

\mathcal{L}_1 Adaptive Optimizer for Online Time-Varying Convex Optimization

Jinrae Kim¹ and Naira Hovakimyan¹

Abstract—We propose an adaptive method for online time-varying (TV) convex optimization, termed \mathcal{L}_1 adaptive optimization (\mathcal{L}_1 -AO). TV optimizers utilize a prediction model to exploit the temporal structure of TV problems, which can be inaccurate in the online implementation. Inspired by \mathcal{L}_1 adaptive control, the proposed method augments an adaptive update law to estimate and compensate for the uncertainty from the prediction inaccuracies. The proposed method provides performance bounds of the error in the optimization variables and cost function, allowing efficient and reliable optimization for TV problems. Numerical simulation results demonstrate the effectiveness of the proposed method for online TV convex optimization.

I. INTRODUCTION

Optimization is critical across a wide range of applications. As streaming data are available in modern engineering systems, it is necessary to solve the optimization problem in real-time in dynamic environments [1], [2], [3], [4]. The challenge is that quite often the optimization objective and constraint functions change over time. These problems are referred to as *time-varying* (TV) optimization.

Solving TV optimization problems at each time instant by conventional time-invariant methods may be computationally burdensome. Instead, one can exploit the underlying temporal structure (dynamics) to predict how the problem varies over time and update the optimization variables accordingly, reducing the computational load significantly [5]. In practice, it is challenging to accurately predict how the cost and constraint functions change over time. The prediction model is employed based on imperfect prior knowledge, making the underlying dynamics *uncertain* in online implementation. The inaccuracy of the prediction model may degrade the performance of the TV algorithm significantly. Therefore, it is crucial to compensate for the uncertainty from the prediction inaccuracies for efficient and reliable TV optimization.

Statement of Contributions: We propose an adaptive method for continuous-time (CT) online TV convex optimization, which recovers the performance of the baseline optimizer by compensating for the uncertainty from the prediction inaccuracies. This allows *robustness certification* w.r.t. the error in optimization variables and optimality gap.

*This work is financially supported by Air Force Office of Scientific Research (AFOSR) grant #FA9550-21-1-0411, National Aeronautics and Space Administration (NASA) grant #80NSSC20M0229, NASA University Leadership Initiative (ULI) grant #80NSSC22M0070, and National Science Foundation (NSF) Information and Intelligent Systems (IIS) grant #23-31878.

¹Jinrae Kim and Naira Hovakimyan are with Department of Mechanical Engineering, The Grainger College of Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801 {jinrae, nhovakim}@illinois.edu

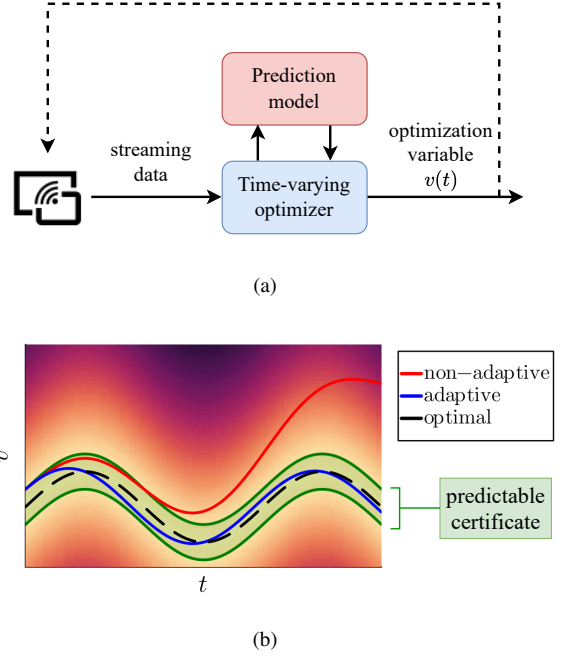


Fig. 1. (a) Time-varying (TV) optimization methods update the optimization variable continuously using a prediction model to track the optimal solution. (b) In the online implementation, the prediction model may be inaccurate, leading to performance degradation. The proposed \mathcal{L}_1 adaptive optimization (\mathcal{L}_1 -AO) compensates for the uncertainty from prediction inaccuracies and provides predictable robustness certification for TV convex optimization problems. The heatmap indicates the TV cost function.

Figure 1 illustrates the proposed method. Let $\Phi(t, v)$ be the smooth and strongly convex cost function such that $\nabla_{vv}\Phi \succeq m_f I$ for some $m_f > 0$. The optimization variable at time t is denoted by $v(t)$. The optimal solution v^* attains zero gradient, i.e., $v^*(t) = \arg \min_v \Phi(t, v) \iff \nabla_v \Phi(t, v^*(t)) = 0$. The gradient dynamics are given as $\dot{\nabla}_v \Phi = \nabla_{vt} \Phi + \nabla_{vv} \Phi \dot{v}$. Consider a baseline TV optimizer \dot{v}_b , designed based on the approximate gradient dynamics $\dot{\nabla}_v \Phi = \hat{\nabla}_{vt} \Phi + \nabla_{vv} \Phi \dot{v}_b$ with a prediction model $\hat{\nabla}_{vt} \Phi$, i.e., $\|\nabla_v \Phi(t, v(t))\| \rightarrow 0$ and $\|v(t) - v^*(t)\| \rightarrow 0$ as $t \rightarrow \infty$ under the approximate dynamics with $\dot{v} = \dot{v}_b$.

Performance: We define the performance by the gradient of the cost function and the tracking error of the optimization variable. First, the proposed method provides a feedback update law $\dot{v} = \dot{v}_b + \dot{v}_a$ such that the gradient is uniformly bounded by $\rho > 0$, i.e., $\|\nabla_v \Phi(t, v(t))\| \leq \rho, \forall t \in [0, t_f]$, where \dot{v}_a is the adaptive update term, $t_f \leq \infty$ is the operation horizon, and $\rho > 0$ is a positive number that is computable *a priori*. Second, the proposed method ensures that the tracking error of the optimization variable is uni-

formly bounded as $\|v(t) - v^*(t)\| \leq \rho/m_f, \forall t \in [0, t_f]$. It implies that the proposed method can provide performance bounds in the presence of prediction inaccuracies.

Optimality: We define the optimality by the gap of the cost function, i.e., $\Phi(t, v(t)) - \Phi(t, v^*(t))$. Similar to the performance, the proposed method guarantees the uniform boundedness of the optimality gap as $0 \leq \Phi(t, v(t)) - \Phi(t, v^*(t)) \leq \rho^2/m_f, \forall t \in [0, t_f]$. It also implies that the proposed method can provide the optimality bound in the presence of prediction inaccuracies.

Related work: TV optimization algorithms have been developed in different time settings including discrete-time (DT) [6] and CT systems [7], [8]. CT algorithms can be considered as the limit case of DT algorithms, desirable for real-time applications such as robotics and control [5], [9]. To improve the tracking performance, a TV optimization algorithm can exploit a model to predict how the TV problem evolves. If such a prediction is not considered, the solution may drift away over time [10]. The prediction model can be data-driven or model-based, and an inaccurate prediction model may lead to performance degradation. Interested readers are referred to a survey paper [5] for more details on TV optimization.

Notation: Let \mathbb{N} and \mathbb{R} be the sets of natural numbers and real numbers, respectively, and $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$. The sets of non-negative and positive real numbers are denoted by $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{> 0}$, respectively. For a natural number $p \in \mathbb{N}$, let $[p] := \{1, \dots, p\}$. Given a matrix A , let $\underline{\lambda}(A)$ be the smallest real part of A 's eigenvalues. Given $n \in \mathbb{N}$, I_n denotes the $n \times n$ identity matrix. The function argument may be simplified or omitted if the context is clear, e.g., $\Phi(t) = \Phi(t, v(t))$.

II. PRELIMINARIES

Consider the following TV optimization problem given by

$$\min_v f_0(t, v) \quad \text{subject to } f_i(t, v) \leq 0, \quad \forall i \in [p], \quad (1)$$

where $t \in \mathbb{R}_{\geq 0}$ denotes the time, $v \in \mathbb{R}^{n_v}$ is the optimization variable, $f_0 : \mathbb{R}_{\geq 0} \times \mathbb{R}^{n_v} \rightarrow \mathbb{R}$ is the cost function, and $f_i : \mathbb{R}_{\geq 0} \times \mathbb{R}^{n_v} \rightarrow \mathbb{R}$ are inequality constraints, $\forall i \in [p]$. Suppose that $f_i(t, v)$ are twice differentiable w.r.t. t and three-times differentiable w.r.t. v , $\forall i \in \{0\} \cup [p]$, and $f_0(t, \cdot)$ is uniformly strongly convex, i.e., $\nabla_{vv} f_0(t, v) \succeq m_f I_{n_v}, \forall (t, v)$ for some $m_f > 0$. The problem can be approximated by an unconstrained problem with log-barrier functions as follows [7]:

$$\min_v \Phi(t, v) := f_0(t, v) - \frac{1}{c(t)} \sum_{i \in [p]} \log(-f_i(t, v)), \quad (2)$$

where $c : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{> 0}$ is the penalty parameter that controls the quality of the approximation. In this regard, only unconstrained problems will be considered in this study. The minimizer can be characterized as $v^*(t) = \arg \min_v \Phi(t, v) \iff \nabla_v \Phi(t, v^*(t)) = 0$.

