Escaping spurious local minimum trajectories in online time-varying nonconvex optimization

Yuhao Ding, Javad Lavaei, and Murat Arcak *

Abstract

Devising efficient online algorithms that can track the optimizers of time-varying nonconvex optimization problems has many real-world applications. A major limitation of the existing online tracking methods is that they only focus on tracking a specific local minimum trajectory, which may lead to finding poor spurious local solutions. In this paper, we study the role of the natural temporal variation in helping simple online tracking methods find and track time-varying global minima for online nonconvex optimization problems. To this end, we investigate the properties of a time-varying gradient flow system with inertia, which can be regarded as the continuous-time limit of (1) the stationary condition for a discretized sequential optimization problem with a proximal regularization and (2) the online tracking scheme. We show that the inherent temporal variation of a time-varying optimization problem could re-shape the landscape by making it one point strongly convex over a large region during some time interval. Sufficient conditions are derived to guarantee that no matter how a local search method is initialized, it will track a time-varying global solution after some time. The results are illustrated in a benchmark example with many local minima.

1 Introduction

In this paper, we study an unconstrained online optimization problem whose objective function varies continuously in time, namely,

$$\min_{x(t) \in \mathbb{R}^n} f(x(t), t)$$

(1)

where $t \geq 0$ denotes the time and $x(t)$ is the optimization variable that depends on $t$. For each time $t$, the function $f(x, t)$ could potentially be nonconvex in $x$ with many local minima. The objective is to solve the above problem in an online fashion under the assumption that at any given time $t$ the function $f(x, t)$ is known while no knowledge about $f(x, t')$ is available for any $t' > t$. The optimization at each time instance could be highly complex due to NP-hardness, which is an impediment to finding its global minima. This paper aims to investigate under what conditions simple local search algorithms can solve the above online optimization problem to almost global optimality after some finite time. More precisely, consider a global solution of $f(x, t)$ as a function of time $t$, which defines a global minimum solution trajectory. It is desirable to devise an algorithm that can track this trajectory with some error at the initial time and a diminishing error after some time.

If $f(x, t)$ does not change over time, the problem reduces to a classic (time-invariant) optimization problem. It is known that simple local search methods, such as stochastic gradient descent (SGD) [19], may be able to find a global minimum of the problem (under certain conditions) for almost all initializations due to the randomness embedded in SGD [23, 14, 27]. The objective of this paper is to significantly extend

---

*The authors are with the University of California, Berkeley (e-mail: yuhao_ding@berkeley.edu; lavaei@berkeley.edu; arccak@berkeley.edu).

†This work was supported by grants from ARO, ONR, AFOSR, and NSF.
the above result from a single optimization problem to infinitely-many problems parametrized by time \( t \). In other words, it is desirable to investigate the following question: **Can the temporal variation in the landscape of time-varying nonconvex optimization problems enable online local search methods to track global trajectories?** To answer this question, we study a first-order time-varying ordinary differential equation (ODE), which is the counterpart of the classic gradient flow system for the time-invariant optimization problem and serves as a continuous-time limit of the discrete online tracking method for (1). This ODE is given as

\[
\dot{x} = -\frac{1}{\alpha} \nabla_x f(x, t), \quad x(0) = x_0 \tag{ODE}
\]

where \( \alpha > 0 \) is a constant parameter named **inertia** due to a **proximal regularization**. A system of the form (ODE) is called a **time-varying gradient system with inertia** \( \alpha \). The behavior of the solutions of this system initialized at different points depends on the value of \( \alpha \). We offer a motivating example below before stating the goals of this paper.

### 1.1 Motivating example

**Example 1.** Consider the objective function \( f(x, t) := g(x - b \sin(t)) \), where

\[
g(y) := \frac{1}{4}y^4 + \frac{2}{3}y^3 - \frac{1}{2}y^2 - 2y
\]

The differential equation (ODE) can be written as

\[
\dot{x} = -\frac{1}{\alpha} \left( (x - b \sin(t))^3 + 2(x - b \sin(t))^2 - (x - b \sin(t)) - 2 \right)
\]

This time-varying objective has a spurious (non-global) local minimum trajectory at \(-2 + b \sin(t)\), a local maximum at \(-1 + b \sin(t)\), and a global minimum at \(1 + b \sin(t)\). In Figure 1\(^1\), we show a bifurcation phenomenon numerically. The red lines are the solutions of (ODE). In the case with \( \alpha = 0.3 \) and \( b = 5 \), a trajectory initialized at a local minimum of \( f(\cdot, 0) \) winds up in the region of attraction of the global minimum trajectory. However, for the case with \( \alpha = 0.1 \) and \( b = 5 \), a trajectory initialized at a local minimum of \( f(\cdot, 0) \) remains in the region of attraction of the same local minimum trajectory. In Figure 1c, we study the scenario with \( \alpha = 0.8 \) and \( b = 5 \), and initialize the trajectory at the spurious local minimum of \( f(\cdot, 0) \). The solution of (ODE) fails to track any local minimum trajectory. In Figure 1d, we take \( \alpha = 0.1, b = 10 \) and initialize the trajectory at the spurious local minimum of \( f(\cdot, 0) \). The solution of (ODE) again winds up in the region of attraction of the global minimum trajectory. The observations in this example can be summarized as:

1. Jumping from one trajectory to another trajectory tends to occur with the help of a relatively large inertia when the local minimum trajectory changes the direction abruptly and there happens to exist another local minimum trajectory in the direction of the inertia.

2. When the inertia \( \alpha \) is relatively small, the solution of (ODE) tends to track a local minimum trajectory closely and converges to a local minimum trajectory quickly.

### 1.2 Our contributions

To mathematically study the observations made in Example 1 for a general online optimization problem, we focus on the aforementioned time-varying gradient flow system with inertia \( \alpha \) as a continuous-time limit.

\footnote{In order to increase visibility, the objective function values are rescaled.}
of an online updating scheme for (1). We show that the time-varying gradient flow system with inertia $\alpha$ is also a continuous-time limit of the stationary condition for a discretized sequential optimization problem with a proximal regularization. The existence and uniqueness of the solution of such ODE is proven.

As a main result of this work, it is proved that the natural temporal variation of the time-varying optimization problem encourages the exploration of the state space and re-shaping the landscape by making it one point strongly convex over a large region during some time interval. We show that if a given spurious local minimum trajectory is a shallow minimum trajectory compared to the global minimum trajectory, then the temporal variation of the time-varying optimization would trigger escaping the spurious local minimum trajectory for free. We develop two sufficient conditions under which a certain local minimum trajectory will jump to another local minimum trajectory. We then derive a sufficient condition on the inertia $\alpha$ to guarantee that the solution of (ODE) can track the local minimum trajectory from an arbitrary point in the region of locally one point strong convexity around this local minimum trajectory. We also provide an ultimate tracking error bound and estimate the time for reaching the tracking error bound. To illustrate the technical results on how the time variation nature of an online optimization problem enables escaping a spurious minimum trajectory, we offer a case study with many shallow minimum trajectories.
1.3 Related work

**Online time-varying optimization problems**: Time-varying optimization problems of the form (1) arise in the real-time optimal power flow problem [43] for which the power loads and renewable generations are time-varying and operational decisions should be made every 5 minutes, as well as in the real-time estimation of the state of a nonlinear dynamic system [35]. Other examples include model predictive control [7], time-varying compressive sensing [36, 3] and online economic optimization [25, 46]. There are many papers on designing efficient online algorithms to track the optimizers of time-varying convex optimization problems [39, 13, 4, 38]. With respect to time-varying nonconvex optimization problems, the work [16] presents a comprehensive theory on the structure and singularity of the KKT trajectories for time-varying optimization problems. On the algorithm side, [43] provides regret-type results in the case where the constraints are lifted to the objective function via penalty functions. [42] develops a running regularized primal-dual gradient algorithm to track a KKT trajectory, and offers asymptotic bounds on the tracking error. [32] obtains an ODE to approximate the KKT trajectory and derives an algorithm based on a predictor-corrector method to track the ODE solution. Recently, [11] asked the question of whether the natural temporal variation in a time-varying nonconvex optimization problem could help a local tracking method escape spurious local minimum trajectories, but lacked theoretical results on this phenomenon.

**Online optimization for machine learning**: A common framework in machine learning for analyzing a time-varying optimization problem is online optimization [19]. In general, the main goal in such online convex optimization is to propose a sequential algorithm and measure its performance through the notion of stationary regret [48] or dynamic regret [48, 5, 22], depending on whether there is any additional condition on the temporal variability of the sequence of objective functions. Most of the literature have focused on online convex optimization, while the main challenges are in the non-convex case. With respect to online nonconvex optimization, [20] proposes to minimize a surrogate notion of local regret, which measures the sub-optimality compared to a local point-wise solution to the problem. Contrary to this line of research, we focus on the global landscape and the ultimate tracking error bound of time-varying nonconvex optimization problems.

**Local search methods for global optimization**: Nonconvexity is inherent in many real-world problems: the classical compressive sensing and matrix completion/sensing [10, 8, 9], training of deep neural networks [31], the optimal power flow problem [29], and others. From the classical complexity theory, this nonconvexity is perceived to be the main contributor to the intractability of these problems. However, it has been recently shown that simple local search methods, such as gradient-based algorithms, have a superb performance in solving nonconvex optimization problems. For example, [30] shows that the gradient descent with a random initialization could avoid the saddle points almost surely and [23, 14] proves that a perturbed gradient descent and SGD could escape the saddle points efficiently. Furthermore, it has been shown that nearly-isotropic classes of problems in matrix completion/sensing [6, 15, 47], robust principle component analysis [12, 24], and dictionary recovery [41] have benign landscape, implying that they are free of spurious local minima (a non-global local minimum is called spurious). The work [27] proves that SGD could help escape sharp local minima of a loss function by taking the alternative view that SGD works on a convolved (thus smoothed) version of the loss function. However, these results are all for time-invariant optimization problems for which the landscape is time-invariant. In contrast, many real-world problems should be solved sequentially over time with time-varying data. Therefore, it is essential to study the effect of the temporal variation on the landscape of time-varying nonconvex optimization problems.

