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

Dan Li, Dariush Fooladivanda, and Sonia Martínez, Fellow, IEEE

Abstract—This article 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 dataset, 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 the dissipativity theory.

Index Terms—Adaptive learning, data-driven modeling, optimization under distributional uncertainties.

I. INTRODUCTION

ONLINE optimization has attracted significant attention from various fields, including machine learning, information theory, robotics, and smart power systems; see [1], [2], [3]. 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 nonstationary stochastic processes [4], [5]. Regret minimization aims to deal with nonstationarity 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], [9], [10], [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 datasets, 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], [15], [16], [17], [18]. The aforementioned works rely on the use of Hankel system representations of the linear time-invariant (LTI) system, and may be subject or not to additional uncertainty. In particular, the authors in [19] leveraged the behavioral theory to obtain sublinear 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, Oymak and Ozay [20], who exploited least squares and the Ho–Kalman algorithm; Tsiamis and Pappas [21], who used subspace identification techniques for LTI systems subject to unknown Gaussian disturbances; and Fattahi et al. [22], who resorted 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 nonrobust 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], [25], [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 authors in [27] and [28] considered 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. Li et al. [29] proposed 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 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, the authors in [30] and [31] solved for alternative stable solutions, or optimal solutions w.r.t. to the distribution they induce. In addition to this, and while the authors in [30] and [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 nonsmooth convex problem optimizations. For each of these categories, and smoothed-out versions of these problems, we propose an online control algorithm dynamics, which extends the Nesterov’s accelerated-gradient method. Adapting the 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 $\mathbb{R}^m, \mathbb{R}_+^m, \mathbb{Z}_+^m, \mathbb{R}^{m \times n}$ 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 $x \in \mathbb{R}^m$ is $x^\top$, and $1_m$ is a shorthand for $(1, \ldots, 1)^\top \in \mathbb{R}^m$. We index vectors with subscripts, i.e., $x_k \in \mathbb{R}^m$ with $k \in \mathbb{Z}_+$, and given $x \in \mathbb{R}^m$ we denote its $i$th component by $x_i$. We denote by $\|x\|_2$ and $\|x\|_\infty$ the 2-norm and $\infty$-norm, respectively. The inner product of $\mathbb{R}^m$ is given as $\langle x, y \rangle := x^\top y, x, y \in \mathbb{R}^m$; thus, $\|x\| := \sqrt{\langle x, x \rangle}$. The gradient of a real-valued function $\ell : \mathbb{R}^m \to \mathbb{R}$ is denoted as $\nabla \ell(x)$ and $\nabla_z \ell(x)$ is the partial derivative w.r.t. $x$. In what follows, $\text{dom}(\ell) := \{x \in \mathbb{R}^m : -\infty < \ell(x) < +\infty\}$. A function $\ell : \text{dom}(\ell) \to \mathbb{R}$ is $M$-strongly convex, if for any $y, z \in \text{dom}(\ell)$ there exists $g \in \mathbb{R}^m$ such that $\ell(y) \geq \ell(z) + g^\top(y-z) + M\|y-z\|^2_2/2$, for some $M > 0$. The function $\ell$ is convex if $M = 0$. We call a vector $g$ a subgradient of $\ell$ at $x$ and denote by $\partial \ell(z)$ the subgradient set. If $\ell$ is differentiable at $x$, then $\partial \ell(x) = \{\nabla \ell(x)\}$. Finally, the operation $\Pi_U(x) : X \to U$ projects the set $X \subseteq \mathbb{R}^m$ onto $U \subseteq \mathbb{R}^n$ under the Euclidean norm. We write $\Pi_U(x) := \text{argmin}\|x-z\|^2_2 + \chi_U(x)$, where $x \in X$, and $\chi_U(x) = 0$ if $z \in U$, otherwise $+\infty$. Endow $\mathbb{R}^n$ with the Borel $\sigma$-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 $\mathcal{M} = \{Q \in \mathcal{P}(\mathbb{R}^n) : \int_{\mathbb{R}^n} \|x\|dQ < +\infty\}$. 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 $d_W(Q_1, Q_2) := \sup_{f \in L} \int f(x)dQ_1 - \int f(x)dQ_2$, where $L$ is the space of all Lipschitz functions with Lipschitz constant 1. We denote a closed Wasserstein ball of radius $\epsilon$ (also called an ambiguity set) centered at a distribution $\mathcal{P} \in \mathcal{M}$ by $B_{\epsilon}(\mathcal{P}) := \{Q \in \mathcal{M} : d_W(\mathcal{P}, Q) \leq \epsilon\}$. The Dirac measure at $x_0 \in \mathbb{R}^n$ is a distribution in $\mathcal{P}(\mathbb{R}^n)$ denoted by $\delta_{x_0}$. Given $A \in \mathcal{B}$, we have $\delta_{x_0}(A) = 1$, if $x_0 \in A$, otherwise 0. A random vector $x \in \mathbb{R}^n$ with probability distribution $Q$ is sub-Gaussian if there are positive constants $C, \nu$ such that $Q(\|x\| > t) \leq Ce^{-\nu t^2}$. Equivalently, a zero-mean random vector $x \in \mathbb{R}^n$ is sub-Gaussian if for any $a \in \mathbb{R}^n$ we have $\mathbb{E}[\exp(a^\top x)] \leq \exp(\|a\|^2\nu^2/2)$ for some $\nu$.

