)}{s_p}\left(\sum_{x\in\mathcal{X}} s_p(x)\sum_{i=1}^m \log(\nu_a^i(x))\frac{\partial\left(\frac{p_i}{\sum_i p_i}\right)}{\partial p_j}\right) \\
&= -\log(s_p) - \frac{(\sum_i p_i)}{s_p}\left(\sum_{x\in\mathcal{X}} s_p(x)\sum_{i=1}^m \log(\nu_a^i(x))\left(\frac{\delta_{ij}(\sum_i p_i) - p_i}{(\sum_i p_i)^2}\right)\right) \\
&= -\log(s_p) - \frac{(\sum_i p_i)}{s_p}\left(\sum_{x\in\mathcal{X}} s_p(x)\left(\frac{\log(\nu_a^j(x))}{(\sum_i p_i)} - \sum_{i=1}^m \frac{p_i\log(\nu_a^i(x))}{(\sum_i p_i)^2}\right)\right) \\
&= -\log(s_p) - \sum_{x\in\mathcal{X}} \frac{s_p(x)}{s_p}\log(\nu_a^j(x)) + \sum_{x\in\mathcal{X}} \frac{s_p(x)}{s_p}\sum_{i=1}^m \frac{p_i}{(\sum_i p_i)}\log(\nu_a^i(x)) \\
&= -\log(s_p) - \sum_{x\in\mathcal{X}} \frac{s_p(x)}{s_p}\log(\nu_a^j(x)) + \sum_{x\in\mathcal{X}} \frac{s_p(x)}{s_p}\log(s_p(x)).
\end{aligned}
$$

We observe that $s_p(x) \geq 0, \forall x \in \mathcal{X}$ and that $\sum_{x\in\mathcal{X}} \frac{s_p(x)}{s_p} = 1$. Denote this probability distribution as $S(p)$. Hence we have that:

$$
\begin{aligned}
\frac{\partial H_a(p)}{\partial p_j} &= \sum_{x\in\mathcal{X}} \frac{s_p(x)}{s_p}\log\left(\frac{s_p(x)/s_p}{\nu_a^j(x)}\right) \\
&= D(S(p)||\nu_a^j),
\end{aligned}
$$

Now under our assumptions we have that $D(S(p)||\nu_a^j) \leq \log\left(\frac{1}{\epsilon}\right)$. This immediately implies $||\nabla H_a|| \leq \sqrt{m}\log\left(\frac{1}{\epsilon}\right)$ and is the Lipschitz constant in the domain $\mathcal{D}_{>0} = \{p : \sum_i p_i > 0\}$.

Now we have found the Lipschitz constant of $H_a(p)$ in the two disjoint domains $\mathcal{D}_{\leq 0}$ and $\mathcal{D}_{>0}$. Further, we have argued that $H_a(p)$ is continuous throughout $\mathbb{R}^m$. Thus, we apply Lemma B.3 to conclude that $H_a(p)$ is Lipschitz throughout $\mathbb{R}^m$ and has a Lipschitz constant of $\sqrt{m}\log\left(\frac{1}{\epsilon}\right)$.

Since we have $H(p) = \max_a H_a(p)$ we can invoke Lemma B.2 to conclude that $H(p)$ is also Lipschitz with Lipschitz constant $\sqrt{m}\log\left(\frac{1}{\epsilon}\right)$.

Note that as $H_a(P)$ when $p \in \mathcal{D}_{\leq 0}$ is expressed as a $\min_{x\in\mathcal{X}}$ with each co-efficient of p_i being positive ($-\log(\nu_a^i(x)) > 0$ under our assumptions) we conclude that $H_a(p)$ is strictly increasing in each co-ordinate p_j for $p \in \mathcal{D}_{\leq 0}$.This concludes the proof of the lemma. In the case $p \in \mathcal{D}_{>0}$, as directly calculated above, $\frac{\partial H_a(p)}{\partial p_j} \geq 0$ and this implies that in $\mathcal{D}_{>0}$, $H_a(p)$ is non-decreasing in p_j. Thus $H_a(p)$ is non-decreasing over the entire domain for each a. From this the non-decreasing nature of $H(p)$ follows. This concludes the proof of the lemma. □