Solving the optimization problem in (2) for each t may be computationally burdensome in real-time applications. Instead, one can exploit the temporal structure to find the

minimizer efficiently [5]. Consider the gradient dynamics as follows:

$$\dot{\nabla}_v \Phi(t) = \nabla_{vt} \Phi(t) + \nabla_{vv} \Phi(t) \dot{v}(t), \quad (3)$$

with $\nabla_v \Phi(0) = \nabla_v \Phi(0, v(0))$. The rate of optimization variables, \dot{v} , can be viewed as the *control input* of the gradient dynamics. Then, an update law $\dot{v}(t)$ can be designed to regulate the gradient to zero.

In the online implementation, the knowledge about $\frac{\partial f_i}{\partial t}$ and $\nabla_{vt} f_i, \forall i \in \{0\} \cup [p]$ may be inaccurate, which appears in the term $\nabla_{vt} \Phi$ in (3). For example, external streaming data $d(t)$, such as system state or environmental data, may determine the optimization problem, i.e., $\Phi(t, v) = \chi(d(t), v)$ for some function χ , while the knowledge about the dynamics of streaming data, $\dot{d}(t)$, is limited. Suppose a nominal prediction model $\tilde{\nabla}_{vt} \Phi$ is available. By rearranging (3), the prediction error $\tilde{\nabla}_{vt} \Phi := \tilde{\nabla}_{vt} \Phi - \nabla_{vt} \Phi$ can be viewed as *uncertainty* of the gradient system as

$$\begin{aligned} \dot{\nabla}_v \Phi(t) &= \tilde{\nabla}_{vt} \Phi(t) + \nabla_{vv} \Phi(t) \dot{v}(t) - \tilde{\nabla}_{vt} \Phi(t) \\ &=: \hat{\nabla}_{vt} \Phi(t) + \nabla_{vv} \Phi(t) (\dot{v}(t) + \sigma(t)), \end{aligned} \quad (4)$$

where $\sigma(t) := -\nabla_{vv} \Phi^{-1}(t) \tilde{\nabla}_{vt} \Phi(t)$.

We define norm balls and tubes w.r.t. the gradient and the error in optimization variables for the problem statement.

Definition 1. Given $\rho > 0$, define the norm balls w.r.t. the gradient and the error in optimization variables as $\mathcal{O}_{\nabla}(t, \rho) := \{v \mid \|\nabla_v \Phi(t, v)\| \leq \rho\}$ and $\mathcal{O}(t, \rho) := \{v \mid \|v - v^*(t)\| \leq \rho\}$, respectively, for all $t \geq 0$. Also, define the tubes w.r.t. the gradient and optimization variables as $\mathcal{T}_{\nabla}(\rho) := \cup_{t \geq 0} \mathcal{O}_{\nabla}(t, \rho)$ and $\mathcal{T}(\rho) := \cup_{t \geq 0} \mathcal{O}(t, \rho)$, respectively.

Problem Statement: Given a time-varying optimization problem in (2), the objective is to design the rate of the optimization variables, $\dot{v}(t)$, that guarantees a performance bound with pre-computable tube width $\rho > 0$, such that $v(t) \in \mathcal{O}_{\nabla}(t, \rho) \subset \mathcal{T}_{\nabla}(\rho), \forall t \geq 0$ in the presence of the prediction error.

In the following, we introduce two methods related to \mathcal{L}_1 -AO: i) continuous-time prediction-correction interior-point method and ii) \mathcal{L}_1 adaptive control.

A. Prediction-correction Interior-point method

In [7], a continuous-time prediction-correction interior-point (PCIP) method was proposed as follows:

$$\dot{v}(t) = -\nabla_{vv} \Phi(t)^{-1} (\nabla_{vt} \Phi(t) + P \nabla_v \Phi(t)), \quad (5)$$

with $v(0) = v_0$, where $P \succeq \beta I_{n_v}$ is the optimizer gain with $\beta > 0$, which determines the convergence rate. The terms $\nabla_{vt} \Phi$ and $\nabla_v \Phi$ correspond to the *prediction* and *correction* terms, which can be interpreted as *feedforward* and *feedback* terms in control [7]. The prediction term tracks the changes in cost function over time, and the correction term pushes the optimization variables to the optimal solution. The update law forms a closed-loop dynamics $\dot{\nabla}_v \Phi = -P \nabla_v \Phi$. Considering the Lyapunov function $V(\nabla_v \Phi) := \frac{1}{2} \nabla_v \Phi^\top \nabla_v \Phi$, it can be shown that $\dot{V} \leq -2\beta V$ under the PCIP update law,

which implies the exponential convergence of the gradient to the origin [7].

For online implementation, a modified PCIP was proposed in [7]: if a prediction model with globally uniformly bounded prediction error is available, i.e., $\|\tilde{\nabla}_{vt}\Phi(t, v)\| \leq \eta, \forall(t, v)$ for some $\eta > 0$, the modified PCIP update law is given by

$$\dot{v}(t) = -\nabla_{vv}^{-1}\Phi(t) \left(\hat{\nabla}_{vt}\Phi(t) + \frac{P\nabla_v\Phi(t)}{\max(\|\nabla_v\Phi(t)\|, \epsilon)} \right), \quad (6)$$

with $v(0) = v_0$. The inequalities $\beta > \eta$ and $\epsilon > 0$ imply $\lim_{t \rightarrow \infty} \|\nabla_v\Phi(t, v)\| \leq \eta\epsilon/\beta$ [7]. The modified PCIP may suffer from several potential issues: i) high gain is required to reduce the asymptotic bound, which is undesirable for numerical stability, and ii) the assumption that the prediction error is globally uniformly bounded may be unrealistic.

B. \mathcal{L}_1 Adaptive Control

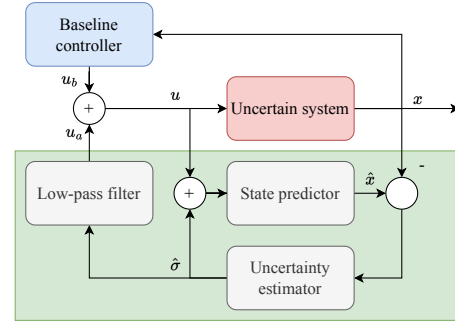
Consider the following system dynamics:

$$\dot{x}(t) = f(x(t)) + B(x(t))(u(t) + \sigma(t, x(t))), \quad (7)$$

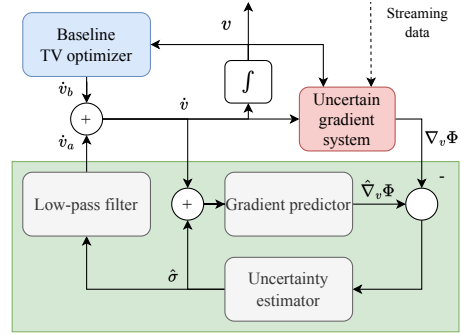
with $x(0) = x_0$, where $x : \mathbb{R}_{\geq 0} \rightarrow X \subset \mathbb{R}^{n_x}$ and $u : \mathbb{R}_{\geq 0} \rightarrow U \subset \mathbb{R}^{n_u}$ denote the state and control input, respectively, $f : X \rightarrow \mathbb{R}^{n_x}$ and $B : X \rightarrow \mathbb{R}^{n_x \times n_u}$ are known, and $\sigma : \mathbb{R}_{\geq 0} \times X \rightarrow \mathbb{R}^{n_u}$ is the matched uncertainty. We assume $B(x)$ has full row rank for all $x \in X$. We further assume the existence of a *baseline control input* $u_b(t)$ such that $0 \in \mathbb{R}^{n_x}$ is a globally exponentially stable equilibrium of (7) with $u(t) = u_b(t)$ and $\sigma \equiv 0$. In the presence of the uncertainty $\sigma \neq 0$, the input $u(t) = u_b(t)$ may not be able to stabilize (7), leading to an unpredictable and unquantifiable divergence of $x(t)$. Hence, we design the input $u(t)$ as $u(t) = u_b(t) + u_a(t)$, where $u_a(t)$ denotes the \mathcal{L}_1 adaptive control input. The \mathcal{L}_1 adaptive control (\mathcal{L}_1 -AC) is a robust adaptive control methodology that decouples the adaptation loop from the control loop, which enables arbitrarily fast adaptation with guaranteed robustness margins [11]. The \mathcal{L}_1 -AC has been successfully implemented on NASA AirSTAR sub-scale aircraft [12] and Calspan's variable-stability Learjet [13]. The \mathcal{L}_1 -AC input $u_a(t)$ is defined through the following three components: i) state predictor, ii) uncertainty estimator (adaptation law), and iii) low-pass filter. Fig. 2(a) illustrates the architecture of the \mathcal{L}_1 -AC. The state predictor is governed by the following equation:

$$\dot{\hat{x}}(t) = A_s \hat{x}(t) + f(x(t)) + B(x(t))(u(t) + \hat{\sigma}(t)), \quad (8)$$

with $\hat{x}(0) = x_0$, where $A_s \in \mathbb{R}^{n_x \times n_x}$ is a Hurwitz matrix, $\hat{x} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^{n_x}$ is the predictor state, $\hat{\sigma} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^{n_u}$ is the estimated uncertainty, and $\tilde{x} := \hat{x} - x$. The estimated uncertainty $\hat{\sigma}(t)$ can be updated using adaptation laws such as projection-based adaptation or piecewise-constant adaptation [11], [14]. The adaptive control input is then given by $u_a(s) = C(s)(-\hat{\sigma}(s))$, where $C(s)$ is a low-pass filter with *bandwidth* ω and satisfies $C(0) = 1$. The \mathcal{L}_1 -AC input ensures the existence of an *a priori* $\rho > 0$ such that the state $x(t)$ of the uncertain system (7) with input $u(t) = u_b(t) + u_a(t)$ satisfies the *uniform* bound $\|x\|_{\mathcal{L}_\infty} \leq \rho$. The value of the bound $\rho > 0$ is determined by the choice



(a) \mathcal{L}_1 -AC



(b) \mathcal{L}_1 -AO (proposed)

Fig. 2. Architecture comparison between \mathcal{L}_1 adaptive control (\mathcal{L}_1 -AC) and the proposed \mathcal{L}_1 adaptive optimization (\mathcal{L}_1 -AO). Both architectures are designed to compensate for the uncertainty to recover the performance of the baseline controller or optimizer. (a) In \mathcal{L}_1 -AC, the state predictor predicts the state of the uncertain (physical) system. (b) In \mathcal{L}_1 -AO, the *gradient system* (a virtual system) is considered. The rate of optimization variables, \dot{v} , is regarded as the control input, and the gradient predictor is used to predict the gradient of the cost function.

of the filter bandwidth and the sampling period, ω and T_s , respectively. Further details on the design of \mathcal{L}_1 -AC can be found in [11], [14], [15], [16].

III. \mathcal{L}_1 ADAPTIVE OPTIMIZATION

The main observation from Section II is that a TV optimization problem with an inaccurate prediction model can be interpreted as an output regulation problem of an uncertain nonlinear system, where the output is the gradient. This interpretation allows us to derive an adaptive TV optimization algorithm based on robust adaptive control.

A. Algorithm and Architecture

We propose an adaptive method for online TV optimization, termed \mathcal{L}_1 adaptive optimization (\mathcal{L}_1 -AO). We assume the existence of a *baseline update law* $\dot{v}_b(t)$ such that the update law $\dot{v}(t) = \dot{v}_b(t)$ regulates the gradient to the origin in the absence of uncertainty ((4) with $\sigma \equiv 0$) as follows.

Assumption 1. Suppose there exists a Lyapunov function V such that the following holds for the nominal gradient dynamics ((4) with $\sigma \equiv 0$) under the baseline update law $\dot{v}(t) = \dot{v}_b(t)$:

$$\alpha \|\nabla_v \Phi\|^2 \leq V(\nabla_v \Phi) \leq \bar{\alpha} \|\nabla_v \Phi\|^2, \quad \dot{V} \leq -2\beta V, \quad (9)$$

for some positive numbers $\underline{\alpha}$, $\bar{\alpha}$, and β .

The \mathcal{L}_1 -AO consists of the following three components: i) gradient predictor, ii) uncertainty estimator (adaptation law), and iii) low-pass filter. Adopting the piece-wise constant adaptation law [14], each component can be described as

$$\begin{aligned}\dot{\hat{\nabla}}_v \Phi(t) &= A_s \tilde{\nabla}_v \Phi(t) + \hat{\nabla}_{vt} \Phi(t) + \nabla_{vv} \Phi(t) \dot{v}(t) + h(t), \\ h(iT_s) &= \mu(A_s, T_s) \tilde{\nabla}_v \Phi(iT_s), \quad \forall i \in \mathbb{N} \\ \hat{\sigma}(iT_s) &= \nabla_{vv} \Phi^{-1}(iT_s) h(iT_s), \quad \forall i \in \mathbb{N} \\ \dot{v}_a(s) &= C(s)(-\hat{\sigma}(s)),\end{aligned}\quad (10)$$

with $\hat{\nabla}_v \Phi(0) = \nabla_v \Phi(0, v_0)$, $h(0) = \hat{\sigma}(0) = 0$, and $\dot{v} = \dot{v}_b + \dot{v}_a$, where

$$\mu(A_s, T_s) := (A_s^{-1}(I - e^{A_s T_s}))^{-1} e^{A_s T_s}, \quad (11)$$

and $h(t) := h(iT_s)$ and $\hat{\sigma}(t) := \hat{\sigma}(iT_s)$, $\forall t \in [iT_s, (i+1)T_s)$ for any $i \in \mathbb{N}_0$, represent the estimation of $\nabla_{vv} \Phi(t) \sigma(t)$ and $\sigma(t)$, respectively. $A_s \in \mathbb{R}^{n_v \times n_v}$ is a diagonal Hurwitz matrix, $T_s > 0$ is the sampling period of the adaptation, and $\tilde{\nabla}_v \Phi := \hat{\nabla}_v \Phi - \nabla_v \Phi$. Consider a low-pass filter $C(s) = \frac{\omega}{s+\omega}$ with a filter parameter $\omega > 0$. From (4) and (10), the prediction error dynamics are:

$$\dot{\tilde{\nabla}}_v \Phi(t) = A_s \tilde{\nabla}_v \Phi(t) + h(t) - \nabla_{vv} \Phi(t) \sigma(t). \quad (12)$$

Figure 2(b) illustrates the architecture of the \mathcal{L}_1 -AO.

B. Performance and Optimality Analysis

In this section, the performance and optimality of the proposed \mathcal{L}_1 -AO algorithm in (10) are analyzed.

For arbitrarily chosen $\epsilon > 0$, let $\rho := \sqrt{\bar{\alpha}/\underline{\alpha}} \|\nabla_v \Phi(0)\| + \epsilon$. Next, we assume the following regularity condition.

Assumption 2. Suppose that there exist positive numbers $\Delta_{(\cdot)}$'s such that the following inequalities hold:

$$\begin{aligned}\|\dot{v}_b(t, v)\| &\leq \Delta_{\dot{v}_b}, \quad \left\| \frac{\partial V(\nabla_v \Phi(t, v))}{\partial \nabla_v \Phi} \right\| \leq \Delta_{\frac{\partial V}{\partial \nabla_v \Phi}}, \quad (13) \\ \|\hat{\nabla}_{vt} \Phi(t, v)\| &\leq \Delta_{\hat{\nabla}_{vt} \Phi}, \quad \|\tilde{\nabla}_{vt} \Phi(t, v)\| \leq \Delta_{\tilde{\nabla}_{vt} \Phi}, \\ \|\tilde{\nabla}_{vtt} \Phi(t, v)\| &\leq \Delta_{\tilde{\nabla}_{vtt} \Phi}, \quad \|\tilde{\nabla}_{vvt} \Phi(t, v)\| \leq \Delta_{\tilde{\nabla}_{vvt} \Phi}, \\ \|\nabla_{vv} \Phi(t, v)\| &\leq \Delta_{\nabla_{vv} \Phi}, \quad \|\nabla_{vvt} \Phi(t, v)\| \leq \Delta_{\nabla_{vvt} \Phi}, \\ \|\nabla_{vvv} \Phi(t, v)\| &\leq \Delta_{\nabla_{vvv} \Phi},\end{aligned}$$

for all $t \geq 0$ and $v \in \mathcal{T}_\nabla(\rho)$.