**Continuous-time interpretation of discrete numerical algorithms**: Many iterative numerical optimization algorithms for time-invariant optimization problems can be interpreted as a discretization of a continuous-time process. Then, several new insights have been obtained due to the known results for continuous-time dynamical systems [26, 18]. Perhaps the simplest and oldest example is the gradient flow.
system for the gradient descent algorithm with an infinitesimally small step size. The recent papers [40, 28, 45] study accelerated gradient methods for convex optimization problems from a continuous-time perspective. In addition, the continuous-time limit of the gradient descent is also employed to analyze various non-convex optimization problems, such as deep linear neural networks [37] and matrix regression [17]. It is natural to analyze the continuous-time limit of an online algorithm for tracking a KKT trajectory of time-varying optimization problem [39, 42, 32, 11].

1.4 Paper organization

This paper is organized as follows. Section 2 presents the preliminaries for time-varying optimization and the derivation of time-varying gradient flow with inertia. Section 3 offers an alternative view on the landscape of time-varying nonconvex optimization problems after a change of variables. Section 4 analyzes the jumping, tracking and escaping behaviors of local minimum trajectories. Section 5 illustrates the phenomenon that the time variation of an online optimization problem can assist with escaping spurious local minimum trajectories, by working on a benchmark example with many shallow minimum trajectories. Concluding remarks are given in Section 6. To streamline the presentation, the proofs are deferred to the appendix.

1.5 Notations

The notation \( \| \cdot \| \) shows the Euclidean norm. The interior of the interval \( \bar{I}_t, 2 \) is denoted by \( \text{int}(\bar{I}_t, 2) \). The symbol \( B_r(h(t)) = \{ x \in \mathbb{R}^n : \| x - h(t) \| \leq r \} \) denotes the region centered around a trajectory \( h(t) \) with radius \( r \) at time \( t \). We denote the solution of \( \dot{x} = f(x, t) \) starting from \( x_0 \) at the initial time \( t_0 \) with \( x(t, t_0, x_0) \) or the short-hand notation \( x(t) \) if the initial condition \((t_0, x_0)\) is clear from the context.

2 Preliminaries and Problem Formulation

In this work, we assume that \( f : \mathbb{R}^n \times [0, \infty) \to \mathbb{R} \) is twice continuously differentiable in \( x \) and continuously differentiable in \( t \geq 0 \). Moreover, suppose that \( f \) is uniformly bounded from below, meaning that there exists a constant \( M \) such that \( f(x, t) \geq M \) for all \( x \in \mathbb{R}^n \) and \( t \geq 0 \).

2.1 Time-varying optimization

The first-order stationary condition for the time-varying optimization (1) is as follows:

\[
0 = \nabla_x f(x(t), t)
\]  

(2)

Since the solution is time-varying, we define the notion of stationary trajectories below.

Definition 1. Given a time interval \( I_t \subseteq [0, \infty) \), a continuous trajectory \( h(t) : I_t \to \mathbb{R}^n \) is said to be a stationary trajectory of the time-varying optimization (1) if \( 0 = \nabla_x f(h(t), t) \) for all \( t \in I_t \).

In this work, we assume that the real roots of (2) are all isolated at each time \( t \).

Definition 2. A stationary trajectory \( h(t) : I_t \to \mathbb{R}^n \) is said to be isolated if, given any other stationary trajectory \( h'(t) : I_t \to \mathbb{R}^n \), it holds that \( h(t) \cap h'(t) = \emptyset \) for all \( t \in I_t \).

An isolated stationary trajectory \( h(t) \) can theoretically be a mix of local minima, local maxima and saddle points of the function \( f(x, t) \) at different times. However, the goal of this work is to study only isolated local minimum trajectories of the time-varying optimization (1).
Definition 4. A continuous trajectory \( h(t) : I_t \rightarrow \mathbb{R}^n \) is said to be a **local \( I_t \)-minimum trajectory** of the time-varying optimization (1) if \( I_t \) is a maximal interval such that each point of \( h(t) \) is the local minimum of the time varying optimization (1) at time \( t \in I_t \). It is said to be a **global \( I_t \)-minimum trajectory** of the time-varying optimization (1) if \( I_t \) is a maximal interval such that each point \( h(t) \) is the global minimum of the time-varying optimization (1) at time \( t \in I_t \). In particular, \( h(t) \) is called a **local (or global) \( \infty \)-minimum trajectory** of the time-varying optimization (1) if \( I_t = [0, \infty) \).

After freezing the time \( t \) in (1) at a particular value, one may use local search methods to minimize \( f(x, t) \). If the initial point is close enough to a local solution and the step size is small enough, the algorithm will converge to the local minimum. This leads to the notion of region of attraction defined by resorting to the continuous-time model of local search algorithms (for which the step size is not important anymore).

**Definition 4.** The **region of attraction** of a local minimum \( h(t) \) of \( f(\cdot, t) \) is defined as:

\[
RA(h(t)) = \{ x_0 \in \mathbb{R}^n \mid \lim_{t \rightarrow \infty} x(\hat{t}) = h(t) \quad \text{where} \quad \frac{d\hat{x}(\hat{t})}{dt} = -\nabla_x f(\hat{x}(\hat{t}), t) \quad \text{and} \quad \hat{x}(0) = x_0 \} \tag{3}
\]

**Definition 5.** Consider arbitrary positive scalars \( c \) and \( r \) together with an interval \( \bar{I}_t \), where \( \bar{I}_t \subset I_t \) if \( I_t \) is finite and \( I_t = I_t = [0, \infty] \) otherwise. The function \( f(x, t) \) is said to be **locally \((\bar{I}_t, c, r)\)-one point strongly convex** around the local \( I_t \)-minimum trajectory \( h(t) \) if

\[
\nabla_x f(e + h(t), t)^\top e \geq c \| e \|^2, \quad \forall e \in D, \quad \forall t \in \bar{I}_t \tag{4}
\]

where \( D = \{ e \in \mathbb{R}^n : \| e \| \leq r \} \). The region \( D = \{ e \in \mathbb{R}^n : \| e \| \leq r \} \) is called the **region of locally \((\bar{I}_t, c, r)\)-one point strong convexity** around \( h(t) \).

Note that (4) resembles the (locally) strong convexity condition for the function \( f(x, t) \), but it is only expressed around the point \( h(t) \). This restriction to a single point constitutes the definition of one-point strong convexity and it does not imply that the function is convex. If the minimum trajectory \( h(t) \) is defined only over a finite maximal time interval \( I_t \), then when \( t \) approaches the upper bound of \( I_t \), the function \( f(x, t) \) may tend to become a saddle point or even local maximum solution. Then, one could expect that the one point strong convexity holds only for a subset of the interval \( I_t \subset I_t \). On the other hand, if the minimum trajectory \( h(t) \) is defined over the entire time horizon, the locally one point strong convexity could hold on the unbounded interval. Figure 2 illustrates the difference between the region of convexity and the region of one point convexity with respect to a certain point using the time-invariant univariate objective function given in example 1. In this paper, we assume any local \( I_t \)-minimum trajectory \( h(t) \) satisfies the following assumptions.

**Assumption 1.** \( h(t) \) is isolated.

**Assumption 2.** The time-varying function \( f(x, t) \) is locally one point \((\bar{I}_t, c, r)\)-strongly convex around \( h(t) \) for some constants \( c \) and \( r \) as well as an interval \( \bar{I}_t \).

**Assumption 3.** \( h(t) \) is continuously differentiable.

### 2.2 Derivation of time-varying gradient flow system

In practice, one can only hope to sequentially solve the time-varying optimization problem (1) at some discrete time instances \( 0 = \tau_0 < \tau_1 < \tau_2 < \tau_3 \cdots \) as follows:

\[
\min_{x \in \mathbb{R}^n} f(x, \tau_i), \quad i = 0, 1, 2, \ldots \tag{5}
\]
\( g(x) = 1/4x^4 + 2/3x^3 - 1/2x^2 - 2x \)

\( g'(x) = x^3 + 2x^2 - x - 2 \)

Figure 2: Illustration of local convexity and one point convexity for a time-invariant function: Points A and E are the local minimum solutions of \( g(x) \); point C is the local maximum solution of \( g(x) \); points B and D have a zero second derivative.

In many real-world applications, it is neither practical nor realistic to have solutions that abruptly change over time. To meet this requirement, one may impose a constraint to ensure that the solution at each time \( \tau_i \) will not be away from the one obtained in the previous time \( \tau_{i-1} \) by a given threshold \( d(\tau_i - \tau_{i-1}) > 0 \). This leads to a modified optimization problem at time \( \tau_i \):

\[
\min_{x \in \mathbb{R}^n} f(x, \tau_i) \\
\text{s.t. } \|x - x^*_{i-1}\| \leq d(\tau_i - \tau_{i-1})
\]  

(6)

where \( x^*_{i-1} \) denotes an arbitrary local minimum of the modified optimization problem obtained using a local search method at time iteration \( i - 1 \). An alternative way is to move the constraint in (6) to the objective function by penalizing the deviation of its solution from the one obtained in the previous time step. More precisely, we employ a regularization term that is proportional to a fixed inertia \( \alpha \geq 0 \) and inversely proportional to the time difference \( \tau_i - \tau_{i-1} \). This leads to the following sequence of optimization problems with \textbf{proximal regularization} (except for the initial optimization problem):

\[
\min_{x \in \mathbb{R}^n} f(x, \tau_0),
\]

(7a)

\[
\min_{x \in \mathbb{R}^n} f(x, \tau_i) + \frac{\alpha}{2(\tau_i - \tau_{i-1})} \|x - x^*_{i-1}\|^2, \quad i = 1, 2, \ldots
\]

(7b)

Note that \( \alpha \) could be time-varying (and adaptively changing) in the analysis of this paper, but we restrict our attention to a fixed regularization term to simplify the presentation. Now, one can formally define the notion of discrete local trajectories for the above sequential regularized optimization problem.

\textbf{Definition 6.} Given evenly spaced-out time steps \( 0 = \tau_0 < \tau_1 < \tau_2 < \cdots \), the sequence \( x^*_0, x^*_1, x^*_2, \ldots \) is said to be a \textbf{discrete local trajectory} of the sequential regularized optimization (7) if the following holds:

1. \( x_0 = x^*_0 \), where \( x^*_0 \) is a local minimum of (7a).

