Online Optimization and Ambiguity-based Learning of Distributionally Uncertain Dynamic Systems

Dan Li[1], Dariush Fooladivanda[1] and Sonia Martínez[1]

***Abstract*—This paper proposes a novel approach to construct data-driven online solutions to optimization problems (P) subject to a class of distributionally uncertain dynamical systems. The introduced framework allows for the simultaneous learning of distributional system uncertainty via a parameterized, control-dependent ambiguity set using a finite historical data set, and its use to make online decisions with probabilistic regret function bounds. Leveraging the merits of Machine Learning, the main technical approach relies on the theory of Distributional Robust Optimization (DRO), to hedge against uncertainty and provide less conservative results than standard Robust Optimization approaches. Starting from recent results that describe ambiguity sets via parameterized, and control-dependent empirical distributions as well as ambiguity radii, we first present a tractable reformulation of the corresponding optimization problem while maintaining the probabilistic guarantees. We then specialize these problems to the cases of 1) optimal one-stage control of distributionally uncertain nonlinear systems, and 2) resource allocation under distributional uncertainty. A novelty of this work is that it extends DRO to online optimization problems subject to a distributionally uncertain dynamical system constraint, handled via a control-dependent ambiguity set that leads to online-tractable optimization with probabilistic guarantees on regret bounds. Further, we introduce an online version of the Nesterov's accelerated-gradient algorithm, and analyze its performance to solve this class of problems via dissipativity theory.**

I. Introduction

Online optimization has attracted significant attention from various fields, including Machine Learning, Information Theory, Robotics and Smart Power Systems; see [1]–[3] and references therein. A basic online optimization setting involves the minimization of time-varying convex loss functions, resulting into Online Convex Programming (OCP). Typically, loss objectives in OCP are functions of non-stationary stochastic processes [4], [5]. Regret minimization aims to deal with non-stationarity by reducing the difference between an optimal decision made with information in hindsight, and one made as information is increasingly revealed. Thus, several online algorithms and techniques are aimed at minimizing various types of regret functions [6], [7]. More recently, and with the aim of further reducing the cost, regret-based OCP has integrated prediction models of loss functions [8]–[11]. However, exact models of evolving loss functions may not be available, while alternative data-based approximate models may require large amounts of data that are hard to obtain. This motivates the need of developing new learning algorithms for loss functions that can employ finite data sets, while guaranteeing a precise performance of the corresponding optimization.

Literature Review. Due to recent advances in Data Science and Machine Learning, the question of learning system models as well as distributional uncertainty from data is gaining significant attention. From the early work on Systems Identification [12], Willem's Behavioral Theory and fundamental lemma [13], [14] have been recently leveraged to learn linear, time-invariant system models in predictive control applications [14]–[18]. The aforementioned works rely on the use of Hankel system representations of the LTI system, and may be subject or not to additional uncertainty. In particular, the work [19] leverages the behavioral theory to obtain sub-linear regret bounds for the online optimization of discrete-time unknown but deterministic linear systems. Other approaches to learn LTI systems from input-output data employ concentration inequalities and finite samples, and include, for example, [20], exploiting least squares and the Ho-Kalman algorithm, [21], using subspace identification techniques for LTI systems subject to unknown Gaussian disturbances, and [22], resorting to Lasso-like methods that exploit the sparse representation of LTI systems.

On the other hand, classical online optimization relies on Sample Averaging Approximation (SAA) (with bootstrap) to derive optimal value and/or policy approximations. However, SAA usually requires large amounts of data to provide good approximations of the stochastic cost, which leads to non-robust solutions to unseen data. In contrast, recent developments on measure-of-concentration results [23] have lead to a new type of Distributionally Robust Optimization (DRO) [24]–[26], which aims to bridge this gap. Particularly, the DRO framework enables finite-sample, performance-guaranteed optimization under distributional uncertainty [24], [25], and paves the way to dealing with the control and estimation of system dynamics subject to distributional uncertainty. Motivated by this, the works [27], [28] consider the time evolution of Wasserstein ambiguity sets and their updates under streaming data for estimation. However, the nominal dynamic constraints defined in these problems are assumed to be known, while in practice, these models also need to be identified. The previous work [29] proposes a method for integrating the learning of an unknown and nominal parameterized system dynamics with Wasserstein ambiguity sets. These ambiguity sets are given by a parameter and control-dependent ambiguity ball center as well as a corresponding radius. Taking this as a starting point, and motivated by the direct use of these ambiguity sets in a type of "distributionally robust control", here we further extend this setup in connection with online

* This research was developed with funding from ONR N00014-19-1-2471, and AFOSR FA9550-19-1-0235.

[1] D. Li and S. Martínez are with the Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla, CA 92092, USA. D. Fooladivanda is with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720, USA. dal027.me@gmail.com; dfooladi@berkeley.edu; soniamd@ucsd.edu

optimization problems. Precisely, what distinguishes this work from other approaches is the focus on learning the transition system dynamics itself via control-dependent ambiguity sets. The control method is derived from an online optimization method [6], and, therefore, it does not aim to calculate exactly an optimal control, but to find an approximate solution that leads to a low instantaneous regret function value w.r.t. standard, online and regret-based optimization problems. Finally, this manuscript connects with the topic online optimization using decision-dependent distributions [30], [31], where the uncertainty distribution changes with the decision variable. As these problems are intractable, [30], [31] solve for alternative *stable solutions*, or optimal solutions wrt to the distribution they induce. In addition to this, and while [30], [31] can handle dynamic systems, a main difference with this work is that a dynamic system structure that is being learned is not exploited, which can help reduce uncertainty more effectively.

Statement of Contributions. In this work, we propose a novel approach to solve a class of online optimization problems subject to distributionally uncertain dynamical systems. Our end goal is to produce an online controller that results in bounded instantaneous regrets with high confidence. Our proposed framework is unique in that it enables the online learning of the underlying nominal system, maintains online-problem tractability, and simultaneously provides finite-sample, probabilistic guarantee bounds on the resulting regret. This is achieved by considering a worst-case-system formulation that employs novel parameterized and control-dependent, Wasserstein ambiguity sets. Our learning method precisely consists of updating this ambiguity set. The proposed formulation is valid for a wide class of problems, including but not limited to 1) a class of optimal control problems subject to distributionally uncertain dynamical system, and 2) online resource allocation under distributional uncertainty. To do this, we first obtain tractable problem reformulations for these two cases, which results in online and non-smooth convex problem optimizations. For each of these categories, and smoothed-out versions of these problems, we propose an online control algorithm dynamics, which extends Nesterov's accelerated-gradient method. Adapting dissipativity theory, we prove optimal first-order convergence rate for these algorithms under smoothness and convexity assumptions. This result is crucial to guarantee that the online controller can provide probabilistic guarantees on their regret bounds via the control-dependent ambiguity set. We thus finish our work by quantifying these dynamic regret bounds, and by explicitly characterizing the effect of learning parameters with finite historical samples.

II. Notations

We denote by $\mathbb{R}^m$, $\mathbb{R}^m_{\geq 0}$, $\mathbb{Z}^m_{\geq 0}$ and $\mathbb{R}^{m\times n}$ the m-dimensional real space, nonnegative orthant, nonnegative integer-orthant space, and the space of $m\times n$ matrices, respectively. The transpose of a column vector $\boldsymbol{x}\in\mathbb{R}^m$ is $\boldsymbol{x}^\top$, and $\mathbf{1}_m$ is a shorthand for $(1,\cdots,1)^\top\in\mathbb{R}^m$. We index vectors with subscripts, i.e., $\boldsymbol{x}_k\in\mathbb{R}^m$ with $k\in\mathbb{Z}_{\geq 0}$, and given $\boldsymbol{x}\in\mathbb{R}^m$ we denote its i^{th} component by x_i. We denote by $\|\boldsymbol{x}\|$ and $\|\boldsymbol{x}\|_\infty$ the 2-norm and ∞-norm, respectively. The inner product of $\mathbb{R}^m$ is given as $\langle \boldsymbol{x},\boldsymbol{y}\rangle := \boldsymbol{x}^\top\boldsymbol{y}$, $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^m$; thus, $\|\boldsymbol{x}\| := \sqrt{\langle \boldsymbol{x},\boldsymbol{x}\rangle}$. The gradient of a real-valued function $\ell:\mathbb{R}^m\to\mathbb{R}$ is denoted as $\nabla\ell(\boldsymbol{x})$ and $\nabla_x\ell(\boldsymbol{x})$ is the partial derivative w.r.t. x. In what follows, $\mathrm{dom}(\ell) := \{\boldsymbol{x}\in\mathbb{R}^m \mid -\infty<\ell(\boldsymbol{x})<+\infty\}$. A function $\ell:\mathrm{dom}(\ell)\to\mathbb{R}$ is M-strongly convex, if for any $\boldsymbol{y},\boldsymbol{z}\in\mathrm{dom}(\ell)$ there exists $\boldsymbol{g}\in\mathbb{R}^m$ such that $\ell(\boldsymbol{y})\geq\ell(\boldsymbol{z})+\boldsymbol{g}^\top(\boldsymbol{y}-\boldsymbol{z})+M\|\boldsymbol{y}-\boldsymbol{z}\|^2/2$, for some $M>0$. The function ℓ is convex if $M\geq 0$. We call a vector $\boldsymbol{g}$ a subgradient of ℓ at $\boldsymbol{z}$ and denote by $\partial\ell(\boldsymbol{z})$ the subgradient set. If ℓ is differentiable at $\boldsymbol{z}$, then $\partial\ell(\boldsymbol{z})=\{\nabla\ell(\boldsymbol{z})\}$. Finally, the operation $\Pi_{\mathcal{U}}(\mathcal{X}):\mathcal{X}\to\mathcal{U}$ projects the set $\mathcal{X}\subseteq\mathbb{R}^m$ onto $\mathcal{U}\subseteq\mathbb{R}^m$ under the Euclidean norm. We write $\Pi_{\mathcal{U}}(\boldsymbol{x}) := \operatorname{argmin}_{\boldsymbol{z}}\|\boldsymbol{x}-\boldsymbol{z}\|^2/2+\chi_{\mathcal{U}}(\boldsymbol{z})$, where $\boldsymbol{x}\in\mathcal{X}$, and $\chi_{\mathcal{U}}(\boldsymbol{z})=0$ if $\boldsymbol{z}\in\mathcal{U}$, otherwise $+\infty$. Endow $\mathbb{R}^n$ with the Borel σ-algebra $\mathcal{B}$, and let $\mathcal{P}(\mathbb{R}^n)$ be the set of probability measures (or distributions) over $(\mathbb{R}^n,\mathcal{B})$. The set of probability distributions with bounded first moments is $\mathcal{M}=\{\mathbb{Q}\in\mathcal{P}(\mathbb{R}^n)\mid\int_{\mathbb{R}^n}\|\boldsymbol{x}\|\mathrm{d}\mathbb{Q}<+\infty\}$. We use the Wasserstein metric [32] to define a distance in $\mathcal{M}$, and the dual version of the 1-Wasserstein metric $d_W:\mathcal{M}\times\mathcal{M}\to\mathbb{R}_{\geq 0}$, is defined by $d_W(\mathbb{Q}_1,\mathbb{Q}_2) := \sup_{f\in\mathcal{L}}\int f(\boldsymbol{x})\mathrm{d}\mathbb{Q}_1-\int f(\boldsymbol{x})\mathrm{d}\mathbb{Q}_2$, where $\mathcal{L}$ is the space of all Lipschitz functions with Lipschitz constant 1. We denote a closed Wasserstein ball of radius ϵ (also called an ambiguity set) centered at a distribution $\mathbb{P}\in\mathcal{M}$ by $\mathbb{B}_\epsilon(\mathbb{P}) := \{\mathbb{Q}\in\mathcal{M}\mid d_W(\mathbb{P},\mathbb{Q})\leq\epsilon\}$. The Dirac measure at $\boldsymbol{x}_0\in\mathbb{R}^n$ is a distribution in $\mathcal{P}(\mathbb{R}^n)$ denoted by $\delta_{\{\boldsymbol{x}_0\}}$. Given $A\in\mathcal{B}$, we have $\delta_{\{\boldsymbol{x}_0\}}(A)=1$, if $\boldsymbol{x}_0\in A$, otherwise 0. A random vector $\boldsymbol{x}\in\mathbb{R}^m$ with probability distribution $\mathbb{Q}$ is sub-Gaussian if there are positive constants C,v such that $\mathbb{Q}(\|\boldsymbol{x}\|>t)\leq C\mathrm{e}^{-vt^2}$. Equivalently, a zero-mean random vector $\boldsymbol{x}\in\mathbb{R}^n$ is sub-Gaussian if for any $a\in\mathbb{R}^n$ we have $\mathbb{E}\left[\exp(a^\top\boldsymbol{x})\right]\leq\exp(\|a\|^2\nu^2/2)$ for some ν.

III. Problem Statement, Motivation, and Approach based on Ambiguity Set Learning

We start by introducing a class of online optimization problems, where the objective function is time-varying according to an unknown dynamical system subject to an unknown disturbance. Consider a dynamical system that evolves according to unknown stochastic dynamics

$$\boldsymbol{x}_{t+1} = f(t,\boldsymbol{x}_t,\boldsymbol{u}_t)+\boldsymbol{w}_t,\quad \text{from a given } \boldsymbol{x}_0\in\mathbb{R}^n, \tag{1}$$

where $\boldsymbol{u}_t\in\mathcal{U}\subset\mathbb{R}^m$ is an online decision or control action at time t, $f:\mathbb{R}_{\geq 0}\times\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^n$ is a measurable, but unknown state transition function, and $\boldsymbol{w}_t\in\mathbb{R}^n$, is an unknown, random, disturbance vector. Due to the Markov assumption, $\boldsymbol{x}_t\in\mathbb{R}^n$ can be described by an unknown transition probability measure $\mathbb{P}_{t|t-1}\in\mathcal{P}(\mathbb{R}^n)$, conditioned on the system state and control at time $t-1$. Denote by $\ell:\mathbb{R}^m\times\mathbb{R}^n\to\mathbb{R}$, $(\boldsymbol{u},\boldsymbol{x})\mapsto\ell(\boldsymbol{u},\boldsymbol{x})$ *an a-priori* selected, measurable loss function. Assume that $\mathcal{U}$ is compact, and we are interested in selecting $\boldsymbol{u}_t\in\mathcal{U}$ that minimizes the loss

$$\min_{\boldsymbol{u}_t\in\mathcal{U}}\quad\left\{\mathbb{E}_{\mathbb{P}_{t+1|t}}\left[\ell(\boldsymbol{u}_t,\boldsymbol{x})\right] := \int_{\mathbb{R}^n}\ell(\boldsymbol{u}_t,\boldsymbol{x})\,\mathbb{P}_{t+1|t}(d\boldsymbol{x})\right\}.$$

This objective value is inaccessible since the state distribution $\mathbb{P}_{t+1|t}$ is unknown, and its evolution is highly dependent on the system, disturbance, and as well as on the decisions taken. In this work, we aim to propose an effective online optimization and learning algorithm which tracks the minimizers of the time-varying objective function with low regret in high

probability. Thus, at each time t, we aim to find $\boldsymbol{u} := \boldsymbol{u}_t$ that minimizes the loss in the immediate future at $t+1$

$$\begin{aligned} &\min_{\boldsymbol{u}\in\mathcal{U}} \mathbb{E}_{\mathbb{P}_{t+1|t}}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right], \\ &\text{s.t. } \boldsymbol{x}\sim\mathbb{P}_{t+1|t}, \text{ evolves according to (1).} \end{aligned} \tag{P}$$

This problem formulation is similar to a one-stage optimization problems with unknown system transitions [33]. The expectation operator with respect to $\mathbb{P}_{t+1|t}$ is conditional on the historical realizations $\hat{\boldsymbol{x}}_k$, $k \le t$, the adopted decisions $\hat{\boldsymbol{u}}_k$, $k \le t-1$, the yet-to-be-learned unknown dynamical system f, and realizations $\hat{\boldsymbol{w}}_k$, $k \le t-1$. We will identify $\mathbb{P}_{t+1|t}(d\boldsymbol{x}) \equiv \mathbb{P}_{t+1}(d\boldsymbol{x}|\boldsymbol{u}_t, \boldsymbol{x}_t = \hat{\boldsymbol{x}}_t, \boldsymbol{x}_k = \hat{\boldsymbol{x}}_k, \boldsymbol{u}_k = \hat{\boldsymbol{u}}_k, k \le t-1)$ which, by the Markovian property, satisfies $\mathbb{P}_{t+1|t}(d\boldsymbol{x}) \equiv \mathbb{P}_{t+1}(d\boldsymbol{x}|\boldsymbol{u}_t, \boldsymbol{x}_t = \hat{\boldsymbol{x}}_t)$. At time t, let $\boldsymbol{u}^\star := \boldsymbol{u}_t^\star$ denote an optimizer of Problem (P) and consider the instantaneous regret

$$R_t := \mathbb{E}_{\mathbb{P}_{t+1|t}}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] - \mathbb{E}_{\mathbb{P}_{t+1|t}}\left[\ell(\boldsymbol{u}^\star,\boldsymbol{x})\right],$$

which is the loss incurred if the selected $\boldsymbol{u}$ is different from an optimal decision. Our goal will be to develop a robust online algorithm which ensures a probabilistic bound on the regret. That is, with high probability ρ, the regret R_t is upper bounded by a sum of terms, a first one depending on the initial condition $\boldsymbol{x}_0$; a second one depending on the instantaneous variation of the loss of (P); and a third term related to how well the unknown system f and the uncertainty are characterized; please see Theorem V.1. While the second and third terms are inherent to the system, the effect of the second one can be reduced by considering a predicted loss of the system [11]. In this work, we aim to bound the third term and minimize it by estimating the distribution $\mathbb{P}_{t+1|t}$ via an ambiguity set of distributions. We will show that, as historical data are assimilated over time, this third term asymptotically decays to zero. This is achieved under the following assumption

Assumption III.1 (Independent and stationary sub-Gaussian distributions) The vectors $\boldsymbol{w}_t \in \mathbb{R}^n$, $t \in \mathbb{Z}_{\ge 0}$, are i.i.d. with $\boldsymbol{w}_t \sim \mathbb{Q}$ and zero-mean σ sub-Gaussian[1].

Remark III.1 (On sub-Gaussian distributions) Sub-Gaussian distributions include Gaussian random variables and all distributions of bounded support.

Example 1 (Vehicle path planning and tracking): A two-wheeled vehicle moves in an unknown 2D environment. Assume that an accessible path-planner provides a control signal for the vehicle to track a desired reference trajectory under ideal conditions, see Fig. 1. Fig. 2 shows two examples where, first, the vehicle implements a series of lane changes, and, second, navigates through a planned circular/loopy route. Since both the environment and dynamics are uncertain, exact tracking is rare. Our goal is to learn the real-time road conditions, and by solving the online problem (P), derive a control signal that enables path following minimizing the tracking error with high probability.

