Differential Dynamic Programming for Backward Reachability Analysis.

Author Names Omitted for Anonymous Review. Paper-ID [add your ID here]

I. Related Work

a) Dynamic Programming and Two-Person Games.: The formal relationships between the dynamic programming (DP) optimality condition for the *value* in differential two-person zero-sum games, and the solutions to PDEs that solve "min-max" or "max-min" type nonlinearity (the Isaacs' equation) was presented in Isaacs [16]. Essentially, Isaacs' claim was that if the *value* functions are smooth enough, then they solve certain first-order partial differential equations (PDE) problems with "max-min" or "min-max"-type nonlinearity. However, the DP value functions are seldom regular enough to admit a solution in the classical sense. "Weaker" solutions on the other hand Lions [21], Evans and Souganidis [10], Crandall et al. [6], Crandall and Majda [9], Evans and Souganidis [11] provide generalized "viscosity" solutions to HJ PDEs under relaxed regularity conditions; these viscosity solutions are not necessarily differentiable anywhere in the state space, and the only regularity prerequisite in the definition is continuity Crandall and Lions [8]. However, wherever they are differentiable, they satisfy the upper and lower values of HJ PDEs in a classical sense. Thus, they lend themselves well to many real-world problems existing at the interface of discrete, continuous, and hybrid systems Lygeros [22], Osher and Sethian [31], Mitchell [26], Evans and Souganidis [11], Mitchell et al. [27]. Viscosity Solutions to *Cauchy-type* HJ Equations admit usefulness in backward reachability analysis Mitchell et al. [27]. In scope and focus, this is the bulwark upon which we build our formulation in this paper.

b) Reachability for Systems Verification.: Reachability analysis is one of many verification methods that allows us to reason about (cpntrol-affine) dynamical systems. The verification problem may consist in finding a *set of reachable states* that lie along the trajectory of the solution to a first order nonlinear partial differential equation that originates from some initial state $\boldsymbol{x}_0 = \boldsymbol{x}(0)$ up to a specified time bound, $t = t_f$. *From a set of initial and unsafe state sets, and a time bound, the time-bounded safety verification problem is to determine if there is an initial state and a time within the bound that the solution to the PDE enters the unsafe set*. Reachability could be analyzed in a (i) *forward* sense, whereupon system trajectories are examined to determine if they enter certain states from an *initial set*; (ii) *backward* sense, whereupon system trajectories are examined to determine if they enter certain *target sets*; (iii) *reach set* sense, in which they are examined to see if states reach a set at a *particular time*; or (iv) *reach tube* sense, in which they are evaluated that they reach a set at a point *during a time interval*.

Backward reachable sets (BRS) or tubes (BRTs) are popularly analyzed as a game of two vehicles with non-stochastic dynamics Merz [25]. Such BRTs possess discontinuity at cross-over points (which exist at edges) on the surface of the tube, and may be non-convex. Therefore, treating the end-point constraints under these discontinuity characterizations need careful consideration and analysis when switching control laws if the underlying PDE does not have continuous partial derivatives TO-DO: (we discuss this further in § II).

c) Global Mesh-based Methods and Up-Scaling Reachability Analysis: Consider a reachability problem defined in a space of dimension $D = 12$ based on the non-incremental time-space discretization of each space coordinate. For $N = 100$ nodes, the total nodes required is 10^{120} on the volumetric grid[1]. The curse of dimensionality Bellman [5] greatly incapacitates current uniform grid discretization methods for guaranteeing the robustness of backward reachable sets (BRS) and tubes (BRTs) Mitchell et al. [27] of complex systems. Recent works have started exploring scaling up the Cauchy-type HJ problem for guaranteeing safety of higher-dimensional physical systems: the authors of Bajcsy et al. [2] provide local updates to BRS in unknown static environments with obstacles that may be unknown *a priori* to the agent; using standard meshing techniques for time-space uniform discretization over the entire physical space, and only updating points traversed locally, a safe navigation problem was solved in an environment assumed to be static. This makes it non-amenable to *a priori* unknown *dynamic* environments where the optimal value to the min-max HJ problem may need to be adaptively updated based on changing dynamics.

In Herbert et al. [15], the grid was naively refined along the temporal dimension, leveraging local decomposition schemes together with warm-starting optimization of the value function from previous solutions in order to accelerate learning for safety under the assumption that the system is either completely decoupled, or coupled over so-called "self-contained subsystems". While the empirical results of Bansal and Tomlin [3] demonstrate the feasibility of optimizing for the optimal value function in backward reachability analysis for up to ten dimensions for a system of Dubins vehicles, there are no guarantees that are provided. An analysis exists for a 12 dimensional systems Kaynama et al. [19] with up to a billion data points in the state space, that generates ro-

[1] Whereas, there are only 10^{97} baryons in the observable universe (excluding dark matter)!

bustly optimal trajectories. However, this is restricted to linear systems. Other associated techniques scale reachability with function approximators Fisac et al. [12, 13] in a reinforcement learning framework; again these methods lose the hard safety guarantees owing to the approximation in value function space.

A. Trajectory Optimization in a Reachable Differential Game Setting

Consider two agents interacting in an environment, $\mathcal{E}$, over a finite horizon, $[T, 0]$. The states evolve according to the following continuous-time dynamics

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= f(t, \boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{v}(t)) \quad T \leq t \leq 0 \\ \boldsymbol{x}(T) &= \boldsymbol{x} \end{aligned} \tag{1}$$

where $\boldsymbol{x}$ is the state that evolves from some initial negative time T to final time 0, and $\boldsymbol{u}(\cdot)$ and $\boldsymbol{v}(\cdot)$ are respectively the control and disturbance signals. Here $f(t, \cdot, \cdot, \cdot)$ and $\boldsymbol{x}(\cdot)$ are assumed to be bounded and Lipschitz continuous. This bounded Lipschitz continuity property assures uniqueness of the system response $\boldsymbol{x}(\cdot)$ to controls $\boldsymbol{u}(\cdot)$ and $\boldsymbol{v}(\cdot)$ Evans and Souganidis [11].

For a state $\boldsymbol{x} \in \Omega$ and a fixed time t: $T \leq t < 0$, suppose that the set of all controls for players $\boldsymbol{P}$ and $\boldsymbol{E}$ are respectively

$$\bar{\mathcal{U}} \equiv \{\boldsymbol{u} : [t, 0] \to \mathcal{U} | \boldsymbol{u} \text{ measurable}, \mathcal{U} \in \mathbb{R}^m\}, \tag{2}$$

$$\bar{\mathcal{V}} \equiv \{\boldsymbol{v} : [t, 0] \to \mathcal{V} | \boldsymbol{v} \text{ measurable}, \mathcal{V} \subset \mathbb{R}^p\}. \tag{3}$$

For any admissible control-disturbance pair $(\boldsymbol{u}(\cdot), \boldsymbol{v}(\cdot))$ and initial phase $(\boldsymbol{x}, T)$, there exists a unique Crandall and Lions [8], Evans and Souganidis [10] claim is that

$$\boldsymbol{\xi}(t) = \boldsymbol{\xi}(t; T, \boldsymbol{x}, \boldsymbol{u}(\cdot), \boldsymbol{v}(\cdot)) \tag{4}$$

that satisfies (2) a.e. with the property that

$$\boldsymbol{\xi}(T) = \boldsymbol{\xi}(T; T, \boldsymbol{x}, \boldsymbol{u}(\cdot), \boldsymbol{v}(\cdot)) = \boldsymbol{x}. \tag{5}$$

Read (4): the motion of (2) passing through phase $(\boldsymbol{x}, T)$ under the action of control $\boldsymbol{u}$, and disturbance $\boldsymbol{v}$, and observed at a time t afterwards.