2. \( x^*_k \) is a local minimum of (7b) for \( k = 1, 2, \ldots \).
Due to the first-order optimality condition, the local minimum $x^*_i$ of (7) at time step $\tau_i$ satisfies the equation:

$$\nabla_x f(x^*_i, \tau_i) + \frac{x^*_i - x^*_{i-1}}{\tau_i - \tau_{i-1}} = 0$$

(8)

Since the function $f(x, \tau_i)$ is nonconvex, the problem (7b) may not have a unique solution $x^*_i$. In order to cope with this issue, we study the continuous-time limit of (8) as the time step $\tau_{i+1} - \tau_i$ attenuates to zero. This yields a time-varying ordinary differential equation:

$$\alpha \dot{x}(t) = -\nabla_x f(x(t), t), \quad x(0) = x^*_0$$

(9)

When $\alpha = 0$, the differential equation (9) reduces to the algebraic equation (2), which is indeed the first-order stationary condition for the unregularized time-varying optimization (1). When $\alpha > 0$, we will show that (9) has a unique solution defined for all $t \geq 0$ under the assumption that the solutions of (9) lies in a compact set.

**Theorem 1** (Existence and uniqueness). Assume that $f(x, t)$ is piecewise continuous in $t$, and that its gradient is locally Lipschitz in $x$ for all $t \geq 0$ and $x \in \mathbb{R}^n$. Let $\alpha > 0$ and $D$ be a compact subset of $\mathbb{R}^n$ containing $x_0$ such that every solution of

$$\dot{x} = -\frac{1}{\alpha} \nabla_x f(x(t), t), \quad x(0) = x^*_0$$

(ODE)

lies entirely in $D$. Then, the above differential equation has a unique solution and is defined for all $t \geq 0$.

Note that (8) can be written in the form of the backward Euler method:

$$x^*_i = x^*_{i-1} - \frac{\tau_i - \tau_{i-1}}{\alpha} \nabla_x f(x^*_i, \tau_i)$$

(10)

A direct application of the classical results on convergence of the backward Euler method [21] immediately shows that the solution of (ODE) starting at a local minimum of (7a) is the continuous limit of the discrete local trajectory of the sequential regularized optimization (7).

**Proposition 1** (Convergence of backward Euler). Given $x^*_0$ as a local minimum of (7a), as the time difference $\Delta \tau = \tau_{i+1} - \tau_i$ approaches zero, any sequence of discrete local trajectories $(x^*_k)$ introduced in Definition 6 converges to the (ODE) in the sense that

$$\lim_{\Delta \tau \to 0} \max_{0 \leq k \leq T} \left\| x^*_k - x(\tau_k, x^*_0) \right\| = 0$$

(11)

for all fixed $T > 0$.

In online optimization, it is desirable to predict the solution at a future time (namely, $\tau_i$) only based on the information at the current time (namely, $\tau_{i-1}$). This can be achieved by implementing the forward Euler method to obtain a numerical approximation to the solutions of (ODE):

$$\bar{x}^*_i = \bar{x}^*_{i-1} - \frac{\tau_i - \tau_{i-1}}{\alpha} \nabla_x f(\bar{x}^*_{i-1}, \tau_{i-1})$$

(12)

(note that $\bar{x}^*_0, \bar{x}^*_1, \bar{x}^*_2, ...$ show the approximate solutions). Similar to Theorem 1, classic results on the convergence of the forward Euler method [21] imply that the solution of (ODE) starting at a local minimum of (7a) is the continuous limit of the sequence updated by (12).

---

2 Checking the compactness assumption can be done via the Lyapunov's method without solving the differential equation.
Proposition 2 (Convergence of forward Euler). Given $\bar{x}_0^*$ as a local minimum of (7a), as the time difference $\Delta_r = \tau_{i+1} - \tau_i$ approaches zero, any sequence of $(\bar{x}_k^\Delta)$ updated by (12) converges to the (ODE) in the sense that
\[
\lim_{\Delta_r \to 0} \max_{0 \leq k \leq \frac{T}{\Delta_r}} \|\bar{x}_k^\Delta - x(\tau_k, \tau_0, x_0^*)\| = 0
\]
for all fixed $T > 0$.

Propositions 1 and 2 together guarantee that the solution of (ODE) is a reasonable approximation in the sense that it is the continuous-time limit of both the solution of the sequential regularized optimization problem (7) and the solution of the online updating scheme (12). For this reason, we only study the continuous-time problem (ODE) in the remainder of this paper.

2.3 Jumping, tracking and escaping

In this section, the objective is to study the case where there are at least two local minima at different time instances of the online optimization problem. Consider a local $I_{t,1}$-minimum trajectory $h_1(t)$ and a local $I_{t,2}$-minimum trajectory $h_2(t)$. Suppose that the time-varying function $f(x, t)$ is locally $(\bar{I}_{t,1}, c_1, r_1)$-one point strongly convex around $h_1(t)$ and locally $(\bar{I}_{t,2}, c_2, r_2)$-one point strongly convex around $h_2(t)$. Let $[t_1, t_2] \subset \bar{I}_{t,1} \cap \bar{I}_{t,2}$ be a non-empty interval. We provide the definitions of jumping, tracking and escaping below.

Definition 7. It is said that the solution of (ODE) $(v,u)$-jumps from $h_1(t)$ to $h_2(t)$ over the time interval $[t_1, t_2]$ if there exist $u > 0$ and $v > 0$ such that
\[
\begin{align*}
B_v(h_1(t_1)) &\subseteq RA(h_1(t_1)) \quad (14a) \\
B_u(h_2(t_2)) &\subseteq RA(h_2(t_2)) \quad (14b) \\
\forall x_1 \in B_u(h_1(t_1)) &\Rightarrow x(t_2, t_1, x_1) \in B_u(h_2(t_2)) \quad (14c)
\end{align*}
\]

If $I_{t,2}$ is a finite time interval and $\bar{I}_{t,2} \subset I_{t,2}$, then $h_2(t)$ will disappear after the time interval $I_{t,2}$. In this case, one could only expect that the solution of (ODE) temporarily tracks $h_2(t)$.

Definition 8. It is said that $x(t, t_0, x_0)$ temporarily $u$-tracks $h_2(t)$ if there exist a constant $u > 0$ and a finite time $T$ in the interior of $\bar{I}_{t,2}$ and such that
\[
\begin{align*}
x(t, t_0, x_0) &\in B_u(h_2(t)), \quad \forall t \in \{ t \in \bar{I}_{t,2} : t \geq T \} \quad (15a) \\
B_u(h_2(t)) &\subseteq RA(h_2(t)), \quad \forall t \in \{ t \in \bar{I}_{t,2} : t \geq T \} \quad (15b)
\end{align*}
\]

The term “temporary” in the above definition emphasizes tracking over a possibly finite interval. The counterpart of this definition for an infinite interval will be given below.

Definition 9. It is said that $x(t, t_0, x_0)$ $u$-tracks $h_2(t)$ if there exist a finite time $T > 0$ and a constant $u > 0$ such that
\[
\begin{align*}
x(t, t_0, x_0) &\in B_u(h_2(t)), \quad \forall t \geq T \quad (16a) \\
B_u(h_2(t)) &\subseteq RA(h_2(t)), \quad \forall t \geq T
\end{align*}
\]

In this paper, the objective is to study the scenario where a solution $x(t, t_0, x_0)$ tracking a poor solution $h_1(t)$ at the beginning ends up tracking a better solution $h_2(t)$ after some time. This needs the notion of “escaping” introduced below.
Definition 10. It is said that the solution of (ODE) \((v,u)\)-escapes from \(h_1(t)\) to \(h_2(t)\) if there exist \(T > 0, u > 0\) and \(v > 0\) such that

\[
\begin{align*}
B_v(h_1(t_0)) & \subseteq RA(h_1(t_0)) \\
B_u(h_2(t)) & \subseteq RA(h_2(t)), \quad \forall t \in \{t \in I_{t,2} : t \geq T\} \\
\forall x_0 \in B_v(h_1(t_0)) & \implies x(t, t_0, x_0) \in B_u(h_2(t)), \quad \forall t \in \{t \in I_{t,2} : t \geq T\}
\end{align*}
\]  

(17a, 17b, 17c)

Figure 3: Illustration of jumping and tracking

Figure 3 illustrates the definitions of jumping and tracking for Example 1 with \(\alpha = 0.3\) and \(b = 5\). The objective of this paper is to study when the solution of (ODE) started at a poor local minimum at the initial time jumps to and tracks a better (or global) minimum of the problem after some time. In other words, it is desirable to investigate the escaping property from \(h_1(t)\) and \(h_2(t)\).

3 Optimization landscape after a change of variables

Given two isolated local minimum trajectories \(h_1(t)\) and \(h_2(t)\), one may use the change of variables \(x(t, t_0, x_0) = e(t, t_0, e_0) + h_2(t)\) to transform (ODE) into the form

\[
\dot{e}(t) = -\frac{1}{\alpha} \nabla_x f(e(t) + h_2(t), t) - \dot{h}_2(t), \quad \forall t \geq t_0
\]

(18)

We use \(e(t, t_0, e_0)\) to denote the solution of this differential equation starting at time \(t = t_0\) with the initial point \(e_0 = x_0 - h_2(t_0)\). Note that \(h_1(t)\) and \(h_2(t)\) are local solutions of \(f(x, t)\) and as long as \(f(x, t)\) is time-varying, these functions cannot satisfy (ODE) in general.

3.1 Inertia encouraging the exploration

The first term \(\nabla_x f(e + h_2(t), t)\) in (18) can be understood as a time-varying gradient term that encourages the solution of (18) to track the local minimum \(h_2(t)\) while the second term \(\dot{h}_2(t)\) represents the inertia from this trajectory. In particular, if \(\dot{h}_2(t)\) points toward outside of the region of attraction of \(h_2(t)\) during some time interval, the term \(\dot{h}_2(t)\) acts as an exploration term that encourages the solution of (ODE) to
leave the region of attraction of \( h_2(t) \). The parameter \( \alpha \) balances the roles of the gradient and the inertia. In the extreme case where \( \alpha \) goes to infinity, \( e(t) \) converges to \(-h_2(t)\) and \( x(t) \) approaches a constant trajectory determined by the initial point \( x_0 \); when \( \alpha \) is sufficiently small, the time-varying gradient term dominates the inertia term and the solution of (ODE) would track \( h_2(t) \) closely. With the appropriate proximal regularization \( \alpha \) that balances the time-varying gradient term and the inertia term, the solution of (ODE) could temporarily track a local minimum trajectory while keeping the potential of exploring other local minimum trajectories.