Example 2 (Online resource allocation in the stock market): An agent aims to achieve a target profit of $r_0 = 130\%$ in a highly-fluctuating trading market. Thus, it actively allocates wealth to multiple risky assets while trying to balance resources among assets. As asset-prices are uncertain, modeling the return rate of each asset is specially challenging. To solve this, an agent can aim to learn the real-time returns responsively, estimate the distributions of immediate returns, and then allocate wealth wisely to maximize the expected profit with high probability. This problem fits in the proposed formulation, resulting in online, balanced resource allocation with low regrets.

[1]That is, for all unit vector $\boldsymbol{v}$, we have $\mathbb{E}[e^{\lambda \boldsymbol{v}^\top \boldsymbol{w}_t}] \le e^{\lambda^2\sigma^2/2}$, $\forall \lambda \in \mathbb{R}$. Equivalently, $\mathbb{Q}(\|\boldsymbol{w}_t\| > \lambda) \le e^{-\lambda^2/(4\sigma^2)}$, $\forall \lambda$.

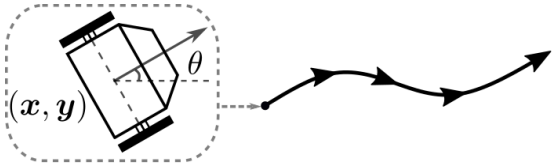


Fig. 1: A two-wheeled vehicle model with $(x, y) \in \mathbb{R}^2$ the position of the center and θ the direction.

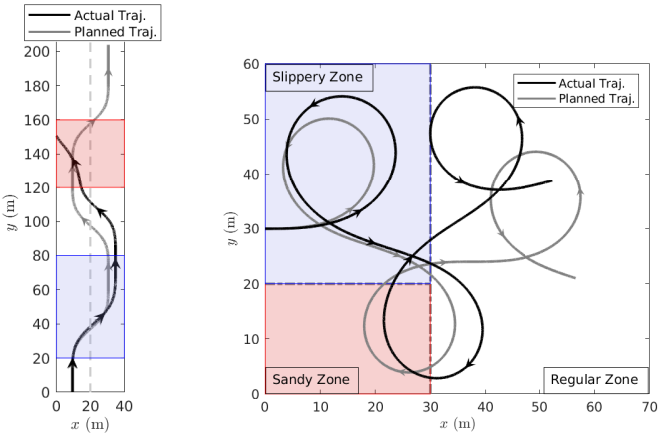


Fig. 2: The (gray) planned trajectory and (black) actual system trajectory in various road zones, with the system state $\boldsymbol{x} = (x, y, \theta) \in \mathbb{R}^2 \times [-\pi, \pi)$. The red region indicates sandy zone while the blue region indicates the slippery zone. Due to unknown road conditions, the actual system trajectories deviate from planned trajectories.

A. Online Constructions of Ambiguity Sets

Our main approach to obtain a suitable control signal is based on learning a set of distributions or ambiguity set that characterizes system uncertainty. More precisely, we employ the dynamic ambiguity set $\mathcal{P}_{t+1}$ proposed in [29]. The set $\mathcal{P}_{t+1}$ contains a class of distributions, which is, in high probability, large enough to include the unknown $\mathbb{P}_{t+1|t}$ under certain conditions. Thus, we can use it to formulate a robust version of the problem at each time instant t. Such characterization enables an online-tractable reformulation of (P) later. We summarize next the construction of these ambiguity sets $\mathcal{P}_{t+1}$. First, we assume the following on the unknown f.

Assumption III.2 (System parametrization) Given $p \in \mathbb{Z}_{>0}$, the system f can be expressed as

$$f(t, \boldsymbol{x}, \boldsymbol{u}) = \sum_{i=1}^{p} \alpha_i^\star f^{(i)}(t, \boldsymbol{x}, \boldsymbol{u}),$$

where $\boldsymbol{\alpha}^\star := (\alpha_1^\star, \ldots, \alpha_p^\star)^\top \in \mathbb{R}^p$ is an unknown parameter, and $f^{(i)} : \mathbb{R}_{\ge 0} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $(t, \boldsymbol{x}, \boldsymbol{u}) \mapsto f^{(i)}(t, \boldsymbol{x}, \boldsymbol{u})$,

$i \in \{1, \dots, p\}$ is a set of p linearly independent known basis functions or *predictors* chosen *a priori*.

Now, given arbitrary $(\boldsymbol{\alpha}, \boldsymbol{u})$, the set $\mathcal{P}_{t+1}$ is a Wasserstein ball centered at a parametric-dependent distribution $\hat{\mathbb{P}}_{t+1|t}$ for each t; that is,

$$\mathcal{P}_{t+1} := \mathbb{B}_{\hat{\epsilon}}(\hat{\mathbb{P}}_{t+1|t}) = \{\mathbb{Q} \mid d_W(\mathbb{Q}, \hat{\mathbb{P}}_{t+1|t}) \leq \hat{\epsilon}\}.$$

Here, $\hat{\epsilon}$ will be a time-varying function $\hat{\epsilon} \equiv \hat{\epsilon}(t, T, \beta, \boldsymbol{\alpha}, \boldsymbol{u})$ which depends on a number of T measurements, and a confidence $\beta \in (0, 1)$. More precisely,

$$\hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha}, \boldsymbol{u}) := \frac{1}{T} \sum_{k \in \mathcal{T}} \delta_{\{\sum_{i=1}^{p} \alpha_i \xi_k^{(i)}(\boldsymbol{\alpha}, \boldsymbol{u})\}}; \tag{2}$$

see the footnote[2], where $\mathcal{T} = \{t-T, \dots, t\}$, for $t \geq T+1$. If $\boldsymbol{\alpha} = \boldsymbol{\alpha}^\star$, then $\sum_{i=1}^{p} \alpha_i \xi_k^{(i)}(\boldsymbol{\alpha}, \boldsymbol{u}) \in \mathbb{R}^n$ provides an outcome $\boldsymbol{x}_{t+1}^{(k)} := f(t, \boldsymbol{x}_t, \boldsymbol{u}) + \boldsymbol{w}_k = \sum_{i=1}^{p} \alpha_i^\star f^{(i)}(t, \boldsymbol{x}_t, \boldsymbol{u}) + \boldsymbol{w}_k$, for each k. For a general $\boldsymbol{\alpha} \approx \boldsymbol{\alpha}^\star$, the value $\sum_{i=1}^{p} \alpha_i \xi_k^{(i)}(\boldsymbol{\alpha}, \boldsymbol{u})$ provides "approximated" outcomes $\boldsymbol{x}_{t+1}^{(k)}$, for each $k = 1, \dots, T$. Then, we claim the probabilistic guarantee of $\mathcal{P}_{t+1}$ by a selection of the parameter $\boldsymbol{\alpha}$ and $\hat{\epsilon}$ for any $\boldsymbol{u}$.

Theorem III.1 (Online probabilistic guarantee [29, Application of Theorem 1]) *Let Assumptions III.1 and III.2 hold. For a given $T \in \mathbb{Z}_{>0}$, historical data $\{\hat{\boldsymbol{x}}_k\}_{k \in \mathcal{T}}$ and $\{\boldsymbol{u}_k\}_{k \in \mathcal{T} \setminus \{t\}}$, $\mathcal{T} = \{t-T, \dots, t\}$, we select $\hat{\mathbb{P}}_{t+1|t}$ as in (2) where $\boldsymbol{\alpha}$ is selected in [29, Theorem 2 (Learning of $\boldsymbol{\alpha}^\star$)][3]. Then, for given $\boldsymbol{u}$ and a confidence-related value $\beta \in (0, 1)$, a radius $\hat{\epsilon} := \hat{\epsilon}(t, T, \beta, \boldsymbol{\alpha}, \boldsymbol{u})$ can be chosen such that*

$$\text{Prob}\left(\mathbb{P}_{t+1|t} \in \mathcal{P}_{t+1}\right) \geq \rho(t). \tag{3}$$

Here, the left-hand-side expression is a shorthand for the probability of the event $\{(\boldsymbol{x}_{t+1}^{(1)}, \dots, \boldsymbol{x}_{t+1}^{(T)}) \in \mathbb{R}^n \times \cdots \times \mathbb{R}^n \mid \mathbb{P}_{t+1|t} \in \mathbb{B}_{\hat{\epsilon}}(\hat{\mathbb{P}}_{t+1|t})\}$ and $\text{Prob} := \mathbb{P}_{t+1|t}^T$ denotes the probability measure defined on the T-fold product of $\mathbb{P}_{t+1|t}$, which evaluates the probability that the selection of samples define an ambiguity ball which contains the true distribution. In particular, the confidence value is

$$\rho(t) := (1-\beta)\left(1 - \exp\left(-\frac{(\gamma^2 - \sqrt{2}c\gamma)T}{2\sqrt{2}\left(c\gamma + \sqrt{2}c^2\right)}\right)\right),$$

where c is a data-dependent positive constant and $\gamma > \sqrt{2}c$ is a user selected parameter. Further, the radius is

$$\hat{\epsilon} := \sqrt{\frac{2nM\sigma^2}{T} \ln(\frac{1}{\beta}) + C_1 T^{-1/\max\{n,2\}}} + \gamma H(t, T, \boldsymbol{u}), \tag{4}$$

where M and C_1 are positive constants, and

$$H(t, T, \boldsymbol{u}) := \frac{1}{T} \sum_{i=1}^{p} \sum_{k \in \mathcal{T}} \|f^{(i)}(k, \hat{\boldsymbol{x}}_k, \boldsymbol{u}_k) - f^{(i)}(t, \hat{\boldsymbol{x}}_t, \boldsymbol{u})\|,$$

which bounds the variation of predicted system trajectories.

Idea of the Proof. The probabilistic guarantees (3) are a consequence of Lemma 1, Theorem 1, Theorem 2 and Eqn. (7) in [29] with Assumptions III.1 and III.2. Precisely, we achieve this by upper bounding the metric $d_W(\mathbb{P}_{t+1|t}, \hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha}, \boldsymbol{u}))$ using $d_W(\mathbb{P}_{t+1|t}, \hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha}^\star, \boldsymbol{u}))$ plus $d_W(\hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha}^\star, \boldsymbol{u}), \hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha}, \boldsymbol{u}))$. Then, the first distance is handled via [29, Lemma 1] using standard measure of concentration results[4], contributing to the first two terms of the radius $\hat{\epsilon}$ in (4). Next, the second distance $d_W(\hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha}^\star, \boldsymbol{u}), \hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha}, \boldsymbol{u}))$ can be bounded in terms of the difference $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star\|$ via [29, Theorem 1], contributing to the third term in $\hat{\epsilon}$. Notice that the third term depends on Assumption III.2 and the selected parameter γ which relies on the selection of $\boldsymbol{\alpha}$ via [29, Theorem 2 (Learning of $\boldsymbol{\alpha}^\star$)]. The confidence value $\rho(t)$ is achieved by Assumption III.1 applying to the same procedure as in [29, Theorem 2], which essentially bounds $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star\|_\infty$ in probability. Precisely, by Assumption III.1, we have $\mathbb{Q}(\|\boldsymbol{w}_t\|_\infty > \eta) \leq \mathrm{e}^{-\eta^2/(4\sigma^2)}$, $\forall\ \eta$, resulting in $\mathbb{E}\left[\|\boldsymbol{w}_t\|_\infty^l\right] \leq 2^{\frac{l}{2}-1}\sigma^l l^{\frac{l}{2}+1}$, $\forall\, l \in \mathbb{Z}_{\geq 0}$, analogous to [34, Lemma 2]. Then, with the proof similar to [34, Theorem IV.2], we achieve

$$\text{Prob}\left(\frac{1}{T}\sum_{k \in \mathcal{T}} (\|\boldsymbol{w}_k\|_\infty) \geq \gamma\right) \leq \exp\left(-\gamma\lambda + \frac{T\sqrt{2}\sigma e\lambda}{2T - 2\sqrt{2}\sigma e\lambda}\right).$$

By selecting

$$\lambda = \begin{cases} \frac{T}{2\sqrt{2}\sigma e} - \frac{T}{2\gamma}, & \text{if } \gamma \geq \sqrt{2}\sigma e, \\ 0, & \text{if } \gamma < \sqrt{2}\sigma e, \end{cases}$$

we follow the proof [34, Theorem IV.2] to achieve

$$\text{Prob}\left(\frac{1}{T}\sum_{k \in \mathcal{T}} (\|\boldsymbol{w}_k\|_\infty) \geq \gamma\right) \leq \begin{cases} \exp\left(-\frac{T(\gamma^2 - \sqrt{2}\sigma e\gamma)}{2\sqrt{2}\sigma e(\gamma + \sqrt{2}\sigma e)}\right), & \text{if } \gamma \geq \sqrt{2}\sigma e, \\ 1, & \text{if } \gamma < \sqrt{2}\sigma e. \end{cases}$$

By bound propagation, we have

$$\text{Prob}\left(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star\|_\infty \leq \gamma\right) \geq 1 - \exp\left(-\frac{(\gamma^2 - \sqrt{2}c\gamma)T}{2\sqrt{2}\left(c\gamma + \sqrt{2}c^2\right)}\right),$$

with $\gamma > \sqrt{2}c$ and c is selected as in [29, Theorem 2]. Finally, the combination of all the above considerations complete the proof. ■

Theorem III.1 provides a methodology to construct online ambiguity sets with guarantees in probability. In general, $\rho(t)$ is strictly smaller than 1 unless there is a way of making $\boldsymbol{\alpha}(t) \to \boldsymbol{\alpha}^\star$. This is implemented in [29] via an online learning algorithm which leads to $\rho(t) \to 1 - \beta$ via Eqn. (7) in the same work. Notice how these constructions are related to the decision variable $\boldsymbol{u}$ and, in the following, we leverage the probabilistic characterization $\mathcal{P}_{t+1} := \mathcal{P}_{t+1}(\boldsymbol{\alpha}, \boldsymbol{u})$ of the distribution $\mathbb{P}_{t+1|t}$ for solutions to (P).

IV. A Tractable Problem Reformulation and Its Specialization to Two Problem Classes

In this section, we start by describing how to deal with the unknown $\mathbb{P}_{t+1|t}$ in Problem (P), via ambiguity sets, which

[2] $\xi_k^{(i)}(\boldsymbol{\alpha}, \boldsymbol{u}) := f^{(i)}(t, \hat{\boldsymbol{x}}_t, \boldsymbol{u}) + \hat{\boldsymbol{x}}_{k+1}/(\boldsymbol{\alpha}^\top \mathbf{1}_p) - f^{(i)}(k, \hat{\boldsymbol{x}}_k, \boldsymbol{u}_k)$, with $\hat{\boldsymbol{x}}_t$, $\hat{\boldsymbol{x}}_{k+1}$, $\hat{\boldsymbol{x}}_k$ being the state measurements at time t, $k+1$, k and $\boldsymbol{u}_k$ being the past input at k, $k \in \mathcal{T}$.

[3] In [29, Theorem 2], the value $\boldsymbol{d}$ plays the role of $\boldsymbol{u}$ in this work.

[4] Lemma 1 in [29] makes use of a stronger Assumption III.1, which requires $\boldsymbol{w}_k$ to be white. However, this can be relaxed to the current assumption by multiplying the upper bound in the lemma with a constant $M > 0$ associated with noise whitening via an appropriate linear transformation.

results in (P1). By doing this, the solution of (P1) provides guarantees on the performance of (P). Unfortunately, this results into an online intractable problem. Thus, we find a tractable reformulation (P2) which is equivalent to (P1) under certain conditions. After this, we focus the rest of our work on two problem sub-classes, which allows us to present and analyze the online algorithms for these problems in the following section. Formally, let us consider

$$\min_{\boldsymbol{u}\in\mathcal{U}} \sup_{\mathbb{Q}\in\mathcal{P}_{t+1}(\boldsymbol{\alpha},\boldsymbol{u})} \mathbb{E}_{\mathbb{Q}}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right], \tag{P1}$$

where, for a fixed $\boldsymbol{\alpha} := \boldsymbol{\alpha}_t$ and $\boldsymbol{u} := \boldsymbol{u}_t \in \mathcal{U}$, it holds that $\mathbb{P}_{t+1|t} \in \mathcal{P}_{t+1}(\boldsymbol{\alpha}, \boldsymbol{u})$ with high probability. This results in

$$\text{Prob}\left(\mathbb{E}_{\mathbb{P}_{t+1|t}}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] \leq \sup_{\mathbb{Q}\in\mathcal{P}_{t+1}} \mathbb{E}_{\mathbb{Q}}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right]\right) \geq \rho(t).$$

Observe that, the probability measure Prob and the bound $\rho(t)$ coincides with that in (3) and notice how the value $\rho(t)$ changes for various data-set sizes T in Theorem III.1.

The solution $\boldsymbol{u}$ and the objective value of (P1) ensure that, when we select $\boldsymbol{u}$ to be the decision for (P), the expected loss of (P) is no worse than that from (P1) with high probability. The formulation (P1) still requires expensive online computations due to its semi-infinite inner optimization problem. Thus, we propose an equivalent reformulation of (P1) for a class of loss functions as in the following assumption.

Assumption IV.1 (Lipschitz loss functions) Consider the loss function $\ell : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, $(\boldsymbol{u},\boldsymbol{x}) \mapsto \ell(\boldsymbol{u},\boldsymbol{x})$. There exists a Lipschitz function $L : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ such that for each $\boldsymbol{u} \in \mathbb{R}^m$, it holds that $\|\ell(\boldsymbol{u},\boldsymbol{x}) - \ell(\boldsymbol{u},\boldsymbol{y})\| \leq L(\boldsymbol{u})\|\boldsymbol{x}-\boldsymbol{y}\|$ for any $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$.

With this, we obtain the following upper bound:

Lemma IV.1 (An upper bound of (P1)) *Let Assumption IV.1 hold. Then, for each $\boldsymbol{u}$, $\boldsymbol{\alpha}$, β, T and t, we have*

$$\sup_{\mathbb{Q}\in\mathcal{P}_{t+1}(\boldsymbol{\alpha},\boldsymbol{u})} \mathbb{E}_{\mathbb{Q}}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] \leq \mathbb{E}_{\hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha},\boldsymbol{u})}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] + \hat{\epsilon}(t,T,\beta,\boldsymbol{\alpha},\boldsymbol{u})L(\boldsymbol{u}),$$

where the empirical distribution $\hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha},\boldsymbol{u})$ and scalar $\hat{\epsilon}(t,T,\beta,\boldsymbol{\alpha},\boldsymbol{u})$ are described as in Section III-A.

Hereafter, see the appendix for all proofs.

Next, we claim that the upper bound in Lemma IV.1 is tight if the following assumption holds.