For any optimal control problem, a value function is constructed based on the optimal cost (or payoff) of any input phase $(\boldsymbol{x}, T)$. In reachability analysis, typically this is defined using a terminal cost function $g(\cdot) : \mathbb{R}^n \to \mathbb{R}$ that satisfies

$$|g(\boldsymbol{x})| \leq k \tag{6a}$$

$$|g(\boldsymbol{x}) - g(\hat{\boldsymbol{x}})\ | \leq k|\boldsymbol{x} - \hat{\boldsymbol{x}}\ | \tag{6b}$$

for constant k and all $T \leq t \leq 0$, $\hat{\boldsymbol{x}}$, $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{u} \in \mathcal{U}$ and $\boldsymbol{v} \in \mathcal{V}$. The zero sublevel set of $g(\boldsymbol{x})$ i.e.

$$\mathcal{L}_0 = \{\boldsymbol{x} \in \bar{\Omega} \,|\, g(\boldsymbol{x}) \leq 0\}, \tag{7}$$

is the *target set* in the phase space $\Omega \times \mathbb{R}$ (proof in Mitchell et al. [27]). This target set can represent the failure set (to avoid) or a goal set (to reach) in the state space. Note that the target set, $\mathcal{L}_0$, is a closed subset of $\mathbb{R}^n$ and is in the closure of Ω. Typically $\mathcal{L}_0$ is user-defined, and $g(x)$ is a signed distance function, that is negative inside the target set and positive elsewhere.

Reachability analysis seeks to capture all initial conditions from which trajectories of the system may enter the target set. This could be desirable (in the case where the target set is a goal) or undesirable (where the target set represents the failure set). Frequently in reachability analysis one seeks to measure the *minimum cost* over time of trajectories of the system:

$$\min_{t} g(\boldsymbol{\xi}(t; T, \boldsymbol{x}_0, \boldsymbol{u}(\cdot), \boldsymbol{v}(\cdot))). \tag{8}$$

If this minimum cost is negative, then the trajectory entered the target set *at some time* $t \in [T, 0]$ over the time horizon. If the minimum cost is positive, then the trajectory will never enter the target set in the time horizon.

B. Hamilton-Jacobi Reachability: Construction of the Value Function

Rather than computing the minimum cost for every possible trajectory of the system, in safety analysis it is sufficient to consider the minimum cost under optimal behavior from both players. The optimal behavior of each player depends on whether the target set represents a goal or a failure set. For a safety (avoiding a failure set) problem setup, the evader $\boldsymbol{E}$ is seeking to maximize the minimum cost (keeping the system out of the target set) and the pursuer $\boldsymbol{P}$ seeks to minimize it. Suppose that the pursuer's mapping strategy (starting at t) is $\beta : \bar{\mathcal{U}}(t) \to \bar{\mathcal{V}}(t)$ provided for each $t \leq \tau \leq T$ and $\boldsymbol{u}, \hat{\boldsymbol{u}} \in \bar{\mathcal{U}}(t)$; then $\boldsymbol{u}(\bar{t}) = \hat{\boldsymbol{u}}(\bar{t})$ a.e. on $t \leq \bar{t} \leq \tau$ implies $\beta[\boldsymbol{u}](\bar{t}) = \beta[\hat{\boldsymbol{u}}](\bar{t})$ a.e. on $t \leq \bar{t} \leq \tau$. The differential game's (lower) value for a solution $\boldsymbol{x}(t)$ that solves (2) for $\boldsymbol{u}(t)$ and $\boldsymbol{v}(t) = \beta[\boldsymbol{u}](\cdot)$ is

$$V(\boldsymbol{x}, t) = \inf_{\beta \in \mathcal{B}(t)} \sup_{\boldsymbol{u} \in \mathcal{U}(t)} \min_{t \in [T, 0]} g\left(\boldsymbol{x}(T)\right). \tag{9}$$

For a goal-satisfaction (liveness) problem setup, the behavior of the evader and pursuer are reversed.

Optimal trajectories emanating from initial phases $(\boldsymbol{x}, T)$ where the value function is non-negative will maintain non-negative cost over the entire time horizon, and therefore will never enter the target set. Optimal trajectories from initial phases where the value function is negative will enter the target set at some point within the time horizon. For the safety problem setup in (9) we can define the corresponding *robustly controlled backward reachable tube* for $\tau \in [T, 0]$[2] as the closure of the open set

$$\begin{aligned} \mathcal{L}([\tau, 0], \mathcal{L}_0) = \{\boldsymbol{x} \in \Omega \,|\, \exists\, \beta \in \bar{\mathcal{V}}(t) \,\forall\, \boldsymbol{u} \in \mathcal{U}(t), \exists\, \bar{t} \in [T, 0], \\ \boldsymbol{\xi}(\bar{t}) \in \mathcal{L}_0\}, \; \bar{t} \in [T, 0]\,. \end{aligned} \tag{10}$$

Read: the set of states from which the strategies of $\boldsymbol{P}$, and for all controls of $\boldsymbol{E}$ imply that we reach the target set within the interval $[T, 0]$. More specifically, following Lemma 2 of Mitchell et al. [27], the states in the reachable set admit the following properties w.r.t the value function $\boldsymbol{V}$

$$\boldsymbol{x} \in \mathcal{L} \implies \boldsymbol{V}^-(\boldsymbol{x}, t) \leq 0 \tag{11a}$$

$$\boldsymbol{V}^-(\boldsymbol{x}, t) \leq 0 \implies \boldsymbol{x} \in \mathcal{L}. \tag{11b}$$

[2]The (backward) horizon T is negative.

The goal of $\boldsymbol{P}$ is to drive the system's trajectories into the unsafe set i.e., $\boldsymbol{P}$ has $\boldsymbol{u}$ at will and aims to minimize the termination time of the game (c.f. (7)); and $\boldsymbol{E}$ seeks to avoid the unsafe set i.e., $\boldsymbol{E}$ has controls $\boldsymbol{v}$ at will and seeks to maximize the termination time of the game (c.f. (7)). For goal-satisfaction (or *liveness*) problem setups, the strategies are flipped and the backward reachable tube instead marks the states from which the evader $\boldsymbol{E}$ can successfully reach the target set despite worst-case efforts of the pursuer $\boldsymbol{P}$.