Based on the above assumption, let

$$\Delta_\sigma := m_f^{-1} \Delta_{\tilde{\nabla}_{vt} \Phi}, \quad (14)$$

$$\Delta_{\hat{\sigma}} := \sqrt{n_v} m_f^{-1} \Delta_{\nabla_{vv} \Phi} \Delta_\sigma, \quad (15)$$

$$\Delta_{\dot{v}} := \Delta_{\dot{v}_b} + \Delta_{\hat{\sigma}}, \quad (16)$$

$$\Delta_{\dot{\nabla}_{vv} \Phi} := \Delta_{\nabla_{vvt} \Phi} + \Delta_{\nabla_{vvv} \Phi} \Delta_{\dot{v}}, \quad (17)$$

$$\Delta_{\dot{\nabla}_{vt} \Phi} := \Delta_{\tilde{\nabla}_{vtt} \Phi} + \Delta_{\tilde{\nabla}_{vvt} \Phi} \Delta_{\dot{v}}, \quad (18)$$

$$\Delta_{\dot{\sigma}} := m_f^{-1} \left(\Delta_{\tilde{\nabla}_{vv} \Phi} m_f^{-1} \Delta_{\tilde{\nabla}_{vt} \Phi} + \Delta_{\dot{\nabla}_{vv} \Phi} \right), \quad (19)$$

$$\zeta_1(\omega) := \Delta_{\frac{\partial V}{\partial \nabla_v \Phi}} \Delta_{\nabla_{vv} \Phi} \left(\frac{\Delta_\sigma}{|2\beta - \omega|} + \frac{\Delta_{\dot{\sigma}}}{2\beta\omega} \right), \quad (20)$$

$$\begin{aligned}\zeta_2(A_s) &:= m_f^{-1} \sqrt{n_v} \left((2\Delta_{\hat{\sigma}} + \lambda(-A_s) \Delta_\sigma) \Delta_{\nabla_{vv} \Phi} \right. \\ &\quad \left. + \Delta_\sigma \Delta_{\dot{\nabla}_{vv} \Phi} \right),\end{aligned}\quad (21)$$

$$\zeta_3(\omega) := \Delta_\sigma \omega, \quad (22)$$

$$\zeta_4(A_s, \omega) := \Delta_{\frac{\partial V}{\partial \nabla_v \Phi}} \Delta_{\nabla_{vv} \Phi} (\zeta_2(A_s) + \zeta_3(\omega)) / (2\beta). \quad (23)$$

Choose the sampling time $T_s > 0$ and the filter bandwidth $\omega > 0$ such that the following inequalities hold:

$$\begin{aligned}\underline{\alpha} \rho^2 &> \zeta_1(\omega) + V_0, \\ T_s &\leq (\underline{\alpha} \rho^2 - \zeta_1(\omega) - V_0) / \zeta_4,\end{aligned}\quad (24)$$

where $V_0 := V(\nabla_v \Phi(0, v(0)))$. Note that $\zeta_1(\omega) \rightarrow +\infty$ as $\omega \rightarrow +\infty$. Therefore, from the definition of ρ and Assumption 1, the conditions in (24) can be satisfied by choosing sufficiently large ω and small T_s .

We now present the main result.

Theorem 1. Suppose Assumption 1, Assumption 2, and (24) hold. Then, under the \mathcal{L}_1 adaptive optimization algorithm in (10), we have

$$v(t) \in \mathcal{O}_\nabla(t, \rho) \subset \mathcal{O}(t, \rho/m_f), \quad \forall t \geq 0. \quad (25)$$

Furthermore, for any $t_1 > 0$, we have

$$\begin{aligned}v(t) &\in \mathcal{O}_\nabla \left(t, \sqrt{(e^{-2\beta t_1} V_0 + \zeta_1 + \zeta_4 T_s) / \underline{\alpha}} \right) \\ &\subset \mathcal{O} \left(t, m_f^{-1} \sqrt{(e^{-2\beta t_1} V_0 + \zeta_1 + \zeta_4 T_s) / \underline{\alpha}} \right),\end{aligned}\quad (26)$$

for all $t \geq t_1$.

Proof. See Appendix B. \square

The main theorem implies that the gradient and the error in optimization variables are uniformly bounded for all $t \geq 0$ and uniformly ultimately bounded for all $t \geq t_1$ for any $t_1 > 0$. The following corollary shows the boundedness of the optimality gap.

Corollary 1. Under the assumptions of Theorem 1, for all $t \geq 0$ we have:

$$0 \leq \Phi(t, v(t)) - \Phi(t, v^*(t)) \leq \rho^2 / m_f. \quad (27)$$

Also, for any $t_1 > 0$, we have for all $t \geq t_1$:

$$0 \leq \Phi(t, v(t)) - \Phi(t, v^*(t)) \leq \frac{e^{-2\beta t_1} V_0 + \zeta_1 + \zeta_4 T_s}{m_f \underline{\alpha}}. \quad (28)$$

Proof. See Appendix C. \square

Remark 1. Assumption 1, and (13) in Assumption 2 hold for the PCIP method in (5) with $\underline{\alpha} = \bar{\alpha} = 1/2$, $\beta = \lambda(P)$, $\Delta_{\dot{v}_b} = m_f^{-1} (\Delta_{\tilde{\nabla}_{vt} \Phi} + \|P\| \rho)$, and $\Delta_{\frac{\partial V}{\partial \nabla_v \Phi}} = \rho$. The Assumption 2 is for $v \in \mathcal{T}_\nabla(\rho)$, not assuming $v(t) \in \mathcal{T}_\nabla(\rho)$, $\forall t \geq 0$.

IV. NUMERICAL SIMULATIONS

In this section, numerical simulations are performed for time-varying optimization problems to demonstrate the effectiveness of the proposed method. Simulation settings can be found in Appendix D.

A. Example 1: Constrained TV optimization

Consider the following TV optimization problem:

$$\min_v f_0(t, v) = \frac{1}{2}v^2 \quad (29)$$

$$\text{subject to } f_1(t, v) = v + d(t) \leq 0, \quad (30)$$

with external streaming data $d(t) = 3 \sin(3t)$. The problem is approximated by a log-barrier function as follows:

$$\min_v \Phi(t, v) = \frac{1}{2}v^2 - \log(-(v + d(t))), \quad (31)$$

which implies $\nabla_{vt}\Phi(t, v) = \dot{d}(t)/(v + d(t))^2$. The nominal prediction model assumes that $d(t)$ slowly varies over time, i.e., $\dot{d}(t) \approx 0$, which implies $\hat{\nabla}_{vt}\Phi(t, v) = 0$.

Results: The simulation results are illustrated in Figure 3. As shown in Fig. 3(a), the modified PCIP method with gain $P = 10$ could not compensate for the prediction inaccuracies effectively and violated the inequality constraint, causing early termination of the simulation (denoted by a magenta star). The same simulation with gain $P = 10^3$ did not violate the constraint and tracked the optimal solution. However, this induced an extremely large update rate at the beginning phase, undesirable for numerical stability. Note that the modified PCIP method with gains larger than $P = 10^3$ caused failures of the numerical integration solvers (immediate termination after one time step). On the other hand, as shown in Fig. 3(b), the proposed \mathcal{L}_1 -AO successfully tracks the optimal solution without any excessive update rate. The optimization variable is confined within a green tube $\mathcal{T}(\rho)$ with $\rho = 0.28$, satisfying the constraint in the presence of prediction inaccuracies.

Table I shows each method's mean elapsed times (total elapsed time divided by the number of time steps). ECOS [17], a time-invariant solver, results in the mean elapsed time larger than the time step $\Delta t = 10^{-3}$, which is not applicable to real-time scenarios. On the other hand, modified PCIP and \mathcal{L}_1 -AO, which are time-varying methods, show significantly small mean elapsed times. The mean elapsed time of \mathcal{L}_1 -AO is larger than that of modified PCIP due to extra computation for adaptation. The mean elapsed time of \mathcal{L}_1 -AO is approximately 1/120 of that of ECOS and 1/50 of that of time step Δt , making it desirable for real-time scenarios.