**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

$$x_{t+1} = f(t, x_t, u_t) + w_t,$$

from a given $x_0 \in \mathbb{R}^n$ (1)

where $w_t \in \mathcal{U} \subseteq \mathbb{R}^m$ 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 $w_t \in \mathcal{R}^n$, is an unknown, random, disturbance vector. Due to the Markov assumption, $x_t \in R^n$ can be described by an unknown transition probability measure $P_{t|t-1} \in \mathcal{P}(\mathcal{R}^n)$, conditioned on the system state and control at time $t-1$. Denote by $\ell : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, $(u, x) \to$...
\( \ell(u, x) \) an a priori selected, measurable loss function. Assume that \( \ell \) is compact, and we are interested in selecting \( u_t \in \ell \) that minimizes the loss

\[
\min_{u_t \in \ell} \mathbb{E}_{P_{t+1}|t}} [\ell(u_t, x)] := \int \mathbb{E}_{P_{t+1}|t}} [\ell(u_t, x) \mid P_{t+1}|t}(dx)
\]

This objective value is inaccessible since the state distribution \( P_{t+1}|t} \) 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 \( u := u_t \) that minimizes the loss in the immediate future at \( t + 1 \)

\[
\min_{u \in \ell} \mathbb{E}_{P_{t+1}|t}} [\ell(u, x)]
\]

s.t. \( x \sim P_{t+1}|t} \), evolves according to (1). ~\( \text{(P)} \)

This problem formulation is similar to one-stage optimization problems with unknown system transitions [33]. The expectation operator with respect to \( P_{t+1}|t} \) is conditional on the historical realizations \( \tilde{x}_k \), \( k \leq t \), the adopted decisions \( \tilde{u}_k \), \( k \leq t - 1 \), the yet-to-be-learned unknown dynamical system \( f \), and realizations \( \tilde{w}_k \), \( k \leq t - 1 \). We will identify \( P_{t+1}|t}(dx) \equiv P_{t+1}|t}(dx|u_t, x_t = \tilde{x}_t, x_k = \tilde{x}_k, u_k = \tilde{u}_k, k \leq t - 1) \) which, by the Markovian property, satisfies

\[
\mathbb{P}_{t+1}|t}(dx) = \mathbb{P}_{t+1}|t}(dx|u_t, x_t = \tilde{x}_t).
\]

At time \( t \), let \( u^* := u_t^* \) denote an optimizer of problem \( \text{(P)} \) and consider the instantaneous regret

\[
R_t := \mathbb{E}_{P_{t+1}|t}} [\ell(u, x)] - \mathbb{E}_{P_{t+1}|t}} [\ell(u^*, x)]
\]

which is the loss incurred if the selected \( u \) 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 \( \rho \), the regret \( R_t \) is upper bounded by a sum of terms, a first one depending on the initial condition \( x_0 \); a second one depending on the instantaneous variation of the loss of \( \text{(P)} \); and a third term related to how well the unknown system \( f \) and the uncertainities 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 \( P_{t+1}|t} \) 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 \( w_t \in \mathbb{R}^n \), \( t \in \mathbb{Z}_{\geq 0} \), are i.i.d. with \( w_t \sim Q \) and zero-mean \( \sigma \) 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 2-D 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 \( \text{(P)} \), 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.