Computing the value function is in general challenging and non-convex. Additionally, the value function is hardly smooth throughout the state space, so it lacks classical solutions even for smooth Hamiltonian and boundary conditions. However, the value function is a "viscosity" (generalized) solution Lions [21], Crandall and Lions [8] of the associated HJ-Isaacs (HJI) PDE, i.e. solutions which are *locally Lipschitz* in $\Omega \times [0, T]$, and with at most first-order partial derivatives in the Hamiltonian. The HJI PDE is as follows:

$$\frac{\partial V}{\partial t}(\boldsymbol{x}, t) + \min\{0, \boldsymbol{H}^-(t; \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{V}_{\boldsymbol{x}}^-)\} = 0 \tag{12}$$

$$V(\boldsymbol{x}, 0) = g(\boldsymbol{x}) \tag{13}$$

where the vector field $V_{\boldsymbol{x}}$ is known in terms of the game's terminal conditions so that the overall game is akin to a two-point boundary-value problem. For more details on the construction of this PDE, see Mitchell et al. [27].

Instead of using state space discretization methods and employing Lax-Friedrichs schemes to resolve the value function over a global mesh, we resolve to classical successive sweep optimal control algorithms Mitter [28], McReynolds [24], in particular a differential dynamic programming (DDP) variant Mayne [23], Jacobson [17], Jacobson and Mayne [18], which are iterative algorithms for obtaining solutions to optimal control problems. Similar to the monotone characteristics of Crandall and Majda [9, 7], these methods (under appropriate conditions Ogunmolu et al. [29]) assure a monotone solution on the state space without resorting to finite differencing schemes. TO-DO: LM: Add further spice here.

Once computed, the value function provides a safety certificate (defined by its zero level set) and corresponding safety controller (defined by the spatial gradients along the zero level set). SH: don't know if we should go into the online control portion of this or not

In addition to producing a value function, the dynamic programming process can be used to generate and minimum time-to-reach (TTR) function. This is a function that maps initial conditions to the minimum time horizon required to reach the target set. This can be computed by "stacking" the zero level sets of the value function as it propagates backwards in time Mitchell [26], Basar and Olsder [4], Athans and Falb [1]. SH: cite Ian Mitchell, maybe Insoon; LM: Is this the right citation you wanted, Sylvia?

II. Differential Approximation of the Terminal Cost

We now introduce the quadratic approximation scheme of the value function. We seek a pair of *saddle point equilibrium* policies, $(\boldsymbol{u}^*, \boldsymbol{v}^*)$ that satisfy the following inequalities for a cost $\boldsymbol{V}$ at an initial time T:

$$\begin{aligned}\boldsymbol{V}(T; \boldsymbol{x}, \boldsymbol{u}^*, \boldsymbol{v}) \le \boldsymbol{V}(T; \boldsymbol{x}, \boldsymbol{u}^*, \boldsymbol{v}^*) \le \boldsymbol{V}(T; \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}^*),\\ \forall\, \boldsymbol{u} \in \mathcal{U}, \boldsymbol{v} \in \mathcal{V} \text{ and } \boldsymbol{x}(T).\end{aligned} \tag{14}$$

A successive approximation to $\boldsymbol{V}(t; \cdot, \cdot, \cdot)$ consists in maintaining local approximations to the global system dynamics at every iteration. First, we apply local controls $\boldsymbol{u}_r(t)$ and $\boldsymbol{v}_r(t)$ on (2) so that the nominal value is $\boldsymbol{V}(t; \cdot, \cdot, \cdot)$ for a resulting nominal state $\boldsymbol{x}_r(\tau)$; $\tau \in [-T, 0]$. The local system dynamics becomes

$$\dot{\boldsymbol{x}}_r(\tau) = f(t; \boldsymbol{x}_r(\tau), \boldsymbol{u}_r(\tau), \boldsymbol{v}_r(\tau));\;\; \boldsymbol{x}_r(0) = \boldsymbol{x}_{r_0}. \tag{15}$$

Our dynamics now describe variations from the nonlinear system c.f. (2) with state and control pairs $\delta\boldsymbol{x}(t)$, $\delta\boldsymbol{u}(t)$, $\delta\boldsymbol{v}(t)$ respectively[3]. Therefore, we write

$$\boldsymbol{x}(t) = \boldsymbol{x}_r(t) + \delta\boldsymbol{x}(t),\;\; \boldsymbol{u}(t) = \boldsymbol{u}_r(t) + \delta\boldsymbol{u}(t), \tag{16a}$$

$$\boldsymbol{v}(t) = \boldsymbol{v}_r(t) + \delta\boldsymbol{v}(t),\;\; t \in [-T, 0]\,. \tag{16b}$$

Abusing notation, we drop the templated time arguments in (16) so that the canonical problem is now

$$\frac{d}{dt}(\boldsymbol{x}_r + \delta\boldsymbol{x}) = f(t; \boldsymbol{x}_r + \delta\boldsymbol{x}, \boldsymbol{u}_r + \delta\boldsymbol{u}, \boldsymbol{v}_r + \delta\boldsymbol{v}), \tag{17}$$

$$\boldsymbol{x}_r(0) + \delta\boldsymbol{x}(0) = \boldsymbol{x}(0), \tag{18}$$

with the associated terminal value

$$\begin{aligned}-\frac{\partial \boldsymbol{V}}{\partial t}(\boldsymbol{x}_r + \delta\boldsymbol{x}, t) = \min\Big\{0, \max_{\delta\boldsymbol{u}\in\mathcal{U}} \min_{\delta\boldsymbol{v}\in\mathcal{V}} \langle f(t; \boldsymbol{x}_r + \delta\boldsymbol{x},\\ \boldsymbol{u}_r + \delta\boldsymbol{u}, \boldsymbol{v}_r + \delta\boldsymbol{v}), \frac{\partial \boldsymbol{V}}{\partial \boldsymbol{x}}(\boldsymbol{x}_r + \delta\boldsymbol{x}, t)\Big\rangle\Big\},\\ \boldsymbol{V}(\boldsymbol{x}_r, 0) = g(0; \boldsymbol{x}_r(0) + \delta\boldsymbol{x}(0));\end{aligned} \tag{19}$$

and state trajectory

$$\boldsymbol{\xi}(t) = \boldsymbol{\xi}(t; t_0, \boldsymbol{x}_r + \delta\boldsymbol{x}, \boldsymbol{u} + \delta\boldsymbol{u}, \boldsymbol{v} + \delta\boldsymbol{v}). \tag{20}$$

For $-T \le t \le 0$ and a $\tau \in [t, 0]$, let the optimal cost for using the optimal control $\boldsymbol{u}^\star(\tau) = \boldsymbol{u}_r(\tau) + \delta\boldsymbol{u}^\star(\tau)$ be $\boldsymbol{V}^\star(\boldsymbol{x}_r, \tau)$; and the optimal cost for using $\boldsymbol{u}_r(\tau)$ be $\boldsymbol{V}_r(\boldsymbol{x}_r, t)$. Suppose further that the difference between these two costs on the phase $(\boldsymbol{x}_r, t)$ is $\tilde{\mathbf{V}}^\star$, then

$$\tilde{\mathbf{V}}^\star = \boldsymbol{V}^\star(\boldsymbol{x}_r, t) - \boldsymbol{V}_r(\boldsymbol{x}_r, t). \tag{21}$$