TABLE I

EXAMPLE 1: MEAN ELAPSED TIME FOR A TIME STEP

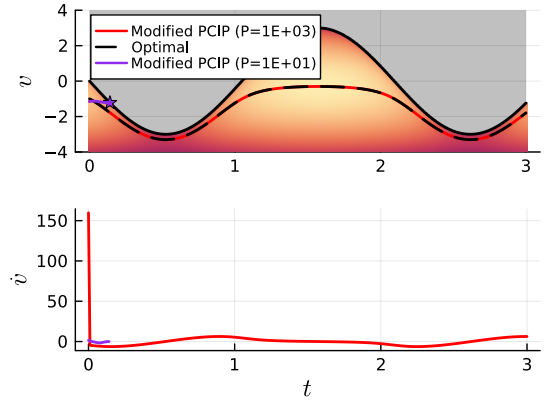
ECOS	Modified PCIP* ($P = 10^3$)	\mathcal{L}_1 -AO (baseline: PCIP)
2.2837 ms	0.0063 ms	0.0187 ms

*: Modified PCIP with $P = 10$ violated constraint

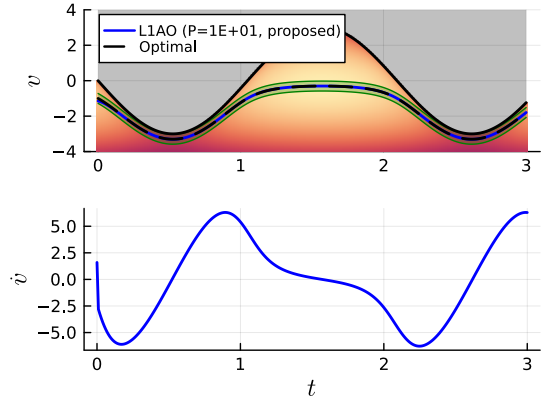
B. Example 2: Robot navigation

We consider a robot navigation problem borrowed from [7, Section IV]. In the robot navigation task, a projected goal is used for collision avoidance, which minimizes the following cost function:

$$\Phi(t, x) = \frac{1}{2}\|x - x_d\|^2 - \frac{1}{c(t)} \sum_{i=1}^m \log(b_i(x_c) - a_i(x_c)^\top x),$$



(a) Modified PCIP



(b) \mathcal{L}_1 -AO (baseline: PCIP)

Fig. 3. Example 1: Simulation result. The gray area denotes the TV constraint. The heatmap indicates the cost function. (a) The modified PCIP with $P = 10$ violated the TV constraint (magenta star). The modified PCIP with a high gain $P = 10^3$ induced a huge update rate at the initial phase, undesirable for numerical stability. (b) The \mathcal{L}_1 -AO updates the optimization variable without an abrupt update. The optimization variable is confined within a green tube, preventing constraint violation.

where $x \in \mathbb{R}^2$ is the optimization variable. The positions of robot and target at time t , $x_c(t) \in \mathbb{R}^2$ and $x_d(t) \in \mathbb{R}^2$, respectively, can be regarded as streaming data. Other details can be found in [7], [2]. The nominal prediction model assumes that the target is static, i.e., $\dot{x}_d(t) \approx 0$. Optimization algorithms produce the estimate $\hat{x}(t)$ of the projected goal, corresponding to the optimization variable $v(t)$ in this paper.

Results: The simulation results are illustrated in Figure 4. As shown in Fig. 4, PCIP exhibited a large tracking error due to the prediction inaccuracies. Modified PCIP showed relatively low tracking error due to the high gain. For both PCIP and modified PCIP, the augmentation of \mathcal{L}_1 -AO significantly improved the tracking performance. The simulation results demonstrate that the \mathcal{L}_1 -AO can effectively compensate for the prediction inaccuracies in online TV optimization.

V. CONCLUSIONS

In this study, we proposed \mathcal{L}_1 adaptive optimization (\mathcal{L}_1 -AO), an adaptive method for online time-varying (TV) convex optimization. The \mathcal{L}_1 -AO has a baseline TV optimizer,

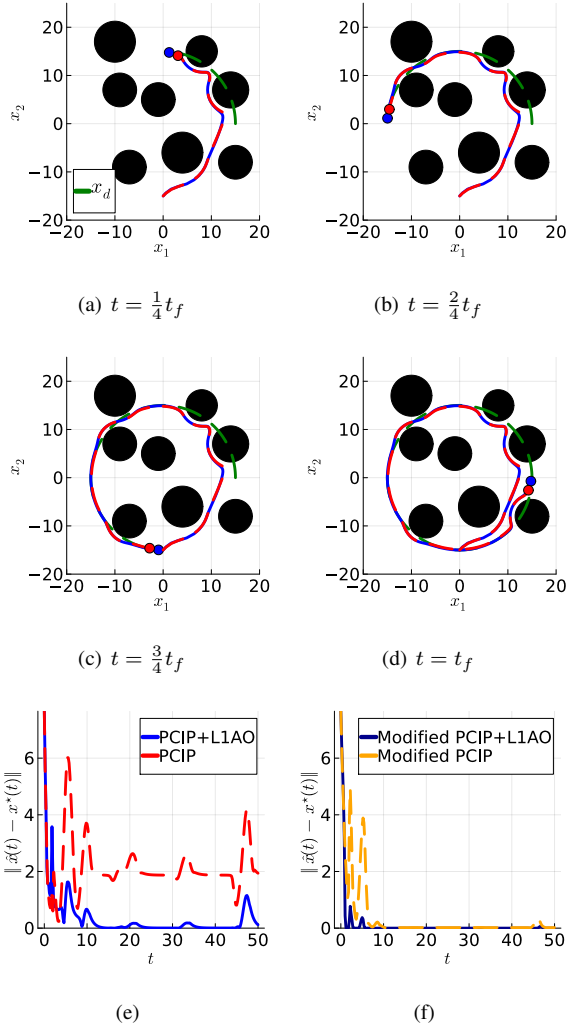


Fig. 4. Example 2: (a)-(d) 2D trajectories of a robot. For visualization, only the results of PCIP and \mathcal{L}_1 -AO with the baseline of PCIP are depicted. (e)-(f) Error in optimization variable: (e) PCIP vs \mathcal{L}_1 -AO (baseline: PCIP), (f) Modified PCIP vs \mathcal{L}_1 -AO (baseline: Modified PCIP).

which updates the optimization variables with a prediction model. The prediction model leverages the underlying temporal structure of TV problems, allowing efficient TV optimization. However, the prediction model can be inaccurate in online implementation, degrading the performance of TV optimizers. The \mathcal{L}_1 -AO estimates and compensates for the uncertainty from the prediction inaccuracies, providing robustness certification with a pre-computable tube width. Numerical simulation results showed that the proposed \mathcal{L}_1 -AO can solve the TV optimization efficiently and reliably in the presence of prediction inaccuracies compared to existing time-invariant and non-adaptive TV methods.

In TV optimization, the streaming data may be influenced by the decision (optimization variable) through a physical dynamical system, which forms a closed-loop feedback with the gradient dynamics. Future work includes analysis with such feedback, relaxing strong convexity for a wider class of problems, and more realistic applications of real-time

TV optimization including robot perception and motion planning.

REFERENCES

- [1] E. F. Camacho and C. Bordons, *Model Predictive Control*. Advanced Textbooks in Control and Signal Processing, London, UK: Springer, 2nd ed., 2007.
- [2] O. Arslan and D. E. Koditschek, "Exact Robot Navigation Using Power Diagrams," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, (Stockholm, Sweden), pp. 1–8, May 2016.
- [3] O. Arslan and D. E. Koditschek, "Sensor-based Reactive Navigation in Unknown Convex Sphere Worlds," *The International Journal of Robotics Research*, vol. 38, no. 2-3, pp. 196–223, 2019.
- [4] Y. Yang, M. Pesavento, M. Zhang, and D. P. Palomar, "An Online Parallel Algorithm for Recursive Estimation of Sparse Signals," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 3, pp. 290–305, 2016.
- [5] A. Simonetto, E. Dall'Anese, S. Paternain, G. Leus, and G. B. Giannakis, "Time-Varying Convex Optimization: Time-Structured Algorithms and Applications," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 2032–2048, 2020.
- [6] A. Simonetto and E. Dall'Anese, "Prediction-Correction Algorithms for Time-Varying Constrained Optimization," *IEEE Transactions on Signal Processing*, vol. 65, no. 20, pp. 5481–5494, 2017.
- [7] M. Fazlyab, S. Paternain, V. M. Preciado, and A. Ribeiro, "Prediction-Correction Interior-Point Method for Time-Varying Convex Optimization," *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 1973–1986, 2018.
- [8] S. Rahili and W. Ren, "Distributed Continuous-Time Convex Optimization With Time-Varying Cost Functions," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1590–1605, 2017.
- [9] J. Kim, J. L. Bullock, S. Cheng, and N. Hovakimyan, "Smooth Reference Command Generation and Control for Transition Flight of VTOL Aircraft Using Time-Varying Optimization," in *AIAA SciTech 2025 Forum*, (Orlando, Florida), Jan. 2025.
- [10] A. Y. Popkov, "Gradient Methods for Nonstationary Unconstrained Optimization Problems," *Automation and Remote Control*, vol. 66, no. 6, pp. 883–891, 2005.
- [11] N. Hovakimyan and C. Cao, *\mathcal{L}_1 Adaptive Control Theory: Guaranteed Robustness with Fast Adaptation*. SIAM, 2010.
- [12] I. Gregory, C. Cao, E. Xargay, N. Hovakimyan, and X. Zou, " \mathcal{L}_1 Adaptive Control Design for NASA AirSTAR Flight Test Vehicle," in *AIAA Guidance, Navigation, and Control Conference*, (Chicago, Illinois), American Institute of Aeronautics and Astronautics, Aug. 2009.
- [13] K. A. Ackerman, E. Xargay, R. Choe, N. Hovakimyan, M. C. Cotting, R. B. Jeffrey, M. P. Blackstun, T. P. Fulkerson, T. R. Lau, and S. S. Stephens, "Evaluation of an \mathcal{L}_1 Adaptive Flight Control Law on Calspan's Variable-Stability Learjet," *Journal of Guidance, Control, and Dynamics*, vol. 40, no. 4, pp. 1051–1060, 2017.
- [14] Z. Wu, S. Cheng, P. Zhao, A. Gahlawat, K. A. Ackerman, A. Lakshmanan, C. Yang, J. Yu, and N. Hovakimyan, " \mathcal{L}_1 quad: \mathcal{L}_1 Adaptive Augmentation of Geometric Control for Agile Quadrotors with Performance Guarantees," *IEEE Transactions on Control Systems Technology*, vol. 33, no. 2, pp. 597–612, 2025.
- [15] X. Wang and N. Hovakimyan, " \mathcal{L}_1 Adaptive Controller for Nonlinear Time-varying Reference Systems," *Systems & Control Letters*, vol. 61, no. 4, pp. 455–463, 2012.
- [16] A. Lakshmanan, A. Gahlawat, and N. Hovakimyan, "Safe Feedback Motion Planning: A Contraction Theory and \mathcal{L}_1 -Adaptive Control Based Approach," in *2020 59th IEEE Conference on Decision and Control (CDC)*, (Jeju, Korea (South)), pp. 1578–1583, IEEE, Dec. 2020.
- [17] A. Domahidi, E. Chu, and S. Boyd, "ECOS: An SOCP Solver for Embedded Systems," in *European Control Conference (ECC)*, (Zurich, Switzerland), July 2013.
- [18] H. K. Khalil, *Nonlinear Systems*. Upper Saddle River, NJ: Prentice Hall, 3rd ed., 2002.