### 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 \( P_{t+1} \) proposed in [29]. The set \( P_{t+1} \) contains a class of distributions, which is, in high probability, large enough to include the unknown \( P_{t+1}|t} \) 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 \( \text{(P)} \) later. We summarize next the construction of these ambiguity sets \( P_{t+1} \). First, we assume the following on the unknown \( f \).

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

\[
f(t, x, u) = \sum_{i=1}^{P} \alpha_i f^{(i)}(t, x, u)
\]

1That is, for all unit vector \( v \), we have \( \mathbb{E}[e^{v^\top w_t}] \leq e^{v^\top \sigma^2}/2 \), \( \forall \lambda \in \mathbb{R} \).

Equivalently, \( Q(\|w_t\| > \lambda) \leq e^{-\lambda^2/(4\sigma^2)} \), \( \forall \lambda \).
where $\alpha^* := (\alpha_1^*, \ldots, \alpha_p^*)^T \in \mathbb{R}^p$ is an unknown parameter, and $f^{(i)} : \mathbb{R}_{>0} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $(t, x, u) \mapsto f^{(i)}(t, x, u)$, $i \in \{1, \ldots, p\}$ is a set of $p$ linearly independent known basis functions or predictors chosen a priori.

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

$$
\mathcal{P}_{t+1} := \mathbb{B}_\varepsilon(\tilde{\mathbb{P}}_{t+1}|_{\alpha, \sigma}) := \{ Q \mid d(W(Q, \tilde{\mathbb{P}}_{t+1}|_{\alpha, \sigma}) \leq \varepsilon \}.
$$

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

$$
\tilde{\mathbb{P}}_{t+1}|_{\alpha, \sigma} := \frac{1}{T} \sum_{k \in T} \delta_{\hat{\alpha} \xi^{(k)}_{\alpha, \sigma}}(\alpha, u)
$$

see the footnote, where $T = \{ t, T, T \}$, for $t \geq T + 1$. If $\alpha = \alpha^*$, then $\sum_{k \in T} \alpha \xi^{(k)}_{\alpha, \sigma} \in \mathbb{R}^n$ provides an outcome $x^{(k)}_{t+1} := f(t, x_t, u) + w_k = \sum_{k \in T} \alpha \xi^{(k)}_{\alpha, \sigma} + f^{(i)}(t, x_t, u) + w_k$, for each $k$. For a general $\alpha \approx \alpha^*$, the value $\sum_{k \in T} \alpha \xi^{(k)}_{\alpha, \sigma} \in \mathbb{R}^n$ provides “approximated” outcomes $x^{(k)}_{t+1}$, for each $k = 1, \ldots, T$. Then, we claim the probabilistic guarantee of $\mathcal{P}_{t+1}$ by a selection of the parameter $\alpha$ and $\varepsilon$ for any $u$.

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

$$
\text{Prob} (\mathbb{P}_{t+1}|_{\alpha, \sigma} \in \mathcal{P}_{t+1}) \geq \rho(t).
$$

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

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

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

$$
\varepsilon := \sqrt{\frac{2nM \beta^2 T}{T}} \ln \left( \frac{1}{\beta} \right) + C_1 T^{-1/\max\{n, 2\}} + \gamma H(t, T, u)
$$

where $M$ and $C_1$ are positive constants, and

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

which bounds the variation of predicted system trajectories.