Theorem 1. *Suppose that $\boldsymbol{V}$ is smooth enough, the HJI variational inequality c.f. (13) admits the following approximated expansion in the state variation $\delta\boldsymbol{x}$ about the nominal*

[3]Note that $\delta\boldsymbol{x}(t)$, $\delta\boldsymbol{u}(t)$, and $\delta\boldsymbol{v}(t)$ are respectively measured with respect to $\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{v}(t)$ and are not necessarily small.

trajectory $\boldsymbol{x}_r$:

$$-\frac{\partial \boldsymbol{V}_r}{\partial t}-\frac{\partial \tilde{\mathbf{V}}}{\partial t}-\left\langle\frac{\partial \boldsymbol{V}_{\boldsymbol{x}}}{\partial t}, \delta \boldsymbol{x}\right\rangle-\frac{1}{2}\left\langle\delta \boldsymbol{x}, \frac{\partial \boldsymbol{V}_{\boldsymbol{x} \boldsymbol{x}}}{\partial t} \delta \boldsymbol{x}\right\rangle= \min \left\{\mathbf{0}, \max_{\delta \boldsymbol{u} \in \mathcal{U}} \min_{\delta \boldsymbol{v} \in \mathcal{V}}\left\langle f^T(t ; \boldsymbol{x}_r+\delta \boldsymbol{x}, \boldsymbol{u}_r+\delta \boldsymbol{u}, \boldsymbol{v}_r+\delta \boldsymbol{v}),\right.\right. \left.\left.\boldsymbol{V}_{\boldsymbol{x}}+\boldsymbol{V}_{\boldsymbol{x} \boldsymbol{x}} \delta \boldsymbol{x}\right\rangle\right\} . \tag{22}$$

Furthermore, this expansion is bounded by $O(\delta \boldsymbol{x}^3)$.

Proof: For the moment, let us focus on the l.h.s. of (19). Our derivations closely follow that of Jacobson [17]. The major difference is that our choice of $\boldsymbol{x}_r$ is guaranteed to be close to that of $\boldsymbol{x}$ so that we need not prescribe stringent conditions for when local control laws are valid on the nonlinear system. In addition, we are working within a game theory framework and are solving a terminal value function on the variational HJI problem. Suppose the optimal terminal cost, $\boldsymbol{V}^\star$, is sufficiently smooth to allow a power series expansion in the state variation $\delta\boldsymbol{x}$ about reduced state, $\boldsymbol{x}_r$, we find that

$$\boldsymbol{V}^\star(\boldsymbol{x}_r+\delta \boldsymbol{x}, t)=\boldsymbol{V}^\star(\boldsymbol{x}_r, t)+\langle\boldsymbol{V}_{\boldsymbol{x}}, \delta \boldsymbol{x}\rangle+\frac{1}{2}\langle\delta \boldsymbol{x}, \boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}^\star \delta \boldsymbol{x}\rangle + \text{h.o.t.} \tag{23}$$

Here, h.o.t. signifies higher order terms. This expansion scheme is consistent with Volterra-series model order reduction methods [14] or differential dynamic programming schemes that decompose nonlinear systems as a summation of Taylor series expansions [18]. Using (21), (23) becomes

$$\boldsymbol{V}^\star(\boldsymbol{x}_r+\delta \boldsymbol{x}, t)=\boldsymbol{V}_r(\boldsymbol{x}_r, t)+\tilde{\mathbf{V}}^\star+\langle\boldsymbol{V}_{\boldsymbol{x}}, \delta \boldsymbol{x}\rangle+ \frac{1}{2}\langle\delta \boldsymbol{x}, \boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}^\star \delta \boldsymbol{x}\rangle + \text{h.o.t.} \tag{24}$$

The expansion in (24) may be more costly than solving for the original value function owing to the large dimensionality of the states as higher order terms are expanded. However, consider:

- $\boldsymbol{V}_r(\boldsymbol{x}_r, t)$ already contains the dominant modes of $\boldsymbol{V}(\boldsymbol{x}, t)$ as a result of the singular value decomposition scheme; therefore w.l.o.g. states in the reduced order basis (ROB), $\boldsymbol{V}_r(\boldsymbol{x}_r, t)$, will be sufficiently close to those that originate in (2);
- If the above is true, the state variation $\delta\boldsymbol{x}$ will be sufficiently small owing to the fact that $\boldsymbol{x} \approx \boldsymbol{x}_r$ c.f. (16).

Therefore, we can avoid the infinite data storage requirement by truncating the expansion in (24) at, say, the quadratic (second-order) terms in $\delta\boldsymbol{x}$. Seeing that $\delta\boldsymbol{x}$ is sufficiently small, the second-order cost terms will dominate higher order terms, and this new cost will result in an $O(\delta\boldsymbol{x}^3)$ approximation error, affording us realizable control laws that can be executed on the system (2). From (23), we have

$$\boldsymbol{V}^\star(\boldsymbol{x}_r+\delta \boldsymbol{x}, t)=\boldsymbol{V}_r+\tilde{\mathbf{V}}^\star+\langle\boldsymbol{V}_{\boldsymbol{x}}, \delta \boldsymbol{x}\rangle+\frac{1}{2}\langle\delta \boldsymbol{x}, \boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}^\star \delta \boldsymbol{x}\rangle . \tag{25}$$

Denoting by $\boldsymbol{V}_{\boldsymbol{x}}^\star$ the co-state on the r.h.s of (19), we can similarly expand it up to second order terms as follows

$$\boldsymbol{V}_{\boldsymbol{x}}^\star(\boldsymbol{x}_r+\delta \boldsymbol{x}, t)=\frac{\partial \boldsymbol{V}_r^\star}{\partial \boldsymbol{x}}(\boldsymbol{x}_r, t)+\langle\boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}^\star(\boldsymbol{x}_r, t), \delta \boldsymbol{x}\rangle. \tag{26}$$

Note that *the co-state in* (26) *and parameters on the r.h.s. of* (25) *are evaluated on the reduced model, specifically at the phase* $(\boldsymbol{x}_r, t)$. Substituting (25) and (26) into (19), abusing notation by dropping the superscripts and the templated phase arguments, we find that

$$-\frac{\partial \boldsymbol{V}_r}{\partial t}-\frac{\partial \tilde{\mathbf{V}}}{\partial t}-\left\langle\frac{\partial \boldsymbol{V}_{\boldsymbol{x}}}{\partial t}, \delta \boldsymbol{x}\right\rangle-\frac{1}{2}\left\langle\delta \boldsymbol{x}, \frac{\partial \boldsymbol{V}_{\boldsymbol{x} \boldsymbol{x}}}{\partial t} \delta \boldsymbol{x}\right\rangle= \min \left\{0, \max_{\delta \boldsymbol{u}} \min_{\delta \boldsymbol{v}}\left\langle f^T(t ; \boldsymbol{x}_r+\delta \boldsymbol{x}, \boldsymbol{u}_r+\delta \boldsymbol{u}, \boldsymbol{v}_r+\delta \boldsymbol{v}),\right.\right. \left.\left.\boldsymbol{V}_{\boldsymbol{x}}+\boldsymbol{V}_{\boldsymbol{x} \boldsymbol{x}} \delta \boldsymbol{x}\right\rangle\right\} . \tag{27}$$