Assumption IV.2 (Convex and gradient-accessible functions) The loss function ℓ is convex in $\boldsymbol{x}$ for each $\boldsymbol{u}$. Further, for each time t with given $\boldsymbol{u}(=\boldsymbol{u}_t) \in \mathcal{U}$ and $\boldsymbol{\alpha}(=\boldsymbol{\alpha}_t) \in \mathbb{R}^p$, there is a system prediction $\sum_{i=1}^p \alpha_i \xi_k^{(i)}(\boldsymbol{\alpha},\boldsymbol{u})$ for some $k \in \mathcal{T}$ such that $\nabla_{\boldsymbol{x}}\ell$ exists and $L(\boldsymbol{u})$ is equal to $\|\nabla_{\boldsymbol{x}}\ell\|$ at $(\boldsymbol{u}, \sum_{i=1}^p \alpha_i \xi_k^{(i)}(\boldsymbol{\alpha},\boldsymbol{u}))$.

The above statement enables the following theorem.

Theorem IV.1 (Equivalent reformulation of (P1)) *Let Assumptions IV.1 and IV.2 hold. Let Ξ_{t+1} denote the support of the distribution $\mathbb{P}_{t+1|t}$. Then, if $\Xi_{t+1} = \mathbb{R}^n$, (P1) is equivalent to the following problem*

$$\min_{\boldsymbol{u}\in\mathcal{U}} \mathbb{E}_{\hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha},\boldsymbol{u})}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] + \hat{\epsilon}(t,T,\beta,\boldsymbol{\alpha},\boldsymbol{u})L(\boldsymbol{u}). \tag{P2}$$

Remark IV.1 (Effects of Assumptions IV.1 and IV.2) We note that Assumption IV.1 on the Lipschitz requirement of loss function is mild. In fact, many engineering problems take state values in a compact set, which then only requires the loss ℓ to be continuous. Assumption IV.2 essentially requires accessible partial gradients (in $\boldsymbol{x}$) of loss functions ℓ. For simple loss functions ℓ, e.g. linear, quadratic, etc, its partial gradient can be readily evaluated. Notice that when Assumption IV.2 fails, Problem (P2) still serves as a relaxation problem of (P1), providing a solution with a valid upper bound.

Notice that the tractability of solutions to (P2) now depend on: 1) the choice of the loss function ℓ and the associated Lipschitz function L, and 2) the decision space $\mathcal{U}$. To be able to further analyze (P2) and further evaluate Assumption IV.2 on gradient-accessible functions, we will impose further structure on the system as follows:

Assumption IV.3 (Locally Lipschitz, control-affine system and basis functions) *The system f is locally Lipschitz in $(t,\boldsymbol{x},\boldsymbol{u})$ and affine in $\boldsymbol{u}$, i.e.,*

$$f(t,\boldsymbol{x},\boldsymbol{u}) := f_1(t,\boldsymbol{x}) + f_2(t,\boldsymbol{x})\boldsymbol{u},$$

for some unknown $f_1 : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}^n$, $f_2 : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$, $\boldsymbol{u} \in \mathcal{U}$ and $t \in \mathbb{Z}_{\geq 0}$. Similarly, for each $i \in \{1,\ldots,p\}$, the basis function $f^{(i)}$ is selected to be

$$f^{(i)}(t,\boldsymbol{x},\boldsymbol{u}) := f_1^{(i)}(t,\boldsymbol{x}) + f_2^{(i)}(t,\boldsymbol{x})\boldsymbol{u},$$

for some known locally Lipschitz functions $f_1^{(i)}$ and $f_2^{(i)}$.

Assumption IV.4 (Convex decision oracle) *The set $\mathcal{U}$ is convex and compact. Furthermore, the projection operation of $\boldsymbol{u} \in \mathbb{R}^m$ onto $\mathcal{U}$, $\Pi_{\mathcal{U}}(\boldsymbol{u})$, admits $O(1)$ computation complexity.*

For simplicity of the discussion, we rewrite (P2) as

$$\min_{\boldsymbol{u}\in\mathcal{U}} G(t,\boldsymbol{u}) := G(t,\boldsymbol{u}|\ell,L,T,\beta,\boldsymbol{\alpha},\hat{\mathbb{P}}_{t+1|t},\hat{\epsilon}),$$

where G represents the objective function of (P2), depending on variables ℓ, L, β, $\boldsymbol{\alpha}$, and $\mathcal{P}_{t+1}$, which are kept fixed in the optimization. Then, Assumption IV.3 allows an explicit expression of G w.r.t. $\boldsymbol{u} := \boldsymbol{u}_t$ and Assumption IV.4 characterizes the convex feasible set of (P2). Note that $G(t,\boldsymbol{u})$ is locally Lipschitz in t.[5]

In the following, we consider two classes of general problems in the form of (P2): 1) an optimal control problem under the uncertainty; 2) an online resource allocation problem with a switch. These problems leverage the probabilistic characterization of the system and common loss functions ℓ. Then, we propose an online algorithm to achieve tractable solutions with a probabilistic regret bound in the next section.

Problem 1: (Optimal control under uncertainty) We consider a problem in form (P), where the system is unknown and is to be optimally controlled. In particular, we employ the following separable loss function

$$\ell(\boldsymbol{u},\boldsymbol{x}) := \ell_1(\boldsymbol{u}) + \ell_2(\boldsymbol{x}), \quad \ell_1 : \mathbb{R}^m \to \mathbb{R},\ \ell_2 : \mathbb{R}^n \to \mathbb{R},$$

with ℓ_1 the cost for the immediate control and ℓ_2 the optimal cost-to-go function. We assume that both ℓ_1 and ℓ_2 are convex,

[5]This can be verified by the local Lipschitz condition on $f^{(i)}$, ℓ, and finite composition of local Lipschitz functions are locally Lipschitz.

and in addition, ℓ_2 is Lipschitz continuous with a constant $\mathrm{Lip}(\ell_2) \in \mathbb{R}_{\geq 0}$, resulting in $L(\boldsymbol{u}) \equiv \mathrm{Lip}(\ell_2)$. Then, by selecting the ambiguity radius $\hat{\epsilon}$ and center $\hat{\mathbb{P}}_{t+1|t}$ of $\mathcal{P}_{t+1}$ as in Section III-A, the objective function of (P2) becomes

$$G(t,\boldsymbol{u}) = \ell_1(\boldsymbol{u}) + \frac{1}{T}\sum_{k\in\mathcal{T}} \ell_2(\boldsymbol{p}_{k,t}) + \mathrm{Lip}(\ell_2)\epsilon + \frac{\gamma\,\mathrm{Lip}(\ell_2)}{T}\sum_{i=1}^{p}\sum_{k\in\mathcal{T}} \|\boldsymbol{H}_k^{(i)}\|,$$

where $\boldsymbol{p}_{k,t}$, $\boldsymbol{H}_k^{(i)} \in \mathbb{R}^n$ are affine in $\boldsymbol{u}$, for each i, k, as

$$\boldsymbol{p}_{k,t} := \sum_{i=1}^{p}\alpha_i\left(f_1^{(i)}(t,\hat{\boldsymbol{x}}_t) - f^{(i)}(k,\hat{\boldsymbol{x}}_k,\boldsymbol{u}_k)\right) + \hat{\boldsymbol{x}}_{k+1} + \left(\sum_{i=1}^{p}\alpha_i f_2^{(i)}(t,\hat{\boldsymbol{x}}_t)\right)\boldsymbol{u},$$

$$\boldsymbol{H}_k^{(i)}(\boldsymbol{u}) := f^{(i)}(k,\hat{\boldsymbol{x}}_k,\boldsymbol{u}_k) - f_1^{(i)}(t,\hat{\boldsymbol{x}}_t) - f_2^{(i)}(t,\hat{\boldsymbol{x}}_t)\boldsymbol{u},$$

and parameters $\boldsymbol{\alpha} \in \mathbb{R}^p$, $\epsilon \in \mathbb{R}_{\geq 0}$ and $\gamma \in \mathbb{R}_{\geq 0}$ are selected as in [29, Section IV]. Intuitively, $\boldsymbol{p}_{k,t}$ is the k^{th} projected outcome of the random variable $\boldsymbol{x}_{t+1}$ and $\boldsymbol{H}_k^{(i)}$ quantifies the variation of predictor $f^{(i)}$ with respect to its previous k^{th} value. Notice that the objective function G is convex in $\boldsymbol{u}$ and therefore online problems (P2) are tractable. In addition, if ℓ_2 has a constant gradient almost everywhere, then Assumption IV.2 on accessible gradients holds and (P2) is equivalent to (P1).

Problem 2: (Online resource allocation) We consider an online resource allocation problem with a switch, where a decision maker aims to make online resource allocation decisions in an uncertain environment. This problem is in form (P) and its objective is

$$\ell(\boldsymbol{u},\boldsymbol{x}) = \max\{0, 1 - \langle \boldsymbol{u}, \phi(\boldsymbol{x})\rangle\}, \quad \phi : \mathbb{R}^n \to \mathbb{R}^m,$$

where ϕ is an affine feature map selected in advance. The decision maker updates the decision $\boldsymbol{u}$ online when $\langle \boldsymbol{u}, \phi(\boldsymbol{x})\rangle < 1$, otherwise switches off. Notice that this type of objective functions appears in many classification problems. In particular, we assume that the system f is independent from the allocation variable, i.e., $f_2 \equiv 0$. See Section VI-B for a more explicit problem formulation involving resource allocation with an assignment switch.

Then, problem (P2) has the objective function

$$G(t,\boldsymbol{u}) = \frac{1}{T}\sum_{k\in\mathcal{T}} \max\{0, 1 - \langle \boldsymbol{u}, \phi(\boldsymbol{p}_{k,t})\rangle\} + q_t L(\boldsymbol{u}),$$

where time-dependent parameters $\boldsymbol{p}_{k,t} \in \mathbb{R}^n$, $q_t \in \mathbb{R}$ are

$$\boldsymbol{p}_{k,t} = \hat{\boldsymbol{x}}_{k+1} + \sum_{i=1}^{p}\alpha_i\left(f_1^{(i)}(t,\hat{\boldsymbol{x}}_t) - f_1^{(i)}(k,\hat{\boldsymbol{x}}_k)\right),\ \forall\, k,\ t,$$

$$q_t = \epsilon + \frac{\gamma}{T}\sum_{i=1}^{p}\sum_{k\in\mathcal{T}} \|f_1^{(i)}(k,\hat{\boldsymbol{x}}_k) - f_1^{(i)}(t,\hat{\boldsymbol{x}}_t)\|,\ \forall\, t,$$

with $\boldsymbol{\alpha} \in \mathbb{R}^p$, $\epsilon \in \mathbb{R}_{\geq 0}$ and $\gamma \in \mathbb{R}_{\geq 0}$ as in [29, Section IV]. We characterize the function $L(\boldsymbol{u})$ by subgradients of the loss function ℓ.

Lemma IV.2 (Quantification of L) *Consider $\ell(\boldsymbol{u},\boldsymbol{x}) := \max\{0, 1 - \langle \boldsymbol{u}, \phi(\boldsymbol{x})\rangle\}$, where $\phi(\boldsymbol{x})$ is differentiable in $\boldsymbol{x}$. Then, the function $L(\boldsymbol{u})$ is*

$$L(u) = \sup_{\boldsymbol{g}\in\partial_{\boldsymbol{x}}\ell(\boldsymbol{u},\boldsymbol{x}),\ \boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{g}\|,$$

where the set $\partial_{\boldsymbol{x}}\ell(\boldsymbol{u},\boldsymbol{x})$ contains all the subgradients of ℓ at $\boldsymbol{x}$, given any $\boldsymbol{u}$ in advance, i.e.,

$$\partial_{\boldsymbol{x}}\ell(\boldsymbol{u},\boldsymbol{x}) := h(\boldsymbol{x},\boldsymbol{u}) \cdot \frac{\partial\phi}{\partial\boldsymbol{x}}(\boldsymbol{x})\boldsymbol{u},$$

where

$$h(\boldsymbol{x},\boldsymbol{u}) = \begin{cases} -1, & \text{if } \langle \boldsymbol{u}, \phi(\boldsymbol{x})\rangle < 1 \\ 0, & \text{if } \langle \boldsymbol{u}, \phi(\boldsymbol{x})\rangle > 1 \\ [-1, 0], & \text{otherwise} \end{cases}.$$

In particular, if $\phi(\boldsymbol{x}) := J\boldsymbol{x}$ for some matrix $J \in \mathbb{R}^{m\times n}$, then $L(u) = \|J^\top \boldsymbol{u}\|$. If $\boldsymbol{x}$ is contained in a compact set X, then $L(u) = \mathrm{Lip}(\phi)\|\boldsymbol{u}\|$,where $\mathrm{Lip}(\phi) \in \mathbb{R}_{\geq 0}$ is the Lipschitz constant of ϕ on X.

Lemma IV.2 indicates that, given a properly selected feature mapping ϕ, the objective G is convex in $\boldsymbol{u}$ and therefore online problems (P2) are convex and tractable. In addition, if ϕ is a linear map almost everywhere, then Assumption IV.2 on accessible gradients holds and (P2) is equivalent to (P1).

V. Online Algorithms

Online convex problems (P2) are non-smooth due to the normed regularization terms in G. To achieve fast, online solutions, we propose a two-step procedure. First, we follow [35], [36] to obtain a smooth version of (P2), called (P2′). Then, we extend the *Nesterov's accelerated-gradient* method [37]—known to achieve an optimal first-order convergence rate for smooth and offline convex problems—to solve the problem (P2′). Finally, we quantify the dynamic regret [4] of online decisions w.r.t. solutions of (P1) in probability.

Step 1: (Smooth approximation of (P2)) To simplify the discussion, let us use the generic notation $F : \mathcal{U} \to \mathbb{R}$ for a convex and potentially non-smooth function, which can represent any particular component of the objective function $G(t,\boldsymbol{u})$ of (P2) at time t.

Definition V.1 (Smoothable function [35]) *We call a convex function $F(\boldsymbol{u})$ smoothable on $\mathcal{U}$ if there exists $a > 0$ such that, for every $\mu > 0$, there is a continuously differentiable convex function $F_\mu : \mathcal{U} \to \mathbb{R}$ satisfying*
(1) $F_\mu(\boldsymbol{u}) \leq F(\boldsymbol{u}) \leq F_\mu(\boldsymbol{u}) + a\mu$, for all $\boldsymbol{u} \in \mathcal{U}$.
(2) There exists $b > 0$ such that F_μ has a Lipschitz gradient over $\mathcal{U}$ with Lipschitz constant b/μ, i.e.,

$$\|\nabla F_\mu(\boldsymbol{u}_1) - \nabla F_\mu(\boldsymbol{u}_2)\| \leq \frac{b}{\mu}\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|,\ \forall\, \boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathcal{U}.$$

To obtain a smooth approximation F_μ of F, we follow the *Moreau proximal approximation* technique [35], described as in the following lemma.

Lemma V.1 (Moreau-Yosida approximation) *Given a convex function $F : \mathcal{U} \to \mathbb{R}$ and any $\mu > 0$, let us denote by $\partial F(\boldsymbol{u})$ the set of subgradients of F at $\boldsymbol{u}$, respectively. Let $D := \sup_{g\in\partial F(\boldsymbol{u}),\boldsymbol{u}\in\mathcal{U}} \|g\|^2 < +\infty$. Then, F is smoothable*

with parameters $(a,b) := (D/2,1)$*, where the smoothed version* $F_\mu : \mathcal{U} \to \mathbb{R}$ *is the Moreau approximation:*

$$F_\mu(\boldsymbol{u}) := \inf_{\boldsymbol{z}\in\mathcal{U}} \left\{ F(\boldsymbol{z}) + \frac{1}{2\mu}\|\boldsymbol{z}-\boldsymbol{u}\|^2 \right\},\ \boldsymbol{u}\in\mathcal{U}.$$

In addition, if F *is* M*-strongly convex with some* $M > 0$*, then* F_μ *is* $M/(1+\mu M)$*-strongly convex. And further, the minimization of* $F(\boldsymbol{u})$ *over* $\boldsymbol{u}\in\mathcal{U}$ *is equivalent to that of* $F_\mu(\boldsymbol{u})$ *over* $\boldsymbol{u}\in\mathcal{U}$ *in the sense that the set of minimizers of two problems are the same.*

From the definition of the smoothable function, we know that: 1) a positive linear combination of smoothable functions is smoothable[6], and 2) the composition of a smoothable function with a linear transformation is smoothable[7]. These properties enable us to smooth each component of G, i.e., ℓ_1, ℓ_2, h and $\|\cdot\|$, which results in a smooth approximation of (P2) via the corresponding G_μ as follows

$$\min_{\boldsymbol{u}\in\mathcal{U}} G_\mu(t,\boldsymbol{u}). \tag{P2$'$}$$

Note that G_μ is locally Lipschitz and minimizers of (P2$'$) are that of (P2). We provide in the following lemma explicit expressions of (P2$'$) for the two problem classes.