**Idea of the Proof:** The probabilistic guarantees (3) are a consequence of [29, Lemma 1, Theorem 1, Theorem 2, and (7)] with Assumptions III.1 and III.2. Precisely, we achieve this by bounding the metric $d_W(\mathbb{P}_{t+1}|_{\alpha, \sigma}, \mathbb{P}_{t+1}|_{\alpha^*, \sigma})$ using $d_W(\mathbb{P}_{t+1}|_{\alpha, \sigma}, \mathbb{P}_{t+1}|_{\alpha^*, \sigma})$ plus $d_W(\mathbb{P}_{t+1}|_{\alpha^*, \sigma}, \mathbb{P}_{t+1}|_{\alpha^*, \sigma})$. Then, the first distance is handled via [29, Lemma 1] using standard measure of concentration results, contributing to the first two terms of the radius $\varepsilon$ in (4). Next, the second distance $d_W(\mathbb{P}_{t+1}|_{\alpha^*, \sigma}, \mathbb{P}_{t+1}|_{\alpha^*, \sigma})$ can be bounded in terms of the difference $\| \alpha - \alpha^* \|$ via [29, Th. 1], contributing to the third term in $\varepsilon$. Notice that the third term depends on Assumption III.2 and the selected parameter $\gamma$ which relies on the selection of $\alpha$ via [29, Th. 2 (Learning of $\alpha^*$)]. The confidence value $\rho(t)$ is achieved by Assumption III.1 applying to the same procedure, as in [29, Th. 2], which essentially bounds $\| \alpha - \alpha^* \|$ in probability. Precisely, by Assumption III.1, we have $\mathbb{Q}(\| \mathbb{w} \|_\infty > \eta) \leq e^{-\eta^2/(4\sigma^2)} \forall \eta$, resulting in $\mathbb{E}(\| w \|_\infty) \leq 2 \exp(-\eta^2/4\sigma^2)$ 4

In [29, Th. 2], the value $d$ plays the role of $u$ in this work.

2Lemma 1 in [29] makes use of a stronger Assumption III.1, which requires $w_k$ 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.
IV. TRACTABLE PROBLEM REFORMULATION AND ITS
SPECIALIZATION TO TWO PROBLEM CLASSES

In this section, we start by describing how to deal with the
unknown \( P_{t+1} \) in Problem (P), via ambiguity sets, which 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 subclasses, which
allows us to present and analyze the online algorithms for these
problems in the following section. Formally, let us consider

\[
\min_{u \in \mathcal{U}} \sup_{Q \in \mathcal{P}_{t+1}(\alpha, u)} E_Q [\ell(u, x)]
\]  
(P1)

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

\[
\text{Prob} \left( E_{P_{t+1}} [\ell(u, x)] \leq \sup_{Q \in \mathcal{P}_{t+1}} E_Q [\ell(u, x)] \right) \geq \rho(t).
\]

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 dataset sizes \( T \) in Theorem III.1.

The solution \( u \) and the objective value of (P1) ensure that,
when we select \( u \) 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 seminfinite 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 \( \ell : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R} \), \( (u, x) \to \ell(u, x) \). There exists a
Lipschitz function \( L : \mathbb{R}^m \to \mathbb{R} \geq 0 \) such that for each \( u \in \mathbb{R}^m \), it holds that \( \| \ell(u, x) - \ell(u, y) \| \leq L(u) \| x - y \| \) for any \( x, y \in \mathbb{R}^n \).

With this, we obtain the following upper bound.

**Lemma IV.1** (An upper bound of (P1)): Let Assumption IV.1 hold.
Then, for each \( u, \alpha, \beta, T, \) and \( t \), we have

\[
\sup_{Q \in \mathcal{P}_{t+1}(\alpha, u)} E_Q [\ell(u, x)] 
\leq E_{\mathcal{P}_{t+1}} [\ell(u, x)] + \tilde{\ell}(t, T, \beta, \alpha, u) L(u)
\]

where the empirical distribution \( \mathcal{P}_{t+1}(\alpha, u) \), \( \ell(u, x) \) and \( L(u) \) 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 \( \ell \) is convex in \( x \) for each \( u \). Further, for each time \( t \) with given \( u = u_i \in \mathcal{U} \), \( \alpha = \alpha_i \in \mathcal{P}_t \), there is a system prediction \( \sum_{k=1}^{p} \alpha_k \hat{L}_k(u, x, u) \) for some \( k \in T \) such that \( \nabla_x \ell \) exists and \( L(u) \) is equal to \( \| \nabla_x \ell \| \) at \( u = \sum_{k=1}^{p} \alpha_k \hat{L}_k(u, x, u) \).