Observe that $\boldsymbol{V}_r+\tilde{\mathbf{V}}$, $\boldsymbol{V}_{\boldsymbol{x}}$, and $\boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}$ are all functions of the phase $(\boldsymbol{x}, r)$ so that

$$\frac{d}{dt}\left(\boldsymbol{V}_r+\tilde{\mathbf{V}}\right)=\frac{\partial}{\partial t}\left(\boldsymbol{V}_r+\tilde{\mathbf{V}}\right)+\langle f^T(t ; \boldsymbol{x}_r, \boldsymbol{u}_r, \boldsymbol{v}_r), \boldsymbol{V}_{\boldsymbol{x}}\rangle \tag{28a}$$

$$\dot{\boldsymbol{V}}_{\boldsymbol{x}}=\frac{\partial \boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}}{\partial t}+\langle f^T(t ; \boldsymbol{x}_r, \boldsymbol{u}_r, \boldsymbol{v}_r), \boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}\rangle \tag{28b}$$

$$\dot{\boldsymbol{V}}_{\boldsymbol{x}\boldsymbol{x}}=\frac{\partial \boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}}{\partial t}. \tag{28c}$$

■

Corollary 1. *If the solution to* (22) *converges to a local optimum, then the backward reachable set(tube) will converge to a local extrema. In addition, if we overapproximate the resulting numerical solution, the reachable set or tube will locally converge to an optimal region in the state space.*

Proof: This corollary is a statement of the local quadratic convergence guarantees of second order gradient methods in differential dynamic programming [20]. ■

The approximation scheme proceeds as follows:

- approximate the nonlinear system dynamics (c.f. (2)), starting with the pursuer and evader's local control schedules, $\{\boldsymbol{u}_r(t)\}_{t=T}^0$ and $\{\boldsymbol{v}_r(t)\}_{t=T}^0$, assumed to be available;
- run the system's passive dynamics with $\{\boldsymbol{u}_r(t)\}_{t=T}^0, \{\boldsymbol{v}_r(t)\}_{t=0}^T$ to generate nominal state trajectories $\{\boldsymbol{x}_r(t)\}_{t=T}^0$, with neighboring trajectories $\{\boldsymbol{x}(t)\}_{t=0}^T$;
- choose a small neighborhood, $\{\delta\boldsymbol{x}(t)\}_{t=T}^0$ of $\{\boldsymbol{x}(t)\}_{t=T}^0$, which provides an optimal reduction in cost as the dynamics no longer represent those of $\{\boldsymbol{x}(t)\}_{t=T}^0$;
- new state and control sequence pairs become $\delta\boldsymbol{x}(t) = \boldsymbol{x}(t)-\boldsymbol{x}_r(t)$, $\delta\boldsymbol{u}(t)=\boldsymbol{u}(t)-\boldsymbol{u}_r(t)$, $\delta\boldsymbol{v}(t)=\boldsymbol{v}(t)-\boldsymbol{v}_r(t)$.

The left hand side of (27) admits a quadratic form, so that we can regress a quadratic form to fit the functionals and derivatives of the optimal structure of the ROB. The r.h.s. can be similarly expanded as above. Define

$$\boldsymbol{H}(t ; \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{V}_{\boldsymbol{x}})=\langle\boldsymbol{V}_{\boldsymbol{x}}, f(t ; \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v})\rangle \tag{29}$$

so that (27) becomes

$$-\frac{\partial \boldsymbol{V}_r}{\partial t}-\frac{\partial \tilde{\mathbf{V}}}{\partial t}-\left\langle\frac{\partial \boldsymbol{V}_{\boldsymbol{x}}}{\partial t}, \delta \boldsymbol{x}\right\rangle-\frac{1}{2}\left\langle\delta \boldsymbol{x}, \frac{\partial \boldsymbol{V}_{\boldsymbol{x x}}}{\partial t} \delta \boldsymbol{x}\right\rangle= \min \left\{\mathbf{0}, \max_{\delta \boldsymbol{u}} \min_{\delta \boldsymbol{v}}\left[\boldsymbol{H}\left(t ; \boldsymbol{x}_r+\delta \boldsymbol{x}, \boldsymbol{u}_r+\delta \boldsymbol{u}, \boldsymbol{v}+\delta \boldsymbol{v}, \boldsymbol{V}_{\boldsymbol{x}}\right)+\left\langle\boldsymbol{V}_{\boldsymbol{x x}} \delta \boldsymbol{x}, f\left(t ; \boldsymbol{x}_r+\delta \boldsymbol{x}, \boldsymbol{u}_r+\delta \boldsymbol{u}, \boldsymbol{v}_r+\delta \boldsymbol{v}\right)\right\rangle\right]\right\} . \tag{30}$$

Expanding the r.h.s. about $\boldsymbol{x}_r, \boldsymbol{u}_r, \boldsymbol{v}_r$ up to second-order only[4], we find that

$$\begin{aligned}
&\min \left\{\mathbf{0}, \max_{\delta \boldsymbol{u}} \min_{\delta \boldsymbol{v}}\left[\boldsymbol{H}+\left\langle\boldsymbol{H}_{\boldsymbol{x}}+\boldsymbol{V}_{\boldsymbol{x x}} f, \delta \boldsymbol{x}\right\rangle+\right.\right. \\
&\left\langle\boldsymbol{H}_{\boldsymbol{u}}, \delta \boldsymbol{u}\right\rangle+\left\langle\boldsymbol{H}_{\boldsymbol{v}}, \delta \boldsymbol{v}\right\rangle+\left\langle\delta \boldsymbol{u},\left(\boldsymbol{H}_{\boldsymbol{u x}}+f_{\boldsymbol{u}}^T \boldsymbol{V}_{\boldsymbol{x x}}\right) \delta \boldsymbol{x}\right\rangle \\
&+\left\langle\delta \boldsymbol{v},\left(\boldsymbol{H}_{\boldsymbol{v x}}+f_{\boldsymbol{v}}^T \boldsymbol{V}_{\boldsymbol{x x}}\right) \delta \boldsymbol{x}\right\rangle+\frac{1}{2}\left\langle\delta \boldsymbol{u}, \boldsymbol{H}_{\boldsymbol{u v}} \delta \boldsymbol{v}\right\rangle \\
&+\frac{1}{2}\left\langle\delta \boldsymbol{v}, \boldsymbol{H}_{\boldsymbol{v u}} \delta \boldsymbol{u}\right\rangle+\frac{1}{2}\left\langle\delta \boldsymbol{u}, \boldsymbol{H}_{\boldsymbol{u u}} \delta \boldsymbol{u}\right\rangle+\frac{1}{2}\left\langle\delta \boldsymbol{v}, \boldsymbol{H}_{\boldsymbol{v v}} \delta \boldsymbol{v}\right\rangle \\
&\left.\left.+\frac{1}{2}\left\langle\delta \boldsymbol{x},\left(\boldsymbol{H}_{\boldsymbol{x x}}+f_{\boldsymbol{x}}^T \boldsymbol{V}_{\boldsymbol{x x}}+\boldsymbol{V}_{\boldsymbol{x x}} f_{\boldsymbol{x}}\right) \delta \boldsymbol{x}\right\rangle\right]\right\} .
\end{aligned} \tag{31}$$