Lemma B.2. *Consider a finite index set I, indexing a collection of real valued Lipschitz functions f_i (we assume all f_i are defined over a common domain $\mathcal{D} \subseteq \mathbb{R}^n$), $i \in I$ with respective Lipschitz constants L_i. Then $g := \max_{i\in I} f_i$ and $h = \min_{i\in I} f_i$ are Lipschitz with Lipschitz constant $\max_{i\in I} L_i$.*

Proof. From the Lipschitzness of f_i we have that $\forall i \in I$ and any pair $x, y \in \mathcal{D}$:

$$f_i(x) \leq f_i(y) + L_i||x-y|| \leq f_i(y) + (\max_i L_i)||x-y||.$$

This implies by taking $\max$ amd $\min$ over I on the above inequality that:

$$g(x) \leq g(y) + (\max_{i \in I} L_i)||x-y||$$
$$h(x) \leq h(y) + (\max_{i \in I} L_i)||x-y||.$$

As the role of x and y can be interchanged in the above argument, we conclude that for all $x, y \in \mathcal{D}$:

$$|g(x) - g(y)| \leq (\max_{i \in I} L_i)||x-y||$$
$$|h(x) - h(y)| \leq (\max_{i \in I} L_i)||x-y||.$$

This proves the claim. □

Lemma B.3. *Suppose we have two Lipschitz functions f_A and f_{A^c} defined on a set $A \subsetneq \mathbb{R}^n$ and its complement $A^c \subsetneq \mathbb{R}^n$ with respective Lipschitz constants L_A and L_{A^c}. Define the glued up function:*

$$h(x) = \begin{cases} f_A(x), & x \in A, \\ f_{A^c}(x), & x \in A^c. \end{cases}$$

Suppose further that h is continuous over $\mathbb{R}^n$, then h is Lipschitz continuous with Lipschitz constant $\max\{L_A, L_{A^c}\}$.

Proof. Clearly when both $x, y \in A$ or $x, y \in A^c$ the following statement holds by Lipschitzness of f_A and f_{A^c}:

$$|h(x) - h(y)| \leq \max\{L_A, L_{A^c}\}||x-y||.$$

So we need to analyse only the case $x \in A$ and $y \in A^c$. Consider the closed straight line segment S between x and y (x, y are included in S). There exists atleast one point $s \in \partial A \cap S$. Now because s lies in the line segment S between x and y we have that:

$$||x-y|| = ||x-s|| + ||y-s||.$$

Assume that $s \in A$ (the case $s \in A^c$ can be analysed similarly). Consider a sequence $y_r \in A^c, \forall r \in \mathbb{N}$ such that $\lim_{r\to\infty} y_r = s$ (such a sequence exist because $s \in \partial A$), then we have for each $r \in \mathbb{N}$:

$$\begin{aligned} |h(x) - h(y)| &\leq |h(x) - h(s)| + |h(s) - h(y_r)| + |h(y_r) - h(y)| \\ &\leq |h(x) - h(s)| + |h(s) - h(y_r)| + L_{A^c}||y_r - y||. \end{aligned}$$

Taking the limit $r \to \infty$ and using the continuity of h and the norm function $d_y(z) = ||z-y||$ we have that:

$$|h(x) - h(y)| \leq |h(x) - h(s)| + L_{A^c}||s-y||.$$

As $s, x \in A$ using the Lipschitzness of f_A we have that:

$$\begin{aligned} |h(x) - h(y)| &\leq L_A||x-s|| + L_{A^c}||s-y|| \\ &\leq \max\{L_A, L_{A^c}\}(||x-s|| + ||s-y||) \\ &= \max\{L_A, L_{A^c}\}||x-y||. \end{aligned}$$

This shows the Lipschitzness of h. □

B.4. Consistency of the PDE with results of Hayashi (2009); Nitinawarat et al. (2013)

