# Analysis of Spurious Local Solutions of Optimal Control Problems: One-Shot Optimization Versus Dynamic Programming

Yuhao Ding, Yingjie Bi and Javad Lavaei

***Abstract*— Dynamic programming (DP) has a rich theoretical foundation and a broad range of applications, especially in the classic area of optimal control and the recent area of reinforcement learning (RL). Many optimal control problems can be solved as a single optimization problem, named one-shot optimization, or via a sequence of optimization problems using DP. However, the computation of their global optima often faces the NP-hardness issue due to the non-linearity of the dynamics and non-convexity of the cost, and thus only local optimal solutions may be obtained at best. Furthermore, in many cases arising in machine learning and model-free approaches, DP is the only viable choice, and therefore it is essential to understand when DP combined with a local search solver works. In this work, we introduce the notions of spurious local minimizers for the one-shot optimization and spurious local minimum policies for DP, and show that there is a deep connection between them. In particular, we prove that under mild conditions the DP method using local search can successfully solve the optimal control problem to global optimality if and only if the one-shot optimization is free of spurious solutions. This result paves the way to understand the performance of local search methods in optimal control and RL.**

## I. INTRODUCTION

Dynamic programming (DP) is a simple mathematical technique that has been widely used in a variety of fields. Following Bellman's influential work [1] on demonstrating the broad scope of DP and laying the foundation of its theory, many mathematical and algorithmic aspects of DP have been investigated [2]. One main application of DP is to solve optimal control problems, with applications in communication systems [3], inventory control [4], powertrain control [5], and many more. Furthermore, many recent successes in artificial intelligence, especially in reinforcement learning (RL) [6], [7], are also deeply rooted in DP. For example, in the challenging domain of classic Atari 2600 games, the work [8] has demonstrated that the deep Q-learning method based on the generalized policy iteration together with a deep neural network as the function approximator for the Q-values surpasses the performance of all previous algorithms and achieves a level comparable to that of a professional human games tester.

Although DP has a rich theoretical foundation and a broad range of applications, the exact solutions of large-scale optimal control problems are often impossible to obtain using DP in practice [6]. Apart from suffering the "curse of dimensionality" when the state space is large, solving DP accurately could also be highly complex. The reason is that DP requires solving optimization sub-problems to global optimality, which is NP-hard in general. Therefore, even though the theory of DP relies on global optimization solvers, practitioners use local optimization solvers based on first- and second-order numerical algorithms. As a result, the theoretical guarantee of DP could break down as soon as a non-global local solution is found in any of the sub-problems. Understanding the performance of local search methods for non-convex problems has been a focal area in machine learning in recent years. This is performed under the notion of spurious solution, which refers to a local minimum that is not a global solution. The specific application areas are neural networks, dictionary learning, deep learning, mixed linear regression, phase retrieval and online optimization [9], [10], [11], [12], [13], [14], [15], [16]. Recently, there has been an increasing interest in understanding the global convergence of the approximate DP algorithms for problems with special structures [17], [18], [19], but the literature lacks a rigorous analysis of spurious solutions in DP.

In this paper, we analyze the spurious solutions of the DP method. To streamline the presentation, we focus on the deterministic finite-horizon optimal control problem whose goal is to find an optimal input sequence such that the total cost is minimized while the dynamics and input constraints are satisfied. One approach to solving the problem is by formulating it as a one-shot optimization problem, and another approach is using the DP to formulate it as a sequential decision-making problem and solve it backwardly. Although it is well-known that for the deterministic optimal control problem, the one-shot method and the DP method return the same globally optimal control sequence, it is not yet known what would occur if the global optimizer needed for solving each sub-optimization problem in DP is replaced by a local optimizer. To address this question, we first introduce the notion of locally minimum control policy of DP and prove that under some mild conditions, each local minimizer of the one-shot optimization corresponds to the control input induced by a locally minimum control policy of DP, and vice versa. This result precisely uncovers the connection between the optimization landscapes of the one-shot and DP optimization problems.

Since DP is an integral part of model-free optimal control and RL, the results of this paper explain that the success of DP solely depends on the optimization landscape of a single optimization problem. Although we focus on the deterministic finite-horizon optimal control problem in this work, its generalization to the infinite-horizon stationary optimal control problem is straightforward. The technique developed

Yuhao Ding, Yingjie Bi and Javad Lavaei are with the Department of Industrial Engineering and Operations Research, University of California, Berkeley. Emails: {yuhao_ding, yingjiebi, lavaei}@berkeley.edu

This work was supported by grants from ARO, AFOSR, ONR and NSF.

in the paper can also be used to study the stochastic optimal control problem, but it is not included here due to space restrictions.

*Notations:* Let $\mathbb{R}$ denote the set of real numbers. We use $B(c, r)$ to denote the open ball centered at $c$ with radius $r$ and use $\bar{B}(c, r)$ to denote the closure of $B(c, r)$. The notation $x \in A - B$ means that $x$ is in the set $A$ but not in the set $B$. Let $\|\cdot\|$ denote the Euclidean norm and $\nabla_x f(x, y)$ denote the gradient of $f(x, y)$ with respect to $x$. The notation $\nabla_x^2 f(x) \succ 0$ means that the Hessian of $f(x)$ is positive definite.

## II. Problem Formulation

Consider a general discrete-time finite-horizon optimal control problem with $n$ time steps:

$$\begin{aligned} \min_{\substack{x_1,\ldots,x_n,\\ u_0,\ldots,u_{n-1}}} \quad & \sum_{i=0}^{n-1} c_i(x_i, u_i) + c_n(x_n) \\ \text{s.t.} \quad & x_{i+1} = f_i(x_i, u_i), \quad i = 0, \ldots, n-1, \\ & u_i \in A, \qquad\qquad\quad i = 0, \ldots, n-1, \\ & x_0 \text{ is given,} \end{aligned} \tag{P1}$$

where $x_i \in \mathbb{R}^N$ is the state at time $i$ and $u_i$ is the control input at time $i$ that is constrained to be in an action space $A \subseteq \mathbb{R}^M$. The state transition is governed by the dynamics $f_i : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}^N$. Each time instance $i$ is associated with a stage cost $c_i : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}$ or the terminal cost $c_n : \mathbb{R}^N \to \mathbb{R}$. Given an initial state $x_0$, the goal of the optimal control problem is to find an optimal control input $(u_0, \ldots, u_{n-1})$ minimizing the sum of the stage costs and the terminal cost. In this paper, the dynamics $f_i$ and the cost functions $c_i$ are assumed to be at least twice continuously differentiable over $\mathbb{R}^N \times \mathbb{R}^M$, and the action space $A$ is assumed to be compact.