Let us recall that when capture[5] occurs, we must have the Hamiltonian of the value function be zero as a necessary condition for the players' saddle-point controls [25, 16] i.e.

$$\boldsymbol{H}_{\boldsymbol{u}}\left(t ; \boldsymbol{x}_r, \boldsymbol{u}_r^{\star}, \boldsymbol{v}_r, \boldsymbol{V}_{\boldsymbol{x}}\right)=0 ; \ \boldsymbol{H}_{\boldsymbol{v}}\left(t ; \boldsymbol{x}_r, \boldsymbol{u}_r, \boldsymbol{v}_r^{\star}, \boldsymbol{V}_{\boldsymbol{x}}\right)=0 . \tag{32}$$

where $\boldsymbol{u}_r^{\star}$ and $\boldsymbol{v}_r^{\star}$ respectively represent the optimal control laws for both players at time t.

A state-control relationship of the following form is sought:

$$\delta \boldsymbol{u}=\boldsymbol{k}_{\boldsymbol{u}} \delta \boldsymbol{x}, \quad \delta \boldsymbol{v}=\boldsymbol{k}_{\boldsymbol{v}} \delta \boldsymbol{x} \tag{33}$$

so that (31) in the context of (32) yields

$$\boldsymbol{H}_{\boldsymbol{u}}+\boldsymbol{H}_{\boldsymbol{u u}} \delta \boldsymbol{u}+\left(\boldsymbol{H}_{\boldsymbol{u x}}+f_{\boldsymbol{u}}^T \boldsymbol{V}_{\boldsymbol{x x}}\right) \delta \boldsymbol{x}+\frac{1}{2} \boldsymbol{H}_{\boldsymbol{u v}} \delta \boldsymbol{v}=0 \tag{34a}$$

$$\boldsymbol{H}_{\boldsymbol{v}}+\boldsymbol{H}_{\boldsymbol{v v}} \delta \boldsymbol{v}+\left(\boldsymbol{H}_{\boldsymbol{v x}}+f_{\boldsymbol{v}}^T \boldsymbol{V}_{\boldsymbol{x x}}\right) \delta \boldsymbol{x}+\frac{1}{2} \boldsymbol{H}_{\boldsymbol{v u}} \delta \boldsymbol{u}=0 . \tag{34b}$$

Using (32) and equating like terms in the resulting equation to those in (33), we have the following for the state gains:

$$\begin{aligned}
\boldsymbol{k}_{\boldsymbol{u}} &=-\frac{1}{2} \boldsymbol{H}_{\boldsymbol{u u}}^{-1}\left[\boldsymbol{H}_{\boldsymbol{u v}} \boldsymbol{k}_{\boldsymbol{v}}+2\left(\boldsymbol{H}_{\boldsymbol{u x}}+f_{\boldsymbol{u}}^T \boldsymbol{V}_{\boldsymbol{x x}}\right)\right], \text { and } \\
\boldsymbol{k}_{\boldsymbol{v}} &=-\frac{1}{2} \boldsymbol{H}_{\boldsymbol{v v}}^{-1}\left[\boldsymbol{H}_{\boldsymbol{v u}} \boldsymbol{k}_{\boldsymbol{u}}+2\left(\boldsymbol{H}_{\boldsymbol{v x}}+f_{\boldsymbol{v}}^T \boldsymbol{V}_{\boldsymbol{x x}}\right)\right] .
\end{aligned} \tag{35}$$

[4] This is because the l.h.s. was truncated at second order expansion previously. Ultimately, the $\delta \boldsymbol{u}, \delta \boldsymbol{v}$ terms will be quadratic in $\delta \boldsymbol{x}$ if we neglect h.o.t.

[5] A capture occurs when $\boldsymbol{E}$'s separation from $\boldsymbol{P}$ becomes less than a specified e.g. capture radius.

Putting the maximizing $\delta \boldsymbol{u}$ and the minimizing $\delta \boldsymbol{v}$ into (31), whilst neglecting terms in $\delta \boldsymbol{x}$ beyond second-order, we have

$$\begin{aligned}
\min \{\mathbf{0},[\boldsymbol{H} & +\left\langle\left(\boldsymbol{H}_{\boldsymbol{x}}+\boldsymbol{V}_{\boldsymbol{x x}} f+\boldsymbol{k}_{\boldsymbol{u}}^T \boldsymbol{H}_{\boldsymbol{u}}+\boldsymbol{k}_{\boldsymbol{v}}^T \boldsymbol{H}_{\boldsymbol{v}}\right), \delta \boldsymbol{x}\right\rangle \\
& +\frac{1}{2}\left\langle\delta \boldsymbol{x},\left(\boldsymbol{H}_{\boldsymbol{x x}}+f_{\boldsymbol{x}}^T \boldsymbol{V}_{\boldsymbol{x x}}+\boldsymbol{V}_{\boldsymbol{x x}} f_{\boldsymbol{x}}+\boldsymbol{k}_{\boldsymbol{u}}^T \boldsymbol{H}_{\boldsymbol{u u}} \boldsymbol{k}_{\boldsymbol{u}}\right.\right. \\
& \left.\left.\left.+\boldsymbol{k}_{\boldsymbol{v}}^T \boldsymbol{H}_{\boldsymbol{v v}} \boldsymbol{k}_{\boldsymbol{v}}\right) \delta \boldsymbol{x}\right\rangle\right]\} .
\end{aligned} \tag{36}$$

Now, we can compare coefficients with the l.h.s. of (30) and find the quadratic expansion of the reduced value function admits the following analytical solution on its right hand side:

$$-\frac{\partial \boldsymbol{V}_r}{\partial t}-\frac{\partial \tilde{\boldsymbol{V}}}{\partial t}=\min \{\mathbf{0}, \boldsymbol{H}\} \tag{37a}$$

$$-\frac{\partial \boldsymbol{V}_{\boldsymbol{x}}}{\partial t}=\min \left\{\mathbf{0}, \boldsymbol{H}_{\boldsymbol{x}}+\boldsymbol{V}_{\boldsymbol{x x}} f+\boldsymbol{k}_{\boldsymbol{u}}^T \boldsymbol{H}_{\boldsymbol{u}}+\boldsymbol{k}_{\boldsymbol{v}}^T \boldsymbol{H}_{\boldsymbol{v}}\right\} \tag{37b}$$

$$\begin{aligned}
-\frac{\partial \boldsymbol{V}_{\boldsymbol{x x}}}{\partial t}=\min \{\mathbf{0}, & \boldsymbol{H}_{\boldsymbol{x x}}+f_{\boldsymbol{x}}^T \boldsymbol{V}_{\boldsymbol{x x}}+\boldsymbol{V}_{\boldsymbol{x x}} f_{\boldsymbol{x}} \\
& \left.+\boldsymbol{k}_{\boldsymbol{u}}^T \boldsymbol{H}_{\boldsymbol{u u}} \boldsymbol{k}_{\boldsymbol{u}}+\boldsymbol{k}_{\boldsymbol{v}}^T \boldsymbol{H}_{\boldsymbol{v v}} \boldsymbol{k}_{\boldsymbol{v}}\right\} .
\end{aligned} \tag{37c}$$