Lemma V.2 (Examples of (P2$'$))
Problem 1: *Consider the following loss function*

$$\ell(\boldsymbol{u},\boldsymbol{x}) := \frac{1}{2}\|\boldsymbol{u}\|^2 + F_\mu(\boldsymbol{x}),\quad \textit{given some } \mu>0,$$

where $F_\mu : \mathbb{R}^n \to \mathbb{R}$ *is a smoothed* ℓ_2*-norm function*[8]*, with* $\mathrm{Lip}(F_\mu) = 1$*. Then, the objective function* $G_\mu(t,\boldsymbol{u})$ *is*

$$\frac{1}{2}\|\boldsymbol{u}\|^2 + \frac{1}{T}\sum_{k\in\mathcal{T}} F_\mu(\boldsymbol{p}_{k,t}) + \epsilon + \frac{\gamma}{T}\sum_{i=1}^{p}\sum_{k\in\mathcal{T}} F_\mu(\boldsymbol{H}_k^{(i)}),$$

where $\boldsymbol{p}$*,* $\boldsymbol{H}$ *are affine in* $\boldsymbol{u}$*, defined as in Section IV. In addition, we have the smoothing parameter of* $G_\mu(t,\boldsymbol{u})$*,* $(a,b) := ((1+p\gamma)/2, \mu + s_0 + \gamma\sum_i s_i)$*, where*

$$s_0 = \sigma_{\max}\left(\left(\sum_{i=1}^{p} \alpha_i f_2^{(i)}(t,\hat{\boldsymbol{x}}_t) \right)^\top \left(\sum_{i=1}^{p} \alpha_i f_2^{(i)}(t,\hat{\boldsymbol{x}}_t) \right) \right),$$

with $\sigma_{\max}$ *denoting the maximum singular value of the matrix, and*

$$s_i = \sigma_{\max}\left(f_2^{(i)}(t,\hat{\boldsymbol{x}}_t)^\top f_2^{(i)}(t,\hat{\boldsymbol{x}}_t) \right),\ i\in\{1,\dots,p\}.$$

Problem 2: *Let us select the feature map* ϕ *to be the identity map with the dimension* $m = n$*, and consider*

$$\ell(\boldsymbol{u},\boldsymbol{x}) := \max\{0, 1-\langle \boldsymbol{u},\boldsymbol{x}\rangle\},\quad \textit{with } L(u) = \|\boldsymbol{u}\|,$$

resulting in

$$G_\mu(t,\boldsymbol{u}) = \frac{1}{T}\sum_{k\in\mathcal{T}} F_\mu^{\mathrm{S}}(\langle \boldsymbol{u}, \boldsymbol{p}_{k,t}\rangle) + q_t F_\mu(\boldsymbol{u}),$$

where $\mu > 0$*, parameters* $\boldsymbol{p}$*,* q *are as in Section IV, and functions* $F_\mu^{\mathrm{S}} : \mathbb{R}\to\mathbb{R}$ *and* $F_\mu : \mathbb{R}^n \to \mathbb{R}$ *are the smoothed switch function*[9] *and* ℓ_2*-norm function*[8]*, respectively. Note that* G_μ *has the smoothing parameter* $(a,b) := ((1+q_t)/2, q_t + 1/T\sum_{k\in\mathcal{T}}\|\boldsymbol{p}_{k,t}\|_\infty^2)$*.* ■

Step 2: (Solution to (P2$'$) as a dynamical system) To solve (P2$'$) online, we propose a dynamical system extending the Nesterov's accelerated-gradient method by adapting gradients of the time-varying objective function. In particular, let $\boldsymbol{u}_t$, $t\in\mathbb{Z}_{\geq 0}$, be solutions of (P2$'$) and let us consider the solution system with some $\boldsymbol{u}_0\in\mathcal{U}$ and $\boldsymbol{y}_0 = \boldsymbol{u}_0$, as

$$\begin{aligned} \boldsymbol{u}_{t+1} &= \Pi_{\mathcal{U}}(\boldsymbol{y}_t - \varepsilon_t \nabla G_\mu(t,\boldsymbol{y}_t)),\\ \boldsymbol{y}_{t+1} &= \boldsymbol{u}_{t+1} + \eta_t(\boldsymbol{u}_{t+1} - \boldsymbol{u}_t), \end{aligned} \tag{5}$$

where $\varepsilon_t \leq \mu/b_t$ with positive parameters μ and $b_t := b$ being those define $G_\mu(t,\boldsymbol{u})$. We denote by ∇G_μ the derivative of G_μ w.r.t. its second argument and denote by $\Pi_{\mathcal{U}}(\boldsymbol{y})$ the projection of $\boldsymbol{y}$ onto $\mathcal{U}$ as in Assumption IV.4 on convex decision oracle. Note that, the gradient function ∇G_μ can be computed in closed form for problems of interest, see, e.g., Appendix A for those of the proposed problems. Further, we select the moment coefficient $\eta_t \in \mathbb{R}_{\geq 0}$ as in Appendix B. In the following, we leverage Appendix B on the stability analysis of the solution system (5) for a regret bound between online decisions and optimal solutions of (P1).

[6] If F_1 is smoothable with parameter (a_1,b_1) and F_2 with parameter (a_2,b_2), then $c_1F_1 + c_2F_2$ is smoothable with parameter $(c_1a_1 + c_2a_2, c_1b_1 + c_2b_2)$, for any $c_1,c_2\geq 0$.

[7] Let $A : \mathcal{U}\to\mathcal{X}$ be a linear transformation and let $\boldsymbol{b}\in\mathcal{X}$. Let $\ell : \mathcal{X}\to\mathbb{R}$ be a smoothable function with parameter (a,b). Then, the function $F : \mathcal{U}\to\mathbb{R}$, $\boldsymbol{u}\mapsto \ell(A\boldsymbol{u}+\boldsymbol{b})$ is smoothable with parameter $(a, b\|A\|^2)$, where $\|A\| := \max_{\|\boldsymbol{u}\|=1}\|A\boldsymbol{u}\|$. If $\mathcal{X} = \mathbb{R}$, then $\|A\|$ is the ℓ_∞ norm.

[8] **The ℓ_2-norm function:** Consider $\boldsymbol{x}\in\mathbb{R}^n$, $F : \boldsymbol{x}\mapsto\|\boldsymbol{x}\|$, and $\mu>0$. Clearly, F is differentiable almost everywhere, except at the origin. Then,

$$\begin{aligned} F_\mu(\boldsymbol{x}) &:= \min_{\boldsymbol{z}\in\mathbb{R}^n}\left\{\|\boldsymbol{z}\| + \frac{1}{2\mu}\|\boldsymbol{z}-\boldsymbol{x}\|^2\right\},\\ &= \min_{r\geq 0}\min_{\|\boldsymbol{z}\|=r}\left\{ r + \frac{1}{2\mu}\left(r^2 - 2\boldsymbol{z}^\top\boldsymbol{x} + \|\boldsymbol{x}\|^2\right)\right\},\\ &= \min_{r\geq 0}\left\{ r + \frac{1}{2\mu}\left(r^2 - 2r\|\boldsymbol{x}\| + \|\boldsymbol{x}\|^2\right)\right\},\\ &= \begin{cases} \frac{\|\boldsymbol{x}\|^2}{2\mu}, & \text{if } \|\boldsymbol{x}\|\leq\mu,\\ \|\boldsymbol{x}\| - \frac{\mu}{2}, & \text{otherwise,}\end{cases} \end{aligned}$$

with the smoothing parameter $(1/2,1)$.

[9] **The Switch function:** Consider $u\in\mathbb{R}$, $F^{\mathrm{S}} : u\mapsto\max\{0,1-u\}$, which is differentiable almost everywhere. For a given $\mu>0$, we compute

$$\begin{aligned} F_\mu^{\mathrm{S}}(u) &:= \min_{z\in\mathbb{R}}\left\{\max\{0,1-z\} + \frac{1}{2\mu}\|z-u\|^2\right\},\\ &= \min\left\{\min_{z\leq 1} 1 - z + \frac{1}{2\mu}\|z-u\|^2, \min_{z\geq 1}\frac{1}{2\mu}\|z-u\|^2\right\}. \end{aligned}$$

Given that

$$\min_{z\leq 1} 1 - z + \frac{1}{2\mu}\|z-u\|^2 = \begin{cases}\frac{1}{2\mu}\|1-u\|^2, & \text{if } u > 1-\mu,\\ 1-u-\frac{\mu}{2}, & \text{if } u\leq 1-\mu,\end{cases}$$

and

$$\min_{z\geq 1}\frac{1}{2\mu}\|z-u\|^2 = \begin{cases}\frac{1}{2\mu}\|1-u\|^2, & \text{if } u<1,\\ 0, & \text{if } u\geq 1,\end{cases}$$

resulting in

$$F_\mu^{\mathrm{S}}(u) := \begin{cases} 1-u-\frac{\mu}{2}, & \text{if } u\leq 1-\mu,\\ \frac{1}{2\mu}\|1-u\|^2, & \text{if } 1-\mu\leq u<1,\\ 0, & \text{if } u\geq 1,\end{cases}$$

with the smoothing parameter $(1/2,1)$.

Theorem V.1 (Probabilistic regret bound of (P1)) *Given any $t \geq 2$, let us denote by $\boldsymbol{u}_t$ and $\boldsymbol{u}_t^\star$ the decision generated by (5) and an optimal solution which solves the online Problem (P1), respectively. Consider the dynamic regret to be the difference of the cost expected to incur if we implement $\boldsymbol{u}_t$ instead of $\boldsymbol{u}_t^\star$, defined as*

$$R_t := \mathbb{E}_{\mathbb{P}_{t+1|t}}\left[\ell(\boldsymbol{u}_t, \boldsymbol{x})\right] - \mathbb{E}_{\mathbb{P}_{t+1|t}}\left[\ell(\boldsymbol{u}_t^\star, \boldsymbol{x})\right].$$

Then, the regret R_t is bounded in probability as follows

$$\text{Prob}\left(R_t \leq \frac{4W_t}{(t+2)^2} + TF_t + a\mu + 2L(\boldsymbol{u}_t^\star)\hat{\epsilon}\right) \geq \rho(t),$$

where W_t depends on the system state at time $t-T$, and F_t depends on the variation of the optimal objective values in $\mathcal{T}$, i.e.,

$$F_t = \max_{k \in \mathcal{T}} \left\{|G_{k+1}^\star - G_k^\star|\right\} + \bar{L},$$

where $G_k^\star := G(k, \boldsymbol{u}_k^\star)$ is the optimal objective value of (P2), or equivalently that of (P1). Further, $\bar{L}$ is the variation bound of G w.r.t. time, and the rest of the parameters are the same as before. Furthermore, if all historical data are assimilated for the decision $\boldsymbol{u}_t$, then, we have

$$\liminf_{t\to\infty} \text{Prob}\left(R_t \leq TF_t + a\mu\right) \geq 1 - \beta,$$

with β a given, arbitrary confidence value.

Theorem V.1 quantifies the dynamic regret of online decisions $\boldsymbol{u}$ w.r.t. solutions to (P1) in high probability. Notice that, the regret bound is dominated by terms: TF_t, $a\mu$ and $L(\boldsymbol{u}_t^\star)\hat{\epsilon}$, which mainly depend on three factors: the data-driven parameters ε, η and μ of the solution system (5), the variation F_t over optimal objective values, and the parameters T, β, γ and $\hat{\epsilon}$ that are related to the system and environment learning. In practice, a small regret bound is determined by 1) an effective learning procedure which contributes to small $\hat{\epsilon}$; 2) a proper selection of the loss function ℓ which results in smoothing procedure with a small parameter $a\mu$; and 3) the problem structure leading to small variations F_t of the optimal objectives values. Furthermore, when we use all the historical data for the objective gradients in the solution system (5), the effect of system ambiguity learning is negligible asymptotically.

Online Procedure: Our online algorithm is summarized in the Algorithm 1.

Online Optimization and Learning Algorithm 1 Opal($\mathcal{I}$)

1: Select $\{f^{(i)}\}_i$, ℓ, β, $\mathcal{U}$, $\boldsymbol{u}_0$, μ, and $t = 1$;
2: **repeat**
3: Update data set $\mathcal{I} := \mathcal{I}_t$;
4: Compute $\boldsymbol{\alpha} := \boldsymbol{\alpha}_t$ as in [29];
5: Select $\hat{\mathbb{P}}_{t+1|t}$ in (2) and $\hat{\epsilon} := \hat{\epsilon}(t, T, \beta, \boldsymbol{\alpha}, \boldsymbol{u})$ in (4);
6: Run dynamical system (5) for $\boldsymbol{u} := \boldsymbol{u}_t$;
7: Apply $\boldsymbol{u}$ to (P) with the regret guarantee;
8: $t \leftarrow t + 1$;
9: **until** time t stops.

VI. Implementation

In this section, we apply our algorithm to the introduced motivating examples, resulting in online-tractable, effective system learning with guaranteed, regret-bounded performance in high probability.

A. Optimal control of an uncertain nonlinear system

We consider the two-wheel vehicle driving under various road conditions, and our goal is to learn one-step prediction of the system state distribution and leverage for path tracking under various unknown road zones. In particular, we represent the two-wheel vehicle as a differential-drive robot subject to uncertainty [38]:

$$\begin{aligned} x_{t+1} =& x_t + h\cos(\theta_t)d_{1,t} + hw_{1,t}, \\ y_{t+1} =& y_t + h\sin(\theta_t)d_{1,t} + hw_{2,t}, \\ \theta_{t+1} =& \theta_t - hd_{2,t} + hw_{3,t}, \\ d_{1,t} =& \frac{r}{2}(v_{l,t} + v_{r,t} + e_{1,t}), \\ d_{2,t} =& \frac{r}{2R}(v_{l,t} - v_{r,t} + e_{2,t}), \end{aligned} \tag{6}$$

where components of states $\boldsymbol{x}_t := (x_t, y_t, \theta_t) \in \mathbb{R}^2 \times [-\pi, \pi) \cong \mathbb{R} \times \mathbb{S}^1$ represent vehicle position and orientation on the 2-D plane. We take the discretization parameter $h = 0.01$ and assume subGaussian uncertainty $\boldsymbol{w}_t := (w_{1,t}, w_{2,t}, w_{3,t}) \in \mathbb{R}^3$ to be a zero-mean, mixture of Gaussian and Uniform distributions with $\sigma = 0.5$. The intermediate variable $\boldsymbol{d}_t := (d_{1,t}, d_{2,t})$ depends on the wheel radius $r = 0.15$ m, the distance between wheels $R = 0.4$ m, the controlled left-right wheel speed $\boldsymbol{u}_t := (v_{l,t}, v_{r,t})$ and an unknown parameter $\boldsymbol{e}_t := (e_{1,t}, e_{2,t})$, which depends on the wheel quality and road conditions. For simplicity, we assume that the planner adapts the system (6) with $\boldsymbol{e}_t \equiv (0, 0)$ and $\boldsymbol{w}_t \equiv (0, 0, 0)$, and the vehicle can move over three types of road zones, the regular zone with $\boldsymbol{e}^{(1)} := (0, 0)$, the slippery zone with $\boldsymbol{e}^{(2)} = (4, 0)$, and the sandy zone with $\boldsymbol{e}^{(3)} = (-1.2, -0.2)$, where locations of these zones are described in Fig. 2.

To adapt the proposed approach, we consider Problem (P) with the following loss function

$$\begin{aligned} \ell(\boldsymbol{u}, x, y, \theta) =& \frac{1}{20}\|\boldsymbol{u} - \boldsymbol{u}^{\text{ref}}\|^2 + \frac{1}{14\sqrt{2}}|x - x^{\text{ref}}| + \\ & \frac{1}{4\sqrt{2}}|y - y^{\text{ref}}| + \frac{289}{8}\left(\cos(\theta) - \cos(\theta^{\text{ref}})\right)^2 + \\ & \frac{289}{8}\left(\sin(\theta) - \sin(\theta^{\text{ref}})\right)^2, \end{aligned}$$

where $(\boldsymbol{u}^{\text{ref}}, x^{\text{ref}}, y^{\text{ref}}, \theta^{\text{ref}})$ are signals generated by the planner, and we select the parameter $\mu = 10^{-4}$ for components which are not smooth. In addition, we assume $\mathcal{U} = [-20, 20]^2$ and utilize $p = 3$ basis functions $\{f^{(i)}\}_i$ in form of (6), with $\boldsymbol{w}_t \equiv (0, 0, 0)$, and

$$\begin{aligned} &i = 1, \quad e_1 = 0, \quad e_2 = 0, \\ &i = 2, \quad e_1 = 10, \quad e_2 = 0, \\ &i = 3, \quad e_1 = 0, \quad e_2 = 10. \end{aligned}$$

Note that the ground truth parameter $\boldsymbol{\alpha}^\star := (1, 0, 0)$ in the regular zone, $\boldsymbol{\alpha}^\star := (0.6, 0.4, 0)$ in the slippery zone, and $\boldsymbol{\alpha}^\star := (1.14, -0.12, -0.02)$ in the sandy zone. At each time t, we have access to model sets $\{f^{(i)}\}_i$ and as well as the real-time data set $\mathcal{I}_t$ with size $T_0 = 100$, which corresponds to the moving time window of order 0.1 second. For the system learning algorithm, notions of norm and inner product are those defined on the vector space $T(\mathbb{R}^2 \times \mathbb{S}) \equiv \mathbb{R}^3$. We employ our online optimization and learning algorithm for

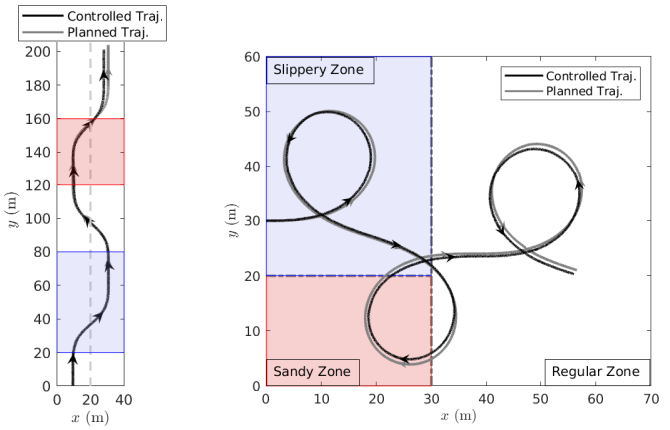
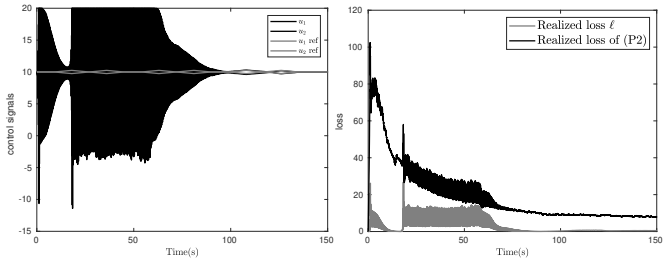


Fig. 3: An example of the (gray) planned trajectory and (black) controlled system trajectory in various road zones, with the system state $\boldsymbol{x} = (x, y, \theta)$. The red region indicates sandy zone while the blue region indicates the slippery zone. With the implemented control, the vehicle follows the planned path with low regrets in high probability.



(a) (b)

Fig. 4: (a) The (gray) control signal provided by the planner and an example of the (black) control signal derived from the proposed approach. (b) The realized loss ℓ and the achieved objective of (P2).

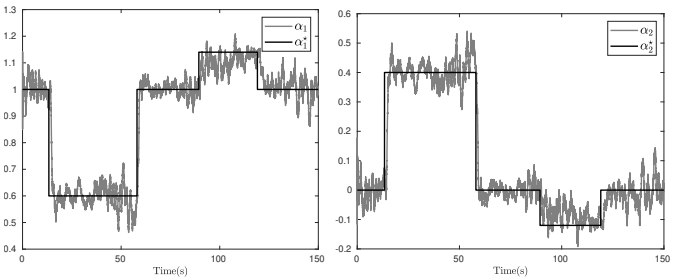


Fig. 5: The component α_1 and α_2 of the real-time parameter $\boldsymbol{\alpha} := (\alpha_1, \alpha_2, \alpha_3)$ in the learning procedure.

the characterization of the uncertain vehicle states, learning of the unknown road-condition parameter $\boldsymbol{e}$, and control towards planned behaviors in real time. The achieved system behaviors are demonstrated in Fig 3, contrasted with the case without the proposed approach, as in Fig. 2. In the following, we analyze each case separately and notice how the proposed approach strikes balance between the given planned control $\boldsymbol{u}^{\text{ref}}$ and the actual control $\boldsymbol{u}$ which reduces the weighted tracking error in road uncertainty.