### 3.2 Inertia creating a one point strongly convex landscape

The differential equation (18) can be written as

\[
\dot{e}(t) = -\frac{1}{\alpha} \nabla e \left( f(e(t) + h_2(t), t) + \alpha \dot{h}_2(t) ^\top e(t) \right)
\]

This can be regarded as a time-varying gradient flow system of the original objective function \( f(e + h_2(t), t) \) plus a time-varying perturbation \( \alpha \dot{h}_2(t) ^\top e \). During some time interval \([t_1, t_2]\), the time-varying perturbation \( \alpha \dot{h}_2(t) ^\top e \) may enable that the time-varying objective function \( f(e + h_2(t), t) + \alpha \dot{h}_2(t) ^\top e \) over the neighborhood of \( h_1(t) \) becomes **one point strongly convexified** with respect to \( h_2(t) \). Under such circumstances, the time-varying perturbation \( \alpha \dot{h}_2(t) ^\top e \) prompts the solution of (19) starting in a neighborhood of \( h_1(t) \) to move to a neighborhood of \( h_2(t) \). Before analyzing this phenomenon, we illustrate the concept in an example.

Consider again Example 1 and recall that \( g(x) \) has 2 local minima at \( x = -2 \) and \( x = 1 \). By taking \( b = 5 \), \( h_1(t) = -2 + 5\sin(t) \) and \( h_2(t) = 1 + 5\sin(t) \), the differential equation (19) can be expressed as

\[
\dot{e}(t) = -\frac{1}{\alpha} \nabla e \left( g(1+e(t))+5\alpha \cos(t)e(t) \right)
\]

The landscape of the new time-varying function \( g(1+e)+5\alpha \cos(t)e \) with the variable \( e \) is shown for two cases \( \alpha = 0.3 \) and \( \alpha = 0.1 \) in Figure 4. The red curves are the solutions of (19) starting from \( e = -3 \). One can observe that when \( \alpha = 0.3 \), the new landscape becomes one point strongly convex around \( h_2(t) \) over the whole region for some time interval, which provides (19) with the opportunity of escaping from the region around \( h_1(t) \) to the region around \( h_2(t) \). However, when \( \alpha = 0.1 \), there are always two locally one point strongly convex regions around \( h_1(t) \) and \( h_2(t) \) and, therefore, (19) fails to escape the region around \( h_1(t) \).

![Figure 4: Illustration of time-varying landscape after change of variables for Example 1](image-url)
To further inspect the case $\alpha = 0.3$, observe in Figure 5a that the landscape of the objective function $g(1 + e) + 1.5\cos(0.9\pi)e$ shows that the region around the spurious local minimum trajectory $h_1(t)$ is one point strongly convexified with respect to $h_2(t)$ at time $t = 0.9\pi$. This is consistent with the fact that the solution of $\dot{e} = \frac{1}{0.3}\nabla_x g(1 + e) - 5\cos(t)$ starting from $e = -3$ jumps to the neighborhood of 0 around time $t = 0.9\pi$, as demonstrated in Figure 5c. Furthermore, if the time interval $[t_1, t_2]$ is relatively large enough to allow transitioning from a neighborhood of $h_1(t)$ to a neighborhood of $h_2(t)$, then the solution of (19) would move to the neighborhood of $h_2(t)$. In contrast, the region around $1 + b\sin(t)$ is never one point strongly convexified with respect to $-2 + b\sin(t)$, as shown in Figure 5b.

From the right-hand side of (19), it can be inferred that if the gradient of $f(\cdot, t)$ is relatively small around some local minimum trajectory, then its landscape is easier to be re-shaped by the time-varying linear perturbation $\alpha h_2(t)^\top e$. The local minimum trajectory with a neighborhood of small gradients usually correspond to a shallow minimum trajectory in which the local minimum trajectory is relatively flat and has a relatively small region of attraction. Thus, the one point strong convexication introduced by the time-varying perturbation could help escape the shallow minimum trajectories.

### 4 Main results

In this section, we study the jumping, tracking and escaping properties for online optimization.

#### 4.1 Jumping

In this part, we derive different sufficient conditions under which the solution of (ODE) jumps from a poor local minimum trajectory to a better (or global) trajectory.

**Theorem 2** (Sufficient conditions for jumping from $h_1(t)$ to $h_2(t)$). Given a local $I_{t,1}$-minimum trajectory $h_1(t)$ and a local $I_{t,2}$-minimum trajectory $h_2(t)$, suppose that the time-varying function $f(x, t)$ is locally $(\bar{I}_{t,1}, c_1, r_1)$-one point strongly convex around $h_1(t)$ and locally $(\bar{I}_{t,2}, c_2, r_2)$-one point strongly convex around $h_2(t)$ in the region $D_1 = \{e \in \mathbb{R}^n : \|e\| \leq r_2\}$. Assume that there exist a nonempty time interval $[t_1, t_2] \subset \bar{I}_{t,1} \cap \bar{I}_{t,2}$, a regularization parameter $\alpha$, a connected subset $D_4$, and a constant $\theta \in (0, 1)$ such that five conditions are satisfied:

1. Trajectory of the real root: For each fixed $t \in [t_1, t_2]$, $\nabla_x f(e + h_2(t), t) + \alpha h_2(t) = 0$ has a real root $\dot{e}(t)$ in the region $D_2 = \{e \in \mathbb{R}^n : \|e\| \leq \rho\}$, where $\rho < r_2$ and $\dot{e}(t)$ is continuously differentiable for all $t \in [t_1, t_2]$.
2. Identifying a positively invariant set: \( D_2 \cup D_3 \subset D_4 \), where \( D_3 = \{ e_1 \in \mathbb{R}^n : e_1 + h_2(t_1) \in \mathcal{B}_e(h_1(t_1)) \} \) and \( D_4 \) is a compact positively invariant subset with respect to (18), i.e.,

\[
e_1 \in D_4 \implies e(t, t_1, e_1) \in D_4, \quad \forall t \in [t_1, t_2]
\]

(20)

3. One point strong convexification: The time-varying function \( f(e + h_2(t), t) + \alpha h_2(t)^\top e \) is locally one point \( w \)-strongly convex around \( \bar{e}(t) \) for all \( e \in D_4 \) and for all \( t \in [t_1, t_2] \), i.e.,

\[
\left( \nabla_x f(e + h_2(t), t) + \alpha h_2(t) \right)^\top (e - \bar{e}(t)) \geq w \| e - \bar{e}(t) \|^2, \quad \forall e \in D_4, \quad \forall t \in [t_1, t_2]
\]

(21)

where \( w > 0 \) is a constant.

4. Upper bound on inertia: \( \alpha \leq \frac{(r_2 - \rho) \theta w}{\sup_{t \in [t_1, t_2]} \| \dot{e}(t) \|} \).

5. Lower bound for the time interval: The following inequality holds:

\[
t_2 - t_1 \geq \frac{\alpha}{w(1 - \theta)} \ln \left( \frac{\| e_1 - \bar{e}(t_1) \|}{r_2 - \rho} \right), \quad \forall e_1 \in D_3
\]

(22)

Then, the solution of (ODE) will \((v, r_2)\)-jump from \( h_1(t) \) to \( h_2(t) \) over the time interval \([t_1, t_2]\).

To avoid directly solving for the real roots of \( \nabla_x f(e + h_2(t), t) + \alpha h_2(t) = 0 \) and checking the condition (21) for all \( t \in [t_1, t_2] \), we propose an approach based on the time-averaged dynamics over a small time interval and named it "small interval averaging". This technique guarantees that the solution of the time-varying differential equation (or system) will converge to a residual set of the origin of (18), provided that:

1. there is a time interval \([t_1, t_2]\) such that the temporal variation makes the averaged objective function during this interval locally one point strongly convex around \( h_2(t) \) not only just over the neighborhood of \( h_2(t) \) but also over the neighborhood of \( h_1(t) \),
2. the original time-varying system is not too distant from the time-invariant averaged system,
3. \([t_1, t_2]\) is relatively large enough to allow the transition of points from the neighborhood of \( h_1(t) \) to the neighborhood of \( h_2(t) \).

This time interval \([t_1, t_2]\) and its time-averaged dynamics over this time interval serve as a certificate for jumping from \( h_1(t) \) to \( h_2(t) \). In what follows, we introduce the notion of averaging a time-varying function over a time interval \([t_1, t_2]\).

**Definition 11.** A function \( f_{av}^h(e) \) is said to be the average function of \( f(e + h_2(t), t) \) over the time interval \([t_1, t_2]\) if

\[
f_{av}^h(e) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f(e + h_2(\tau), \tau) d\tau
\]

(23)

Definition 11 yields that \( \nabla_x f_{av}^h(e) \) is the average function of \( \nabla_x f(e + h_2(\tau), \tau) \). The time-invariant system

\[
\dot{e} = -\frac{1}{\alpha} \nabla_x f_{av}^h(e) - \frac{h_2(t_2) - h_2(t_1)}{t_2 - t_1}
\]

(24)

is said to be a partial interval averaged system of the time-varying system (18) over the time interval \([t_1, t_2]\) with the perturbation term

\[
p(\alpha, e, t) = -\frac{1}{\alpha} \left( \nabla_x f(e + h_2(t), t) - \nabla_x f_{av}^h(e) \right) - \left( \frac{h_2(t) - h_2(t_2) - h_2(t_1)}{t_2 - t_1} \right)
\]

(25)

\(^3\)Our averaging approach distinguishes from classic averaging methods \([18, 26, 44, 2]\) and the partial averaging method \([33]\) in the sense that: (1) it is averaged over a small time interval instead of the entire time horizon, and (2) there is no two time-scale behavior because there is no parameter in (18) that can be taken sufficiently small.
Theorem 3 (Sufficient conditions on jumping from $h_1(t)$ to $h_2(t)$ using averaging). Given a local $I_{1,1}$-minimum trajectory $h_1(t)$ and a local $I_{1,2}$-minimum trajectory $h_2(t)$, suppose that the time-varying function $f(x,t)$ is locally ($I_{1,1},c_1,r_1$)-one point strongly convex around $h_1(t)$ and locally ($I_{1,2},c_2,r_2$)-one point strongly convex around $h_2(t)$ in the region $D_1 = \{ e \in \mathbb{R}^n : \| e \| \leq r_2 \}$. Assume that there exist a nonempty time interval $[t_1,t_2] \subset I_{1,1} \cap I_{1,2}$, a regularization parameter $\alpha$, and a connected subset $D_4$ such that the following five conditions are satisfied:

1. Equilibrium point of the averaged system: The system (24) has an equilibrium point $\bar{e}$ in the region $D_2 = \{ e \in \mathbb{R}^n : \| e \| \leq \rho \}$, where $\rho < r_2$.