The optimal control problem can be solved by two approaches. The first approach is to directly solve (P1) as an one-shot optimization problem that simultaneously solves for all variables. To simplify the following analysis, we eliminate the equality constraints in (P1) via the notation $C(x_k; u_k, \ldots, u_{n-1})$ defined as the cost-to-go started at the time step $k$ with the initial state $x$ and control inputs $u_k, \ldots, u_{n-1}$. In other words,

$$\begin{aligned} C(x) &= c_n(x), \\ C(x; u_k, \ldots, u_{n-1}) &= c_k(x, u_k) \\ &\quad + C(f_k(x, u_k); u_{k+1}, \ldots, u_n), \end{aligned}$$

for $k = 0, \ldots, n-1$. The one-shot optimization problem (P1) can be equivalently written as

$$\begin{aligned} \min \quad & C(x_0; u_0, \ldots, u_{n-1}) \\ \text{s.t.} \quad & u_i \in A, \quad i = 0, \ldots, n-1. \end{aligned} \tag{P2}$$

The second approach to solving the optimal control problem is based on DP. Let $J_k(x_k)$ denote the optimal cost-to-go at the time step $k$ with the initial state $x_k$, i.e.,

$$\begin{aligned} J_k(x_k) = \min \quad & C(x_k; u_k, \ldots, u_{n-1}) \\ \text{s.t.} \quad & u_i \in A, \quad i = k, \ldots, n-1, \end{aligned}$$

Then, $J_k$ can be computed in a backward fashion from the time step $n-1$ to time $0$ through the following recursion:

$$\begin{aligned} J_n(x) &= c_n(x), \\ J_k(x) &= \min_{u \in A} \{ c_k(x, u) + J_{k+1}(f_k(x, u)) \}, \end{aligned} \tag{P3}$$

for $k = 0, \ldots, n-1$. The optimal cost $J_0(x_0)$ equals the optimal objective value of (P1).

However, due to the non-convexity of the look-ahead objective functions, it is generally NP-hard to obtain the globally optimal solution of (P3) for all states and at all times. Specifically, when using the DP to solve the optimal control problem (P1), the first step is to compute

$$\min_{u \in A} \{ c_{n-1}(x_{n-1}, u) + c_n(f_{n-1}(x_{n-1}, u)) \}$$

for every $x_{n-1} \in \mathbb{R}^N$, which requires solving nonconvex optimization problems if the cost function is nonconvex or the dynamics is nonlinear. Since these intermediate problems are normally solved via local search methods, the best expectation is to obtain a local minimizer for $u_{n-1}$ as a function of $x \in \mathbb{R}^N$, denoted by the policy $\pi_{n-1}(x)$. As a result, instead of working with truly optimal cost-to-go functions, one may arrive at an inexact cost-to-go at time $n-1$ as follows:

$$\begin{aligned} J_{n-1}^{\pi}(x_{n-1}) =& c_{n-1}(x_{n-1}, \pi_{n-1}(x_{n-1})) \\ &+ c_n(f_{n-1}(x_{n-1}, \pi_{n-1}(x_{n-1})) \end{aligned}$$

which is obtained based on the local minimizer $\pi_{n-1}(x)$. Subsequently, it is required to solve the optimal decision-making problem

$$\min_{u \in A} \{ c_{n-2}(x_{n-2}, u) + J_{n-1}^{\pi}(f_{n-2}(x_{n-2}, u)) \}$$

for every $x_{n-2} \in \mathbb{R}^N$. By repeating this procedure in a backward fashion toward the time step 0, we obtain a group of policy functions $\pi_k$ and inexact cost-to-go functions $J_k^{\pi}$ for $k = 0, \ldots, n-1$. Given the initial state $x_0$, let

$$\begin{aligned} u_0 = \pi_0(x_0), &\quad x_1 = f_0(x_0, u_0) \\ u_1 = \pi_1(x_1), &\quad x_2 = f_1(x_1, u_1) \\ &\cdots \\ u_{n-1} = \pi_{n-1}(x_{n-1}), &\quad x_n = f_n(x_{n-1}, u_{n-1}), \end{aligned}$$

be the control inputs and the states induced by the policies $\pi_0, \ldots, \pi_{n-1}$. Then, $(u_0, \ldots, u_{n-1})$ is an approximate solution to the original optimal control problem (P1) with the sub-optimal objective value $J_0^{\pi}(x_0)$. This motivates us to define locally minimum control policies based on solving (P3) to local optimality.

*Definition 1:* Given a control policy $\pi = (\pi_0, \ldots, \pi_{n-1})$, the associated Q-functions $Q_k^{\pi}(\cdot, \cdot)$ and cost-to-go functions $J_k^{\pi}(\cdot)$ under the policy $\pi$ are defined backwardly from the time step $n-1$ to the time step $0$ through the following recursion:

$$\begin{aligned} J_n^{\pi}(x) &= c_n(x), \\ Q_k^{\pi}(x, u) &= c_k(x, u) + J_{k+1}^{\pi}(f_k(x, u)), \\ J_k^{\pi}(x) &= Q_k^{\pi}(x, \pi_k(x)), \end{aligned}$$

for $k = 0, \ldots, n-1$.

*Definition 2 (locally minimum control policy):* A control policy $\pi = (\pi_0, \ldots, \pi_{n-1})$ is said to be *a locally minimum control policy of DP* if for all $k \in \{0, \ldots, n-1\}$ and for all $x \in \mathbb{R}^N$, the policy $\pi_k(x)$ is a local minimizer of the Q-function $Q_k^\pi(x, \cdot)$, meaning that there exists $\epsilon_k^*(x) > 0$ such that

$$Q_k^\pi(x, \pi_k(x)) \leq Q_k^\pi(x, \tilde{u}),\ \forall \tilde{u} \in B(\pi_k(x), \epsilon_k^*(x)) \cap A.$$

It is further called a spurious (non-global) locally minimum control policy of DP if $J_0^\pi(x_0) > J_0(x_0)$.

*Definition 3 (local minimizer):* A vector $(u_0^*, \ldots, u_{n-1}^*)$ is said to be a local minimizer of the one-shot optimization problem (P2) if there exists $\epsilon > 0$ such that

$$C(x_0, u_0^*, \ldots, u_{n-1}^*) \leq C(x_0, \tilde{u}_0, \ldots, \tilde{u}_{n-1})$$

for all $\tilde{u}_i \in B(u_i^*, \epsilon) \cap A$ where $i = 0, \ldots, n-1$. It is further called a spurious local minimizer of the one-shot optimization problem if $C(x_0, u_0^*, \ldots, u_{n-1}^*) > J_0(x_0)$.

In the remainder of the paper, we will study the relationship between the (spurious) local minimizers of the one-shot problem and the (spurious) locally minimum control policies of DP. The goal is to show that, under mild conditions, the inexact DP method based on local search algorithms solves the optimal control problem to global optimality if and only if its corresponding one-shot optimization can be successfully solved using local search methods.

## III. Main Results

### A. Local minimizer: From one-shot optimization to DP

In this subsection, we will show that any strict local minimizer of the one-shot problem is induced by a locally minimum control policy $\pi$ of DP. Before proving the theorem, we first provide the following useful lemma.

*Lemma 1:* Given a function $g : \mathbb{R}^N \times A \to \mathbb{R}$, a point $x^* \in \mathbb{R}^N$ and a number $\epsilon > 0$, if $u^* \in A$ is a strict local minimizer of the function $g(x^*, \cdot)$, and $g$ is continuous in a neighborhood of $(x^*, u^*)$, then there exist $\delta > 0$ and a function $h : B(x^*, \delta) \to A$ such that $h(x^*) = u^*$ and that the following holds for all $x \in B(x^*, \delta)$:

1) $h(x)$ is a local minimizer of $g(x, \cdot)$.
2) $h(x) \in B(u^*, \epsilon)$.
3) The function $g(x, h(x))$ is continuous at $x$.