It will be useful to consider the *time-reversed* system of (7), gotten by the transform $\tilde{t} = 1 - t$:

$$\begin{aligned} &\frac{\partial \tilde{V}}{\partial t} + \max_{Q} \min_{w \in \Delta_K} \left\{ \sum_{i=1}^{m} \sum_{a \in [K]} -w_t(a) D(Q_{t,a} || \nu_a^i) \frac{\partial \tilde{V}}{\partial y_i} \right\} = 0 \\ &\tilde{V}(y, 0) = y_{(2)}, \quad \forall y \in R^m. \end{aligned} \tag{29}$$

where the new hamiltonian is $\tilde{H}(p) = -H(p)$ and $\tilde{V}(0,1) = V(0,0) = R_\infty^{go}$. We note that when the dimension $m = 2$, the function $g(y)$ simplifies to be $g(y_1, y_2) = \max\{y_1, y_2\}$ which is convex and Lipschitz (with a Lipschitz constant 1). Now this means we can apply the Hopf formula (24) to the time reversed PDE system to obtain:

$$\tilde{V}(y, \tilde{t}) = \sup_p \{\langle p, y \rangle + \tilde{t} H(p) - g^*(p)\}$$

where g^* is the fenchel conjugate of g. The conjugate g^* (for $m = 2$) is calculated as:

$$g^*(p) = \begin{cases} 0 & \text{if } p \in \Delta_2 \\ +\infty & \text{otherwise.} \end{cases}$$

As we are interested in $\tilde{V}(0,1)$ we have from Lemma 4.2:

$$R_\infty^{go} = \sup_{p \in \Delta_2} \max_a -\log \left(\sum_{x \in \mathcal{X}} (\nu_a^1(x))^{p_1} (\nu_a^2(x))^{p_2} \right)$$

which is equivalent to (Hayashi, 2009)'s characterization of the minimax exponent in the binary case $(m = 2)$.

Remark B.4. The above example also shows the apparent difficulty between $m = 2$ and $m > 2$ cases: the terminal condition $g(y)$ is not convex for $m > 2$ while it is for $m = 2$. The Hopf formula is only valid for convex terminal conditions.

Next, we combine Hopf formula with a convex upper bound for g, by using the comparison principle for viscosity solutions (Theorem A.9) to obtain upper bounds on R_∞^{go}, the minimax exponent. We note that $g(y) \leq g_{i,j}(y) = \max\{y_i, y_j\}$ for all $i \neq j$. Hence from the comparison principle and Hopf formula we have that:

$$\begin{aligned} \tilde{V}(0,1) &\leq \sup_p \{H(p) - g_{i,j}^*(p)\} \\ &= \sup_{p \in \Delta_{\{i,j\}}} \{H(p)\} \\ &= \max_a \; C(\nu_a^i, \nu_a^j) \end{aligned}$$

As this is true for any $(i,j), i \neq j$ we have that:

$$R_\infty^{go} \leq \min_{i \neq j} \max_a \; C(\nu_a^i, \nu_a^j),$$

which is the upper bound result from (Nitinawarat et al., 2013).

B.5. Proof of Proposition 4.3.

Proof. Fix a $\lambda \in \Delta_S$. Consider the the function :

$$g_\lambda(y) = \sum_{(i,j) \in S} \lambda_{i,j} \max\{y_i, y_j\}.$$

We observe that $g(y) \leq g_\lambda(y)$. Further, we note that this is a support function of set A_λ :

$$A_\lambda = \sum_{(i,j) \in S} \lambda_{i,j} \Delta_{\{i,j\}}$$

where $\Delta_{\{i,j\}}$ is viewed as subsets of Δ_m. That is g_λ is the support function of the minkowski sum of $\Delta_{\{i,j\}}$ simplexes. As g_λ is convex and Lipschitz, we can apply the comparison principle (Theorem A.9) or Proposition 1.1 (Crandal & Lions, 1984)) and the Hopf formula (24) to get that :

$$R_\infty^{go} \leq \sup_\beta \max_a -\log\left(\sum_{x\in\mathcal{X}}\prod_{k=1}^m (\nu_a^k(x))^{P(\lambda,\beta)}\right).$$

As the choice of λ was arbitrary the result follows. □

B.6. Proof of Proposition 4.4.