2. Identifying a positively invariant set: $D_2 \cup D_3 \subseteq D_4$, where $D_3 = \{ e_1 \in \mathbb{R}^n : e_1 + h_2(t_1) \in B_r(h_1(t_1)) \subseteq RA(h_1(t_1)) \}$ and $D_4$ is a compact positively invariant subset with respect to (18), i.e.,

$$e_1 \in D_4 \Rightarrow e(t,t_1,e_1) \in D_4, \quad \forall t \in [t_1,t_2] \quad (26)$$

3. One point strongly convexification: The time invariant function $f_{av}^{h_2}(e) + \frac{\alpha(h_2(t_2) - h_2(t_1))}{t_2 - t_1} e$ is locally one point $w$-strongly convex around $\bar{e}$ for all $e \in D_4$, i.e.,

$$\left( \nabla_x f_{av}^{h_2}(e) + \frac{\alpha(h_2(t_2) - h_2(t_1))}{t_2 - t_1} \right) (e - \bar{e}) \geq w \| e - \bar{e} \|^2, \quad \forall e \in D_4 \quad (27)$$

where $w > 0$ is a constant.

4. Bound on perturbation: Suppose that the perturbation $p(\alpha,e,t)$ satisfies the inequality

$$\| p(\alpha,e,t) \| \leq \delta_1(\alpha,t) \| e - \bar{e} \| + \delta_2(\alpha,t), \quad \forall t \in [t_1,t_2] \quad (28)$$

for some positive constants $\delta_1(\alpha)$ and $\delta_2(\alpha)$ such that

$$\int_{t_1}^{t_2} \delta_1(\alpha,\tau) d\tau \leq \eta_1(\alpha)(t_2 - t_1), \quad \forall t \in [t_1,t_2] \quad (29)$$

5. Guarantee of convergence within $[t_1,t_2]$: The following inequality holds:

$$\beta_2(\alpha) \| e_1 - \bar{e} \| e^{-\beta_1(\alpha)(t_2 - t_1)} + \beta_2(\alpha) \int_{t_1}^{t_2} e^{-\beta_1(\alpha)(t_2 - \tau)} \delta_2(\alpha,\tau) d\tau \leq r_2 - \rho, \quad \forall e_1 \in D_3 \quad (30)$$

where $\beta_1(\alpha) = \frac{w}{\alpha} - 2\eta_1(\alpha)$ and $\beta_2(\alpha) = \exp(\eta_2(\alpha))$.

Then, the solution of (ODE) will $(v,r_2)$-jump from $h_1(t)$ to $h_2(t)$ over the time interval $[t_1,t_2]$.

Remark 1. Consider a special case where $f(x,t) = g(x-z(t))$ such that $z(t) : [0,\infty) \rightarrow \mathbb{R}^n$ is a continuous differentiable function. Suppose that $g(.)$ has two local minima $z_1^*$ and $z_2^*$. The online optimization $f(.t)$ has two isolated minimum trajectories $h_1(t) = z_1^* + z(t)$ and $h_2(t) = z_2^* + z(t)$. The time-varying system after the change of variables $x(t) = e + z_2^* + z(t)$ becomes

$$\dot{e} = -\frac{1}{\alpha} \nabla_x g(e + z_2^*) - \dot{z}(t) \quad (31)$$

and its partial interval averaged system over the time interval $[t_1,t_2]$ becomes

$$\dot{e} = -\frac{1}{\alpha} \nabla_x g(e + z_2^*) - \frac{z(t_2) - z(t_1)}{t_2 - t_1} \quad (32)$$

14
By selecting \( \eta_1(\alpha) = 0, \eta_2(\alpha) = 0 \) and \( \delta_2(\alpha, t) = \left\| \tilde{z}(t) - \frac{z(t_2)-z(t_1)}{t_2-t_1} \right\| =: \delta_2(t) \), the condition (30) reduces to

\[
\|e_1 - \bar{e}\| e^{-\frac{\alpha}{w}(t_2-t_1)} + \int_{t_1}^{t_2} e^{-\frac{\alpha}{w}(t_2-\tau)}\delta_2(\tau)d\tau \leq r_2 - \rho \tag{33}
\]

which can be relaxed to the simple condition

\[
\left( \|e_1 - \bar{e}\| - \frac{\alpha}{w}\right) e^{-\frac{\alpha}{w}(t_2-t_1)} + \frac{\alpha}{w} \sup_{\tau \in [t_1,t_2]} \delta_2(\tau) \leq r_2 - \rho \tag{34}
\]

**Remark 2.** In Theorem 3, to ensure that the time-invariant partial interval averaged system is a reasonable approximation of the time-varying system, the time interval \([t_1,t_2]\) should not be very large. On the other hand, to preserve the optimality of the solution with regards to the original time-varying optimization problem without any proximal regularization, it is required to guarantee that the solution of (ODE) is close to \( h_2(t) \). However, to preserve the optimality of the solution with regards to the original time-varying optimization problem without any proximal regularization, it is required to guarantee that the solution of (ODE) is close to \( h_2(t) \). In addition, although the estimation of the convergence time in (30) and (33) may be conservative, the nature of the exponential convergence rate due to the locally one point strongly convex condition would enable a fast jumping of the solution of (18) during \([t_1,t_2]\).

### 4.2 Tracking

In this subsection, we study the tracking property of the local minimum trajectory \( h_2(t) \). First, notice that if \( h_2(t) \) is not constant, the right-hand side of (ODE) is nonzero while the left-hand side is zero. Therefore, \( h_2(t) \) is not a solution of (ODE) in general. This is because the solution of (ODE) approximates the continuous limit of a discrete local trajectory of the sequential regularized optimization problem (7). However, to preserve the optimality of the solution with regards to the original time-varying optimization problem without any proximal regularization, it is required to guarantee that the solution of (ODE) is close to \( h_2(t) \).

First, consider the case when the maximal time interval of \( h_2(t) \) is the entire time horizon \([t_0, \infty] \). If the solution of (18) can be shown to be in a small residual set around 0, then it is guaranteed that \( x(t, t_0, x_0) \) tracks its nearby local minimum trajectory. Notice that (18) can be regarded as a time-varying perturbation of the system

\[
\dot{e} = -\frac{1}{\alpha} \nabla_x f(e + h_2(t), t), \quad \forall t \geq t_0 \tag{35}
\]

Since \( h_2(t) \) is a local minimum trajectory, it is obvious that \( e(t) \equiv 0 \) is an equilibrium point of (35). In addition, since \( f(e + h_2(t), t) \) is locally \((\infty, c_2, r_2)\)-one point around \( h_2(t) \), the stability property of \( e = 0 \) for (35) can be proved, as discussed below.

**Lemma 1.** If \( f(e + h_2(t), t) \) is locally \((\infty, c_2, r_2)\)-one point strongly convex around \( h_2(t) \) in the region \( D = \{ e \in \mathbb{R}^n : \|e\| \leq r_2 \} \), then \( e = 0 \) is a locally exponentially stable equilibrium point of (35).

Since the system (35) has an exponentially stable equilibrium point at \( e = 0 \), one would expect that the solution of the time-varying perturbed system (18) stays in a small residual set of \( e = 0 \) if the perturbation \( \hat{h}_2(t) \) is relatively small. The perturbation \( \hat{h}_2(t) \) being small is equivalent to \( \alpha \) being small. The next theorem shows that every local \( \infty \)-minimum trajectory can be tracked for a sufficiently small \( \alpha \).

**Theorem 4** (Sufficient condition for tracking). Assume that the time-varying function \( f(x, t) \) is locally \((\infty, c_2, r_2)\)-one point strongly convex around \( h_2(t) \). Then, \( h_2(t) \) can be tracked if \( \alpha \) is sufficiently small. In particular, given \( 0 < \theta' < 1, \gamma := \sup_{t \geq 0} \|\hat{h}_2(t)\|, u := \frac{\gamma}{\theta' c_2}, \|x_0 - h_2(0)\| \leq r_2 \), and \( \alpha < \frac{2\theta' r_2}{\gamma} \), the solution
Remark 3. This implies that if the local minimum trajectory $h_2(t)$ is constant, then it will be perfectly tracked with any regularization parameter and cannot be escaped by tuning the regularization parameter.

Remark 4. For a fixed ultimate bound $u$, the convergence rate $(1 - \theta') \frac{c_2}{\alpha}$ shows that $x(t, t_0, x_0)$ converges faster to $B_u(h_2(t))$ as the regularization parameter $\alpha$ reduces.

Remark 5. In the case that the local minimum trajectory $h_2(t)$ is a constant, the upper bound on $\alpha$ simply becomes $\alpha < \infty$. This implies that if the local minimum trajectory $h_2(t)$ is constant, then it will be perfectly tracked with any regularization parameter and cannot be escaped by tuning the regularization parameter.
4.3 Escaping

Combining Theorem 2 or 3 with Theorem 4 or 6 immediately yields a sufficient condition on escaping from one local minimum trajectory to another local minimum trajectory. The proof is omitted for brevity.

**Theorem 7** (Sufficient conditions for escaping from $h_1(t)$ to $h_2(t)$). Given a local $I_{t,1}$-minimum trajectory $h_1(t)$ and a local $I_{t,2}$-minimum trajectory $h_2(t)$, suppose that the time-varying function $f(x,t)$ is locally ($I_{t,1}, c_1, r_1$)-one point strongly convex around $h_1(t)$ and locally ($I_{t,2}, c_2, r_2$)-one point strongly convex around $h_2(t)$ in the region $D_1 = \{e \in \mathbb{R}^n : \|e\| \leq r_2\}$. Let $\gamma = \sup_{t \in I_{t,2}} \|h_2(t)\|$, $0 < \theta' < 1$, $B_{\theta'}(h_1(t)) \subseteq RA(h_1(t))$ and $u = \frac{\alpha}{\theta'^2}$. Under the conditions of Theorem 2 or 3, if $\alpha < \frac{\sqrt{\gamma}}{\theta'^2}$, the solution of (ODE) will $(v, r_2)$-escape from $h_1(t)$ to $h_2(t)$ after $t \geq t_2$.

4.4 Discussions

**Adaptive inertia**: To leverage the potential of the time-varying perturbation $\alpha h_2(t)^{\top}e$ in re-shaping the landscape of the objective function to become locally one point strongly convex over a large region, the regularization parameter $\alpha$ should be selected relatively large. On the other hand, to ensure the solution of (19) will end up tracking a local minimum trajectory (or hopefully, a global minimum trajectory), Theorem 4 prescribes small values for $\alpha$. In practice, especially when the time-varying objective function has many spurious shallow minimum trajectories, this suggests using a relatively large regularization parameter $\alpha$ at the beginning of the time horizon to escape spurious shallow minimum trajectories and then switching to a relative small regularization parameter $\alpha$ for reducing the ultimate tracking error bound.