Example (Lane-changing behavior adaptation) In this scenario, we assume the initial system state $\boldsymbol{x}_0 = (10, 0, \pi/2)$. Further, the vehicle can access path plan in Fig. 2(a) and as well as the suggested wheel speed plan as the gray signal in Fig. 4(a). To demonstrate the learning effect of the algorithm, we show in Fig. 5 components α_1 and α_2 of $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$, where the black lines indicate value of the ground truth $\boldsymbol{\alpha}^\star$ on the planned trajectory and the gray lines represent the learned, real-time estimate of α_1 and α_2 at the actual vehicle position. Notice that $\boldsymbol{\alpha}^\star$ is inaccessible in practice, and from this case study, the proposed approach indeed learns the system dynamics effectively. See, e.g. [29] for more analysis regarding to the effect of the learning behavior and ambiguity sets characterization on the selection of ϵ and γ.

As the proposed loss function ℓ measures the weighted tracking error, the resulting control system trajectory in Fig. 3(a) already reveals the effectiveness of the method and as well as the low regrets in probability. On the other hand, because the system is highly non-linear and uncertain, evaluating the actual optimal objective value of Problem (P) is difficult. Therefore, it's very challenging to evaluate the regret R_t in practice, even though the its probabilistic bounded is proved. Here, we provide in Fig. 4(b) the realized loss ℓ and as well as the realized objective value of Problem (P2), where the loss ℓ reveals one possible objective value of (P), and the objective value of (P2) serves as an upper-bound of that of (P) in high probability. In addition, notice that the derived (black) control signal in Fig. 4(a) has undesirable, high-oscillatory behavior. This is because the chosen loss function ℓ is only locally convex in $\boldsymbol{x}$. When the system disturbances are significant, the proposed approach then revealed certain degradation and control being oscillatory. Nevertheless, a desirable system behavior in Fig. 3(a) is achieved.

Example (Circular route tracking) In this scenario, we consider $\boldsymbol{x}_0 = (0, 30, 0)$. We omit the details as the analysis shares the same spirit as the last lane-changing example.

B. Online resource allocation problem

We consider an online resource allocation problem where an agent or decision maker aims to 1) achieve at least target profit under uncertainty, and 2) allocate resources as uniformly as possible. To do this, the agent distributes available resources, e.g., wealth, time, energy or human resources, to various projects or assets. In particular, for the trading-market motivating example, let us consider that the agent tries to make an online allocation $\boldsymbol{u} \in \mathcal{U}$ of a unit wealth to three assets. At each time t, the agent receives random return rates $\boldsymbol{x}_t \in \mathbb{R}^3_{\geq 0}$ of assets from some unknown and uncertain dynamics

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t + hA(t) + h\boldsymbol{w}_t, \text{ with some } \boldsymbol{x}_0 \in \mathbb{R}^3, \quad (7)$$

where $h = 10^{-3}$ is a stepsize, the vector $A(t)$ is randomly generated, unknown and piecewise constant, and the uncertainty vector $\boldsymbol{w}_t$ is assumed to be sub-Gaussian with $\sigma = 0.1$. Note that this model can serve to characterize a wide class of dynamic (linear and nonlinear) systems. In addition, we assume that the third asset is value preserved, i.e., the third component of $A(t)$ and $\boldsymbol{w}_t$ are zero and $x_3 \equiv 1$. Over time, an example of the resulting unit return rates $\boldsymbol{x}$ is demonstrated in Fig. 6. Then, we denote by $r_0 = 1.3$ and $\langle \boldsymbol{u}, \boldsymbol{x}_{t+1} \rangle$ the target profit and the predicted instantaneous profit, respectively. Note that the decision maker aims to obtain at least a 30% profit and allocate resources online for this purpose. In particular, the decision maker implements an

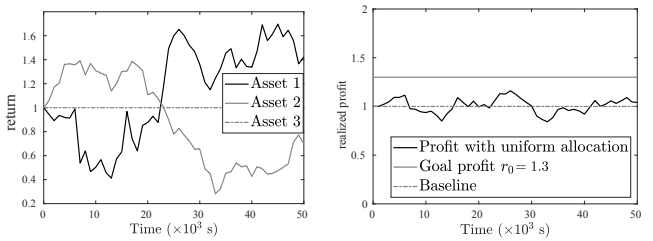


Fig. 6: An example of random returns $\boldsymbol{x} = (x_1, x_2, x_3)$, where returns of the first two assets x_1, $x_2 \in [0, +\infty)$ are highly fluctuating and the third is value-preserving with return $x_3 \equiv 1$. Without asset allocation, agent does not achieve the goal profit $r_0 = 1.3$ and has a chance of losing assets.

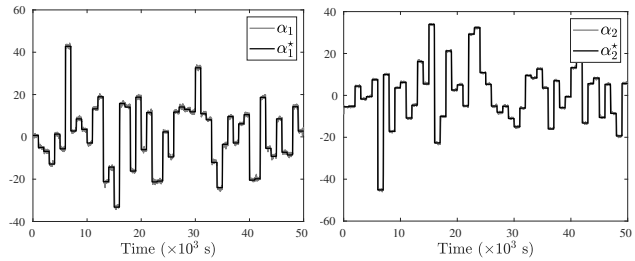


Fig. 7: The component α_1 and α_2 of the real-time parameter $\boldsymbol{\alpha} := (\alpha_1, \alpha_2, \alpha_3)$ in learning, where the values $\alpha_1^\star$ and $\alpha_2^\star$ are the online-inaccessible ground truth. Notice the responsive behavior of the proposed learning algorithm.

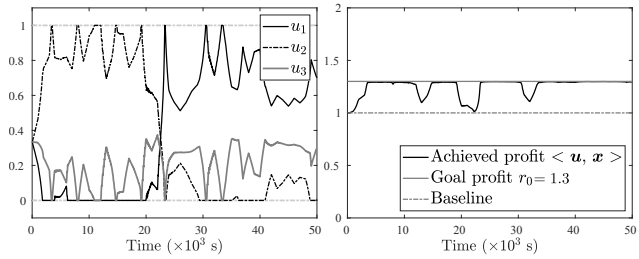


Fig. 8: Real-time resource allocation $\boldsymbol{u}$ and profit $\langle \boldsymbol{u}, \boldsymbol{x} \rangle$. Notice how the decision $\boldsymbol{u} = (u_1, u_2, u_3)$ respects constraints and how the allocation tries to balance the assets when the goal profit r_0 is met.

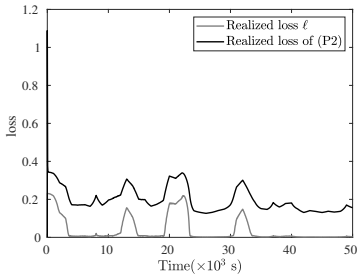


Fig. 9: The realized loss ℓ and the achieved objective of (P2).

allocation online if $\langle \boldsymbol{u}, \boldsymbol{x}_{t+1} \rangle \leq r_0$, otherwise does nothing. This results in (P) with the loss function

$$\ell(\boldsymbol{u}, \boldsymbol{x}) = \max\{0, 1 - \frac{1}{r_0}\langle \boldsymbol{u}, \boldsymbol{x} \rangle\},$$

and set $\mathcal{U}$ a unit simplex. We propose $p = 3$ basis functions

$$f^{(1)} = \boldsymbol{x},\ f^{(2)} = \boldsymbol{x} + 0.1h\boldsymbol{e}_1,\ f^{(3)} = \boldsymbol{x} + 0.1h\boldsymbol{e}_2,$$

where $\boldsymbol{e}_1 = (1, 0, 0)^\top$ and $\boldsymbol{e}_2 = (0, 1, 0)^\top$. At each t, we assume that only historical data are available for online resource allocations. Applying the proposed probabilistic characterization of $\boldsymbol{x}_{t+1}$ as in (P1), we equivalently write it as in form (P2′), where

$$G_\mu(t, \boldsymbol{u}) = \frac{1}{T}\sum_{k\in\mathcal{T}} F_\mu^{\mathrm{S}}(\langle \boldsymbol{u}, \frac{\boldsymbol{p}_{k,t}}{r_0} \rangle) + \frac{q_t}{r_0} F_\mu(\boldsymbol{u}),\ \mu = 0.01,$$

with functions F_μ^{S}[9] and F_μ[8], and real-time data $\boldsymbol{p}_{k,t}$ and q_t determined as in Problem 2. We claim that $G_\mu(t, \boldsymbol{u})$ has a time-dependent Lipschitz gradient constant in $\boldsymbol{u}$ given by $\mathrm{Lip}(G_\mu) = q_t/r_0 + 1/(r_0^2 T)\sum_{k\in\mathcal{T}} \|\boldsymbol{p}_{k,t}\|_\infty^2$, and we use $\varepsilon := 1/\mathrm{Lip}(G_\mu)$ in the solution system (5) to compute the online decisions.

Fig. 7 shows the real-time evolution α_1 and α_2 of the parameter $\boldsymbol{\alpha} := (\alpha_1, \alpha_2, \alpha_3)$, while the behavior of α_3 can be similarly characterized. In these figures, black lines $\alpha_1^\star$ and $\alpha_2^\star$ are determined by the unknown signal $A(t)$ while gray lines α_1 and α_2 are those computed as in [29]. Note that $\boldsymbol{\alpha}^\star$ represents the unknown dynamics f and they are not accessible in reality. It can be seen that the proposed method effectively learns $\boldsymbol{\alpha}^\star$.

Fig. 8 demonstrates the online resource allocation obtained by implementing (5) and the achieved real-time profit $\langle \boldsymbol{u}, \boldsymbol{x} \rangle$. The decision $\boldsymbol{u}$ starts from the uniform allocation $\boldsymbol{u}_0 = (1/3, 1/3, 1/3)$ and is then adjusted to approach the target profit $r_0 = 1.3$. Once the target is achieved, the agent then maintains the profit while trying to balance the allocation if possible. When the return rate $\boldsymbol{x}$ is low/unbalanced, as in Fig. 6, the agent tries to improve and achieve the target profit by allocating resources more aggressively. Though did not appear in the current scenario, in case that the return rate is high and the target profit value is achieved, the agent focuses on balancing the allocation while maintaining the profit. If both the target profit and allocation balance are achieved, then the agent stops re-allocating resources and monitors the return rate $\boldsymbol{x}$ until the switch turns on, e.g., when the near future profit prediction drops below r_0 again. In addition, notice how the target profit was achieved with the proposed control strategy as demonstrated in Fig. 8, which contrasts with uniform allocation case as in Fig. 6.

Fig. 9 demonstrates the evaluation of the time-varying loss ℓ as well as the realized objective value of Problem (P2). Due to the unknown time-varying distributions $\mathbb{P}_{t|t-1}$, the evaluation of the objective values of Problem (P) is intractable, and the realized loss of (P2) serves as a high-confidence upper bound of that of(P). Nevertheless, the target profit is achieved with low regret in high confidence, as revealed in Fig. 8.

VII. Conclusions

In this paper, we proposed a unified solution framework for online learning and optimization problems in form of (P). The proposed method allowed us to learn an unknown and uncertain dynamic system, while providing a characterization of the system with online-quantifiable probabilistic guarantees that certify the performance of online decisions. The approach provided tractable, online convex version of (P), via a series of equivalent reformulation techniques. We explicitly demonstrated the framework via two problem classes conforming to (P): an optimal control problem under uncertainty and an

online resource allocation problem. These two problem classes resulted in explicit, online and non-smooth convex optimization problems. We extended Nesterov's accelerated-gradient method to an online fashion and provided a solution system for online decision generation of (P). The quality of the online decisions were analytically certified via a probabilistic regret bound, which revealed its relation to the learning parameters and ambiguity sets. Two motivating examples applying the proposed framework were empirically tested, demonstrating the effectiveness of the proposed framework with the bounded regret guarantees in probability. We leave the relaxation of assumptions and the comparison of this work with other methods as the future work.

APPENDIX

A. Computation of the objective gradients

Let ℓ, G and G_μ be those in Lemma V.2 on examples of (P2′). We now derive $\nabla G_\mu := \nabla_{\boldsymbol{u}} G_\mu(t, \boldsymbol{u})$ as follows.

Problem 1 (Optimal control under uncertainty):

$$\nabla_{\boldsymbol{u}} G_\mu(t,\boldsymbol{u}) = \frac{1}{\mu}\boldsymbol{u} + \frac{1}{T}\sum_{k\in\mathcal{T}} \nabla_{\boldsymbol{u}} F_\mu(\boldsymbol{p}_{k,t}) + \frac{\gamma}{T}\sum_{i=1}^{p}\sum_{k\in\mathcal{T}} \nabla_{\boldsymbol{u}} F_\mu(\boldsymbol{H}_k^{(i)}),$$

where, for each $k \in \mathcal{T}$, the term $\nabla_{\boldsymbol{u}} F_\mu(\boldsymbol{p}_{k,t})$ is

$$\begin{cases} \frac{1}{\mu}\left(\sum_{i=1}^{p} \alpha_i f_2^{(i)}(t,\hat{\boldsymbol{x}}_t)\right)^\top \boldsymbol{p}_{k,t}, & \text{if } \|\boldsymbol{p}_{k,t}\| \le \mu, \\ \frac{1}{\|\boldsymbol{p}_{k,t}\|}\left(\sum_{i=1}^{p} \alpha_i f_2^{(i)}(t,\hat{\boldsymbol{x}}_t)\right)^\top \boldsymbol{p}_{k,t}, & \text{otherwise,} \end{cases}$$

and, for $k \in \mathcal{T}$, $i \in \{1,\dots,p\}$, the term $\nabla_{\boldsymbol{u}} F_\mu(\boldsymbol{H}_k^{(i)})$ is

$$\begin{cases} -\frac{1}{\mu}(f_2^{(i)}(t,\hat{\boldsymbol{x}}_t))^\top \boldsymbol{H}_k^{(i)}, & \text{if } \|\boldsymbol{H}_k^{(i)}\| \le \mu, \\ -\frac{1}{\|\boldsymbol{H}_k^{(i)}\|}(f_2^{(i)}(t,\hat{\boldsymbol{x}}_t))^\top \boldsymbol{H}_k^{(i)}, & \text{otherwise}. \end{cases}$$

Problem 2 (Online resource allocation):

$$\nabla_{\boldsymbol{u}} G_\mu(t,\boldsymbol{u}) = \frac{1}{T}\sum_{k\in\mathcal{T}} \nabla_{\boldsymbol{u}} F_\mu^{\mathrm{S}}(\langle \boldsymbol{u}, \boldsymbol{p}_{k,t}\rangle) + q_t \nabla_{\boldsymbol{u}} F_\mu(\boldsymbol{u}),$$

where

$$\nabla_{\boldsymbol{u}} F_\mu(\boldsymbol{u}) := \begin{cases} \frac{1}{\mu}\boldsymbol{u}, & \text{if } \|\boldsymbol{u}\| \le \mu, \\ \frac{1}{\|\boldsymbol{u}\|}\boldsymbol{u}, & \text{otherwise,} \end{cases}$$

and, for each $k \in \mathcal{T}$, the gradient $\nabla_{\boldsymbol{u}} F_\mu^{\mathrm{S}}(\langle \boldsymbol{u}, \boldsymbol{p}_{k,t}\rangle)$ is

$$\begin{cases} -\boldsymbol{p}_{k,t}, & \text{if } \langle \boldsymbol{u}, \boldsymbol{p}_{k,t}\rangle \le 1-\mu, \\ -\frac{1-\langle \boldsymbol{u}, \boldsymbol{p}_{k,t}\rangle}{\mu}\boldsymbol{p}_{k,t}, & \text{if } 1-\mu \le \langle \boldsymbol{u}, \boldsymbol{p}_{k,t}\rangle < 1, \\ 0, & \text{if } \langle \boldsymbol{u}, \boldsymbol{p}_{k,t}\rangle \ge 1. \end{cases}$$

These explicit expressions provide ingredients for the solution system. With different selections of the norm, the expression varies accordingly.

B. Stability Analysis of the Solution System

Here, we adapt dissipativity theory to address the performance of the online solution system (5). This part of the work is an online-algorithmic extension of the existing Nesterov's accelerated-gradient method and its convergence analysis in [39]–[41]. Our extension (5) inherits from the work in [40], where the difference is that gradient computations in (5) are from time-varying objective functions in (P2′). To simplify the discussion, the notation we used in this subsection is different from that in the main body of the paper. Consider the online problem, analogous to (P2′), defined as follows

$$\min_{\boldsymbol{x}\in\mathcal{X}} \; f_t(\boldsymbol{x}), \quad t = 0,1,2,\dots \tag{8}$$

where $f_t(\boldsymbol{x})$ is locally Lipschitz in t with the parameter $h(\boldsymbol{x})$ and, at each time t, the objective function f_t are m_t-strongly convex and L_t-smooth, with $m_t \ge 0$ and $L_t > 0$. The convex set $\mathcal{X} \subset \mathbb{R}^n$ is analogous to that in Assumption IV.4 on convex decision oracle. The solution system to (8), analogous to (5), is

$$\begin{aligned} \boldsymbol{x}_{t+1} &= \Pi(\boldsymbol{y}_t - \alpha_t \nabla f_t(\boldsymbol{y}_t)), \\ \boldsymbol{y}_{t+1} &= \boldsymbol{x}_{t+1} + \beta_{t+1}\,(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t), \\ &\text{with some } \boldsymbol{y}_0 = \boldsymbol{x}_0 \in \mathcal{X}, \end{aligned} \tag{9}$$

where $\alpha_t \le 1/L_t$ and β_t is selected iteratively, following

$$\delta_{-1} = 1,\ \delta_{t+1} := \frac{1+\sqrt{1+4\delta_t^2}}{2},\ \beta_t := \frac{\delta_{t-1}-1}{\delta_t}.$$

Note that $\delta_t^2 - \delta_t = \delta_{t-1}^2$, $t = 0,1,2,\dots$. The projection $\Pi(\boldsymbol{x})$ at each time t is equivalently written as

$$\Pi(\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{z}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{x}\|^2 + \alpha_t \ell(\boldsymbol{z}),$$

with $\ell(\boldsymbol{z}) = 0$ if $\boldsymbol{z} \in \mathcal{X}$, otherwise $+\infty$. Note that the projection operation is a convex problem with the objective function being strongly convex. Thus, $\Pi(\boldsymbol{x})$ is a singleton (the unique minimizer) and satisfies the optimality condition [42]

$$\boldsymbol{x} - \Pi(\boldsymbol{x}) \in \alpha_t \partial\ell(\Pi(\boldsymbol{x})),$$

where the r.h.s. is the sub-differential set of ℓ at $\Pi(\boldsymbol{x})$. Equivalently, we write the above condition as

$$\Pi(\boldsymbol{x}) = \boldsymbol{x} - \alpha_t \partial\ell(\Pi(\boldsymbol{x})).$$