The above statement enables the following theorem.

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

\[
\min_{u \in \mathcal{U}} \mathcal{P}_{t+1}(\alpha, u) [\ell(u, x)] + \tilde{\ell}(t, T, \beta, \alpha, u) L(u) \tag{P2}
\]

Remark IV.1 (Effects of Assumptions IV.1 and IV.2): We note
that Assumption IV.1 on the Lipschitz requirement of loss func-
tion is mild. In fact, many engineering problems take state values
in a compact set, which then only requires the loss \( \ell \) to be con-
tinuous. Assumption IV.2 essentially requires accessible partial
gradients (in \( x \)) of loss functions \( \ell \). For simple loss functions
\( \ell \), e.g., linear, quadratic, etc, its partial gradient can be readily
valuated. 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 \( \ell \) 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 \( (t, x) \) and
affine in \( u \), i.e.,

\[
f(t, x, u) := f_1(t, x) + f_2(t, x) u
\]

for some unknown \( f_1 : \mathbb{R}_0^m \times \mathbb{R}^n \to \mathbb{R}^n \), \( f_2 : \mathbb{R}_0^m \times \mathbb{R}^n \to \mathbb{R}^n \), \( u \in \mathcal{U} \) and \( t \in \mathbb{Z}_0^m \). Similarly, for each \( i \in \{1, ..., p\} \),
the basis function \( f_i(t, x, u) := f_i(t, x) + f_i(t, x) u \)
for some known locally Lipschitz functions \( f_1 \) and \( f_2 \).

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

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

\[
\min_{u \in \mathcal{U}} G(t, u) := G(t, u|\ell, T, \beta, \alpha, \mathcal{P}_{t+1} (t, \ell))
\]

where \( G \) represents the objective function of (P2), depending
on variables \( \ell, T, \beta, \alpha \), 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. \( u = u_i \), and Assumption IV.4 characterizes
the convex feasible set of (P2). Note that \( G(t, u) \) is locally Lipschitz
in \( t \).

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 characteri-
sation of the system and common loss functions \( \ell \). 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:

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

with \( \ell_1 \) the cost for the immediate control and \( \ell_2 \) the optimal
cost-to-go function. We assume that both \( \ell_1 \) and \( \ell_2 \) are convex,
and in addition, \( \ell_2 \) is Lipschitz continuous with a constant
\( \text{Lip}(\ell_2) \in \mathbb{R}_0^m \), resulting in \( L(u) \equiv \text{Lip}(\ell_2) \). Then, by selecting