APPENDIX

A. Lemmas for Theorem 1

In this section, we present lemmas for Theorem 1.

Lemma 1. For all $t \geq 0$ and $\rho > 0$, $\mathcal{O}_\nabla(t, \rho) \subset \mathcal{O}(t, \rho/m_f)$ and $\mathcal{T}_\nabla(\rho) \subset \mathcal{T}(\rho/m_f)$.

Proof. Using mean-value theorem with respect to the second argument of $\nabla_v \Phi(t, v)$ around $v^*(t)$ implies

$$\nabla_v \Phi(t, v) - \nabla_v \Phi(t, v^*(t)) = \nabla_{vv} \Phi(t, \eta(t))(v - v^*(t)),$$

where $\eta(t)$ is a convex combination of v and $v^*(t)$. Then,

$$\begin{aligned} \|v - v^*(t)\| &= \|\nabla_{vv}^{-1} \Phi(t, \eta(t)) \nabla_v \Phi(t, v)\| \\ &\leq \|\nabla_{vv}^{-1} \Phi(t, \eta(t))\| \|\nabla_v \Phi(t, v)\| \\ &\leq m_f^{-1} \|\nabla_v \Phi(t, v)\|, \end{aligned}$$

since $\nabla_v \Phi(t, v^*(t)) = 0$ and $\|\nabla_{vv}^{-1} \Phi(t, v)\| \leq m_f^{-1}$ from the uniform strong convexity of $\Phi(t, \cdot)$, which concludes the proof. \square

Lemma 2. If $v(t) \in \mathcal{T}_\nabla(\rho)$, $\forall t \in [0, \tau]$ for some $\tau > 0$, then the following relationships hold:

$$\|\hat{\sigma}\|_{\mathcal{L}_\infty}^{[0, \tau]} \leq \Delta_{\hat{\sigma}}, \quad (32)$$

$$\|\dot{v}_a\|_{\mathcal{L}_\infty}^{[0, \tau]} \leq \Delta_{\hat{\sigma}}. \quad (33)$$

$$\|\dot{\sigma}\|_{\mathcal{L}_\infty}^{[0, \tau]} \leq \Delta_{\hat{\sigma}}. \quad (34)$$

Proof. Proof of (32): It is sufficient to show that $\|\hat{\sigma}(iT_s)\| \leq \Delta_{\hat{\sigma}}$, $\forall i \in \{0, \dots, \lfloor \tau/T_s \rfloor\}$. First, $\hat{\sigma}(0) = 0$. Integrating (12) from iT_s to $(i+1)T_s$ implies

$$\begin{aligned} \tilde{\nabla}_v \Phi((i+1)T_s) &= e^{A_s T_s} \tilde{\nabla}_v \Phi(iT_s) \\ &+ \int_{iT_s}^{(i+1)T_s} e^{A_s((i+1)T_s - t)} (h(iT_s) - \nabla_{vv} \Phi(t) \sigma(t)) dt, \end{aligned} \quad (35)$$

for all $i \in \mathbb{N}_0$. It can be deduced from (10) and (35) that the following holds:

$$\begin{aligned} \tilde{\nabla}_v \Phi((i+1)T_s) &= - \int_{iT_s}^{(i+1)T_s} e^{A_s((i+1)T_s - t)} \nabla_{vv} \Phi(t) \sigma(t) dt, \end{aligned} \quad (36)$$

for all $i \in \mathbb{N}$. Since A_s is a diagonal matrix, applying the mean-value theorem element-wise yields

$$\tilde{\nabla}_v \Phi((i+1)T_s) = A_s^{-1} (e^{A_s T_s} - I) (\nabla_{vv} \Phi \sigma)^*, \quad (37)$$

for all $i \in \mathbb{N}_0$, where $(\nabla_{vv} \Phi \sigma)_j^* := (\nabla_{vv} \Phi(\tau_j) \sigma(\tau_j))_j$ with time indices $\tau_j \in (iT_s, (i+1)T_s)$ induced by the mean-value theorem for j -th component. Also, it is straightforward from the definition of σ that

$$\|\sigma(t, v(t))\| \leq m_f^{-1} \Delta_{\tilde{\nabla}_{vt} \Phi} =: \Delta_\sigma, \quad \forall t \in [0, \tau]. \quad (38)$$

For all $i \in \mathbb{N}_0$, substituting (10) into (37) yields

$$\begin{aligned} \|\hat{\sigma}((i+1)T_s)\| &= \|\nabla_{vv} \Phi^{-1}((i+1)T_s) e^{A_s T_s} (\nabla_{vv} \Phi \sigma)^*\| \\ &\leq m_f^{-1} \|e^{A_s T_s}\| \|(\nabla_{vv} \Phi \sigma)^*\| \leq m_f^{-1} \|(\nabla_{vv} \Phi \sigma)^*\| \\ &\leq m_f^{-1} \sqrt{n_v} \|(\nabla_{vv} \Phi \sigma)^*\|_\infty = m_f^{-1} \sqrt{n_v} |(\nabla_{vv} \Phi \sigma)_{j^*}^*| \\ &\leq m_f^{-1} \sqrt{n_v} \|\nabla_{vv} \Phi(\tau_{j^*}) \sigma(\tau_{j^*})\| \\ &\leq m_f^{-1} \sqrt{n_v} \|\nabla_{vv} \Phi(\tau_{j^*})\| \|\sigma(\tau_{j^*})\| \\ &\leq m_f^{-1} \sqrt{n_v} \Delta_{\nabla_{vv} \Phi} \Delta_\sigma =: \Delta_{\hat{\sigma}}, \end{aligned} \quad (39)$$

where $j^* := \arg \max_j |(\nabla_{vv} \Phi \sigma)_j^*|$. Therefore, $\|\hat{\sigma}(iT_s)\| \leq \Delta_{\hat{\sigma}}$, $\forall i \in \{0, \dots, \lfloor \tau/T_s \rfloor\}$.

Proof of (33): By [11, Lemma A.7.1], we have

$$\|\dot{v}_a\|_{\mathcal{L}_\infty}^{[0, \tau]} \leq \|C(s)\|_{\mathcal{L}_1} \|\hat{\sigma}\|_{\mathcal{L}_\infty}^{[0, \tau]} = \|\hat{\sigma}\|_{\mathcal{L}_\infty}^{[0, \tau]} \leq \Delta_{\hat{\sigma}}, \quad (40)$$

with the filter C with unity DC gain.