We apply this equivalent representation to the solution system (9), resulting in

$$\begin{aligned} \boldsymbol{x}_{t+1} &= \boldsymbol{y}_t - \alpha_t \nabla f_t(\boldsymbol{y}_t) - \alpha_t \partial\ell(\boldsymbol{w}_t), \\ \boldsymbol{y}_{t+1} &= \boldsymbol{x}_{t+1} + \beta_{t+1}\,(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t), \\ \boldsymbol{w}_t &= \boldsymbol{x}_{t+1}. \end{aligned} \tag{10}$$

Note that (10) is not an explicit online algorithm, as the state $\boldsymbol{x}_{t+1}$ is yet to be determined. However, we leverage this equivalent reformulation for the convergence analysis of solutions to (9) to a sequence of optimizers of (8), denoted by $\{\boldsymbol{x}_t^\star\}$. To do this, let $\boldsymbol{z}_t := (\boldsymbol{x}_t - \boldsymbol{x}_t^\star,\ \boldsymbol{x}_{t-1} - \boldsymbol{x}_{t-1}^\star)$ denote the tracking error vector and represent (10) as the error dynamical system

$$\begin{aligned} \boldsymbol{z}_{t+1} &= A_t \boldsymbol{z}_t + B_t^{\boldsymbol{u}} \boldsymbol{u}_t + B_t^{\boldsymbol{v}} \boldsymbol{v}_t, \\ &\text{with } \boldsymbol{z}_1 = (\boldsymbol{x}_1 - \boldsymbol{x}_1^\star,\ \boldsymbol{x}_0 - \boldsymbol{x}_0^\star), \end{aligned} \tag{11}$$

with the gradient input $\boldsymbol{u}_t := \nabla f_t(\boldsymbol{y}_t) + \partial\ell(\boldsymbol{w}_t)$, the reference signal $\boldsymbol{v}_t := (\boldsymbol{x}_t^\star - \boldsymbol{x}_{t-1}^\star,\ \boldsymbol{x}_{t+1}^\star - \boldsymbol{x}_t^\star)$, the matrices

$$A_t = \begin{bmatrix} 1+\beta_t & -\beta_t \\ 1 & 0 \end{bmatrix},\ B_t^{\boldsymbol{u}} = \begin{bmatrix} -\alpha_t \\ 0 \end{bmatrix},\ B_t^{\boldsymbol{v}} = \begin{bmatrix} \beta_t & -1 \\ 0 & 0 \end{bmatrix},$$

and the auxiliary variables

$$\begin{aligned} \boldsymbol{y}_t - \boldsymbol{x}_t^\star &= \begin{bmatrix} 1+\beta_t & -\beta_t \end{bmatrix} \boldsymbol{z}_t + \begin{bmatrix} \beta_t & 0 \end{bmatrix} \boldsymbol{v}_t, \\ \boldsymbol{w}_t - \boldsymbol{x}_t^\star &= \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{z}_{t+1} + \begin{bmatrix} 0 & 1 \end{bmatrix} \boldsymbol{v}_t. \end{aligned}$$

We provide the following stability analysis of the system.

Theorem A.1 (Stability of (9)) *Consider the solution algorithm (9), or equivalently (10).*
(1) For each $t \geq 1$, we have the following

$$f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_{t+1}) \geq \boldsymbol{\xi}_t{}^\top X_{1,t}\, \boldsymbol{\xi}_t,$$

$$f_t(\boldsymbol{x}_t^\star) - f_t(\boldsymbol{x}_{t+1}) \geq \boldsymbol{\xi}_t{}^\top X_{2,t}\, \boldsymbol{\xi}_t.$$

Here, $\boldsymbol{\xi}_t := (\boldsymbol{z}_t,\ \boldsymbol{u}_t,\ \boldsymbol{v}_t)$, and

$$X_{1,t} := \frac{1}{2}\begin{pmatrix} m\beta^2 & -m\beta^2 & -\beta & m\beta^2 & 0 \\ -m\beta^2 & m\beta^2 & \beta & -m\beta^2 & 0 \\ -\beta & \beta & \alpha(2-L\alpha) & -\beta & 0 \\ m\beta^2 & -m\beta^2 & -\beta & m\beta^2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$X_{2,t} := \frac{1}{2}\begin{pmatrix} m(1+\beta)^2 & -\eta & -(1+\beta) & \eta & 0 \\ -\eta & m\beta^2 & \beta & -m\beta^2 & 0 \\ -(1+\beta) & \beta & \alpha(2-L\alpha) & -\beta & 0 \\ \eta & -m\beta^2 & -\beta & m\beta^2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

with $\eta = m(1+\beta)\beta$ and the parameters (m, L, α, β) are a short-hand notation for $(m_t, L_t, \alpha_t, \beta_t)$.

(2) Given the horizon parameter $T_0 \in \mathbb{Z}_{>0}$ with $T = \min\{t-1, T_0\}$. Then, for any $t \geq 2$, the solution $\boldsymbol{x}_t$ from (9) achieves

$$f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_t^\star) \leq \frac{4G_t}{(t+2)^2} + TF_t + TK_t + \frac{4(t-T-1+\delta_0)^2}{(t+2)^2}(f_{t-T}(\boldsymbol{x}_{t-T}) - f_{t-T}(\boldsymbol{x}_{t-T}^\star)).$$

where the time-dependent parameters G_t, F_t and K_t are determined by f_t, α_t and β_t.

Proof of Theorem A.1: (1) By the m-strong convexity and L-smoothness of f, we have

$$f(\boldsymbol{x}) - f(\boldsymbol{y}) \geq \nabla f(\boldsymbol{y})^\top(\boldsymbol{x}-\boldsymbol{y}) + \frac{m}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2, \quad (12)$$

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \geq \nabla f(\boldsymbol{y})^\top(\boldsymbol{y}-\boldsymbol{x}) - \frac{L}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2. \quad (13)$$

(1a) Consider (12) with $(\boldsymbol{x}, \boldsymbol{y}) \equiv (\boldsymbol{x}_t, \boldsymbol{y}_t)$. We leverage $\boldsymbol{y}_t = \boldsymbol{x}_t + \beta\,(\boldsymbol{x}_t - \boldsymbol{x}_{t-1})$ and the distributive law [10] for

$$\begin{aligned} &f(\boldsymbol{x}_t) - f(\boldsymbol{y}_t) \\ &\geq \beta\nabla f(\boldsymbol{y}_t)^\top(\boldsymbol{x}_{t-1}-\boldsymbol{x}_t) + \frac{m\beta^2}{2}\|\boldsymbol{x}_{t-1}-\boldsymbol{x}_t\|^2, \\ &= \beta(\nabla f(\boldsymbol{y}_t) + \partial\ell(\boldsymbol{w}_t))^\top(\boldsymbol{x}_{t-1}-\boldsymbol{x}_t-\boldsymbol{x}_{t-1}^\star+\boldsymbol{x}_t^\star) \\ &+ \frac{m\beta^2}{2}\|\boldsymbol{x}_{t-1}-\boldsymbol{x}_t-\boldsymbol{x}_{t-1}^\star+\boldsymbol{x}_t^\star\|^2 \\ &+ \beta(\nabla f(\boldsymbol{y}_t) + \partial\ell(\boldsymbol{w}_t))^\top(\boldsymbol{x}_{t-1}^\star-\boldsymbol{x}_t^\star) \\ &- \beta\partial\ell(\boldsymbol{w}_t)^\top(\boldsymbol{x}_{t-1}-\boldsymbol{x}_t) \\ &+ m\beta^2(\boldsymbol{x}_{t-1}-\boldsymbol{x}_t-\boldsymbol{x}_{t-1}^\star+\boldsymbol{x}_t^\star)^\top(\boldsymbol{x}_{t-1}^\star-\boldsymbol{x}_t^\star) \\ &+ \frac{m\beta^2}{2}\|\boldsymbol{x}_{t-1}^\star-\boldsymbol{x}_t^\star\|^2. \end{aligned}$$

[10] Apply 1) $a^\top c = (a+b)^\top(c-d) + (a+b)^\top d - b^\top c$ and 2) $c^\top c = (c-d)^\top(c-d) + 2(c-d)^\top d + d^\top d$, with $a = \nabla f(\boldsymbol{y}_t)$, $b = \partial\ell(\boldsymbol{w}_t)$, $c = \boldsymbol{x}_{t-1} - \boldsymbol{x}_t$, $d = \boldsymbol{x}_{t-1}^\star - \boldsymbol{x}_t^\star$,

We re-organize the the right-hand-side into the matrix form as

$$\frac{1}{2}\boldsymbol{\delta}_t{}^\top\begin{pmatrix} m\beta^2 & -m\beta^2 & -\beta & m\beta^2 \\ -m\beta^2 & m\beta^2 & \beta & -m\beta^2 \\ -\beta & \beta & 0 & -\beta \\ m\beta^2 & -m\beta^2 & -\beta & m\beta^2 \end{pmatrix}\boldsymbol{\delta}_t - \beta\partial\ell(\boldsymbol{w}_t)^\top(\boldsymbol{x}_{t-1}-\boldsymbol{x}_t),$$

with $\boldsymbol{\delta}_t{}^\top := (\boldsymbol{x}_t - \boldsymbol{x}_t^\star, \boldsymbol{x}_{t-1} - \boldsymbol{x}_{t-1}^\star, \nabla f(\boldsymbol{y}_t) + \partial\ell(\boldsymbol{w}_t), \boldsymbol{x}_t^\star - \boldsymbol{x}_{t-1}^\star)$.

(1b) Consider (13) with $(\boldsymbol{x}, \boldsymbol{y}) \equiv (\boldsymbol{x}_{t+1}, \boldsymbol{y}_t)$. We leverage $\boldsymbol{x}_{t+1} = \boldsymbol{y}_t - \alpha\nabla f_t(\boldsymbol{y}_t) - \alpha\partial\ell(\boldsymbol{w}_t)$ and the distribution law, resulting in

$$\begin{aligned} f(\boldsymbol{y}_t) - f(\boldsymbol{x}_{t+1}) &\geq \alpha\nabla f(\boldsymbol{y}_t)^\top(\nabla f(\boldsymbol{y}_t) + \partial\ell(\boldsymbol{w}_t)) \\ &\quad - \frac{L\alpha^2}{2}\|\nabla f(\boldsymbol{y}_t) + \partial\ell(\boldsymbol{w}_t)\|^2, \\ &= \frac{\alpha(2-L\alpha)}{2}\|\nabla f(\boldsymbol{y}_t) + \partial\ell(\boldsymbol{w}_t)\|^2 \\ &\quad - \alpha\partial\ell(\boldsymbol{w}_t)^\top(\nabla f(\boldsymbol{y}_t) + \partial\ell(\boldsymbol{w}_t)). \end{aligned}$$

Now, we sum the terms involving $\partial\ell(\boldsymbol{w}_t)$ in the r.h.s. of inequalities in (1a) and (1b), leverage (10), and then apply the convexity of ℓ, $\boldsymbol{x}_t \in \mathcal{X}$ and $\boldsymbol{w}_t = \boldsymbol{x}_{t+1} \in \mathcal{X}$, to obtain the following

$$\begin{aligned} &-\beta\partial\ell(\boldsymbol{w}_t)^\top(\boldsymbol{x}_{t-1}-\boldsymbol{x}_t) - \alpha\partial\ell(\boldsymbol{w}_t)^\top(\nabla f(\boldsymbol{y}_t) + \partial\ell(\boldsymbol{w}_t)) \\ &= -\partial\ell(\boldsymbol{w}_t)^\top(\boldsymbol{x}_t - \boldsymbol{w}_t) \geq \ell(\boldsymbol{w}_t) - \ell(\boldsymbol{x}_t) = 0, \end{aligned}$$

which results in $f(\boldsymbol{x}_t) - f(\boldsymbol{x}_{t+1}) \geq \boldsymbol{\xi}_t{}^\top X_{1,t}\, \boldsymbol{\xi}_t$.

Note that we have identified (f, m, L, α, β) with $(f_t, m_t, L_t, \alpha_t, \beta_t)$, and note that $\nabla f_t(\boldsymbol{x}_t^\star) + \partial\ell(\boldsymbol{x}_t^\star) = 0$.

(1c) Similarly, consider (12) with $(\boldsymbol{x}, \boldsymbol{y}) \equiv (\boldsymbol{x}_t^\star, \boldsymbol{y}_t)$. From $\boldsymbol{y}_t = \boldsymbol{x}_t + \beta\,(\boldsymbol{x}_t - \boldsymbol{x}_{t-1})$ and the distributive law,

$$\begin{aligned} &f(\boldsymbol{x}_t^\star) - f(\boldsymbol{y}_t) \\ &\geq \nabla f(\boldsymbol{y}_t)^\top(\boldsymbol{x}_t^\star - \boldsymbol{y}_t) + \frac{m}{2}\|\boldsymbol{x}_t^\star - \boldsymbol{y}_t\|^2, \\ &= (\nabla f(\boldsymbol{y}_t) + \partial\ell(\boldsymbol{w}_t))^\top(\boldsymbol{x}_t^\star - \boldsymbol{y}_t + \beta\boldsymbol{x}_t^\star - \beta\boldsymbol{x}_{t-1}^\star) \\ &+ \frac{m}{2}\|-(1+\beta)(\boldsymbol{x}_t - \boldsymbol{x}_t^\star) + \beta(\boldsymbol{x}_{t-1} - \boldsymbol{x}_{t-1}^\star)\|^2 \\ &- \beta(\nabla f(\boldsymbol{y}_t) + \partial\ell(\boldsymbol{w}_t))^\top(\boldsymbol{x}_t^\star - \boldsymbol{x}_{t-1}^\star) - \partial\ell(\boldsymbol{w}_t)^\top(\boldsymbol{x}_t^\star - \boldsymbol{y}_t) \\ &- m\beta[-(1+\beta)(\boldsymbol{x}_t - \boldsymbol{x}_t^\star) + \beta(\boldsymbol{x}_{t-1} - \boldsymbol{x}_{t-1}^\star)]^\top(\boldsymbol{x}_t^\star - \boldsymbol{x}_{t-1}^\star) \\ &+ \frac{m\beta^2}{2}\|\boldsymbol{x}_t^\star - \boldsymbol{x}_{t-1}^\star\|^2, \\ &= \frac{1}{2}\boldsymbol{\delta}_t{}^\top\begin{pmatrix} m(1+\beta)^2 & -\eta & -(1+\beta) & \eta \\ -\eta & m\beta^2 & \beta & -m\beta^2 \\ -(1+\beta) & \beta & 0 & -\beta \\ \eta & -m\beta^2 & -\beta & m\beta^2 \end{pmatrix}\boldsymbol{\delta}_t \\ &- \partial\ell(\boldsymbol{w}_t)^\top(\boldsymbol{x}_t^\star - \boldsymbol{y}_t), \end{aligned}$$

with $\eta = m(1+\beta)\beta$. We add this inequality to that in (1b) and leverage

$$\begin{aligned} &-\partial\ell(\boldsymbol{w}_t)^\top(\boldsymbol{x}_t^\star - \boldsymbol{y}_t) - \alpha\partial\ell(\boldsymbol{w}_t)^\top(\nabla f(\boldsymbol{y}_t) + \partial\ell(\boldsymbol{w}_t)) \\ &= -\partial\ell(\boldsymbol{w}_t)^\top(\boldsymbol{x}_t^\star - \boldsymbol{w}_t) \geq \ell(\boldsymbol{w}_t) - \ell(\boldsymbol{x}_t^\star) = 0, \end{aligned}$$

resulting in $f(\boldsymbol{x}_t^\star) - f(\boldsymbol{x}_{t+1}) \geq \boldsymbol{\xi}_t{}^\top X_{2,t}\, \boldsymbol{\xi}_t$.

(2) Let us define the time varying function

$$V_t(\boldsymbol{z}_t) := \begin{bmatrix} \boldsymbol{z}_t \\ \boldsymbol{x}_t^\star - \boldsymbol{x}_{t-1}^\star \end{bmatrix}^\top H_t \begin{bmatrix} \boldsymbol{z}_t \\ \boldsymbol{x}_t^\star - \boldsymbol{x}_{t-1}^\star \end{bmatrix},$$

where we take

$$H_t := \frac{1}{2\alpha_{t-1}} \begin{bmatrix} \delta_{t-1} \\ 1-\delta_{t-1} \\ \delta_{t-1} \end{bmatrix} \begin{bmatrix} \delta_{t-1}, & 1-\delta_{t-1}, & \delta_{t-1} \end{bmatrix},$$

with $\{\alpha_t\}_t$ those in the solution system (9) and $\{\delta_t\}_t$ the sequence of scalars which defines $\{\beta_t\}_t$. Now, verify

$$V_{t+1}(\boldsymbol{z}_{t+1}) - \frac{\alpha_{t-1}}{\alpha_t} V_t(\boldsymbol{z}_t) = {\boldsymbol{\xi}_t}^\top J_t \boldsymbol{\xi}_t,$$

where $\boldsymbol{\xi}_t := (\boldsymbol{z}_t,\ \boldsymbol{u}_t,\ \boldsymbol{v}_t)$, which are those define (11), resulting in $\boldsymbol{\xi}_t := (\boldsymbol{x}_t - \boldsymbol{x}_t^\star,\ \boldsymbol{x}_{t-1} - \boldsymbol{x}_{t-1}^\star, \nabla f_t(\boldsymbol{y}_t) + \partial \ell(\boldsymbol{w}_t), \boldsymbol{x}_t^\star - \boldsymbol{x}_{t-1}^\star,\ \boldsymbol{x}_{t+1}^\star - \boldsymbol{x}_t^\star)$ and

$$J_t = \frac{1}{2\alpha_t} \begin{pmatrix} 0 & 0 & -\alpha_t\delta_t\delta_{t-1} & -\delta_{t-1} & 0 \\ 0 & 0 & \alpha_t\beta_t\delta_t^2 & \beta_t\delta_t & 0 \\ -\alpha_t\delta_t\delta_{t-1} & \alpha_t\beta_t\delta_t^2 & \alpha_t^2\delta_t^2 & -\alpha_t\beta_t\delta_t^2 & 0 \\ -\delta_{t-1} & \beta_t\delta_t & -\alpha_t\beta_t\delta_t^2 & 1-2\delta_{t-1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Let us compute