*Proof:* By assumption, there exist $\delta_1 > 0$ and $0 < \epsilon_1 < \epsilon$ such that the function $g$ is continuous on the set $\bar{B}(x^*, \delta_1) \times (B(u^*, \epsilon_1) \cap A)$ and

$$g(x^*, u) > g(x^*, u^*) \tag{1}$$

for every $u \in B(u^*, \epsilon_1) \cap A$ with $u \neq u^*$. If $u^*$ is an isolated point of $A$, then one can simply choose $\delta = \delta_1$ and $h(x) = u^*$ for all $x \in B(x^*, \delta)$. Otherwise, there exists $u' \in B(u^*, \epsilon_1) \cap A$ with $u' \neq u^*$. Let

$$\epsilon_2 = \|u' - u^*\| \in (0, \epsilon_1)$$

and consider the optimization problem

$$\begin{aligned} \min_u \quad & g(x^*, u) \\ \text{s.t.} \quad & \|u - u^*\| = \epsilon_2, \quad u \in A. \end{aligned}$$

The feasible set of the above problem is nonempty and compact, and therefore its optimal value is attained by some point $\hat{u} \in A$. In light of (1), it holds that

$$\Delta = g(x^*, \hat{u}) - g(x^*, u^*) > 0.$$

Since $g$ is uniformly continuous on the compact set $\bar{B}(x^*, \delta_1) \times (\bar{B}(u^*, \epsilon_2) \cap A)$, there exists $0 < \delta \leq \delta_1$ such that

$$|g(x, u) - g(x^*, u)| < \frac{\Delta}{2}$$

for all $x \in B(x^*, \delta)$ and $u \in \bar{B}(u^*, \epsilon_2) \cap A$. For every $x \in B(x^*, \delta)$, define $h(x)$ to be an arbitrary global minimizer of $g(x, \cdot)$ over the compact set $\bar{B}(u^*, \epsilon_2) \cap A$. Then, $h(x) \in B(u^*, \epsilon)$ with $h(x^*) = u^*$. Moreover,

$$g(x, h(x)) = \min_{u \in \bar{B}(u^*, \epsilon_2) \cap A} g(x, u), \quad \forall x \in B(x^*, \delta),$$

which is a continuous function of $x$ due to the Berge maximum theorem [20]. It remains to show that $h(x)$ is a local minimizer of $g(x, \cdot)$ over the entire space $A$. If $\|h(x) - u^*\| = \epsilon_2$, then

$$\begin{aligned} g(x, h(x)) &> g(x^*, h(x)) - \frac{\Delta}{2} \geq g(x^*, \hat{u}) - \frac{\Delta}{2} \\ &= g(x^*, u^*) + \frac{\Delta}{2} > g(x, u^*), \end{aligned}$$

which contradicts the fact that $h(x)$ is a global minimizer of $g(x, \cdot)$ on the set $\bar{B}(u^*, \epsilon_2) \cap A$. As a result, $h(x) \in B(u^*, \epsilon_2) \cap A$, which implies that $h(x)$ is a global minimizer of $g(x, \cdot)$ over the set $B(u^*, \epsilon_2) \cap A$; thus, it is a local minimizer of $g(x, \cdot)$ over the entire space $A$. ■

*Theorem 1:* If the one-shot problem has a (spurious) strict local minimizer $(u_0^*, \ldots, u_{n-1}^*)$, then there exists a (spurious) locally minimum control policy $\pi$ of DP with the property that $\pi_k(x_k^*) = u_k^*$ for all $k \in \{0, \ldots, n-1\}$, where $(x_0^*, \ldots, x_n^*)$ is the state sequence associated with the (spurious) solution of the one-shot problem.

*Proof:* Assume that $(u_0^*, \ldots, u_{n-1}^*)$ is a strict local minimizer of the one-shot problem. There exists $\epsilon > 0$ such that

$$C(x_0; u_0^*, \ldots, u_{n-1}^*) < C(x_0; u_0, \ldots, u_{n-1}), \tag{2}$$

for every control sequence $(u_0, \ldots, u_{n-1}) \neq (u_0^*, \ldots, u_{n-1}^*)$ with the property that $u_i \in B(u_i^*, \epsilon) \cap A$ for $i = 0, \ldots, n-1$. In what follows, we will prove by a backward induction that there exist policies $\pi_0, \ldots, \pi_{n-1}$, and positive numbers $\delta_0, \ldots, \delta_n$, such that they jointly satisfy the following properties:

1) $\pi_k(x_k)$ is a local minimizer of the function $Q_k^\pi(x_k, \cdot)$ for all $x_k \in \mathbb{R}^N$.
2) $\pi_k(x_k^*) = u_k^*$.
3) For all $x_k \in B(x_k^*, \delta_k)$, it holds that

$$\pi_k(x_k) \in B(u_k^*, \epsilon),\ f_k(x_k, \pi_k(x_k)) \in B(x_{k+1}^*, \delta_{k+1}).$$

4) $J_k^\pi$ is lower semi-continuous on $\mathbb{R}^N$ and continuous on $B(x_k^*, \delta_k)$.

For the base step $k = n$, we choose an arbitrary $\delta_n > 0$ and notice that $J_n^\pi(x) = c_n(x)$ , implying that $J_n^\pi$ is always continuous. For $k < n$, assume that $\pi_{k+1}, \dots, \pi_{n-1}$ and $\delta_{k+1}, \dots, \delta_n$ with the above properties have been found. First, by the continuity of $f_k$, there exist $\delta_k' > 0$ and $0 < \epsilon_k < \epsilon$ such that

$$f_k(x_k, u_k) \in B(x_{k+1}^*, \delta_{k+1}), \quad \forall (x_k, u_k) \in S_k, \tag{3}$$

where

$$S_k = B(x_k^*, \delta_k') \times (B(u_k^*, \epsilon_k) \cap A).$$

Since

$$Q_k^\pi(x_k, u_k) = c_k(x_k, u_k) + J_{k+1}^\pi(f_k(x_k, u_k))$$

and $J_{k+1}^\pi$ is continuous on $B(x_{k+1}^*, \delta_{k+1})$, $Q_k^\pi$ is continuous on $S_k$. Next, for every $\tilde{u}_k \in B(u_k^*, \epsilon_k) \cap A$, if we define

$$\begin{aligned} \tilde{x}_{k+1} &= f_k(x_k^*, \tilde{u}_k), \quad \tilde{u}_{k+1} = \pi_{k+1}(\tilde{x}_{k+1}), \\ \tilde{x}_{k+2} &= f_{k+1}(\tilde{x}_{k+1}, \tilde{u}_{k+1}), \quad \tilde{u}_{k+2} = \pi_{k+2}(\tilde{x}_{k+2}), \\ &\cdots \\ \tilde{x}_{n-1} &= f_{n-2}(\tilde{x}_{n-2}, \tilde{u}_{n-2}), \quad \tilde{u}_{n-1} = \pi_{n-1}(\tilde{x}_{n-1}), \end{aligned}$$

by applying (3) and then the third property above repeatedly, we arrive at

$$\tilde{u}_i \in B(u_i^*, \epsilon) \cap A, \quad \forall i \in \{k+1, \dots, n-1\}.$$

When $\tilde{u}_k \neq u_k^*$, it follows from (2) and the second property above that

$$\begin{aligned} Q_k^\pi(x_k^*, \tilde{u}_k) &= C(x_k^*; \tilde{u}_k, \dots, \tilde{u}_{n-1}) \\ &= C(x_0; u_0^*, \dots, u_{k-1}^*, \tilde{u}_k, \dots, \tilde{u}_{n-1}) - \sum_{i=0}^{k-1} c_i(x_i^*, u_i^*) \\ &> C(x_0; u_0^*, \dots, u_{n-1}^*) - \sum_{i=0}^{k-1} c_i(x_i^*, u_i^*) \\ &= C(x_k^*; u_k^*, \dots, u_{n-1}^*) = Q_k^\pi(x_k^*, u_k^*). \end{aligned}$$