Proof. **Monotonicity**: From the representation of V in (18) and the monotonicity in g we know that for $t = 1$ the result is true. Now if $a \succeq b$ then there will always exist some strategy starting from a that will dominate terminal vector y_1 co-ordinate wise any strategy from b. This implies the monotonicity.
Additive Homogenity From the same representation (18) we have that since all co-ordinates of y are just translated by c for the vector $y + c\mathbb{1}$, all strategies terminal values are also similarly translated by c. We then observe the terminal loss g is additive homogenous and hence V is also additive homogenous.
Scale invariance Consider the following change of variables :

$$r = t + s(1-t)$$
$$\frac{y(r)}{(1-t)} = \tilde{y}(s),$$

with $s \in [0,1]$ and hence $r \in [t,1]$ and the initial condition $y(t) = y$ and $\tilde{y}(0) = \frac{y}{1-t}$. Now consider the following derivative:

$$\begin{aligned}\frac{d\tilde{y}(s)}{ds} &= \frac{1}{(1-t)}\frac{dy(r)}{ds}\\ &= \frac{1}{(1-t)}\frac{dy(r)}{dr}\frac{dr}{ds}\\ &= \frac{dy(r)}{dr}.\end{aligned}$$

Thus the dynamics for the game in the new co-ordinates $(s, \tilde{y}(s))$ is the same as the old co-ordinates $(r, y(r))$. In the new co-ordinates it begin with $s = 0$ and an initial condition of $\frac{y}{1-t}$. For any trajectory the final payoff in the new co-ordinates will $g(\tilde{y}(1)) = g\left(\frac{y(1)}{1-t}\right) = \frac{1}{1-t}g(y_1)$. From this relation the result immediately follows. □

B.7. Proof of Theorem 4.5

Proof. From the scale invariance property we have that if we set:

$$\hat{V}(y) = V(y, 0)$$

Then we have that

$$\frac{\partial V(y,t)}{\partial t} = -\hat{V}(y/(1-t), 0) + \langle \nabla\hat{V}, y/(1-t)\rangle$$

and that:

$$\frac{\partial V(y,t)}{\partial y_i} = \frac{\partial \hat{V}(y/(1-t))}{\partial y_i}.$$

Substituting this into HJI-PDE (7) and choosing $y = y/(1-t)$ gives the result. Another proof of the same result can also be seen in (Glimm et al., 1997). □

B.8. Proof of Theorem 4.7

Proof. To show that the sublevel set $S_0 := \{\bar{y} \mid \hat{V}(-\bar{y}) \leq 0\}$ is approachable under the time-averaged dynamics, we equivalently show that S_0 is *weakly invariant* (or viable) for the Agent. By Nagumo's Theorem for differential inclusions

(see, e.g., Theorem 2.4 in Chapter 4 of Clarke et al. (1998)), S_0 is flow-invariant if and only if the velocity vector can always be forced into the Bouligand tangent cone, or equivalently, it forms a non-positive inner product with any outward proximal normal.

The time-averaged dynamics are given by $\dot{\bar{y}} = \frac{1}{t}(f(w,Q) - \bar{y})$, where $f(w,Q)_i = \sum_{a\in[K]} w(a)D(Q_a\|\nu_a^i)$. For S_0 to be approachable, the Agent must be able to choose an action $w \in \Delta_K$ such that for any response $Q \in \Delta_{\mathcal{X}}^K$ by Nature, the flow points inwards. For any boundary point $\bar{y} \in \partial S_0$ and any proximal normal $\zeta \in N_{S_0}^P(\bar{y})$, we require:

$$\min_{w\in\Delta_K} \max_{Q\in\Delta_{\mathcal{X}}^K} \left\langle \zeta, \frac{1}{t}(f(w,Q) - \bar{y}) \right\rangle \leq 0. \tag{30}$$

Since $t > 0$, we can drop the $1/t$ factor. Recall the definition of the Hamiltonian under Isaacs' condition: $H(p) = \max_w \min_Q \langle p, f(w,Q)\rangle$. We can rewrite the min-max term in (30) as:

$$\begin{aligned}\min_w \max_Q \langle \zeta, f(w,Q)\rangle &= -\max_w \min_Q \langle -\zeta, f(w,Q)\rangle \\ &= -H(-\zeta).\end{aligned}$$

Thus, the approachability condition (30) simplifies to:

$$-H(-\zeta) - \langle \zeta, \bar{y}\rangle \leq 0 \iff H(-\zeta) + \langle \zeta, \bar{y}\rangle \geq 0. \tag{31}$$