**Sequential jumping**: When the time-varying objective function $f(x, t)$ has many local minimum trajectories, the solution of (ODE) may sequentially jump from one local minimum trajectory to another local minimum trajectory. To illustrate this concept, consider the local minimum trajectories $h_1(t), h_2(t), ..., h_m(t)$, where $h_m(t)$ is a global target trajectory. Assume that there exists a sequence of time intervals $[t_i^1, t_i^2]$ for $i = 1, 2, \ldots, m - 1$ such that the conditions of Theorem 2 or 3 are satisfied for $h_i(t)$ and $h_{i+1}(t)$ during each time interval. Then, by sequentially deploying Theorem 2 or 3, it can be concluded that the solution of (ODE) will jump from $h_1(t)$ to $h_m(t)$ after $t \geq t_m^m$. Furthermore, if $h_m(t)$ is tractable with the given $\alpha$, the solution of (ODE) will escape from $h_1(t)$ to $h_m(t)$ after $t \geq t_m^m$.

5 Numerical Example

**Example 2**. Consider the two-dimensional non-convex function

$$g(x) = -20e^{-\sqrt{0.5(x_1^2 + x_2^2)} + d^2} - 0.5e^{0.5(\cos(2\pi x_1) + \cos(2\pi x_2))} + 0.5e + 20e^{-d}$$

(41)

This function has a global minimum at $(0, 0)$ with the optimal value $0$ and many spurious local minima. Its landscape is shown in Figure 6a. When $d = 0$, this function is called the Ackley function [1], which is a benchmark function for global optimization algorithms. To make this function twice continuously differentiable, we take $d = 0.01$. Consider the time-varying objective function $f(x, t) = g(x - z(t))$, where

$$z(t) = \begin{bmatrix} 7 \sin(t) \\ 7 \cos(t) \end{bmatrix}$$

Two local $\infty$-minimum trajectories of this online optimization problem are $h_1(t) = [1.95, 0.97]^\top + z(t)$ and $h_2(t) = [0, 0]^\top + z(t)$. It can be observed in Figures 6b and 6c that, around time $t = 0$, the time-varying objective function around a neighborhood of $h_1(0)$ is one point strongly convexified with respect to $h_2(0)$.
Thus, one could expect that the solution of (ODE) would jump from $h_1(t)$ to $h_2(t)$. More formally, it can be shown that $g(x)$ is locally $(3.3,1.1)$-one point strongly convex with respect to the origin, which implies that $f(x,t)$ is locally $(\infty,3.3,1.1)$-one point strongly convex around $h_2(t)$. To ensure that the solution of (ODE) will track $h_2(t)$, we need to take $\alpha < \frac{e^2 e_t}{\sup_{t \in [0,1]} \| \dot{x}(t) \|}$ for $0 < \theta < 1$. In this case, $\alpha = 0.5$ simply satisfies the tracking condition. Then, by the change of variables $x = e + h_2(t)$, the differential equation (18) can be written as

$$
\dot{e}(t) = -2 \left[ \begin{array}{c} 10e^{-\sqrt{0.5(e_1^2+e_2^2)+d^2}} \frac{e_1}{\sqrt{0.5(e_1^2+e_2^2)+d^2}} + 0.5\pi e^{(0.5(\cos(2\pi e_1)+\cos(2\pi e_2)))} \sin(2\pi e_1) \\ 10e^{-\sqrt{0.5(e_1^2+e_2^2)+d^2}} \frac{e_2}{\sqrt{0.5(e_1^2+e_2^2)+d^2}} + 0.5\pi e^{(0.5(\cos(2\pi e_1)+\cos(2\pi e_2)))} \sin(2\pi e_2) \end{array} \right] - \left[ \begin{array}{c} 7\cos(t) \\ -7\sin(t) \end{array} \right]$$

(42)

By selecting the time interval $[0, \frac{\pi}{3}]$, the averaged system can be obtained as

$$
\dot{e}(t) = -2 \left[ \begin{array}{c} 10e^{-\sqrt{0.5(e_1^2+e_2^2)+d^2}} \frac{e_1}{\sqrt{0.5(e_1^2+e_2^2)+d^2}} + 0.5\pi e^{(0.5(\cos(2\pi e_1)+\cos(2\pi e_2)))} \sin(2\pi e_1) \\ 10e^{-\sqrt{0.5(e_1^2+e_2^2)+d^2}} \frac{e_2}{\sqrt{0.5(e_1^2+e_2^2)+d^2}} + 0.5\pi e^{(0.5(\cos(2\pi e_1)+\cos(2\pi e_2)))} \sin(2\pi e_2) \end{array} \right] - \left[ \begin{array}{c} \frac{56}{\pi} \sin\left(\frac{\pi}{3}\right) \\ \frac{56}{\pi} \cos\left(\frac{\pi}{3}\right) - 1 \end{array} \right]
$$

(43)

This system has an equilibrium point at $[-0.0034, 0.0007]^\top$. Then Condition 1 in Theorem 3 is met with $\rho = 0.01$. Let $D_1 = B_{1,1}(0)$, $D_2 = B_{0.01}(0)$, $D_3 = \{e \in \mathbb{R}^n : e_1 + h_2(t_1) \in B_{0.1}(h_1(t_1))\}$ and $D_4 = [-0.2, 2.1] \times [-0.1, 1.1]$. It follows that $D_2 \cup D_3 \subseteq D_4$. In addition, on the boundary points $e_1 = 2.1$ and $e_1 = -2$, the derivative of $e_1$ along the trajectory of (42) is negative and positive, respectively, for all $e_2 \in [-0.1, 1.1]$ and $t \in [0, \frac{\pi}{3}]$. Similarly, on the boundary points $e_2 = 1.1$ and $e_2 = -0.1$, the derivative of $e_2$ along the trajectory of (42) is negative and positive, respectively, for all $e_1 \in [-0.2, 2.1]$ and $t \in [0, \frac{\pi}{3}]$. This implies that $D_4$ is a positively invariant set with respect to (42) for $t \in [0, \frac{\pi}{3}]$. This shows that Condition 2 in Theorem 3 is also met. Furthermore, (27) and (33) are satisfied for $w = 1.3$. Thus, the conditions of Theorem 6 are all met, and therefore the solution of (42) will $(0,1,1.1)$-escape from $h_1(t)$ to $h_2(t)$. Furthermore, we have verified for 1000 runs of random initialization over $x_0 - z(0) \in [-5.5] \times [-5.5]$ that all solutions of (42) will sequentially jump over the local minimum trajectories and end up tracking the global trajectory $[0,0]^\top + z(t)$ after $t \geq 10\pi$.

6 Conclusion

In this work, we study the landscape of time-varying nonconvex optimization problems. The objective is to understand when simple local search algorithms can find (and track) time-varying global solutions
of the problem over time. We introduce a time-varying gradient flow system with controllable inertia as a continuous-time limit of the stationary condition for discretized sequential optimization problems with proximal regularization and online updating scheme. Via a change of variables, the time-varying gradient flow system is regarded as a composition of a time-varying gradient term and a time-varying perturbation term due to the inertia. We show that the time-varying perturbation term due to the inertia encourages the exploration of the state space and re-shapes the landscape by potentially making it one point strongly convex over a large region during some time interval. We introduce the notions of jumping and escaping, and use them to develop sufficient conditions under which the time-varying solution jumps from a poor local trajectory to a better (or global) minimum trajectory over a finite time interval. We illustrate in a two-dimensional benchmark example with many shallow minimum trajectories that the natural time variation of the problem enables escaping spurious local minima over time. Avenues for future work include the extension of the current work to constrained time-varying nonconvex optimization problems. Furthermore, it is useful to study how to systematically introduce an exogenous temporal variation to a time-invariant nonconvex optimization problem in order to find its global minimum using an online-optimization-based local search method.

References


Appendix

6.1 Proof of Theorem 1

Proposition 3. [26, Theorem 3.1] Let \( f(t, x) \) be piecewise continuous in \( t \) and satisfy the Lipschitz condition
\[
\|f(t, x) - f(t, y)\| \leq L\|x - y\|, \quad \forall x, y \in D = \{x \in \mathbb{R}^n : \|x - x_0\| \leq r\}, \quad \forall t \in [t_0, t_1]
\] (44)

Then, there exists some \( \delta > 0 \) such that the state equation \( \dot{x} = f(t, x) \) with \( x(t_0) = x_0 \) has a unique solution over \( [t_0, t_0 + \delta] \).

Proposition 4. [33, Corollary 3.2] Under the conditions of Proposition 3, there exists a maximal interval \( [t_0, T) \) over which the unique solution starting at \( (t_0, x_0) \) exists.

Lemma 2. Under the conditions of Proposition 3, let \( [t_0, T) \) be the maximal interval over which the unique solution starting at \( (t_0, x_0) \) exists with \( T < \infty \). Let \( W \) be any compact subset of \( D \). There exists some \( t \in [t_0, T) \) with the property that \( x(t) \notin W \).