$$\begin{aligned} M_t &:= \delta_{t-1}^2 X_{1,t} + \delta_t X_{2,t} \\ &= \frac{1}{2} \begin{pmatrix} m_t(\delta_t^2-1) & -m_t\beta_t\delta_t\delta_{t-1} & -\delta_t\delta_{t-1} & m_t\beta_t\delta_t\delta_{t-1} & 0 \\ -m_t\beta_t\delta_t\delta_{t-1} & m_t\beta_t^2\delta_t^2 & \beta_t\delta_t^2 & -m_t\beta_t^2\delta_t^2 & 0 \\ -\delta_t\delta_{t-1} & \beta_t\delta_t^2 & \alpha_t(2-L_t\alpha_t)\delta_t^2 & -\beta_t\delta_t^2 & 0 \\ m_t\beta_t\delta_t\delta_{t-1} & -m_t\beta_t^2\delta_t^2 & -\beta_t\delta_t^2 & m_t\beta_t^2\delta_t^2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \end{aligned}$$

and then achieve

$$\begin{aligned} {\boldsymbol{\xi}_t}^\top (J_t - M_t)\boldsymbol{\xi}_t = &\begin{bmatrix} \boldsymbol{z}_t \\ \boldsymbol{x}_t^\star - \boldsymbol{x}_{t-1}^\star \end{bmatrix}^\top N_{1,t} \begin{bmatrix} \boldsymbol{z}_t \\ \boldsymbol{x}_t^\star - \boldsymbol{x}_{t-1}^\star \end{bmatrix} \\ &+ \begin{bmatrix} \boldsymbol{z}_t \\ \boldsymbol{x}_t^\star - \boldsymbol{x}_{t-1}^\star \end{bmatrix}^\top N_{2,t} \begin{bmatrix} \boldsymbol{z}_t \\ \boldsymbol{x}_t^\star - \boldsymbol{x}_{t-1}^\star \end{bmatrix} \\ &- \alpha_t(1-L_t\alpha_t){\boldsymbol{u}_t}^\top \boldsymbol{u}_t, \end{aligned}$$

with, for each $t \geq 1$,

$$\begin{aligned} N_{1,t} &:= \frac{1}{2} \begin{pmatrix} -m_t(\delta_t^2-1) & m_t\beta_t\delta_t\delta_{t-1} & -m_t\beta_t\delta_t\delta_{t-1} \\ m_t\beta_t\delta_t\delta_{t-1} & -m_t\beta_t^2\delta_t^2 & m_t\beta_t^2\delta_t^2 \\ -m_t\beta_t\delta_t\delta_{t-1} & m_t\beta_t^2\delta_t^2 & -m_t\beta_t^2\delta_t^2 \end{pmatrix}, \\ &\cong \frac{m_t}{2} \begin{pmatrix} -(\delta_t^2-1) & \beta_t\delta_t\delta_{t-1} & 0 \\ \beta_t\delta_t\delta_{t-1} & -\beta_t^2\delta_t^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \preceq 0, \end{aligned}$$

and, using the fact that $\delta_t > (t+1)/2,\ \forall\, t \geq 0$, we have

$$N_{2,t} := \frac{1}{2} \begin{pmatrix} 0 & 0 & -\delta_{t-1} \\ 0 & 0 & \beta_t\delta_t \\ -\delta_{t-1} & \beta_t\delta_t & 1-2\delta_{t-1} \end{pmatrix} \preceq 0.$$

Then, if we select $\alpha_t \leq 1/L_t$, it results in

$${\boldsymbol{\xi}_t}^\top (J_t - M_t)\boldsymbol{\xi}_t \leq 0.$$

We rewrite it as

$$\begin{aligned} &V_{t+1}(\boldsymbol{z}_{t+1}) - \frac{\alpha_{t-1}}{\alpha_t} V_t(\boldsymbol{z}_t) \leq {\boldsymbol{\xi}_t}^\top M_t \boldsymbol{\xi}_t, \\ &\leq \delta_{t-1}^2(f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_{t+1})) + \delta_t(f_t(\boldsymbol{x}_t^\star) - f_t(\boldsymbol{x}_{t+1})), \\ &= -\delta_t^2(f_t(\boldsymbol{x}_{t+1}) - f_t(\boldsymbol{x}_t^\star)) + \delta_{t-1}^2(f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_t^\star)). \end{aligned}$$

As f_t being locally Lipschitz in t, there exists a non-negative function $h(\boldsymbol{x})$ such that

$$f_{t+1}(\boldsymbol{x}_{t+1}) - f_t(\boldsymbol{x}_{t+1}) \leq h(\boldsymbol{x}_{t+1}),$$

resulting in

$$\begin{aligned} &V_{t+1}(\boldsymbol{z}_{t+1}) - \frac{\alpha_{t-1}}{\alpha_t} V_t(\boldsymbol{z}_t) \\ &\leq -\delta_t^2(f_{t+1}(\boldsymbol{x}_{t+1}) - f_{t+1}(\boldsymbol{x}_{t+1}^\star)) + \delta_{t-1}^2(f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_t^\star)) \\ &- \delta_t^2(f_{t+1}(\boldsymbol{x}_{t+1}^\star) - f_t(\boldsymbol{x}_t^\star)) + \delta_t^2 h(\boldsymbol{x}_{t+1}),\ \forall\, t \end{aligned}$$

Summing up the above set of inequalities over the moving horizon window $t \in \mathcal{T} = \{t-1, \ldots, t-T\}$, where $T = \min\{t-1, T_0\}$ with some $T_0 \in \mathbb{Z}_{>0}$, we obtain

$$\begin{aligned} &V_t(\boldsymbol{z}_t) + \sum_{k\in\mathcal{T}} (1 - \frac{\alpha_{k-1}}{\alpha_k}) V_k(\boldsymbol{z}_k) - V_{t-T}(\boldsymbol{z}_{t-T}) \\ &\leq -\delta_{t-1}^2(f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_t^\star)) \\ &\quad + \delta_{t-T-1}^2(f_{t-T}(\boldsymbol{x}_{t-T}) - f_{t-T}(\boldsymbol{x}_{t-T}^\star)) \\ &\quad - \sum_{k\in\mathcal{T}} \delta_k^2(f_{k+1}(\boldsymbol{x}_{k+1}^\star) - f_k(\boldsymbol{x}_k^\star)) + \sum_{k\in\mathcal{T}} \delta_k^2 h(\boldsymbol{x}_{k+1}). \end{aligned}$$

Let us denote by G_t, K_t, and F_t, respectively, the horizon accumulated potential, the bound of the locally Lipschitz function h, and the variation bound of the optimal objective values. That is,

$$\begin{aligned} G_t :=& V_{t-T}(\boldsymbol{z}_{t-T}) - V_t(\boldsymbol{z}_t) - \sum_{k\in\mathcal{T}} (1 - \frac{\alpha_{k-1}}{\alpha_k}) V_k(\boldsymbol{z}_k), \\ K_t :=& \max_{k\in\mathcal{T}} \{h(\boldsymbol{x}_{k+1})\}, \\ F_t :=& \max_{k\in\mathcal{T}} \{|f_{k+1}(\boldsymbol{x}_{k+1}^\star) - f_k(\boldsymbol{x}_k^\star)|\}. \end{aligned}$$

Then, using the fact that **(1)** $\delta_{t-1} \geq (t+2)/2$, for all $t \geq 0$; **(2)** $\delta_{t-T-1} \leq t - T - 1 + \delta_0$ with $\delta_0 = (1+\sqrt{5})/2$, and **(3)** δ_t is monotonically increasing, we have

$$\begin{aligned} f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_t^\star) \leq& \frac{4G_t}{(t+2)^2} + TF_t + TK_t \\ &+ \frac{4(t-T-1+\delta_0)^2}{(t+2)^2}(f_{t-T}(\boldsymbol{x}_{t-T}) - f_{t-T}(\boldsymbol{x}_{t-T}^\star)). \end{aligned}$$

Note that, when $t \leq T_0 + 1$, we have $T = t - 1$. This gives

$$\begin{aligned} f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_t^\star) \leq& \frac{4G_t}{(t+2)^2} + (t-1)F_t + (t-1)K_t \\ &+ \frac{4\delta_0^2}{(t+2)^2}(f_1(\boldsymbol{x}_1) - f_1(\boldsymbol{x}_1^\star)). \quad \blacksquare \end{aligned}$$

C. Proofs of lemmas and theorems

Proof of Lemma IV.1: By the definition of the ambiguity set, we have that, for any distribution $\mathbb{Q} \in \mathcal{P}_{t+1}(\boldsymbol{\alpha}, \boldsymbol{u})$

$$d_{\mathrm{W}}(\mathbb{Q}, \hat{\mathbb{P}}_{t+1|t}) \leq \hat{\epsilon},$$

which, by Kantorovich-Rubinstein Theorem, is equivalent to

$$\int_{\mathcal{Z}} h(\boldsymbol{x})\mathbb{Q}(d\boldsymbol{x}) - \int_{\mathcal{Z}} h(\boldsymbol{x})\hat{\mathbb{P}}_{t+1|t}(d\boldsymbol{x}) \leq \hat{\epsilon}, \quad \forall h \in \mathcal{L},$$

where $\mathcal{L}$ is the set of functions with Lipschitz constant 1 and $\mathcal{Z}$ is the support of the random variable $\boldsymbol{x}$. For a given $\boldsymbol{u}$, let us select h to be

$$h(\boldsymbol{x}) := \frac{\ell(\boldsymbol{u}, \boldsymbol{x})}{L(\boldsymbol{u})},$$

where L is the positive Lipschitz function as in Assumption IV.1. Substituting h to the above inequality, we have

$$\int_{\mathcal{Z}} \ell(\boldsymbol{u},\boldsymbol{x})\mathbb{Q}(d\boldsymbol{x}) - \int_{\mathcal{Z}} \ell(\boldsymbol{u},\boldsymbol{x})\hat{\mathbb{P}}_{t+1|t}(d\boldsymbol{x}) \leq \hat{\epsilon}L(\boldsymbol{u}),$$

or equivalently

$$\mathbb{E}_{\mathbb{Q}}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] \leq \mathbb{E}_{\hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha},\boldsymbol{u})}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] + \hat{\epsilon}L(\boldsymbol{u}).$$

As the inequality holds for every $\mathbb{Q} \in \mathcal{P}_{t+1}$, therefore

$$\sup_{\mathbb{Q}\in\mathcal{P}_{t+1}(\boldsymbol{\alpha},\boldsymbol{u})} \mathbb{E}_{\mathbb{Q}}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] \leq \mathbb{E}_{\hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha},\boldsymbol{u})}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] + \hat{\epsilon}(t,T,\beta,\boldsymbol{\alpha},\boldsymbol{u})L(\boldsymbol{u}). \quad \blacksquare$$

Proof of Theorem IV.1: We show this by constructing a distribution in the ambiguity set. By Assumption IV.2 on convex and gradient-accessible functions, there exists an index $j \in \mathcal{T}$ such that the derivative $\nabla_{\boldsymbol{x}}\ell(\boldsymbol{u},\boldsymbol{x})$ at $(\boldsymbol{u},\bar{\boldsymbol{x}}^{(j)})$, $\bar{\boldsymbol{x}}^{(j)} := \sum_{i=1}^{p} \alpha_i \xi_j^{(i)}(\boldsymbol{\alpha},\boldsymbol{u})$, satisfies

$$\|\nabla_{\boldsymbol{x}}\ell(\boldsymbol{u},\bar{\boldsymbol{x}}^{(j)})\| = L(\boldsymbol{u}).$$

Now using this index j, we construct a parameterized distribution as follows

$$\mathbb{Q}(\Delta\boldsymbol{x}) = \frac{1}{T}\sum_{k\in\mathcal{T},k\neq j} \delta_{\{\sum_{i=1}^{p}\alpha_i\xi_k^{(i)}(\boldsymbol{\alpha},\boldsymbol{u})\}} + \frac{1}{T}\delta_{\{\bar{\boldsymbol{x}}^{(j)}+\Delta\boldsymbol{x}\}},$$

where $\Delta\boldsymbol{x} \in \mathbb{R}^n$ with $\|\Delta\boldsymbol{x}\| \leq T\hat{\epsilon}$. By the definition of the ambiguity set and, since the support of the distribution $\mathbb{P}$ is $\Xi_{t+1} = \mathbb{R}^n$, we have $\mathbb{Q}(\Delta\boldsymbol{x}) \in \mathcal{P}_{t+1}(\boldsymbol{\alpha},\boldsymbol{u})$.

Next, we quantify the lower bound of the following term

$$\mathbb{E}_{\mathbb{Q}(\Delta\boldsymbol{x})}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] - \mathbb{E}_{\hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha},\boldsymbol{u})}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] = \frac{1}{T}\left(\ell(\boldsymbol{u},\bar{\boldsymbol{x}}^{(j)}+\Delta\boldsymbol{x}) - \ell(\boldsymbol{u},\bar{\boldsymbol{x}}^{(j)})\right).$$

By Assumption IV.2 on the convexity of ℓ on $\boldsymbol{x}$, we have

$$\ell(\boldsymbol{u},\bar{\boldsymbol{x}}^{(j)}+\Delta\boldsymbol{x}) - \ell(\boldsymbol{u},\bar{\boldsymbol{x}}^{(j)}) \geq \nabla_{\boldsymbol{x}}\ell(\boldsymbol{u},\bar{\boldsymbol{x}}^{(j)})^{\top}\Delta\boldsymbol{x}.$$

Then, by selecting

$$\Delta\boldsymbol{x} := \frac{T\hat{\epsilon}\nabla_{\boldsymbol{x}}\ell(\boldsymbol{u},\bar{\boldsymbol{x}}^{(j)})}{\|\nabla_{\boldsymbol{x}}\ell(\boldsymbol{u},\bar{\boldsymbol{x}}^{(j)})\|},$$

we have

$$\nabla_{\boldsymbol{x}}\ell(\boldsymbol{u},\bar{\boldsymbol{x}}^{(j)})^{\top}\Delta\boldsymbol{x} = T\hat{\epsilon}L(\boldsymbol{u}).$$

These bounds result in

$$\mathbb{E}_{\mathbb{Q}(\Delta\boldsymbol{x})}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] - \mathbb{E}_{\hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha},\boldsymbol{u})}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] \geq \hat{\epsilon}L(\boldsymbol{u}).$$

As $\mathbb{Q}(\Delta\boldsymbol{x}) \in \mathcal{P}_{t+1}(\boldsymbol{\alpha},\boldsymbol{u})$, therefore

$$\sup_{\mathbb{Q}\in\mathcal{P}_{t+1}(\boldsymbol{\alpha},\boldsymbol{u})} \mathbb{E}_{\mathbb{Q}}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] \geq \mathbb{E}_{\hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha},\boldsymbol{u})}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] + \hat{\epsilon}L(\boldsymbol{u}).$$

Finally, with Assumption IV.1 on Lipschitz loss functions and Lemma IV.1 on an upper bound of (P1), we equivalently write Problem (P1) as

$$\inf_{\boldsymbol{u}\in\mathcal{U}} \mathbb{E}_{\hat{\mathbb{P}}_{t+1|t}(\boldsymbol{\alpha},\boldsymbol{u})}\left[\ell(\boldsymbol{u},\boldsymbol{x})\right] + \hat{\epsilon}(t,T,\beta,\boldsymbol{\alpha},\boldsymbol{u})L(\boldsymbol{u}),$$

which is the Problem (P2). $\blacksquare$

Proof of Lemma IV.2: This is the direct application of the definition of the local Lipschitz condition. $\blacksquare$

Proof of Lemma V.1: First, we have

$$F_\mu(\boldsymbol{u}) \leq F(\boldsymbol{u}) + \frac{1}{2\mu}\|\boldsymbol{u}-\boldsymbol{u}\|^2 = F(\boldsymbol{u}),\ \forall\ \boldsymbol{u}\in\mathcal{U}.$$

Then, we compute

$$\begin{aligned} F(\boldsymbol{u}) - F_\mu(\boldsymbol{u}) &= \sup_{\boldsymbol{z}\in\mathcal{U}}\left\{F(\boldsymbol{u}) - F(\boldsymbol{z}) - \frac{1}{2\mu}\|\boldsymbol{z}-\boldsymbol{u}\|^2\right\}, \\ &\leq \sup_{\boldsymbol{z}\in\mathcal{U}}\left\{\boldsymbol{g}(\boldsymbol{u})^{\top}(\boldsymbol{u}-\boldsymbol{z}) - \frac{1}{2\mu}\|\boldsymbol{z}-\boldsymbol{u}\|^2\right\}, \\ &\leq \sup_{\boldsymbol{z}}\left\{\boldsymbol{g}(\boldsymbol{u})^{\top}(\boldsymbol{u}-\boldsymbol{z}) - \frac{1}{2\mu}\|\boldsymbol{z}-\boldsymbol{u}\|^2\right\}, \\ &\leq \frac{\mu}{2}\boldsymbol{g}(\boldsymbol{u})^{\top}\boldsymbol{g}(\boldsymbol{u}) \leq \frac{D}{2}\mu, \end{aligned}$$

where the equality comes from the definition of $F_\mu(\boldsymbol{u})$, the first inequality leverages the convexity of F, the second one relaxes the constraint set, the third one applies the achieved optimizer $\boldsymbol{z}^\star = \boldsymbol{u} - \mu\boldsymbol{g}(\boldsymbol{u})$, and the last one is from the boundedness of subgradients.