5This can be verified by the local Lipschitz condition on \( f_i(t, x, u) \), and finite
composition of local Lipschitz functions are locally Lipschitz.
the ambiguity radius $\ell$ and center $\bar{P}_{t+1}$ of $P_{t+1}$, as in Section II-I-A, the objective function of (P2) becomes
\begin{align*}
G(t, u) = \ell_1(u) + \frac{1}{T} \sum_{k \in T} \ell_2(p_{k,t}) + \text{Lip}(\ell_2)\epsilon + \frac{\gamma \text{Lip}(\ell_2)}{T} \sum_{i=1}^{P} \| H^{(i)}_k \|
\end{align*}
where $p_{k,t}, H^{(i)}_k \in \mathbb{R}^n$ are affine in $u$, for each $i$, $k$, as
\begin{align*}
p_{k,t} := \sum_{i=1}^{P} \alpha_i \left( f^{(i)}_1(t, \hat{x}_t) - f^{(i)}(k, \hat{x}_k, u_k) \right) + \hat{x}_{k+1} + \left( \sum_{i=1}^{P} \alpha_i f^{(i)}_2(t, \hat{x}_t) \right) u
\end{align*}
$H^{(i)}_k(u) := f^{(i)}(k, \hat{x}_k, u_k) - f^{(i)}_1(t, \hat{x}_t) - f^{(i)}_2(t, \hat{x}_t) u$ and parameters $\alpha \in \mathbb{R}^P$, $\epsilon \in \mathbb{R}_{\geq 0}$, and $\gamma \in \mathbb{R}_{\geq 0}$ are selected, as in [36, Sec. IV]. Intuitively, $p_{k,t}$ is the $k$th projected outcome of the random variable $x_{t+1}$ and $H^{(i)}_k$ quantifies the variation of predictor $f^{(i)}$ with respect to its previous $k$th value. Notice that the objective function $G$ is convex in $u$ and therefore online problems (P2) are tractable. In addition, if $\ell_2$ 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
\begin{align*}
\ell(u, x) = \max \{ 0, 1 - \langle u, \phi(x) \rangle \}, \quad \phi : \mathbb{R}^n \rightarrow \mathbb{R}^m
\end{align*}
where $\phi$ is an affine feature map selected in advance. The decision maker updates the decision $u$ online when $\langle u, \phi(x) \rangle < 1$, 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
\begin{align*}
G(t, u) = \frac{1}{T} \sum_{k \in T} \max \{ 0, 1 - \langle u, \phi(p_{k,t}) \rangle \} + q_t L(u)
\end{align*}
where time-dependent parameters $p_{k,t} \in \mathbb{R}^n$, $q_t \in \mathbb{R}$ are
\begin{align*}
p_{k,t} = \hat{x}_{k+1} + \sum_{i=1}^{P} \alpha_i \left( f^{(i)}_1(t, \hat{x}_t) - f^{(i)}(k, \hat{x}_k) \right) \quad \forall \ k, \ t
\end{align*}
\begin{align*}
q_t = \epsilon + \frac{\gamma}{T} \sum_{i=1}^{P} \| f^{(i)}_2(k, \hat{x}_k) - f^{(i)}_2(t, \hat{x}_t) \| \quad \forall \ t
\end{align*}
with $\alpha \in \mathbb{R}^P$, $\epsilon \in \mathbb{R}_{\geq 0}$, and $\gamma \in \mathbb{R}_{\geq 0}$ as in [36, Sec. IV]. We characterize the function $L(u)$ by subgradients of the loss function $\ell$.

Lemma IV.2 (Quantification of $L$): Consider $\ell(u, x) := \max \{ 0, 1 - \langle u, \phi(x) \rangle \}$, where $\phi(x)$ is differentiable in $x$. Then, the function $L(u)$ is
\begin{align*}
L(u) = \sup_{g \in \partial_\ell \ell(u, x)} \| g \|
\end{align*}
where the set $\partial_\ell \ell(u, x)$ contains all the subgradients of $\ell$ at $x$, given any $u$ in advance, i.e.,
\begin{align*}
\partial_\ell \ell(u, x) := h(x, u) \cdot \frac{\partial \phi(x)}{\partial x} u
\end{align*}
where
\begin{align*}
h(x, u) = \begin{cases} 
-1, & \text{if } \langle u, \phi(x) \rangle < 1 \\
0, & \text{if } \langle u, \phi(x) \rangle \geq 1 \\
[-1, 0], & \text{otherwise}.
\end{cases}
\end{align*}

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

Lemma IV.2 indicates that, given a properly selected feature mapping $\phi$, the objective $G$ is convex in $u$ and therefore online problems (P2) are convex and tractable. In addition, if $\phi$ 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 nonsmooth due to the normed regularization terms in $G$. To achieve fast, online solutions, we propose a two-step procedure. First, we follow [35] and [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 $F : U \rightarrow \mathbb{R}$ for a convex and potentially nonsmooth function, which can represent any particular component of the objective function $G(t, u)$ of (P2) at time $t$.

Definition VI.1 (Smoothable function [35]): We call a convex function $F(u)$ smoothable on $U$ if there exists $\mu > 0$ such that, for every $\mu > 0$, there is a continuously differentiable convex function $F'_\mu : U \rightarrow \mathbb{R}$ satisfying
\begin{align*}
1) \quad F'_\mu(u) \leq F(u) \leq F'_\mu(u) + \mu u, \forall u \in U.
2) \quad F(u) \geq 0, \forall u \in U.
\end{align*}

To obtain a smooth approximation $F'_\mu$ of $F$, we follow the Moreau proximal approximation technique [35], described as in the following lemma.