Proof. To prove by contradiction, suppose that there is no time \( t \) satisfying the stated property. Then, it holds that \( x(t) \in W \) for all \( t \in [t_0, T) \). It suffices to show that \( [t_0, T) \) is not the maximal interval of existence. The solution of \( \dot{x} = f(t, x) \) relative to \( x(t_1) \) can be written as
\[
x(t) = x(t_1) + \int_{t_1}^{t} f(\tau, x(\tau))d\tau, \quad \forall t_1, t \in [t_0, T)
\] (45)

Since \( f(t, x) \) is piecewise continuous in \( t \) and continuous in \( x \), there exists a constant \( M > 0 \) such that \( \|f(\tau, x(\tau))\| \leq M \) for all \( \tau \in [t_0, T) \). Thus,
\[
\|x(t) - x(t_1)\| = \left\|\int_{t_1}^{t} f(\tau, x(\tau))d\tau\right\| \leq \int_{t_1}^{t} M d\tau = M(t - t_1)
\] (46)
which implies that \( x(t) \) is uniformly continuous on \( [t_0, T) \). Then, by the continuous extension theorem, \( f(t, x) \) can be defined at the endpoint \( T \) in such a way that \( f(t, x) \) becomes continuous on \( [t_0, T) \). In other words,
\[
x(T) = x(t_0) + \lim_{t \to T} \int_{t_0}^{t} f(\tau, x(\tau))d\tau = x(t_0) + \int_{t_0}^{T} f(\tau, x(\tau))d\tau
\] (47)
Therefore, the solution \( x(T) \) is defined and since \( W \) is closed, it holds that \( x(T) \in W \). Then, it follows from Proposition 3 (applied to the point \((T, x(T))\)) that there is a \( \delta > 0 \) with the property that the solution can be extended to \([t_0, T + \delta]\). This contradicts the fact that \([t_0, T]\) is the maximal interval of existence, and completes the proof. \( \square \)

**Proof of Theorem 1.** If \( f(x, t) \) is piecewise continuous in \( t \) and its gradient is locally Lipschitz in \( x \) for all \( t \geq 0 \) and \( x \in D \subset \mathbb{R}^n \), then \( \nabla_x f(t, x) \) satisfies the conditions of Propositions 3-4. It results from Propositions 3-4 that there exists a unique solution for (ODE) over \([t_0, T]\) that is the maximal interval of unique existence. It is enough to show that \( T = \infty \). Due to Lemma 2, if the time \( T \) is finite, the solution must leave every compact subset of \( D \). However, the solution never leaves the compact set \( W \). This implies that \( T = \infty \).

### 6.2 Proof of Theorem 2

**Proposition 5** (Comparison lemma). \([26, \text{Lemma 3.4}]\) Consider the scalar differential equation

\[
\dot{u} = f(t, u), \quad u(t_0) = u_0
\]  

where \( f(t, u) \) is continuous in \( t \) and locally Lipschitz in \( u \) for all \( t \geq t_0 \) and \( u \in J \subset \mathbb{R} \). Let \([t_0, T)\) \( (T \) could be infinity) be the maximal interval of existence of the solution \( u(t) \), and suppose that \( u(t) \in J \) for all \( t \in [t_0, T) \). Let \( v(t) \) be a continuous function whose upper right-hand derivative \( D^+ v(t) \) satisfies the differential inequality

\[
D^+ v(t) \leq f(t, v(t)), \quad v(t_0) \leq u_0
\]  

with \( v(t) \in J \) for all \( t \in [t_0, T) \). Then, it holds that \( v(t) \leq u(t) \) for all \( t \in [t_0, T) \).

**Proof of Theorem 2.** First, notice that since \( D_4 \) is a compact positively invariant set with respect to the dynamics (18), it follows from Theorem 1 that (18) has a unique solution defined for \( t \in [t_1, t_2] \) whenever \( e_1 \in D_4 \). We take a positive semi-definite time-varying function \( V(e, t) = \frac{1}{2} \| e - \bar{e}(t) \|^2 : D_4 \to \mathbb{R} \) as the Lyapunov function for the system (18). The derivative of \( V(e) \) along the trajectories of (18) can be expressed as

\[
\dot{V} = (e - \bar{e}(t))^\top \left( -\frac{1}{\alpha} \nabla_x f(e + \bar{h}_2(t), t) - \dot{\bar{h}}_2(t) \right) + (e - \bar{e}(t))^\top \dot{\bar{e}}(t), \quad \forall e \in D_4
\]

\[
\leq -\frac{w}{\alpha} \| e - \bar{e}(t) \|^2 + \| \dot{\bar{e}}(t) \| \| e - \bar{e}(t) \|, \quad \forall e \in D_4
\]

\[
\leq -\frac{w}{\alpha} \| e - \bar{e}(t) \|^2 - \frac{\theta w}{\alpha} \| e - \bar{e}(t) \|^2 + \sup_{t \in [t_1, t_2]} \left( \| \dot{\bar{e}}(t) \| \right) \| e - \bar{e}(t) \|, \quad \forall e \in D_4
\]

\[
\leq -\frac{(\alpha \sup_{t \in [t_1, t_2]} \| \dot{\bar{e}}(t) \|)}{\theta w} \| e - \bar{e}(t) \|^2, \quad \forall e \in \left\{ e \in D_4 : \| e - \bar{e}(t) \| \geq \frac{\alpha \sup_{t \in [t_1, t_2]} \| \dot{\bar{e}}(t) \|}{\theta w} \right\}
\]  

(50)

By taking \( e_1 \in D_3 \subset D_4 \), since \( D_4 \) is a positively invariant set with respect to the dynamics (18) for \( t \in [t_1, t_2] \), any trajectory of (18) starting from \( D_3 \) will stay in \( D_4 \). Thus, the bound in (50) is valid. Let \( \delta := \sup_{t \in [t_1, t_2]} \| \dot{\bar{e}}(t) \| \) and \( \nu := \frac{\alpha \theta w}{\sup_{t \in [t_1, t_2]} \| \dot{\bar{e}}(t) \|} \). To ensure that the trajectory of (18) enters the time-varying set \( B_{r_2 - \rho} = \{ e \in \mathbb{R}^n : \| e - \bar{e}(t) \| \leq r_2 - \rho \} \), it is required to have \( \frac{\alpha \theta w}{\sup_{t \in [t_1, t_2]} \| \dot{\bar{e}}(t) \|} \leq r_2 - \rho \) or \( \alpha \leq \frac{(r_2 - \rho) \theta w}{\sup_{t \in [t_1, t_2]} \| \dot{\bar{e}}(t) \|} \).

Now, it is desirable to show that if the finite time interval \([t_1, t_2]\) is large enough, the solution of (18) will enter the time-varying set \( B_{r_2 - \rho} = \{ e \in \mathbb{R}^n : \| e - \bar{e}(t) \| \leq r_2 - \rho \} \) with an exponential convergence rate. Since \( \dot{V} \) is negative in \( \Gamma = \{ e \in D_4 : \| e - \bar{e}(t) \| \geq v \} \) and \( D_4 \) is a positively invariant set for all \( t \in [t_1, t_2] \), a trajectory starting from \( \Gamma \) must stay in \( D_4 \) and move in a direction of decreasing \( V(e) \). The
function \( V(e) \) will continue decreasing until the trajectory enters the set \( \{ e \in D_4 : \| e - \bar{e}(t) \| \leq v \} \) or until time \( t_2 \). Let us show that the trajectory enters \( B_{r_2 - \rho} \) before \( t_2 \) if \( t_2 - t_1 \geq \frac{a}{w(1-\theta)} \ln \left( \frac{\| e_1 - \bar{e}(t_1) \|}{r_2 - \rho} \right) \). Since \( V(e(t), t) = \frac{1}{2} \| e - \bar{e}(t) \|^2 \), (50) can be written as

\[
\dot{V} \leq -(1 - \theta) \frac{2w}{\alpha} V, \quad \forall e \in \left\{ e \in D_4 : \| e - \bar{e}(t) \| \geq v \right\}, \quad \forall t \in [t_1, t_2]
\]

By the comparison lemma, \( V \) satisfies

\[
V(e(t), t) \leq \exp \left\{ - (1 - \theta) \frac{2w}{\alpha} (t - t_1) \right\} V(e_1, t_1)
\]

Hence,

\[
\| e(t) - \bar{e}(t) \| \leq \exp \left\{ - (1 - \theta) \frac{w}{\alpha} (t - t_1) \right\} \| e_1 - \bar{e}(t_1) \|
\]

The inequality \( \| e(t_2) - \bar{e}(t_2) \| \leq r_2 - \rho \) holds if \( t_2 - t_1 \geq \frac{a}{w(1-\theta)} \ln \left( \frac{\| e_1 - \bar{e}(t_1) \|}{r_2 - \rho} \right) \).

### 6.3 Proof of Theorem 3

**Proof.** As shown in the proof of Theorem 2, the differential equation (18) has a unique solution defined for \( t \in [t_1, t_2] \) whenever \( e_1 \in D_4 \). By using the positive semi-definite function \( V(e) = \frac{1}{2} \| e - \bar{e} \|^2 : D_4 \rightarrow \mathbb{R} \) as the Lyapunov function for the system (18), the derivative of \( V(e) \) along the trajectories of (18) can be obtained as

\[
\dot{V} = (e - \bar{e})^T \left( - \frac{1}{\alpha} \nabla_x f_{av}(e) - \frac{h_2(t_2) - h_2(t_1)}{t_2 - t_1} \right) + p(\alpha, e, t), \quad \forall e \in D_4
\]

\[
\leq - \frac{w}{\alpha} \| e - \bar{e} \|^2 + \delta_1(\alpha, t) \| e - \bar{e} \|^2 + \delta_2(\alpha, t) \| e - \bar{e} \|, \quad \forall e \in D_4
\]

Since \( V = \frac{1}{2} \| e - \bar{e} \|^2 \), one can derive an upper bound on \( \dot{V} \) as

\[
\dot{V} \leq - \left[ \frac{2w}{\alpha} - 2\delta_1(\alpha, t) \right] V + \delta_2(\alpha, t) \sqrt{2V}
\]

To obtain a linear differential inequality, we consider \( W(t) = \sqrt{V(e(t))} \). When \( V(e(t)) \neq 0 \), it holds that \( \dot{W} = \dot{V}/2\sqrt{V} \) and

\[
\dot{W} \leq - \left[ \frac{w}{\alpha} - \delta_1(\alpha, t) \right] W + \frac{\delta_2(\alpha, t)}{\sqrt{2}}
\]

When \( V(e(t)) = 0 \), we have \( e(t) = \bar{e} \). Writing the Tylor expansion of \( e(t + \epsilon) \) for a sufficiently small \( \epsilon \) yields that

\[
e(t + \epsilon) = e(t) + \epsilon \left( - \frac{1}{\alpha} \nabla_x f_{av}(e) - \frac{h_2(t_2) - h_2(t_1)}{t_2 - t_1} + p(\alpha, \bar{e}, t) \right) + o(\epsilon)
\]

\[
e(t + \epsilon) = \bar{e} + \epsilon p(\alpha, \bar{e}, t) + o(\epsilon)
\]

This implies that

\[
\| e(t + \epsilon) - \bar{e} \|^2 = \epsilon^2 \| p(\alpha, \bar{e}, t) \|^2 + o(\epsilon^2)
\]

Therefore,

\[
V(e(t + \epsilon)) = \frac{\epsilon^2}{2} \| p(\alpha, \bar{e}, t) \|^2 + o(\epsilon^2)
\]
Proposition 6. [26, Theorem 4.10] Let \( e = 0 \) be an equilibrium point for (35) and \( D = \{ e \in \mathbb{R}^n : \| e \| \leq r_2 \} \).
Let $V : [0, \infty) \times D \rightarrow \mathbb{R}$ be a continuously differentiable function such that

$$k_1 \|e\|^p \leq V(t, e) \leq k_2 \|e\|^p$$

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} f(t, x) \leq -k_3 \|e\|^p$$

for all $t \geq 0$ and $e \in D$, where $k_1$, $k_2$ and $k_3$ are positive constants. Then, $e = 0$ is exponentially stable.