As a result, $u_k^*$ is a strict local minimizer of $Q_k^\pi(x_k^*, \cdot)$. Applying Lemma 1 to the function $Q_k^\pi$ with $x_k^*$ and $\epsilon_k$, one can find $0 < \delta_k < \delta_k'$ and a function $h_k : B(x_k^*, \delta_k) \to A$ such that $h_k(x_k^*) = u_k^*$ and that the following holds for every $x_k \in B(x_k^*, \delta_k)$:

1) $h_k(x_k)$ is a local minimizer of $Q_k^\pi(x_k, \cdot)$.
2) $h_k(x_k) \in B(u_k^*, \epsilon_k) \subseteq B(u_k^*, \epsilon)$, which together with (3) implies that $f_k(x_k, h_k(x_k)) \in B(x_{k+1}^*, \delta_{k+1})$.
3) The function $Q_k^\pi(x_k, h_k(x_k))$ is continuous at $x_k$.

Let $\pi_k$ be the extension of the function $h_k$ by setting $\pi_k(x_k)$ to be any global minimizer of the lower semi-continuous function $Q_k^\pi(x_k, \cdot)$ over the compact set $A$ if $x_k \notin B(x_k^*, \delta_k)$. Obviously, $\pi_k$ satisfies the first three properties. To verify the last property, observe that

$$J_k^\pi(x_k) = \begin{cases} Q_k^\pi(x_k, h_k(x_k)), & \text{if } x_k \in B(x_k^*, \delta_k), \\ H_k(x_k), & \text{otherwise}, \end{cases}$$

in which

$$H_k(x_k) = \min_{u \in A} Q_k^\pi(x_k, u),$$

and therefore $J_k^\pi$ is continuous on the set $B(x_k^*, \delta_k)$. In addition, note that $J_{k+1}^\pi$ and thus $Q_k^\pi$ is lower semi-continuous, while $A$ is compact. Hence, it follows from the Berge maximum theorem [20] that $H_k$ is also lower semi-continuous on $\mathbb{R}^N$, which implies that $J_k^\pi$ is lower semi-continuous on $\mathbb{R}^N - \bar{B}(x_k^*, \delta_k)$. For every point $\bar{x}_k$ on the boundary of $B(x_k^*, \delta_k)$, since $H_k$ is lower semi-continuous at $\bar{x}_k$, for every $\bar{\epsilon} > 0$ there exists $\bar{\delta} > 0$ such that

$$J_k^\pi(x_k) \geq H_k(x_k) > H_k(\bar{x}_k) - \bar{\epsilon} = J_k^\pi(\bar{x}_k) - \bar{\epsilon}$$

holds for all $x_k \in B(\bar{x}_k, \bar{\delta})$. Therefore, $J_k^\pi$ is also lower semi-continuous at $\bar{x}_k$.

By the first and second properties, $\pi = (\pi_0, \dots, \pi_{n-1})$ will be a locally minimum control policy of DP. Furthermore, if $(u_0^*, \dots, u_{n-1}^*)$ is a spurious local minimizer of the one-shot problem, then

$$J_0^\pi(x_0) = C(x_0; u_0^*, \dots, u_{n-1}^*) > J_0(x_0),$$

which implies that $\pi$ is also a spurious locally minimum control policy of DP. ■

*Remark 1:* DP can be viewed as a reformulation of the optimal control problem from a single one-shot optimization problem to a sequence of optimization problems. When a non-convex problem is reformulated, its local minimizers could change and for example convexification serves as a reformulation in a higher dimensional space that eliminates spurious solutions. However, Theorem 1 shows that, under mild conditions, DP is a reformulation of one-shot optimization problem that preserves local minimizers. Note that a spurious solution of DP is a set of functions, where a spurious solution of the one-shot optimization is a vector.

By taking the contrapositive of Theorem 1, one can immediately obtain the result that the one-shot problem has no spurious strict local minimizers as long as DP has no spurious locally minimum control policies.

### *B. Local minimizer: From DP to one-shot optimization*

Although the input sequence induced by the globally minimal control policy is the global minimizer of the one-shot problem, it turns out the induced input sequence under a spurious locally minimum control policy of DP does not generally imply a spurious local minimizer of the one-shot problem. However, this implication is indeed the case if some mild conditions are satisfied.

Before presenting the theorem, we first show that the differences between two state sequences starting from the same initial state are small if the the differences between the corresponding control sequences are small.

*Lemma 2:* Consider the system under an input sequence $(u_0, \dots, u_{n-1})$ with associated state sequence $(x_0, \dots, x_n)$. Then, there exist continuous and non-decreasing functions $L_k(\delta_0, \dots, \delta_k)$, $k = 0, \dots, n-1$, satisfying $L_k(0, \dots, 0) = 0$ and the following property: for any input sequence

$(\tilde{u}_0, \ldots, \tilde{u}_{n-1})$ with $\tilde{u}_i \in B(u_i, \delta_i) \cap A$ for all $i \in \{0, \ldots, n-1\}$, it holds that

$$\|x_{k+1} - \tilde{x}_{k+1}\| \leq L_k(\delta_0, \ldots, \delta_k),$$

where $(\tilde{x}_0, \ldots, \tilde{x}_n)$ is the state sequence correponding to $(\tilde{u}_0, \ldots, \tilde{u}_{n-1})$ starting from the initial state $x_0$.

*Proof:* For $i = 1$, we have

$$\begin{aligned}\|x_1 - \tilde{x}_1\| =& \|f_0(x_0, u_0) - f_0(\tilde{x}_0, \tilde{u}_0)\| \\ \leq& L_{u,0}(\delta_0)\|u_0 - \tilde{u}_0\| \leq L_0(\delta_0),\end{aligned}$$

where

$$L_{u,0}(\delta_0) = \max_{u \in \bar{B}(u_0, \delta_0)} \|\nabla_u f_0(x_0, u)\|,$$
$$L_0(\delta_0) = L_{u,0}(\delta_0)\delta_0.$$

By the above definition, $L_0$ is continuous and non-decreasing with $L_0(0) = 0$. Then, for $i = 2$, we have

$$\begin{aligned}&\|x_2 - \tilde{x}_2\| = \|f_1(x_1, u_1) - f_1(\tilde{x}_1, \tilde{u}_1)\| \\ &\quad \leq \|f_1(x_1, u_1) - f_1(\tilde{x}_1, u_1) + f_1(\tilde{x}_1, u_1) - f_1(\tilde{x}_1, \tilde{u}_1)\| \\ &\quad \leq L_{x,1}(\delta_0)\|x_1 - \tilde{x}_1\| + L_{u,1}(\delta_0, \delta_1)\|u_1 - \tilde{u}_1\| \\ &\quad \leq L_1(\delta_0, \delta_1),\end{aligned}$$

where

$$\begin{aligned}L_{x,1}(\delta_0) =& \max_{x \in \bar{B}(x_1, L_0(\delta_0))} \|\nabla_x f_1(x, u_1)\|, \\ L_{u,1}(\delta_0, \delta_1) =& \max_{x \in \bar{B}(x_1, L_0(\delta_0)), u \in \bar{B}(u_1, \delta_1)} \|\nabla_u f_1(x, u)\|, \\ L_1(\delta_0, \delta_1) =& L_{x,1}(\delta_0)L_0(\delta_0) + L_{u,1}(\delta_0, \delta_1)\delta_1.\end{aligned}$$