Further, given F as described, it is well-known (see, e.g., [43, Proposition 12.15] for details) that F_μ is convex and continuously differentiable where its gradient ∇F_μ is Lipschitz continuous with constant $1/\mu$. In addition, the minimizer $\boldsymbol{z}^\star(\boldsymbol{u})$ of F_μ is achievable and unique, resulting in an explicit gradient expression of F_μ as follows

$$\nabla F_\mu(\boldsymbol{u}) = \frac{1}{\mu}(\boldsymbol{u} - \boldsymbol{z}^\star(\boldsymbol{u})).$$

In addition, we claim that, if F is M-strongly convex, F_μ is $M/(1+\mu M)$-strongly convex, following [44, Theorem 2.2]. Finally, we equivalently write the minimization problem as follows

$$\begin{aligned} \min_{\boldsymbol{u}\in\mathcal{U}} F_\mu(\boldsymbol{u}) &= \min_{\boldsymbol{u}\in\mathcal{U}}\min_{\boldsymbol{z}\in\mathcal{U}}\left\{F(\boldsymbol{z}) + \frac{1}{2\mu}\|\boldsymbol{z}-\boldsymbol{u}\|^2\right\} \\ &= \min_{\boldsymbol{z}\in\mathcal{U}}\min_{\boldsymbol{u}\in\mathcal{U}}\left\{F(\boldsymbol{z}) + \frac{1}{2\mu}\|\boldsymbol{z}-\boldsymbol{u}\|^2\right\} \\ &= \min_{\boldsymbol{z}\in\mathcal{U}} F(\boldsymbol{z}), \end{aligned}$$

where the first line applies the achievability of the minimizer of the problem that defines F_μ, the second switches the minimization operators, the third applies the fact that $\boldsymbol{u} = \boldsymbol{z}$ solves the inner problem. This concludes that any $\boldsymbol{u}$ that minimizes F_μ also minimizes F, and vice versa. $\blacksquare$

Proof of Theorem V.1: Let us consider the solution system (5). At each time t, let us select $\varepsilon := \varepsilon_t = 1/\operatorname{Lip}(G_\mu)$, or equivalently, μ/b with $b = \max_{k\in\mathcal{T}} b_k$. Let η_t satisfy

$$\delta_{-1} = 1,\ \delta_{t+1} := \frac{1+\sqrt{1+4\delta_t^2}}{2},\ \eta_t := \frac{\delta_{t-1}-1}{\delta_t}.$$

Then, by Theorem A.1 with $t \geq 2$, the following holds

$$G_\mu(t,\boldsymbol{u}_t) - G_\mu(t,\boldsymbol{u}_t^\star) \leq \frac{4W_t}{(t+2)^2} + TF_t, \tag{14}$$

where $\boldsymbol{u}_t^\star$ is a solution to (P2′), $T = \min\{t-1, T_0\}$ with some horizon parameter $T_0 \in \mathbb{Z}_{>0}$. Notice that T_0 is the length of the used historical data whenever such data are available. The time-varying parameter W_t depends on the initial objective discrepancy and the accumulated energy storage in the considered time horizon $\mathcal{T}$, and F_t is the variation bound of the optimal objective values in $\mathcal{T}$. Specifically, we have

$$F_t = \max_{k\in\mathcal{T}} \left\{|G_\mu(k+1, \boldsymbol{u}_{k+1}^\star) - G_\mu(k, \boldsymbol{u}_k^\star)|\right\} + \bar{L},$$

with $\bar{L}$ the variation bound of $G_\mu(t, \boldsymbol{u}_t)$ w.r.t. time t. Let us consider the storage function $V_t(\boldsymbol{z}_t) := \boldsymbol{z}_t{}^\top H_t \boldsymbol{z}_t$, where $\boldsymbol{z}_t := (\boldsymbol{u}_t - \boldsymbol{u}_t^\star,\ \boldsymbol{u}_{t-1} - \boldsymbol{u}_{t-1}^\star,\ \boldsymbol{u}_t^\star - \boldsymbol{u}_{t-1}^\star)$ and

$$H_t := \frac{1}{2\varepsilon_{t-1}} \begin{bmatrix} \delta_{t-1} \\ 1-\delta_{t-1} \\ \delta_{t-1} \end{bmatrix} \begin{bmatrix} \delta_{t-1} & 1-\delta_{t-1} & \delta_{t-1} \end{bmatrix} \succeq 0.$$

Then we have

$$\begin{aligned} W_t = V_{t-T}(\boldsymbol{z}_{t-T}) - V_t(\boldsymbol{z}_t) - \sum_{k\in\mathcal{T}} (1 - \frac{\varepsilon_{k-1}}{\varepsilon_k}) V_k(\boldsymbol{z}_k) \\ + (t - T - 1 + \delta_0)^2 (f_{t-T}(\boldsymbol{x}_{t-T}) - f_{t-T}(\boldsymbol{x}_{t-T}^\star)), \end{aligned}$$

where the first two term is the energy decrease in the horizon $\mathcal{T}$; the third sum term indicates the instantaneous energy change, which depends on the online, estimated Lipschitz constant; the last term depends on the goodness of the initial decision at the beginning of the current $\mathcal{T}$. Note how the selection of ε_t and T affect W_t (or G_t in Theorem A.1). In the most conservative scenario, we select $\varepsilon_t := \min\{\varepsilon_{t-1}, \mu/b_t\}$ and $T_0 = \infty$, which results in a constant upper bound of W_t as follows

$$W_t \le V_1(\boldsymbol{z}_1) + \delta_0^2 (f_1(\boldsymbol{x}_1) - f_1(\boldsymbol{x}_1^\star)),$$

therefore, in this case, the bound (14) essentially depends on the growing term $(t-1)F_t$. A less conservative way is to use moving horizon strategy, with $\varepsilon_t := \min\{\varepsilon_{t-1}, \mu/b_t\}$ but a finite T_0. Then, as t is sufficiently large, we have

$$W_t \le V_{t-T}(\boldsymbol{z}_{t-T}) + t^2 (f_{t-T}(\boldsymbol{x}_{t-T}) - f_{t-T}(\boldsymbol{x}_{t-T}^\star)),$$

where, in this case, the bound (14) essentially depends on F_t and $f_{t-T}(\boldsymbol{x}_{t-T}) - f_{t-T}(\boldsymbol{x}_{t-T}^\star)$.

Now, we consider for any $t \ge 2$. By Definition V.1, there exists a constant $a > 0$ such that

$$G(t, \boldsymbol{u}_t) - a\mu \le G_\mu(t, \boldsymbol{u}_t),$$

and by Lemma V.1, we have that $\boldsymbol{u}_t^\star$ is a minimizer of (P2′) if and only if it is that of (P2), and

$$G_\mu(t, \boldsymbol{u}_t^\star) \equiv G(t, \boldsymbol{u}_t^\star).$$

This results in

$$G(t, \boldsymbol{u}_t) - G(t, \boldsymbol{u}_t^\star) \le \frac{4W_t}{(t+2)^2} + TF_t + a\mu, \tag{15}$$

with an equivalent expression of F_t as

$$F_t = \max_{k\in\mathcal{T}} \left\{|G_{k+1}^\star - G_k^\star|\right\} + \bar{L},$$

where $G_k^\star := G(k, \boldsymbol{u}_k^\star)$ is the optimal objective value of (P2) or, later we see, equivalent to that of (P1).

Next, by Theorem IV.1 on the equivalence of (P1) and (P2), $\boldsymbol{u}_t^\star$ is a minimizer of (P2) if and only if it is also that of (P1), and

$$G(t, \boldsymbol{u}_t^\star) \equiv \sup_{\mathbb{Q}\in\mathcal{P}_{t+1}(\boldsymbol{\alpha}, \boldsymbol{u}_t^\star)} \mathbb{E}_{\mathbb{Q}} \left[\ell(\boldsymbol{u}_t^\star, \boldsymbol{x})\right]. \tag{16}$$

Further, as in Section IV, we claim that Problem (P1) provides a probabilistic bound for the objective of (P), resulting in

$$\text{Prob}\left(\mathbb{P}_{t+1|t} \in \mathcal{P}_{t+1}\right) \ge \rho(t), \text{ or equivalently,} \tag{17}$$

$$\text{Prob}\left(\mathbb{E}_{\mathbb{P}_{t+1|t}} \left[\ell(\boldsymbol{u}_t, \boldsymbol{x})\right] \le G(t, \boldsymbol{u}_t)\right) \ge \rho(t), \tag{18}$$

with $\rho(t)$ as in Theorem III.1. Then by (17), we know that $\mathbb{P}_{t+1|t} \in \mathcal{P}_{t+1}$ if and only if $d_{\mathrm{W}}(\mathbb{P}_{t+1|t}, \hat{\mathbb{P}}_{t+1|t}) \le \hat{\epsilon}$ where $\hat{\epsilon}$ is selected as in Theorem III.1. Further, since d_{W} is a metric, for any $\mathbb{Q} \in \mathcal{P}_{t+1}$, we claim

$$\begin{aligned} d_{\mathrm{W}}(\mathbb{Q}, \mathbb{P}_{t+1|t}) &\le d_{\mathrm{W}}(\mathbb{Q}, \hat{\mathbb{P}}_{t+1|t}) + d_{\mathrm{W}}(\mathbb{P}_{t+1|t}, \hat{\mathbb{P}}_{t+1|t}), \\ &\le \hat{\epsilon} + \hat{\epsilon} \le 2\hat{\epsilon}. \end{aligned}$$

By Assumption IV.1 and the same proof procedure of Lemma IV.1 on the above inequality, we have, for every $\boldsymbol{u}$, the following:

$$\sup_{\mathbb{Q}\in\mathcal{P}_{t+1}(\boldsymbol{\alpha}, \boldsymbol{u})} \mathbb{E}_{\mathbb{Q}} \left[\ell(\boldsymbol{u}, \boldsymbol{x})\right] \le \mathbb{E}_{\mathbb{P}_{t+1|t}} \left[\ell(\boldsymbol{u}, \boldsymbol{x})\right] + 2L(\boldsymbol{u})\hat{\epsilon}.$$

By taking $\boldsymbol{u} := \boldsymbol{u}_t^\star$ and using (16), we have

$$G(t, \boldsymbol{u}_t^\star) \le \mathbb{E}_{\mathbb{P}_{t+1|t}} \left[\ell(\boldsymbol{u}_t^\star, \boldsymbol{x})\right] + 2L(\boldsymbol{u}_t^\star)\hat{\epsilon}. \tag{19}$$

We combine the inequality (15), (18) and (19), resulting in

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}_{t+1|t}} \left[\ell(\boldsymbol{u}_t, \boldsymbol{x})\right] - \mathbb{E}_{\mathbb{P}_{t+1|t}} \left[\ell(\boldsymbol{u}_t^\star, \boldsymbol{x})\right] \\ &\qquad \le \frac{4W_t}{(t+2)^2} + TF_t + a\mu + 2L(\boldsymbol{u}_t^\star)\hat{\epsilon}, \end{aligned}$$

with the probability at least $\rho(t)$, holds for any $t \ge 2$. Furthermore, if all historical data are assimilated for the decision $\boldsymbol{u}_t$, i.e., we select $T_0 = \infty$ with $\varepsilon_t := \min\{\varepsilon_{t-1}, \mu/b_t\}$, then, the term W_t is upper bound by a constant and, the radius $\hat{\epsilon}$ asymptotically goes to zero due to the selection as in [29, Section IV]. Consequently, this results in

$$\liminf_{t\to\infty} \text{Prob}\left(R_t \le TF_t + a\mu\right) \ge 1 - \beta. \qquad \blacksquare$$

References

[1] S. Shalev-Shwartz, "Online learning and online convex optimization," Foundations and Trends in Machine Learning, vol. 4, no. 2, pp. 107–194, 2012.

[2] A. Rakhlin, K. Sridharan, and A. Tewari, "Online learning: Random averages, combinatorial parameters, and learnability," in Advances in Neural Information Processing Systems, pp. 1984–1992, 2010.

[3] 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, pp. 2032–2048, November 2020.

[4] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in Int. Conf. on Machine Learning, pp. 928–936, 2003.

[5] E. Hazan, "Introduction to online convex optimization," Foundations and Trends in Optimization, vol. 2, no. 3-4, pp. 157–325, 2016.

[6] A. Mokhtari, S. Shahrampour, A. Jadbabaie, and A. Ribeiro, "Online optimization in dynamic environments: Improved regret rates for strongly convex problems," in IEEE Int. Conf. on Decision and Control, pp. 7195–7201, 2016.

[7] A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan, "Online optimization: Competing with dynamic comparators," in Artificial Intelligence and Statistics, pp. 398–406, 2015.

[8] A. Rakhlin and K. Sridharan, "Online learning with predictable sequences," Journal of Machine Learning Research, 2013.

[9] E. Hall and R. Willett, "Online convex optimization in dynamic environments," IEEE Journal of Selected Topics in Signal Processing, vol. 9, no. 4, pp. 647–662, 2015.

[10] N. Chen, A. Agarwal, A. Wierman, S. Barman, and L. Andrew, "Online convex optimization using predictions," in ACM Int. Conf. on Measurement and Modeling of Computer Systems, pp. 191–204, 2015.

[11] Y. Li, G. Qu, and N. Li, "Online optimization with predictions and switching costs: Fast algorithms and the fundamental limit," arXiv preprint arXiv:1801.07780, 2018.

[12] L. Ljung, System identification. Prentice Hall, 1999.

[13] J. C. Willems, P. Rapisarda, I. Markovsky, and B. D. Moor, "A note on persistency of excitation," Systems and Control Letters, vol. 54, no. 4, pp. 325–329, 2005.

[14] T. Maupong and P. Rapisarda, "Data-driven control: A behavioral approach," Systems and Control Letters, vol. 101, pp. 37–43, 2017.

[15] C. D. Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," IEEE Transactions on Automatic Control, vol. 65, no. 3, pp. 909–924, 2019.

[16] J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control: In the shallows of the DeePC," in European Control Conference, pp. 307–312, 2019.

[17] J. Berberich, J. Köhler, M. Muller, and F. Allgower, "Data-driven model predictive control with stability and robustness guarantees," IEEE Transactions on Automatic Control, 2020.

[18] A. Allibhoy and J. Cortés, "Data-based receding horizon control of linear network systems," IEEE Control Systems Letters, vol. 5, no. 4, pp. 1207–1212, 2020.

[19] M. Nonhoff and M. A. Muller, "Online convex optimization for data-driven control of dynamical systems," vol. 1, pp. 180–193, 2022.

[20] S. Oymak and N. Ozay, "Non-asymptotic identification of LTI systems from a single trajectory," in American Control Conference, pp. 5655–5661, 2019.

[21] A. Tsiamis and G. J. Pappas, "Finite-sample analysis of stochastic system identification," in IEEE Int. Conf. on Decision and Control, pp. 3648–3654, 2019.

[22] S. Fattahi, N. Matni, and S. Sojoudi, "Learning sparse dynamical systems from a single sample trajectory," in IEEE Int. Conf. on Decision and Control, pp. 2682–2689, 2019.

[23] N. Fournier and A. Guillin, "On the rate of convergence in Wasserstein distance of the empirical measure," Probability Theory and Related Fields, vol. 162, no. 3-4, p. 707–738, 2015.

[24] R. Gao and A. Kleywegt, "Distributionally robust stochastic optimization with Wasserstein distance," arXiv preprint arXiv:1604.02199, 2016.

[25] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations," Mathematical Programming, vol. 171, no. 1-2, pp. 115–166, 2018.

[26] S. Shafieezadeh-Abadeh, D. Kuhn, and P. Mohajerin Esfahani, "Regularization via mass transportation," Journal of Machine Learning Research, vol. 20, no. 103, pp. 1–68, 2019.

[27] D. Boskos, J. Cortés, and S. Martínez, "Data-driven ambiguity sets with probabilistic guarantees for dynamic processes," IEEE Transactions on Automatic Control, vol. 66, no. 7, pp. 2991–3006, 2021.

[28] D. Boskos, J. Cortés, and S. Martinez, "Data-driven ambiguity sets for linear systems under disturbances and noisy observations," in American Control Conference, (Denver, CO), pp. 4491–4496, July 2020.

[29] D. Li, D. Fooladivanda, and S. Martínez, "Online learning of parameterized uncertain dynamical environments with finite-sample guarantees," IEEE Control Systems Letters, vol. 5, no. 4, pp. 1351–1356, 2021.

[30] D. Dursvyatskiy and L. Xiao, "Stochastic optimization with decision-dependent distributions," Mathematics of Operations Research, 2022.

[31] K. Wood, G. Bianchin, and E. Dall-Anese, "Online projected gradient descent for stochastic optimization with decision-dependent distributions," IEEE Control Systems Letters, vol. 6, pp. 1646–1651, 2022.

[32] L. V. Kantorovich and G. S. Rubinstein, "On a space of completely additive functions," Vestnik Leningrad. Univ, vol. 13, no. 7, p. 52–59, 1958.

[33] D. Bertsekas, Dynamic programming and optimal control: Volume I, vol. 4. Athena scientific, 2012.

[34] D. Li, D. Fooladivanda, and S. Martínez, "Online learning of parameterized uncertain dynamical environments with finite-sample guarantees," ArXiv. Preprint arXiv:2009.02390, 2020.

[35] A. Beck and M. Teboulle, "Smoothing and first order methods: A unified framework," SIAM Journal on Optimization, vol. 22, no. 2, pp. 557–580, 2012.

[36] Y. Nesterov, "Smooth minimization of non-smooth functions," Mathematical Programming, vol. 103, no. 1, pp. 127–152, 2005.

[37] Y. Nesterov, Introductory lectures on convex optimization: A basic course, vol. 87. Springer Science & Business Media, 2013.

[38] S. M. LaValle, Planning algorithms. Cambridge University Press, 2006.

[39] B. Hu and L. Lessard, "Dissipativity theory for Nesterov's accelerated method," arXiv preprint arXiv:1706.04381, 2017.

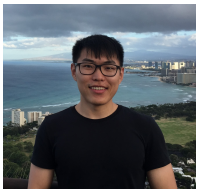
[40] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," SIAM Journal on Optimization, vol. 26, no. 1, pp. 57–95, 2016.

[41] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM Journal on Imaging Sciences, vol. 2, no. 1, pp. 183–202, 2009.

[42] R. T. Rockafellar and R. J.-B. Wets, Variational analysis. Springer, 1998.

[43] H. H. Bauschke and P. L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces, vol. 408. Springer, 2011.

[44] C. Lemaréchal and C. Sagastizábal, "Practical aspects of the Moreau-Yosida regularization: Theoretical preliminaries," SIAM Journal on Optimization, vol. 7, no. 2, pp. 367–385, 1997.



Dan Li receieved the B.E. degree in automation from the Zhejiang University, Hangzhou, China, in 2013, the M.Sc. degree in chemical engineering from Queen's University, Kingston, Canada, in 2016. He is currently a Ph.D. student at University of California, San Diego, CA, USA. His current research interests include data-driven systems and optimization, dynamical systems and control, optimization algorithms, applied computational methods, and stochastic systems. He received Outstanding Student Award from Zhejiang University in 2012, Graduate Student Award from Queen's University in 2014, and Fellowship Award from University of California, San Diego, in 2016.

Dariush Fooladivanda received the Ph.D. degree from the University of Waterloo, in 2014, and a B.S. degree from the Isfahan University of Technology, all in electrical engineering. He is currently a Postdoctoral Research Associate in the Department of Electrical Engineering and Computer Sciences at the University of California Berkeley. His research interests include theory and applications of control and optimization in large scale dynamical systems.

Sonia Martínez is a Professor at the Department of Mechanical and Aerospace Engineering at the University of California, San Diego. She received her Ph.D. degree in Engineering Mathematics from the Universidad Carlos III de Madrid, Spain, in May 2002. Following a year as a Visiting Assistant Professor of Applied Mathematics at the Technical University of Catalonia, Spain, she obtained a Postdoctoral Fulbright Fellowship and held appointments at the Coordinated Science Laboratory of the University of Illinois, Urbana-Champaign during 2004, and at the Center for Control, Dynamical systems and Computation (CCDC) of the University of California, Santa Barbara during 2005.

Her research interests include networked control systems, multi-agent systems, and nonlinear control theory with applications to robotics and cyber-physical systems. For her work on the control of underactuated mechanical systems she received the Best Student Paper award at the 2002 IEEE Conference on Decision and Control. She co-authored with Jorge Cortés and Francesco Bullo "Motion coordination with Distributed Information" for which they received the 2008 Control Systems Magazine Outstanding Paper Award. She is a Senior Editor of the IEEE Transactions on Control of Networked Systems and an IEEE Fellow.