Furthermore, comparing the above with (28) and noting that $-\dot{\boldsymbol{V}}_r=0$[6], we find that

$$-\dot{\tilde{\boldsymbol{V}}}=-\frac{\partial \tilde{\boldsymbol{V}}}{\partial t} \triangleq \min \left\{\mathbf{0}, \boldsymbol{H}-\boldsymbol{H}\left(t ; \boldsymbol{x}_r, \boldsymbol{u}_r, \boldsymbol{v}_r, \boldsymbol{V}_{\boldsymbol{x}}\right)\right\} \tag{38a}$$

$$-\dot{\boldsymbol{V}}_{\boldsymbol{x}}=\min \left\{\mathbf{0}, \boldsymbol{H}_{\boldsymbol{x}}+\boldsymbol{V}_{\boldsymbol{x x}}\left(f-f\left(t ; \boldsymbol{x}_r, \boldsymbol{u}_r, \boldsymbol{v}_r\right)\right)\right. \tag{38b}$$

$$\left.+\boldsymbol{k}_{\boldsymbol{u}}^T \boldsymbol{H}_{\boldsymbol{u}}+\boldsymbol{k}_{\boldsymbol{v}}^T \boldsymbol{H}_{\boldsymbol{v}}\right\} \tag{38c}$$

$$\begin{aligned}
-\frac{\partial \boldsymbol{V}_{\boldsymbol{x x}}}{\partial t}=\min \{\mathbf{0}, & \boldsymbol{H}_{\boldsymbol{x x}}+f_{\boldsymbol{x}}^T \boldsymbol{V}_{\boldsymbol{x x}}+\boldsymbol{V}_{\boldsymbol{x x}} f_{\boldsymbol{x}} \\
& \left.+\boldsymbol{k}_{\boldsymbol{u}}^T \boldsymbol{H}_{\boldsymbol{u u}} \boldsymbol{k}_{\boldsymbol{u}}+\boldsymbol{k}_{\boldsymbol{v}}^T \boldsymbol{H}_{\boldsymbol{v v}} \boldsymbol{k}_{\boldsymbol{v}}\right\}
\end{aligned} \tag{38d}$$

where $\boldsymbol{k}_{\boldsymbol{u}}$ and $\boldsymbol{k}_{\boldsymbol{v}}$ are as defined in (33). Note that at a saddle point, the first-order necessary condition for optimality c.f. (32) implies

$$-\dot{\tilde{\boldsymbol{V}}}=\min \left\{0, \boldsymbol{H}-\boldsymbol{H}\left(t ; \boldsymbol{x}_r, \boldsymbol{u}_r, \boldsymbol{v}_r, \boldsymbol{V}_{\boldsymbol{x}}\right)\right\} \tag{39a}$$

$$-\dot{\boldsymbol{V}}_{\boldsymbol{x}}=\min \left\{0, \boldsymbol{H}_{\boldsymbol{x}}+\boldsymbol{V}_{\boldsymbol{x x}}\left(f-f\left(\boldsymbol{x}_r, \boldsymbol{u}_r, \boldsymbol{v}_r\right)\right)\right\} \tag{39b}$$

$$\begin{aligned}
-\frac{\partial \boldsymbol{V}_{\boldsymbol{x x}}}{\partial t}=\min \{\mathbf{0}, & \boldsymbol{H}_{\boldsymbol{x x}}+f_{\boldsymbol{x}}^T \boldsymbol{V}_{\boldsymbol{x x}}+\boldsymbol{V}_{\boldsymbol{x x}} f_{\boldsymbol{x}} \\
& \left.+\boldsymbol{k}_{\boldsymbol{u}}^T \boldsymbol{H}_{\boldsymbol{u u}} \boldsymbol{k}_{\boldsymbol{u}}+\boldsymbol{k}_{\boldsymbol{v}}^T \boldsymbol{H}_{\boldsymbol{v v}} \boldsymbol{k}_{\boldsymbol{v}}\right\}
\end{aligned} \tag{39c}$$

whereupon every quantity in (39) is evaluated at $\boldsymbol{x}_r, \boldsymbol{u}^{\star}$.

The boundary conditions for (39) at $t=0$ is

$$\boldsymbol{V}\left(\boldsymbol{x}_r, 0\right)=\boldsymbol{g}\left(0 ; \boldsymbol{x}_r(0)\right) ; \tag{40}$$

so that

$$\tilde{\boldsymbol{V}}(0)=0 \tag{41a}$$

$$\boldsymbol{V}_{\boldsymbol{x}}(0)=\boldsymbol{g}_{\boldsymbol{x}}\left(0 ; \boldsymbol{x}_r(0)\right) \tag{41b}$$

$$\boldsymbol{V}_{\boldsymbol{x x}}(0)=\boldsymbol{g}_{\boldsymbol{x x}}\left(0 ; \boldsymbol{x}_r(0)\right) . \tag{41c}$$

The following control laws are then applied

$$\boldsymbol{u}=\boldsymbol{u}_r+\boldsymbol{k}_{\boldsymbol{u}} \delta \boldsymbol{x}, \tag{42}$$

$$\boldsymbol{v}=\boldsymbol{v}_r+\boldsymbol{k}_{\boldsymbol{v}} \delta \boldsymbol{x} . \tag{43}$$

[6] The stage cost is zero from (??).

Therefore, at any time on a ROB of the value function, a local approximation of $\boldsymbol{V}$ consists in employing the TO-DO: Lax-Friedrichs scheme on the following system

$$\begin{aligned} -\left[\boldsymbol{E} + \boldsymbol{F}\delta\boldsymbol{x} + \frac{1}{2}\delta\boldsymbol{x}\boldsymbol{G}\delta\boldsymbol{x}\right] = \min\{\boldsymbol{0}, \boldsymbol{H} \\ -\boldsymbol{H}(t;\boldsymbol{x}_r,\boldsymbol{u}_r,\boldsymbol{v}_r,\boldsymbol{V}_{\boldsymbol{x}}) + \boldsymbol{H}_{\boldsymbol{x}} + \boldsymbol{V}_{\boldsymbol{xx}}\left(f - f(t;\boldsymbol{x}_r,\boldsymbol{u}_r,\boldsymbol{v}_r)\right) \\ +\boldsymbol{H}_{\boldsymbol{xx}} + f_{\boldsymbol{x}}^T\boldsymbol{V}_{\boldsymbol{xx}} + \boldsymbol{V}_{\boldsymbol{xx}}f_{\boldsymbol{x}} + \boldsymbol{k}_{\boldsymbol{u}}^T\boldsymbol{H}_{\boldsymbol{uu}}\boldsymbol{k}_{\boldsymbol{u}} + \boldsymbol{k}_{\boldsymbol{v}}^T\boldsymbol{H}_{\boldsymbol{vv}}\boldsymbol{k}_{\boldsymbol{v}}\} \end{aligned} \tag{44}$$

where $\boldsymbol{E}, \boldsymbol{F},$ and $\boldsymbol{G}$ are appropriately defined.