Similarly, $L_1$ is continuous and non-decreasing in $\delta_0$ and $\delta_1$, and $L_0(0) = 0$ further implies $L_1(0, 0) = 0$. Repeating this procedure yields that for every $k \in \{0, \ldots, n-1\}$, there exists a continuous and non-decreasing function $L_k(\delta_0, \ldots, \delta_k)$ with $L_k(0, \ldots, 0) = 0$ such that

$$\|x_{k+1} - \tilde{x}_{k+1}\| \leq L_k(\delta_0, \ldots, \delta_k).$$

This completes the proof. ∎

*Theorem 2:* Consider a (spurious) locally minimum control policy $\pi = (\pi_0, \ldots, \pi_{n-1})$, and let the corresponding input and state sequences associated with the initial state $x_0$ be denoted as $(u_0^*, \ldots, u_{n-1}^*)$ and $(x_0^*, \ldots, x_n^*)$. If $\pi_k$ is Lipschitz continuous in a neighborhood of $x_k^*$ and $\epsilon_k^*$ (see Definition 2) is continuous at $x_k^*$ for $k = 0, \ldots, n-1$, then $(u_0^*, \ldots, u_{n-1}^*)$ is also a (spurious) local minimizer of the one-shot problem.

*Proof:* We first show that there exist positive constants $\delta_0, \ldots, \delta_{n-1}$ such that for every $L > 0$ and $i = 0, \ldots, n-1$, the following holds:

$$\delta_i + Ld_i \leq \inf_{x \in \bar{B}(x_i^*, d_i)} \epsilon_i^*(x), \tag{4}$$

where $d_0 = 0$ and $d_i = L_{i-1}(\delta_0, \ldots, \delta_{i-1})$ for $i > 0$ with $L_{i-1}$ given in the statement of Lemma 2. In the latter case, because $L_{i-1}$ is continuous with $L_{i-1}(0, \ldots, 0) = 0$, and $\epsilon_i^*$ is continuous at $x_i^*$, if $\delta_0, \ldots, \delta_{n-1}$ are sufficiently small,

$$\inf_{x \in \bar{B}(x_i^*, d_i)} \epsilon_i^*(x) \geq \frac{1}{2}\epsilon_i^*(x_i^*) > 0.$$

Thus, at step $n-1$, there must exist $\delta_0, \ldots, \delta_{n-1}$ for which (4) holds. Then, at step $n-2$, if $\delta_0, \ldots, \delta_{n-2}$ do not make (4) hold, since $L_{n-2}$ is non-decreasing, we can further reduce $\delta_0, \ldots, \delta_{n-2}$ to satisfy (4) for $i = n-2$ without breaking (4) for $i = n-1$. By repeating this procedure, one can show that there exist positive constants $\delta_0, \ldots, \delta_{n-1}$ such that (4) holds for all $i \in \{0, \ldots, n-1\}$. Moreover, we can again reduce $\delta_0, \ldots, \delta_{n-1}$ such that each $\pi_i$ is Lipschitz continuous over the set $\bar{B}(x_i^*, d_i)$.

Now, it is desirable to show that for all $\tilde{u}_i \in B(u_i^*, \delta_i) \cap A$ where $i \in \{0, \ldots, n-1\}$, we have

$$J_0^\pi(x_0) = C(x_0; u_0^*, \ldots, u_{n-1}^*) \leq C(x_0; \tilde{u}_0, \ldots, \tilde{u}_{n-1}).$$

Since (4) implies that $\delta_0 \leq \epsilon_0^*(x_0)$, one can write

$$\begin{aligned}&J_0^\pi(x_0) \leq c_0(x_0, \tilde{u}_0) + J_1^\pi(\tilde{x}_1), \\ &\text{where } \tilde{x}_1 = f_0(x_0, \tilde{u}_0),\ \forall \tilde{u}_0 \in B(u_0^*, \delta_0) \cap A.\end{aligned}$$

For every $i \in \{1, \ldots, n-1\}$, by the definition of local optimality of $\pi_i(\tilde{x}_i)$,

$$\begin{aligned}&J_i^\pi(\tilde{x}_i) \leq c_i(\tilde{x}_i, \tilde{u}_i) + J_{i+1}^\pi(\tilde{x}_{i+1}), \\ &\text{where } \tilde{x}_{i+1} = f_i(\tilde{x}_i, \tilde{u}_i),\ \forall \tilde{u}_i \in B(\pi_i(\tilde{x}_i), \epsilon_i^*(\tilde{x}_i)) \cap A.\end{aligned} \tag{5}$$

Now, we aim to show that (5) also holds for all $\tilde{u}_i \in B(\pi_i(x_i^*), \delta_i) \cap A$; or equivalently,

$$B(\pi_i(x_i^*), \delta_i) \cap A \subseteq B(\pi_i(\tilde{x}_i), \epsilon_i^*(\tilde{x}_i)) \cap A.$$

It suffices to prove that

$$\delta_i + \|\pi_i(x_i^*) - \pi_i(\tilde{x}_i)\| \leq \epsilon_i^*(\tilde{x}_i). \tag{6}$$

Because of the Lipschitz continuity of $\pi_i$ in $\bar{B}(x_i^*, d_i)$, there exists a positive constant $L_\pi$ such that

$$\|\pi_i(x_i^*) - \pi_i(\tilde{x}_i)\| \leq L_\pi\|x_i^* - \tilde{x}_i\|, \quad i = 0, \ldots, n-1.$$

Then, the inequality (6) must hold because of (4).

Now, assume that $\pi$ is a spurious locally minimum control policy of DP, i.e., $J_0^\pi(x_0) > J_0(x_0)$. Then, because of

$$C(x_0; u_0^*, \ldots, u_{n-1}^*) = J_0^\pi(x_0) > J_0(x_0),$$

$(u_0^*, \ldots, u_{n-1}^*)$ is also a spurious local minimizer of the one-shot problem. ∎

One situation where $\epsilon_i^*(x)$ is continuous in a neighborhood of $x_i^*$ is that the local minimizers of $Q_i^\pi(x, \cdot)$ do not bifurcate at $x_i^*$. As shown later in Example 2, if the local minimizers of $Q_i^\pi(x, \cdot)$ bifurcates at $x_i^*$, then $\epsilon_i^*(x)$ is discontinuous at $x_i^*$ and the infimum of $\epsilon_i^*(x)$ may be zero. In this case, there is no guarantee that the induced control input of the locally optimal control policy of DP will also be a local minimizer of the one-shot problem. Thus, the assumption of the continuity of $\epsilon_i^*(x)$ is necessary for the results in Theorem 2 to be true. Another situation where the results in Theorem 2 hold true is that the Hessian of $Q_i^\pi(x, \cdot)$ is positive definite. We present this idea below.