**Proof of Lemma 1.** We take a positive semi-definite function $V(e) = \frac{1}{2} \|e\|^2 : D \rightarrow \mathbb{R}$ as the Lyapunov function. The derivative of $V(e)$ along the trajectories of (ODE) satisfies

$$\dot{V} = e^\top \left(- \frac{1}{\alpha} \nabla_x f(e + h_2(t), t) \right)$$

$$\leq -\frac{c}{\alpha} \|e\|^2$$

Then the conditions in Proposition 6 are satisfied for $V = \frac{1}{2} \|e\|^2$, $p = 2$, $k_1 = k_2 = \frac{1}{2}$ and $k_3 = \frac{c}{\alpha}$. As a result, $e = 0$ is a locally exponentially stable equilibrium point of (35).

### 6.5 Proof of Theorem 4

**Proof.** Consider the positive semi-definite function $V(e) = \frac{1}{2} \|e\|^2 : D \rightarrow \mathbb{R}$ as the Lyapunov function for the system (18), where $D = \{e \in \mathbb{R}^n : \|e\| \leq r_2\}$. The derivative of $V(e)$ along the trajectories of (18) can be written as

$$\dot{V} = e^\top \left(- \frac{1}{\alpha} \nabla_x f(e + h_2(t), t) - \dot{h}_2(t) \right), \quad \forall t \geq t_0$$

$$\leq -\frac{c_2}{\alpha} \|e\|^2 + \gamma \|e\|, \quad \forall t \geq t_0$$

$$= -(1 - \theta')\frac{c_2}{\alpha} \|e\|^2 - \theta'\frac{c_2}{\alpha} \|e\|^2 + \gamma \|e\|, \quad \forall t \geq t_0$$

$$\leq -(1 - \theta')\frac{c_2}{\alpha} \|e\|^2, \quad \forall \|e\| \geq \frac{\alpha\gamma}{\theta'c_2}, \quad \forall t \geq t_0$$

We aim to show that if $u := \left(\frac{\alpha\gamma}{\theta'c_2}\right) < r_2$ or $\alpha < \frac{c_2\theta' r_2}{\gamma}$, the set $D$ has the property that any trajectory starting in $D$ at $t_0$ enters the set $\mathcal{B}_u(0) = \{e \in \mathbb{R}^n : \|e\| \leq u\}$ with an exponential convergence rate. Since the derivative $\dot{V}$ is negative on the boundaries $\partial D$ and $\partial \mathcal{B}_u(0)$, (72) implies that the sets $D$ and $\mathcal{B}_u(0)$ are positively invariant. Since $D$ is also a compact set, it follows from Theorem 1 that (18) has a unique solution defined for all $t \geq t_0$ whenever $e_0 \in D$.

Since $\dot{V}$ is negative in $\Gamma = \{e \in \mathbb{R}^n : u \leq \|e\| \leq r_2\}$, any trajectory starting in $\Gamma$ must move in a direction of decreasing $V(e)$, leading to the property that the function $V(e)$ will continue decreasing until the trajectory enters the set $\mathcal{B}_u(0)$ in finite time and stays therein for all future times. Let us show that the trajectory enters $\mathcal{B}_u(0)$ with an exponential convergence rate. Since $V(e) = \frac{1}{2} \|e\|^2$, (72) can be written as

$$\dot{V} \leq -(1 - \theta')\frac{2c_2}{\alpha} V, \quad \|e\| \geq u, \quad \forall t \geq t_0$$

By the comparison lemma, $V$ satisfies

$$V(e(t)) \leq \exp \left\{-(1 - \theta')\frac{2c_2}{\alpha} (t - t_0)\right\} V(e(0))$$

26
Hence,
\[ \|e(t)\| \leq \exp\left\{ -(1 - \theta') \frac{c_2}{\alpha} (t - t_0) \right\} \|e(0)\| \]  
(75)
This inequality holds over the interval \([t_0, t_0 + \frac{\alpha}{c_2(1 - \theta')} \ln\left(\frac{r_2}{u}\right)]\) during which \(\|e\| \geq u\). Since \(B_u(0)\) is a positively invariant set, \(e(t, t_0, e_0)\) will stay in \(B_u(0)\) for all future times. By the change of variables \(x(t, t_0, e_0) = e(t, t_0, e_0) + h_2(t)\), we have
\[ x(t, t_0, x_0) \in B_u(h_2(t)) \subseteq RA(h_2(t)), \quad \forall t \geq t_0 + \frac{\alpha}{c_2(1 - \theta')} \ln\left(\frac{r_2}{u}\right) \]  
(76)
if \(x_0 \in B_{\bar{r}_2}(h_2(t_0))\). This completes the proof. \(\square\)

### 6.6 Proof of Theorem 5

**Proof.** Under the conditions of Theorem 4 and the condition that \(\|x^1_0 - h_2(0)\| \leq r_2, \|x^2_0 - h_2(0)\| \leq r_2\), it holds that \(\|x(t, t_0, x_0) - h_2(t)\| \leq u\) and \(\|x(t, t_0, x_0^0) - h_2(t)\| \leq u\) for \(t > t' := t_0 + \frac{\alpha}{c_2(1 - \theta')} \ln\left(\frac{r_2}{u}\right)\). If \(u \leq \tilde{r}_2 \leq r_2\) or \(\alpha \geq \frac{c_2(1 - \theta')}{\tilde{r}_2}\), we obtain \(\nabla_x f(x, t) \leq -\frac{c_2}{\alpha} < 0\) for \(x \in B_u(h(t))\) and \(t \geq t'\). By denoting \(x(t) = x(t, t_0, x_0)\) and \(z(t) = x(t, t_0, x_0^2)\), the system (ODE) governing these two solutions can be written as
\[
\dot{x}(t) = -\frac{1}{\alpha} \nabla_x f(x, t) \tag{77a}
\]
\[
\dot{z}(t) = -\frac{1}{\alpha} \nabla_z f(z, t) \tag{77b}
\]
Applying the mean value theorem to the above equations yields that
\[
\dot{x}(t) - \dot{z}(t) = -\frac{1}{\alpha} \left( \nabla_x f(x, t) - \nabla_z f(z, t) \right) \tag{78a}
\]
\[
= -\frac{1}{\alpha} \nabla_{yy} f(y, t)(x(t) - z(t)) \tag{78b}
\]
where \(y(t) = \lambda(t)x(t) + (1 - \lambda(t))z(t)\) for some \(0 \leq \lambda(t) \leq 1\). By multiplying \((x(t) - z(t))\) to the both sides of (78b), we arrive at
\[
\frac{d}{dt} \|x(t) - z(t)\|^2 = -\frac{2}{\alpha} \nabla_{yy} f(y, t) \|x(t) - z(t)\|^2 \tag{79}
\]
Since after \(t > t'\), \(x(t) \in B_u(h_2(t))\), \(z(t) \in B_u(h_2(t))\), and \(B_u(h_2(t))\) is a convex set for each fixed \(t\), we have \(y(t) \in B_u(h_2(t))\) for \(t > t'\). This implies that \(\nabla_{yy} f(y, t) \leq c_2\). Then, the solution of (79) satisfies
\[
\|x(t) - z(t)\|^2 \leq e^{-\frac{2x_2}{\alpha t}} \|x(t') - z(t')\|^2 \tag{80}
\]
Therefore, \(\lim_{t \to \infty} \|x(t) - z(t)\| = 0\). \(\square\)

### 6.7 Proof of Theorem 6

**Proof.** Consider the positive semi-definite function \(V(e) = \frac{1}{2} \|e\|^2 : D \to \mathbb{R}\) as the Lyapunov function for the system (39), where \(D = \{e \in \mathbb{R}^n : \|e\| \leq r_2\}\). Similar to the inequality (72), the derivative of \(V(e)\) along the trajectories of (39) satisfies
\[
\dot{V} \leq - (1 - \theta') \frac{c_2}{\alpha} \|e\|^2, \quad \forall \|e\| \geq \frac{\alpha c_2}{\theta' c_2}, \quad \forall t \in \tilde{I}_t \tag{81}
\]
First, we show that if $u := \frac{\alpha \gamma}{r_2} < r_2$ or $\alpha < \frac{c_2 \theta'}{\gamma}$, the set $D$ has the property that any trajectory starting in $D$ at $t_1$ stays in the set $D$ for all $t \in \bar{I}_{t,2}$. Notice that since the derivative $\dot{V}$ is negative on the boundary $\partial D$, (72) implies that the set $D$ is positively invariant. Since $D$ is also a compact set, it follows from Theorem 1 that (18) has a unique solution defined for all $t \in \bar{I}_{t,2}$ whenever $\epsilon_1 := x_1 - h(t_1) \in D$. Then, the set $D$ being positively invariant implies that

$$x(t, t_1, x_1) \in B_{r_2}(h_2(t)) \subseteq RA(h_2(t)), \quad \forall t \in \bar{I}_{t,2}$$

By choosing $x(t_1, t_0, x_0) = x_1$, one can conclude that $x(t, t_0, x_0)$ will temporarily $r_2$-track $h_2(t)$. Next, we show that if the finite time interval $\bar{I}_{t,2}$ is large enough, the solution of (39) will enter the set $B_u(0) = \{ e \in \mathbb{R}^n : \| e \| \leq u \}$ with an exponential convergence rate and stays in $B_u(0)$ for all future times. Since $\dot{V}$ is negative in $\Gamma = \{ e \in \mathbb{R}^n : u \leq \| e \| \leq r_2 \}$ for all $t \in \bar{I}_{t,2}$, a trajectory starting from $\Gamma$ must move in a direction of decreasing $V(e)$ and the function $V(e)$ will continue decreasing until the trajectory enters the set $B_u(0)$ or until time $t_2$. The fact that the trajectory enters $B_u(0)$ before $t_2$ if $t_2 - t_1 > \frac{\alpha}{c_2(1-\theta')} \ln(\frac{r_2}{u})$ is based on the same argument used in the proof of Theorem 4. 

\[\square\]