III. Results and Discussion.

We now provide results and analysis of the proposed numerical algorithm on benchmark control problems.

TO-DO: If we decide to do bang-bang optimal control, then we need a new formulation in our power series expansion function. Whence why I think we should work with e.g. a Dubins car model for a PoC since its control law (the turn orientation) admits continuity in the state space.

IV. Conclusion.

References

[1] Michael Athans and Peter L Falb. *Optimal Control: An Introduction to the Theory and its Applications.* Courier Corporation, 2013. 3

[2] Andrea Bajcsy, Somil Bansal, Eli Bronstein, Varun Tolani, and Claire J Tomlin. An Efficient Reachability-based Framework for Provably Safe Autonomous Navigation in Unknown Environments. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 1758–1765. IEEE, 2019. 1

[3] Somil Bansal and Claire J Tomlin. DeepReach : A Deep Learning Approach to High-Dimensional Reachability. 1

[4] Tamer Basar and Geert Jan Olsder. *Dynamic noncooperative game theory*, volume 23. Siam, 1999. 3

[5] Richard Bellman. *Dynamic programming*. Princeton University Press, 1957. ISBN 0-486-42809-5. 1

[6] M. G. Crandall, L. C. Evans, and P. L. Lions. Some Properties of Viscosity Solutions of Hamilton-Jacobi Equations. *Transactions of the American Mathematical Society*, 282(2):487, 1984. ISSN 00029947. doi: 10.2307/1999247. 1

[7] Michael Crandall and Andrew Majda. The method of fractional steps for conservation laws. *Numerische Mathematik*, 34(3):285–314, 1980. ISSN 0029599X. doi: 10.1007/BF01396704. 3

[8] Michael G Crandall and Pierre-Louis Lions. Viscosity solutions of hamilton-jacobi equations. *Transactions of the American mathematical society*, 277(1):1–42, 1983. 1, 2, 3

[9] Michael G Crandall and Andrew Majda. Monotone difference approximations for scalar conservation laws. *Mathematics of Computation*, 34(149):1–21, 1980. 1, 3

[10] L.C. Evans and Panagiotis E. Souganidis. Differential games and representation formulas for solutions of Hamilton-Jacobi-Isaacs equations. *Indiana Univ. Math. J*, 33(5):773–797, 1984. ISSN 0022-2518. 1, 2

[11] L.C. Evans and Panagiotis E. Souganidis. Differential games and representation formulas for solutions of Hamilton-Jacobi-Isaacs equations. *Indiana Univ. Math. J*, 33(5):773–797, 1984. ISSN 0022-2518. 1, 2

[12] Jaime F. Fisac, Anayo K. Akametalu, Melanie N. Zeilinger, Shahab Kaynama, Jeremy Gillula, and Claire J. Tomlin. A General Safety Framework for Learning-Based Control in Uncertain Robotic Systems. *IEEE Transactions on Automatic Control*, 64(7):2737–2752, 2019. ISSN 15582523. doi: 10.1109/TAC.2018.2876389. 2

[13] Jaime F. Fisac, Neil F. Lugovoy, Vicenc Rubies-Royo, Shromona Ghosh, and Claire J. Tomlin. Bridging hamilton-jacobi safety analysis and reinforcement learning. *Proceedings - IEEE International Conference on Robotics and Automation*, 2019-May:8550–8556, 2019. ISSN 10504729. doi: 10.1109/ICRA.2019.8794107. 2

[14] Chenjie Gu. Qlmor: A projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 30(9):1307–1320, 2011. 4

[15] Sylvia Herbert, Jason J Choi, Suvansh Sanjeev, Marsalis Gibson, Koushil Sreenath, and Claire J Tomlin. Scalable learning of safety guarantees for autonomous systems using hamilton-jacobi reachability. *arXiv preprint arXiv:2101.05916*, 2021. 1

[16] R Isaacs. Differential games 1965. *Kreiger, Huntigton, NY*. 1, 5

[17] David H Jacobson. New second-order and first-order algorithms for determining optimal control: A differential dynamic programming approach. *Journal of Optimization Theory and Applications*, 2(6):411–440, 1968. 3, 4

[18] David H. Jacobson and David Q. Mayne. *Differential Dynamic Programming*. American Elsevier Publishing Company, Inc., New York, NY, 1970. 3, 4

[19] Shahab Kaynama, Ian M. Mitchell, Meeko Oishi, and Guy A. Dumont. Scalable safety-preserving robust control synthesis for continuous-time linear systems. *IEEE Transactions on Automatic Control*, 60(11):3065–3070, 2015. 1

[20] L-Z Liao and Christine A Shoemaker. Convergence in unconstrained discrete-time differential dynamic programming. *IEEE Transactions on Automatic Control*, 36 (6):692–706, 1991. 4

[21] Pierre-Louis Lions. *Generalized solutions of Hamilton-Jacobi equations*, volume 69. London Pitman, 1982. 1, 3

[22] John Lygeros. On reachability and minimum cost optimal control. *Automatica*, 40(6):917–927, 2004. 1

[23] David Mayne. A Second-Order Gradient Method for Determining Optimal Trajectories of Non-linear Discrete-Time Systems. *International Journal of Control*, 3(1):

85–95, 1966. 3

[24] Stephen R. McReynolds. The successive sweep method and dynamic programming. *Journal of Mathematical Analysis and Applications*, 19(3):565–598, 1967. ISSN 10960813. 3

[25] AW Merz. The game of two identical cars. *Journal of Optimization Theory and Applications*, 9(5):324–343, 1972. 1, 5

[26] Ian Mitchell. A Robust Controlled Backward Reach Tube with (Almost) Analytic Solution for Two Dubins Cars (Presentation). 74:242–224, 2020. doi: 10.29007/mx3f. 1, 3

[27] Ian M. Mitchell, Alexandre M. Bayen, and Claire J. Tomlin. A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games. *IEEE Transactions on Automatic Control*, 50(7):947–957, 2005. ISSN 00189286. doi: 10.1109/TAC.2005.851439. 1, 2, 3

[28] S. K. Mitter. Successive approximation methods for the solution of optimal control problems. *Automatica*, 3(3-4):135–149, 1966. ISSN 00051098. 3

[29] Olalekan Ogunmolu, Nicholas Gans, and Tyler Summers. Minimax iterative dynamic game: Application to nonlinear robot control tasks. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6919–6925. IEEE, 2018. 3

[30] S Osher and R Fedkiw. Level Set Methods and Dynamic Implicit Surfaces. *Applied Mechanics Reviews*, 57(3):B15–B15, 2004. ISSN 0003-6900.

[31] Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988. 1