*Theorem 3:* Assume that $A$ is convex. Consider a (spurious) locally minimum control policy $\pi = (\pi_0, \ldots, \pi_{n-1})$, and let the corresponding input and state sequences associated with the initial state $x_0$ be denoted as $(u_0^*, \ldots, u_{n-1}^*)$

and $(x_0^*, \dots, x_n^*)$. If $\pi_k$ is twice continuously differentiable in a neighborhood of $x_k^*$ and

$$\nabla_u^2 Q_k^\pi(x_k^*, u_k^*) \succ 0, \quad \forall k \in \{0, \dots, n-1\},$$

then $(u_0^*, \dots, u_{n-1}^*)$ is also a (spurious) local minimizer of the one-shot problem.

*Proof:* First, we will use induction to find positive numbers $\delta_0, \dots, \delta_n$ and $\epsilon_0, \dots, \epsilon_{n-1}$ such that

$$\nabla_u^2 Q_k^\pi(x, u) \succ 0, \tag{7}$$
$$\pi_k(x) \in B(u_k^*, \epsilon_k), \tag{8}$$
$$f_k(x, u) \in B(x_{k+1}^*, \delta_{k+1}), \tag{9}$$

for every $x \in B(x_k^*, \delta_k)$, $u \in B(u_k^*, \epsilon_k) \cap A$, and $k \in \{0, \dots, n-1\}$. At the base step $k = n$, we choose an arbitrary $\delta_n > 0$. At the induction step, since $f_k$ is continuous and $\nabla_u^2 Q_k^\pi$ is continuous at $(x_k^*, u_k^*)$, there exist $\delta_k > 0$ and $\epsilon_k > 0$ such that both (7) and (9) are satisfied for all $x \in B(x_k^*, \delta_k)$ and $u \in B(u_k^*, \epsilon_k) \cap A$. Moreover, as $\pi_k$ is continuous at $x_k^*$, (8) will be satisfied by further reducing $\delta_k$.

For every $(\tilde{u}_0, \dots, \tilde{u}_{n-1})$ with $\tilde{u}_k \in B(u_k^*, \epsilon_k) \cap A$, let $(\tilde{x}_0, \dots, \tilde{x}_n)$ be its corresponding state sequence (note that $\tilde{x}_0 = x_0$). It follows from (9) that

$$\tilde{x}_k \in B(x_k^*, \delta_k), \quad \forall k \in \{0, \dots, n-1\},$$

which together with (8) implies that

$$\pi_k(\tilde{x}_k) \in B(u_k^*, \epsilon_k), \quad \forall k \in \{0, \dots, n-1\}.$$

In light of (7), $Q_k^\pi(\tilde{x}_k, \cdot)$ is a convex function on the convex set $B(u_k^*, \epsilon_k) \cap A$. Because $\pi_k(\tilde{x}_k) \in B(u_k^*, \epsilon_k) \cap A$ is a local minimizer of the function $Q_k^\pi(\tilde{x}_k, \cdot)$, it must be a global minimizer of this function over $B(u_k^*, \epsilon_k) \cap A$. Thus, for $k \in \{0, \dots, n-1\}$, we have

$$\begin{aligned} c_k(\tilde{x}_k, \tilde{u}_k) + J_{k+1}^\pi(\tilde{x}_{k+1}) &= Q_k^\pi(\tilde{x}_k, \tilde{u}_k) \\ &\geq Q_k^\pi(\tilde{x}_k, \pi_k(\tilde{x}_k)) \\ &= J_k^\pi(\tilde{x}_k). \end{aligned}$$

By adding all of the above inequalities, one can obtain

$$C(x_0; \tilde{u}_0, \dots, \tilde{u}_{n-1}) \geq J_0^\pi(x_0) = C(x_0; u_0^*, \dots, u_{n-1}^*),$$

which shows that $(u_0^*, \dots, u_{n-1}^*)$ is a local minimizer of the one-shot problem. Furthermore, if $\pi$ is a spurious locally minimum control policy of DP, namely, $J_0^\pi(x_0) > J_0(x_0^*)$, then

$$C(x_0; u_0^*, \dots, u_{n-1}^*) = J_0^\pi(x_0) > J_0(x_0).$$

As a result, $(u_0^*, \dots, u_{n-1}^*)$ is also a spurious local minimizer of the one-shot problem. ■

By taking the contrapositive, one can immediately conclude that DP has no spurious locally minimum control policies that satisfy the regularity conditions in either Theorem 2 or Theorem 3 as long as the one-shot problem has no spurious local minima.

### *C. Stationary point: From DP to one-shot optimization*

In Section III-B, we have mentioned that if the assumptions of Theorem 2 or Theorem 3 are not satisfied, an induced controlled input of the locally minimum control policy of DP does not necessarily imply a local minimizer of the one-shot problem. Therefore, it is desirable to discover what property such induced input satisfies for the one-shot optimization problem. In this subsection, we will show that under some conditions the control input induced by a locally minimum control policy of DP is also a stationary point of the one-shot problem.

*Definition 4:* Given a set $S$ and a continuously differentiable function $g$, a point $s^* \in S$ is said to be a stationary point of the optimization problem $\min_{s \in S} g(s)$ if

$$\nabla_s g(s^*) \in \mathcal{N}_S(s^*),$$

where $\mathcal{N}_S(s^*)$ denotes the normal cone of the set $S$ at the point $s^*$ [21].

In the following theorem and its proof, we will regard the gradient of a scalar function as a row vector. Let $\mathbf{D}_k^\pi(x)$ be the Jacobian matrix of $\pi_k(\cdot)$ at point $x$, $\mathbf{D}_k^{f,x}(x, u)$ be the Jacobian matrix of the function $f_k(\cdot, u)$ at point $x$ while viewing $u$ as a constant, and similarly $\mathbf{D}_k^{f,u}(x, u)$ be the Jacobian matrix of $f_k(x, \cdot)$ at point $u$ while viewing $x$ as a constant.

*Theorem 4:* Consider a locally minimum control policy $\pi = (\pi_0, \dots, \pi_{n-1})$, and let the corresponding input and state sequences associated with the initial state $x_0$ be denoted as $(u_0^*, \dots, u_{n-1}^*)$ and $(x_0^*, \dots, x_n^*)$. If for every $k \in \{0, \dots, n-1\}$:

1) $\pi_k$ is continuously differentiable in a neighborhood of $x_k^*$;
2) either $\pi_k(x_k^*)$ is in the interior of $A$ or $\mathbf{D}_k^\pi(x_k^*) = 0$,

then $(u_0^*, \dots, u_{n-1}^*)$ is a stationary point of the one-shot problem satisfying the first-order necessary optimality condition.