Proof of (34): First, $\|\dot{v}(t)\| \leq \|\dot{v}_b(t)\| + \|\dot{v}_a(t)\| \leq \Delta_{\dot{v}_b} + \Delta_{\dot{\sigma}} =: \Delta_{\dot{v}}$, $\forall t \in [0, \tau]$. Then, by definition, we have

$$\begin{aligned} \dot{\sigma}(t, v(t)) &= \frac{d}{dt} (\nabla_{vv} \Phi^{-1} \tilde{\nabla}_{vt} \Phi) \\ &= - \left(\frac{d \nabla_{vv} \Phi^{-1}}{dt} \tilde{\nabla}_{vt} \Phi + \nabla_{vv} \Phi^{-1} \frac{d \tilde{\nabla}_{vt} \Phi}{dt} \right) \\ &= \nabla_{vv} \Phi^{-1} \left(\frac{d \nabla_{vv} \Phi}{dt} \nabla_{vv} \Phi^{-1} \tilde{\nabla}_{vt} \Phi - \frac{d \tilde{\nabla}_{vt} \Phi}{dt} \right), \end{aligned} \quad (41)$$

and

$$\left\| \frac{d \nabla_{vv} \Phi(t, v(t))}{dt} \right\| = \|\nabla_{vvt} \Phi(t) + \nabla_{vvv} \Phi(t) \dot{v}(t)\| \quad (42)$$

$$\leq \Delta_{\nabla_{vvt} \Phi} + \Delta_{\nabla_{vvv} \Phi} \Delta_{\dot{v}} =: \Delta_{\tilde{\nabla}_{vt} \Phi}, \quad (43)$$

$$\left\| \frac{d \tilde{\nabla}_{vt} \Phi(t, v(t))}{dt} \right\| = \|\tilde{\nabla}_{vtt} \Phi(t) + \tilde{\nabla}_{vvt} \Phi(t) \dot{v}(t)\| \quad (44)$$

$$\leq \Delta_{\tilde{\nabla}_{vtt} \Phi} + \Delta_{\tilde{\nabla}_{vvt} \Phi} \Delta_{\dot{v}} =: \Delta_{\dot{\tilde{\nabla}}_{vt} \Phi}. \quad (45)$$

This implies

$$\begin{aligned} \|\dot{\sigma}(t, v(t))\| &\leq m_f^{-1} (\Delta_{\tilde{\nabla}_{vt} \Phi} m_f^{-1} \Delta_{\dot{\tilde{\nabla}}_{vt} \Phi} + \Delta_{\dot{\tilde{\nabla}}_{vt} \Phi}) \\ &=: \Delta_{\dot{\sigma}}, \end{aligned} \quad (46)$$

which concludes the proof. \square

Lemma 3. If $v(t) \in \mathcal{O}_\nabla(t, \rho)$, $\forall t \in [0, \tau]$ for some $\tau > 0$, then the error in uncertainty estimation $\tilde{\sigma}(t) := \hat{\sigma}(t) - \sigma(t, v(t))$ is bounded as

$$\|\tilde{\sigma}(t)\| \leq \begin{cases} \Delta_\sigma, & \forall t \in [0, T_s), \\ \zeta_2(A_s) T_s, & \forall t \in [T_s, \tau]. \end{cases} \quad (47)$$

Proof. First, $\hat{\sigma}(0) = 0$ implies $\|\tilde{\sigma}(t)\| = \|\sigma(t)\| \leq \Delta_\sigma$, $\forall t \in [0, T_s)$. For all $t \in [(i+1)T_s, (i+2)T_s)$, $\forall i \in \mathbb{N}_0$,

$$\begin{aligned} \|\tilde{\sigma}(t)\| &= \|\hat{\sigma}(t) - \sigma(t)\| \\ &= \|\nabla_{vv} \Phi^{-1}((i+1)T_s) e^{A_s T_s} (\nabla_{vv} \Phi \sigma)^* - \sigma(t)\| \\ &\leq m_f^{-1} \|e^{A_s T_s} (\nabla_{vv} \Phi \sigma)^* - \nabla_{vv} \Phi((i+1)T_s) \sigma(t)\|. \end{aligned} \quad (48)$$

Similar to the proof of Lemma 2, using the mean-value theorem, the following inequality can be derived from (48):

$$\begin{aligned} \|\tilde{\sigma}(t)\| &\leq m_f^{-1} \sqrt{n_v} \\ &\|e^{A_s T_s} \nabla_{vv} \Phi(\tau_{j^\#}) \sigma(\tau_{j^\#}) - \nabla_{vv} \Phi((i+1)T_s) \sigma(t)\|, \\ &\leq m_f^{-1} \sqrt{n_v} (\chi_1(t) + \chi_2(t) + \chi_3(t)), \end{aligned} \quad (49)$$

where $j^\# := \arg \max_j |e^{A_s T_s} \nabla_{vv} \Phi(\tau_j) \sigma(\tau_j) - \nabla_{vv} \Phi((i+1)T_s) \sigma(t)|$, and

$$\begin{aligned} \chi_1(t) &:= \|\nabla_{vv} \Phi((i+1)T_s) (\sigma(t) - \sigma(\tau_{j^\#}))\|, \\ \chi_2(t) &:= \|(\nabla_{vv} \Phi((i+1)T_s) - \nabla_{vv} \Phi(\tau_{j^\#})) \sigma(\tau_{j^\#})\|, \\ \chi_3(t) &:= \|(I_{n_v} - e^{A_s T_s}) \nabla_{vv} \Phi(\tau_{j^\#}) \sigma(\tau_{j^\#})\|. \end{aligned} \quad (50)$$

From Lemma 2, one can deduce the following inequalities,

$$\begin{aligned}
\chi_1(t) &\leq \Delta_{\nabla_{vv}\Phi} \|\sigma(\tau_{j\#}) - \sigma(t)\| \leq \Delta_{\nabla_{vv}\Phi} \int_{\tau_{j\#}}^t \|\dot{\sigma}(\tau)\| d\tau \\
&\leq \Delta_{\nabla_{vv}\Phi} \Delta_{\dot{\sigma}} (t - \tau_{j\#}) \leq 2\Delta_{\nabla_{vv}\Phi} \Delta_{\dot{\sigma}} T_s, \\
\chi_2(t) &\leq \Delta_{\sigma} \left\| (\nabla_{vv}\Phi((i+1)T_s) - \nabla_{vv}\Phi(\tau_{j\#})) \right\| \\
&\leq \Delta_{\sigma} \int_{\tau_{j\#}}^{(i+1)T_s} \left\| \frac{d\nabla_{vv}\Phi}{dt} \right\| dt \\
&\leq \Delta_{\sigma} \Delta_{\dot{\nabla}_{vv}\Phi} ((i+1)T_s - \tau_{j\#}) \leq \Delta_{\sigma} \Delta_{\dot{\nabla}_{vv}\Phi} T_s, \\
\chi_3(t) &\leq \Delta_{\sigma} \|(I_{n_v} - e^{A_s T_s}) \nabla_{vv}\Phi(\tau_{j\#})\| \\
&\leq \Delta_{\sigma} \Delta_{\nabla_{vv}\Phi} \|I_{n_v} - e^{A_s T_s}\| \leq \Delta_{\sigma} \Delta_{\nabla_{vv}\Phi} \lambda(-A_s) T_s.
\end{aligned} \tag{51}$$

A fortiori,

$$\begin{aligned}
\|\tilde{\sigma}(t)\| &\leq m_f^{-1} \sqrt{n_v} (\chi_1(t) + \chi_2(t) + \chi_3(t)) \\
&\leq m_f^{-1} \sqrt{n_v} \left((2\Delta_{\dot{\sigma}} + \lambda(-A_s) \Delta_{\sigma}) \Delta_{\nabla_{vv}\Phi} \right. \\
&\quad \left. + \Delta_{\sigma} \Delta_{\dot{\nabla}_{vv}\Phi} \right) T_s =: \zeta_2(A_s) T_s,
\end{aligned}$$

for all $t \in [T_s, \tau]$, which concludes the proof. \square

B. Proof of Theorem 1

The proof is done by contradiction. Suppose $v(t) \in \mathcal{O}_{\nabla}(t, \rho)$, $\forall t \geq 0$ is not true. From $v(0) \in \mathcal{O}_{\nabla}(0, \rho)$ and the continuity of $\nabla_v \Phi$, there exists $t^* \in (0, \infty)$ such that $\|\nabla_v \Phi(t^*, v(t^*))\| = \rho$ and $\|\nabla_v \Phi(t, v(t))\| < \rho$, $\forall t \in [0, t^*]$. Let

$$\begin{aligned}
\eta(t) &:= \mathcal{L}^{-1}[C(s)\sigma(s)], \\
\hat{\eta}(t) &:= \mathcal{L}^{-1}[C(s)\hat{\sigma}(s)] = -\dot{v}_a(t), \\
\nu^\top(t) &:= \frac{\partial V(\nabla_v \Phi(t, v(t)))}{\partial \nabla_v \Phi} \nabla_{vv}\Phi(t, v(t)), \\
\Delta_\nu &:= \Delta_{\nabla} \frac{\partial V}{\partial \nabla_v \Phi} \Delta_{\nabla_{vv}\Phi},
\end{aligned} \tag{52}$$

where \mathcal{L}^{-1} denotes the inverse Laplace transformation. Then, we have

$$\begin{aligned}
\dot{V} &= \frac{\partial V}{\partial \nabla_v \Phi} \dot{\nabla}_v \Phi = \frac{\partial V}{\partial \nabla_v \Phi} (\dot{\nabla}_{vt}\Phi + \nabla_{vv}\Phi(\dot{v}_b + \dot{v}_a + \sigma)) \\
&\leq -2\beta V + \frac{\partial V}{\partial \nabla_v \Phi} \nabla_{vv}\Phi(\sigma(t) + \dot{v}_a(t)) \\
&= -2\beta V + \nu^\top(t)(\sigma(t) - \eta(t)) + \nu^\top(t)(\eta(t) - \hat{\eta}(t)).
\end{aligned} \tag{53}$$