Next, we characterize the proximal normal ζ. By the calculus of proximal subgradients (Clarke et al. (1998), Chapter 1, Section 6), the proximal normal cone to a sublevel set is generated by the proximal subdifferential of its defining function. Thus, if $\zeta \in N_{S_0}^P(\bar{y})$, there exists a scalar $\lambda \geq 0$ and a proximal subgradient $\eta \in \partial_P(\hat{V}(-\cdot))(\bar{y})$ such that $\zeta = \lambda\eta$. By the chain rule, $\eta = -p$ for some $p \in \partial_P\hat{V}(-\bar{y})$. Letting $y = -\bar{y}$, we have $\hat{V}(y) = 0$ (since $\bar{y} \in \partial S_0$), and the outward normal takes the form $\zeta = -\lambda p$, where $p \in \partial_P\hat{V}(y)$.

Substituting $\zeta = -\lambda p$ and $\bar{y} = -y$ into our required condition (31), we obtain:

$$H(\lambda p) + \langle -\lambda p, -y\rangle \geq 0.$$

Since H is positively homogeneous of degree 1 (Proposition 4.4), we can factor out $\lambda \geq 0$. The condition therefore holds if and only if for all $p \in \partial_P\hat{V}(y)$:

$$H(p) + \langle p, y\rangle \geq 0. \tag{32}$$

To prove (32), we utilize the fact that $V(x,t) = (1-t)\hat{V}\left(\frac{x}{1-t}\right)$ is the viscosity solution to the time-dependent HJI PDE, $\partial_t V + H(\nabla_x V) = 0$. The condition $p \in \partial_P\hat{V}(y)$ implies there exists a smooth test function ϕ touching $\hat{V}$ from below at y, such that $\nabla\phi(y) = p$. We construct a space-time test function $\Phi(x,t) = (1-t)\phi\left(\frac{x}{1-t}\right)$. Consequently, $V - \Phi$ has a local minimum at $(y, 0)$.

By the definition of a viscosity *supersolution* (Clarke et al. (1998), Chapter 4, Section 7), at this local minimum, the PDE operator applied to the test function must be non-negative:

$$\partial_t\Phi(y,0) + H(\nabla_x\Phi(y,0)) \geq 0.$$

Evaluating the derivatives of Φ at $(y,0)$ yields $\nabla_x\Phi(y,0) = \nabla\phi(y) = p$, and $\partial_t\Phi(y,0) = -\phi(y) + \langle\nabla\phi(y), y\rangle = -\hat{V}(y) + \langle p, y\rangle$. Substituting these into the supersolution inequality gives:

$$-\hat{V}(y) + \langle p, y\rangle + H(p) \geq 0.$$

Since $y = -\bar{y}$ is on the boundary of the zero-sublevel set, $\hat{V}(y) = 0$. This immediately yields $\langle p, y\rangle + H(p) \geq 0$, verifying condition (32).

Therefore, Nagumo's condition is strictly satisfied everywhere on ∂S_0, proving that the sublevel set S_0 is approachable. □

C. Proof related to Section 5

C.1. Proof of Lemma 5.1

Proof. We will use the fact that :

$$0 \leq \dot{y}_{t,i} \leq \log\left(\frac{1}{\epsilon}\right).$$

This implies that from any starting point y_0, we have that each co-ordinate of $y_{t,i}$, for $t \in [0,1]$, is contained in the hypercube $[y_{0,1}, y_{0,1} + \log\left(\frac{1}{\epsilon}\right) t] \times \cdots \times [y_{0,m}, y_{0,m} + \log\left(\frac{1}{\epsilon}\right) t]$. Now, we are given that when $y_0 = 0$, we have that over the hypercube $[0, \log\left(\frac{1}{\epsilon}\right)]^m$, $g(y) = g^{'}(y)$. Thus, the set of all possible reachable states from $y_0 = 0$ in the time interval $[0,1]$ is contained in the hypercube $[0, \log\left(\frac{1}{\epsilon}\right)]^m$. Similarly one can argue that for an initial condition (y,t) satisfying $0 \leq y_i \leq \log\left(\frac{1}{\epsilon}\right) t$ for any $t \in [0,1)$ the reachable set of states at terminal time $t = 1$ is included within the hypercube $[0, \log\left(\frac{1}{\epsilon}\right)]^m$.
From lemma A.10 eq. (25) and the fact that both PDE formulations (mentioned in the statement of lemma 5.1) have the same dynamics (they differ only in the terminal value function g and $g^{'}$) we have that:

$$\begin{aligned} V(y,t) &= \sup_{w \in \mathcal{A}(t)} \inf_{Q_{[t,1]} \in \mathcal{Q}(t)} g(y(1)) \\ V^{'}(y,t) &= \sup_{w \in \mathcal{A}(t)} \inf_{Q_{[t,1]} \in \mathcal{Q}(t)} g^{'}(y(1)). \end{aligned} \tag{33}$$

Now, since as the $g = g^{'}$ whenever $y(1) \in [0, \log\left(\frac{1}{\epsilon}\right)]^m$ and we have argued that the set of reachable terminal states $y(1)$ are in $[0, \log\left(\frac{1}{\epsilon}\right)]^m$ whenever $0 \leq y_i \leq \log\left(\frac{1}{\epsilon}\right) t$ for any $t \in [0,1)$ and all $i \in [m]$, we conclude that $V(y,t) = V^{'}(y,t)$ from eq. (33). □

C.2. Proof of Lemma 5.2.

Proof. The fact that $g^{'}$ is bounded, continuous and has a compact support $C_{v/2}$ follows directly from the construction of $g^{'}$. Further if $y \in C$, then $\prod_C^{(2)}(y) = y$ and $d_\infty(y, C) = 0$. This implies that $g^{'}(y) = g(y)$ whenever $y \in C$. We now show the Lipschitzness.

We observe that $\prod_C^{(2)}(.)$ is a Lipschitz function with Lipschitz constant 1 which follows from the non-expansiveness of the euclidean projection onto a closed convex set. Further $g(.)$ was also a Lipschitz function with Lipschitz constant 1. Their composition is therefore also Lipschitz. The distance metric $d_\infty(., C)$ is Lipschitz in l_∞ metric and hence also Lipschitz in l_2 distance. Consequently so is $1 - \frac{2d_\infty(.,C)}{v}$. The product of two Lipschitz functions over a compact set is also Lipschitz over the compact set. This implies that over the set $C_{v/2}$ we have that $g\left(\prod_C^{(2)}(.)\right)\left(1 - \frac{2d_\infty(.,C)}{v}\right)$ is also Lipschitz. Clearly the zero function is Lipschitz over the set $\mathbb{R}^m \setminus C_{v/2}$. And clearly $g^{'}$ is continuous over the entire euclidean space $\mathbb{R}^m$. We can then apply Lemma B.3 to conclude that $g^{'}$ is Lipschitz.

Consider any $y \notin [-3v/2, 3v/2]^m$. Using this as the starting point we observe that all reachable positions from this point is the set $[y, y+v]^m$ (see similar argument in the proof of lemma 5.1). But we have that :

$$[y, y+v]^m \cap C_{v/2} = \phi.$$

If not, there exists $z \in C_{v/2}$ which implies:

$$-v/2 \leq z_i \leq 3v/2$$

for each $i \in [m]$ but also $z \in [y, y+v]^m$ that is:

$$y_i \leq z_i \leq y_i + v$$

for each $i \in [m]$. But as $y \notin [-3v/2, 3v/2]^m$, there must exist some $j \in [m]$ such that $y_j < -3v/2$ or $y_j > 3v/2$. If $y_j < -3v/2$, then from above this implies $z_j < v/2$ contradicting that $v/2 \leq z_i$ for all $i \in [m]$. If $y_j > 3v/2$, then $3v/2 < z_j$ and this contradicts the fact that $z_i \leq 3v/2$ for all $i \in [m]$. Thus we have that $[y, y+v]^m \cap C_{v/2} = \phi$. Using the representation eq (25) for $V^{'}(y,t)$, the fact that the reachable set of $y \notin [-3v/2, 3v/2]^m$ doesn't intersect with the support $C_{v/2}$ of $g^{'}$ we conclude that $V^{'}(y,t) = 0$ for such y for all $t \in [0,1]$. This concludes the proof of the lemma. □

Remark C.1. The projection in $g^{'}$ is taken wrt to l_2 norm so that there exists a unique projection point. The distance was taken as d_∞ because the sets $C, C_{v/2}$ were axis aligned hyper-rectangles and the d_∞ distance is easy to compute. For the same reason, it is also true that $\prod_C^{(2)}(y) \subseteq \prod_C^{(\infty)}(y)$ for every y, that is the l_2 projection point is also a projection point under the l_∞ norm.