*Proof:* First, we will apply induction to prove that

$$\nabla_x J_k^\pi(x_k^*) = \nabla_x C(x_k^*; u_k^*, \dots, u_{n-1}^*) \tag{10}$$

holds for $k \in \{0, \dots, n\}$. The base step $k = n$ is obvious. For the induction step, observe that

$$\begin{aligned} \nabla_x Q_k^\pi(x, u) &= \nabla_x c_k(x, u) + \nabla_x J_{k+1}^\pi(f_k(x, u)) \mathbf{D}_k^{f,x}(x, u), \\ \nabla_x J_k^\pi(x) &= \nabla_x [Q_k^\pi(x, \pi_k(x))] \\ &= \nabla_x Q_k^\pi(x, \pi_k(x)) + \nabla_u Q_k^\pi(x, \pi_k(x)) \mathbf{D}_k^\pi(x). \end{aligned}$$

Therefore,

$$\begin{aligned} \nabla_x J_k^\pi(x_k^*) =& \nabla_x c_k(x_k^*, u_k^*) \\ &+ \nabla_x J_{k+1}^\pi(x_{k+1}^*) \mathbf{D}_k^{f,x}(x_k^*, u_k^*) \\ &+ \nabla_u Q_k^\pi(x_k^*, u_k^*) \mathbf{D}_k^\pi(x_k^*). \end{aligned} \tag{11}$$

If $u_k^*$ is in the interior of $A$, since $u_k^*$ is a local minimizer of $Q_k^\pi(x_k^*, \cdot)$, we have $\nabla_u Q_k^\pi(x_k^*, u_k^*) = 0$. Otherwise, by the assumption, it holds that $\mathbf{D}_k^{\pi}(x_k^*) = 0$. In either case, the last term of (11) is zero. On the other hand,

$$
\begin{aligned}
&\nabla_x C(x; u_k^*, \ldots, u_{n-1}^*) \\
&\quad = \nabla_x c_k(x, u_k^*) + \nabla_x [C(f_k(x, u_k^*); u_{k+1}^*, \ldots, u_{n-1}^*)] \\
&\quad = \nabla_x c_k(x, u_k^*) + \nabla_x C(f_k(x, u_k^*); u_{k+1}^*, \ldots, u_{n-1}^*) \\
&\qquad \mathbf{D}_k^{f,x}(x, u_k^*).
\end{aligned}
$$

Now, (10) can be obtained by taking $x = x_k^*$ in the above equality and then combine it with the induction hypothesis and (11). Finally, for $k \in \{0, \ldots, n-1\}$, one can write

$$
\begin{aligned}
\nabla_{u_k} C(x_0; u_0^*, \ldots, u_{n-1}^*) &= \nabla_u c_k(x_k^*, u_k^*) \\
&\quad + \nabla_x C(x_{k+1}^*; u_{k+1}^*, \ldots, u_{n-1}^*) \mathbf{D}_k^{f,u}(x_k^*, u_k^*) \\
&= \nabla_u c_k(x_k^*, u_k^*) + \nabla_x J_{k+1}^{\pi}(x_{k+1}^*) \mathbf{D}_k^{f,u}(x_k^*, u_k^*) \\
&= \nabla_u Q_k^{\pi}(x_k^*, u_k^*),
\end{aligned}
$$

in which the last second equality is due to (10). Since $u_k^*$ is a local minimizer of $Q_k^{\pi}(x_k^*, \cdot)$, $\nabla_u Q_k^{\pi}(x_k^*, u_k^*) \in \mathcal{N}_A(u_k^*)$, which shows that $(u_0^*, \ldots, u_{n-1}^*)$ is a stationary point of the one-shot problem. ■

## IV. Numerical Examples

To be able to effectively demonstrate the results of this paper through visualization, we will provide two low-dimensional examples in this section.

*Example 1:* Consider an optimal control problem with the control constraint $A = [-10, 10]$ and

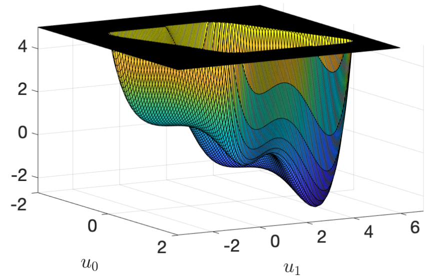
$$
\begin{aligned}
c_0(x, u) &= 0, \\
c_1(x, u) &= \frac{1}{4}u^4 - \frac{3x+4}{3}u^3 + \frac{3x^2+8x+3}{2}u^2 \\
&\quad - x(x+1)(x+3)u + \exp(x^4), \\
c_2(x) &= 0, \;\; f_0(x, u) = x + u, \;\; f_1(x, u) = x + u.
\end{aligned}
$$

At the initial state $x_0 = 0$, the one-shot problem can be written as

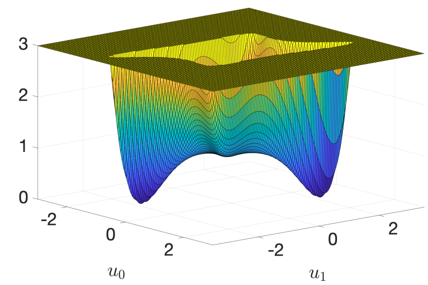
$$
\begin{aligned}
\min_{u_0 \in A, u_1 \in A} \Big\{ &\frac{1}{4}u_1^4 - \frac{3u_0+4}{3}u_1^3 + \frac{3u_0^2+8u_0+3}{2}u_1^2 \\
&- u_0(u_0+1)(u_0+3)u_1 + \exp(u_0^4) \Big\}.
\end{aligned}
$$

This one-shot optimization problem has 3 spurious local minimizers $(-0.523, -0.523), (-0.523, 2.477)$, $(0.938, 0.938)$ and the globally optimal minimizer $(0.938, 3.938)$. The landscape of this objective function is shown in Fig. 1(a). The optimal control problem can also be solved sequentially by DP. At the time step 1, the Q-function is $Q_1^{\pi}(x, u_1) = c_1(x, u_1)$, which has the maximum point $x+1$, the spurious local minimizer $x$ and the global minimizer $x+3$. One can choose a continuous policy

$$
\pi_1(x) = \begin{cases} -10, & x < -10, \\ x, & -10 \le x \le 10, \\ 10, & x > 10, \end{cases}
$$




(a) Example 1




(b) Example 2

Fig. 1: Landscape of the one-shot optimization.

whose associated cost-to-go function is

$$
J_1^{\pi}(x) = \begin{cases} g(-10), & x < -10, \\ g(x), & -10 \le x \le 10, \\ g(10), & x > 10, \end{cases}
$$

where $g(x) = -\frac{1}{12}(3x^4 + 16x^3 + 18x^2) + \exp(x^4)$. At the time step $i = 0$ and at the initial state $x_0 = 0$, the Q-function is $Q_1^{\pi}(0, u_0) = J_1^{\pi}(u_0)$, which has a spurious local minimizer at $-0.523$ and a global minimum at $0.938$. If we choose $\pi_0(0) = -0.523$, then the induced input under $\pi$ of DP is $(-0.523, -0.523)$ and if we choose $\pi_0(0) = 0.938$, then the induced input under $\pi$ of DP is $(0.938, 0.938)$. Both of these input sequences are spurious local minimizers of the one-shot problem.

One can also choose

$$
\pi_1(x) = \begin{cases} -10, & x < -13, \\ x+3, & -13 \le x \le 7, \\ 10, & x > 7, \end{cases}
$$

whose associated cost-to-go function is

$$
J_1^{\pi}(x) = \begin{cases} g(-10), & x < -13, \\ g(x), & -13 \le x \le 7, \\ g(10), & x > 7, \end{cases}
$$

where $g(x) = -\frac{1}{12}(3x^4 + 16x^3 + 18x^2 + 27) + \exp(x^4)$. At the time step $i = 0$ and at the initial state $x_0 = 0$, the Q-function is $Q_1^{\pi}(0, u_0) = J_1^{\pi}(u_0)$, which has a spurious local minimizer at $-0.523$ and a global minimum at $0.938$. If we choose $\pi_0(0) = 0.938$, then the locally minimum control policy $\pi$ is non-spurious and its induced input $(0.938, 3.938)$ is the global minimizer of the one-shot problem. However, if we choose $\pi_0(0) = -0.523$, then the locally minimum control policy $\pi$ is spurious and its induced input $(-0.523, 2.477)$ is the spurious minimizer of the one-shot problem.