By comparison lemma [18], we have

$$V(t) \leq e^{-2\beta t} V_0 + \chi_4(t) + \chi_5(t), \tag{54}$$

where

$$\begin{aligned}
\chi_4(t) &:= \int_0^t e^{-2\beta(t-\tau)} \nu^\top(\tau)(\sigma(\tau) - \eta(\tau)) d\tau, \\
\chi_5(t) &:= \int_0^t e^{-2\beta(t-\tau)} \nu^\top(\tau)(\eta(\tau) - \hat{\eta}(\tau)) d\tau.
\end{aligned} \tag{55}$$

From Assumption 2 and Lemma 2, $\|\nu\|_{\mathcal{L}^\infty}^{[0, t^*]} \leq \Delta_\nu$ and $\|\dot{\sigma}\|_{\mathcal{L}^\infty}^{[0, t^*]} \leq \Delta_{\dot{\sigma}}$. Since χ_4 is the solution to the following equation

$$\dot{\chi}_4(t) = -2\beta \chi_4(t) + \nu^\top(t) \phi(t), \quad \chi_4(0) = 0, \tag{56}$$

$$\phi(s) = (I_{n_v} - C(s))\sigma(s), \tag{57}$$

and $\|\sigma(0)\| \leq \Delta_\sigma$, it can be shown by [16, Lemma A.1] that

$$\|\chi_4\|_{\mathcal{L}^\infty}^{[0, t^*]} \leq \Delta_\nu \left(\frac{\Delta_\sigma}{|2\beta - \omega|} + \frac{\Delta_{\dot{\sigma}}}{2\beta\omega} \right) =: \zeta_1(\omega). \tag{58}$$

Let $\tilde{\eta}(t) := \hat{\eta}(t) - \eta(t)$. The dynamics of $\tilde{\eta}$ are given as $\dot{\tilde{\eta}}(t) = -\omega \tilde{\eta}(t) - \omega \sigma(t)$, $\forall t \in [0, T_s]$, and the solution is $\tilde{\eta}(t) = -\int_0^t e^{-\omega(t-\tau)} \omega \sigma(\tau) d\tau$. Therefore, for all $t \in [0, T_s]$, we have

$$\begin{aligned}
\|\tilde{\eta}(t)\| &\leq \Delta_\sigma \int_0^t \omega e^{-\omega(t-\tau)} d\tau = \Delta_\sigma (1 - e^{-\omega t}) \leq \Delta_\sigma \omega t \\
&\leq \Delta_\sigma \omega T_s =: \zeta_3(\omega) T_s < (\zeta_2(A_s) + \zeta_3(\omega)) T_s.
\end{aligned} \tag{59}$$

For all $t \geq T_s$, the dynamics are given by $\dot{\tilde{\eta}}(t) = -\omega \tilde{\eta}(t) + \omega \tilde{\sigma}(t)$, and the solution is $\tilde{\eta}(t) = e^{-\omega(t-T_s)} \tilde{\eta}(T_s) + \int_{T_s}^t e^{-\omega(t-\tau)} \omega \tilde{\sigma}(\tau) d\tau$. Then, we have

$$\begin{aligned}
\|\tilde{\eta}(t)\| &\leq e^{-\omega(t-T_s)} \|\tilde{\eta}(T_s)\| + \int_{T_s}^t e^{-\omega(t-\tau)} \omega \|\tilde{\sigma}(\tau)\| d\tau \\
&\leq (\zeta_3(\omega) + (1 - e^{-\omega(t-T_s)}) \zeta_2(A_s)) T_s \\
&\leq (\zeta_2(A_s) + \zeta_3(\omega)) T_s.
\end{aligned} \tag{60}$$

Therefore, $\|\tilde{\eta}\|_{\mathcal{L}^\infty}^{[0, t^*]} \leq (\zeta_2(A_s) + \zeta_3(\omega)) T_s$, which implies

$$\begin{aligned}
\|\chi_5(t)\| &\leq \Delta_\nu \int_0^t e^{-2\beta(t-\tau)} \|\tilde{\eta}(\tau)\| d\tau \\
&\leq \Delta_\nu \frac{\zeta_2(A_s) + \zeta_3(\omega)}{2\beta} T_s =: \zeta_4(A_s, \omega) T_s,
\end{aligned} \tag{61}$$

for all $t \in [0, t^*]$.

From (54), (58), and (61), we have

$$\begin{aligned}
V(t) &\leq e^{-2\beta t} V_0 + \zeta_1(\omega) + \zeta_4(A_s, \omega) T_s, \\
&< V_0 + \zeta_1(\omega) + \zeta_4(A_s, \omega) T_s,
\end{aligned} \tag{62}$$

for all $t \in (0, t^*]$, which implies

$$V(t) < V_0 + \zeta_1(\omega) + \zeta_4(A_s, \omega) T_s, \tag{63}$$

for all $t \in [0, t^*]$. Also, $\|\nabla_v \Phi(t^*, v(t^*))\| = \rho$. From Assumption 1,

$$\begin{aligned}
\underline{\alpha} \rho^2 &\leq V(t^*) < V_0 + \zeta_1(\omega) + \zeta_4(A_s, \omega) T_s \\
\Rightarrow T_s &> \frac{\underline{\alpha} \rho^2 - V_0 - \zeta_1(\omega)}{\zeta_4(A_s, \omega)},
\end{aligned} \tag{64}$$

which contradicts (24). Therefore, $v(t) \in \mathcal{O}_{\nabla}(t, \rho)$, $\forall t \geq 0$. The first part of (26) can be deduced from Assumption 1 and (62).

The rest of the proof is straightforward from Lemma 1.

C. Proof of Corollary 1

The left-hand side of the inequality is trivial. For the right-hand side, the convexity of $\Phi(t, \cdot)$ implies

$$\Phi(t, v(t)) - \Phi(t, v^*(t)) \leq \nabla_v \Phi(t, v(t))^\top (v(t) - v^*(t)), \tag{65}$$

and $\nabla_v \Phi(t, v(t))^\top (v(t) - v^*(t)) \leq \|\nabla_v \Phi(t, v(t))\| \|v(t) - v^*(t)\|$, which concludes the proof by Theorem 1.

D. Simulation settings

All simulations are performed on an M1 Macbook Air. Fixed-time Euler integration is used for numerical simulation with the time step $\Delta t = 10^{-3}$.

1) *Example 1:* For the problem in Section IV-A, we compare the following methods: i) modified PCIP in (6) with $\epsilon = 1$ and ii) \mathcal{L}_1 -AO in (10) with the PCIP method in (5) as the baseline optimizer. The optimal solution is obtained using a time-invariant solver, ECOS [17]. The gain of the PCIP method is $P = 10$ by default, and the parameters of the \mathcal{L}_1 -AO are set as $T_s = 10^{-3}$, $\omega = 10^3$, $A_s = -1$.

2) *Example 2:* For the problem in Section IV-B, we compare the following methods: i) PCIP vs PCIP+L1AO

(i.e., \mathcal{L}_1 -AO with the baseline of PCIP), ii) Modified PCIP vs Modified PCIP+L1AO (i.e., \mathcal{L}_1 -AO with the baseline of modified PCIP). The optimal solution is obtained using a time-invariant solver, ECOS [17]. The gain of PCIP is set as $P = I_2$, and the gain of the modified PCIP is set as $P = 10I_2$ with $\epsilon = 0.1$. The parameters of \mathcal{L}_1 -AO are set as $T_s = 10^{-3}$, $\omega = 50$, and $A_s = -0.1I_2$. The simulation time is set as $t_f = 50$. The target position is modeled as $x_d(t) = 15 \left[\cos\left(\frac{2\pi t}{t_f}\right), \sin\left(\frac{2\pi t}{t_f}\right) \right]^\top$. The penalty parameter for interior-point formulation is set as $c(t) = 50e^{t/50}$.