The indices are in the range $\lceil \frac{-3v}{2\Delta h} \rceil \leq i_j \leq \lfloor \frac{3v}{2\Delta h} \rfloor$. For any evaluation of the approximate value function at a spatial point outside this range we set the evaluation to be zero (this is justified from the results of lemma 5.2). The eq. (10) gives us a way to compute the approximate value function in the next time step from the evaluations of the current time step. We set the initial condition as follows:

$$V_i^0 = g^{'}((i_1\Delta h, i_2\Delta h, \ldots, i_m\Delta h)). \tag{34}$$

We now show the relationship between Δh and Δt that must be satisfied for the monotonicity to hold for FD scheme (10):

Lemma C.2. *The explicit upwind scheme proposed in eq.* (10) *is a monotone scheme, provided Δh and Δt are chosen to satisfy the following stability criteria:*

$$1 - \frac{m\Delta t}{\Delta h} \log\left(\frac{1}{\epsilon}\right) \geq 0. \tag{35}$$

Proof. For every $j \in [m]$, define for each $j \in [m]$:

$$i_j = (i_1, \ldots, i_{j+1}, \ldots, i_m).$$

Further, denote the RHS of eq. (10) as:

$$G(V_i^n, V_{i_1}^n, \ldots, V_{i_m}^n) = V_i^n - \Delta t \tilde{H}\left(\left(\frac{\Delta_j V_i^n}{\Delta h}\right)_{j\in[m]}\right). \tag{36}$$

Note that terms like $V_{i_j}^n$ are implicitly present in the difference term $\Delta_j V_i^n$. To show the upwind scheme is monotone, we need to show that $G(.)$ is non-decreasing in each of the aforementioned terms.

First we deal with terms like $V_{i_j}^n$. From monotonicity of hamiltonian proved in lemma 4.2 and the fact that we are using the hamiltonian of the reversed time PDE 29 implies that $\tilde{H}$ is co-ordinate wise non-increasing. From this it immediately follows that the $G(.)$ is non-decreasing in each $V_{i_j}^n$ for each $j \in [m]$.

Next, we prove the non-decreasing behavior of $G(.)$ in V_i^n. Suppose, we perturb V_i^n to $V_i^n + \delta$, with $\delta > 0$. Then we have that:

$$\begin{aligned}
G(V_i^n + \delta, V_{i_1}^n, \ldots, V_{i_m}^n) - G(V_i^n, V_{i_1}^n, \ldots, V_{i_m}^n) &= \delta - \Delta t\Bigg\{\tilde{H}\left(\left(\frac{\Delta_j V_i^n}{\Delta h} - \frac{\delta}{\Delta h}\right)_{j\in[m]}\right) \\
&\quad - \tilde{H}\left(\left(\frac{\Delta_j V_i^n}{\Delta h}\right)_{j\in[m]}\right)\Bigg\} \\
&\geq \delta - \Delta t\left|\left|\left(\frac{\delta}{\Delta h}\right)_{j\in[m]} \sqrt{m}\log\left(\frac{1}{\epsilon}\right)\right|\right| \\
&= \delta\left(1 - \frac{m\Delta t}{\Delta h}\log\left(\frac{1}{\epsilon}\right)\right) \\
&\geq 0,
\end{aligned}$$

where the first inequality follows from the Lipschitzness of the hamiltonian proven in lemma 4.2. The second inequality comes from the stability conditions eq. (35) imposed in the hypothesis of the lemma. This shows that $G(.)$ is non-decreasing in V_i^n as well. This shows the requisite monotonicity of the upwind scheme (10). □