In this example, one can observe that each strictly local minimizer of the one-shot problem corresponds to a locally minimum control policy of DP, which validates the result of Theorem 1. In addition, it can be noticed that since the minimizer of $Q_1^\pi(x,\cdot)$ does not bifurcate with $x$, Theorem 2 also holds.

*Example 2:* Consider the problem in Example 1 but change $c_1(x,u)$ to $\frac{1}{4}u^4 - \frac{x}{3}u^3 - x^2u^2 + \exp(x^4)$. At the initial state $x_0 = 0$, the one-shot problem can be written as

$$\min_{u_0\in A, u_1\in A} \left\{\frac{1}{4}u_1^4 - \frac{u_0}{3}u_1^3 - u_0^2u_1^2 + \exp(u_0^4)\right\}.$$

It has 3 stationary points $(0,0)$ and $((\log(\frac{8}{3}))^{\frac{1}{4}}, 2(\log(\frac{8}{3}))^{\frac{1}{4}})$ and $(-(\log(\frac{8}{3}))^{\frac{1}{4}}, -2(\log(\frac{8}{3}))^{\frac{1}{4}})$. The later two are the global minimizers of this one-shot problem. To understand why $(0,0)$ is not a local minimizer of the one-shot problem, we take $u_0 = u_1 = \epsilon$ and use the Taylor expansion of the exponential function to arrive at:

$$\frac{1}{4}\epsilon^4 - \frac{1}{3}\epsilon^4 - \epsilon^4 + \exp(\epsilon^4) = -\frac{1}{12}\epsilon^4 + 1 + o(\epsilon^4),$$

which is strictly less that 1 for sufficiently small values of $\epsilon$. This implies that $(0,0)$ is not a locally optimal solution of the one-shot problem. The landscape of this objective function is shown in Fig. 1(b). It can also be solved sequentially by DP. For the initial state $x_0$, it has 3 different induced input sequences under the locally minimum control policy: $(\log(\frac{8}{3}))^{\frac{1}{4}}, 2(\log(\frac{8}{3}))^{\frac{1}{4}})$, $(-(\log(\frac{8}{3}))^{\frac{1}{4}}, -2(\log(\frac{8}{3}))^{\frac{1}{4}})$ and $(0,0)$. The first two points are the global minimizers of the one-shot problem but $(0,0)$ is not a local minimizer of the one-shot problem.

In this example, the Q-function $Q_1^\pi(x,\cdot) = c_1(x,\cdot)$ has 3 stationary points $0, -x, 2x$ and all 3 points will merge to a single point when $x = 0$ and $\nabla_u^2 Q_1(0,0) = 0$. Therefore, the assumptions in Theorem 2 and Theorem 3 are violated, and $(0,0)$ is not a local minimizer of the one-shot problem. This clarifies the role of the regularity conditions needed in those theorems. On the other hand, consistent with Theorem 4, $(0,0)$ is a saddle point (which is a stationary point) of the one-shot optimization.

## V. Conclusions

In this paper, we study the (spurious) local solutions of arbitrary optimal control problems through two different formulations: one-shot (single) optimization problem aimed at solving for all input values at the same time, and DP method aimed at finding the input values sequentially. We introduce the notions of spurious (non-global) local minimizers for the one-shot problem and spurious locally minimum control policies for DP. We prove that under some mild conditions, each local minimizer of the one-shot optimization corresponds to an input sequence induced by some locally minimum control policy of DP and vice versa. We also prove that if the control sequence induced by a policy satisfies the first-order optimality condition for DP, then it also satisfies the first-order necessary optimality condition for the one-shot optimization problem. This is the first result in the literature on the connection between the spurious solutions of the one-shot and DP methods. A natural future direction would be to extend this work to stochastic dynamics under a parameterized policy, which helps better understand the quality of the local solutions obtained by reinforcement learning algorithms.

## References


[1] R. E. Bellman, *Dynamic programming*. Princeton university press, 1957.

[2] D. P. Bertsekas, *Dynamic programming and optimal control*. Athena scientific Belmont, MA, 1995, vol. 1, no. 2.

[3] R. Bellman and R. Kalaba, "On the role of dynamic programming in statistical communication theory," *IRE Transactions on Information Theory*, vol. 3, no. 3, pp. 197–203, 1957.

[4] E. A. Feinberg, "Optimality conditions for inventory control," in *Optimization Challenges in Complex, Networked and Risky Systems*. INFORMS, 2016, pp. 14–45.

[5] I. Kolmanovsky, I. Siverguina, and B. Lygoe, "Optimization of powertrain operating policy for feasibility assessment and calibration: Stochastic dynamic programming approach," in *American Control Conference (ACC)*, vol. 2. IEEE, 2002, pp. 1425–1430.

[6] D. P. Bertsekas, *Reinforcement learning and optimal control*. Athena Scientific Belmont, MA, 2019.

[7] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[8] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[9] C. Josz, Y. Ouyang, R. Zhang, J. Lavaei, and S. Sojoudi, "A theory on the absence of spurious solutions for nonconvex and nonsmooth optimization," in *Advances in neural information processing systems*, 2018, pp. 2441–2449.

[10] M. Soltanolkotabi, A. Javanmard, and J. D. Lee, "Theoretical insights into the optimization landscape of over-parameterized shallow neural networks," *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 742–769, 2018.

[11] Z. Allen-Zhu, Y. Li, and Z. Song, "A convergence theory for deep learning via over-parameterization," in *International Conference on Machine Learning*. PMLR, 2019, pp. 242–252.

[12] S. Mei, Y. Bai, and A. Montanari, "The landscape of empirical risk for non-convex losses," *arXiv preprint arXiv:1607.06534*, 2016.

[13] R. Y. Zhang, S. Sojoudi, and J. Lavaei, "Sharp restricted isometry bounds for the inexistence of spurious local minima in nonconvex matrix recovery." *Journal of Machine Learning Research*, vol. 20, no. 114, pp. 1–34, 2019.

[14] Y. Chen, Y. Chi, J. Fan, and C. Ma, "Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval," *Mathematical Programming*, vol. 176, no. 1-2, pp. 5–37, 2019.

[15] I. Molybog, S. Sojoudi, and J. Lavaei. (2019) Role of sparsity and structure in the optimization landscape of non-convex matrix sensing. [Online]. Available: https://lavaei.ieor.berkeley.edu/KSP_2019_1.pdf

[16] S. Fattahi, C. Josz, R. Mohammadi, J. Lavaei, and S. Sojoudi, "Absence of spurious local trajectories in time-varying optimization: A control-theoretic perspective," in *Conference on Control Technology and Applications (CCTA)*. IEEE, 2020.

[17] Q. Cai, Z. Yang, J. D. Lee, and Z. Wang, "Neural temporal-difference learning converges to global optima," in *Advances in Neural Information Processing Systems*, 2019, pp. 11 315–11 326.

[18] B. Liu, Q. Cai, Z. Yang, and Z. Wang, "Neural trust region/proximal policy optimization attains globally optimal policy," in *Advances in Neural Information Processing Systems*, 2019, pp. 10 565–10 576.

[19] J. Bhandari and D. Russo, "Global optimality guarantees for policy gradient methods," *arXiv preprint arXiv:1906.01786*, 2019.

[20] C. Berge, *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*. Courier Corporation, 1997.

[21] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*. Springer Science & Business Media, 2009, vol. 317.