Lemma VI.1 (Moreau–Yosida approximation): Given a convex function $F : U \rightarrow \mathbb{R}$ and any $\mu > 0$, let us denote by $\partial F(u)$ the set of subgradients of $F$ at $u$, respectively. Let $D := \sup_{u \in U} ||\partial F(u)||$. Then, $F$ is smoothable with parameters $(a, b) := (D^2/2, 1)$, where the smoothed version $F'_\mu : U \rightarrow \mathbb{R}$ is the Moreau approximation
\begin{align*}
F'_\mu(u) := \inf_{v \in U} \left\{ F(v) + \frac{1}{2\mu} ||v - u||^2 \right\}, \quad u \in U.
\end{align*}

In addition, if $F$ is $M$-strongly convex with some $M > 0$, then $F'_\mu$ is $M/(1 + \mu M)$-strongly convex. And further, the minimization of $F'(u)$ over $u \in U$ is equivalent to that of $F'_\mu(u)$ over $u \in U$ 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, and 2) the composition of a smoothable function with a linear transformation is smoothable. These properties enable us to smooth each component of $G$, i.e., $\ell_1, \ell_2, h$, and $\| \cdot \|$, which results in a smooth approximation of (P2) via the corresponding $G_\mu$ as follows:

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

Note that $G_\mu$ 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:

$$\ell(u, x) := \frac{1}{2} \| u \|^2 + F_\mu(x),$$

where $F_\mu : \mathbb{R}^n \to \mathbb{R}$ is a smoothable $\ell_2$-norm function, with $\text{Lip}(F_\mu) = 1$. Then, the objective function $G_\mu(t, u)$ is

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

where $p, H$ are affine in $u$, defined as in Section IV. In addition, we have the smoothing parameter of $G_\mu(t, u), (a, b) := ((1 + \mu T)/2, \mu + 8 \sum_i s_i)$, where

$$s_0 = \sigma_{\max} \left( \sum_{i=1}^p \alpha_i f^{(2)} \right),$$

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

$$s_i = \sigma_{\max} \left( f^{(2)} \right), \quad i \in \{1, \ldots, p\}.$$

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

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

with $L(u) = \| u \|$ resulting in

$$G_\mu(t, u) = \frac{1}{T} \sum_{k \in T} F_\mu(\langle u, p_{k,t} \rangle) + \epsilon F_\mu(u)$$

where $\mu > 0$, parameters $p, q$ are as in Section IV, and functions $F_\mu : \mathbb{R} \to \mathbb{R}$ and $F_\mu : \mathbb{R}^n \to \mathbb{R}$ are the smooth switch function and $\ell_2$-norm function, respectively. Note that $G_\mu$ has the smoothing parameter $(a, b) := (\| x \|^2/2, q + 1/T \sum_{k \in T} \| p_{k,t} \|^2)$.

**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 $u_t, t \in \mathbb{Z}_{\geq 0}$, be solutions of (P2') and let us consider the solution system with some $u_0 \in U$ and $y_0 = y_0$, as

$$u_{t+1} = \Pi_U(y_t - \epsilon \nabla G_\mu(t, u_t))$$

$$y_{t+1} = u_{t+1} + \eta \ell(u_{t+1} - u_t)$$ (5)

where $\epsilon \leq \mu/b$ with positive parameters $\mu$ and $b := b$ being those define $G_\mu(t, u)$. We denote by $\nabla G_\mu$ the derivative of $G_\mu$ w.r.t. its second argument and denote by $\Pi_\mu(y)$ the projection of $y$ onto $U$ as in Assumption IV.4 on convex decision oracle. Note that, the gradient function $\nabla G_\mu$ 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 $\eta_\ell \in \mathbb{R}_{\geq 0}$ 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).

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

$$R_t := \mathbb{E}_{p_{t+1}}[\ell(u_t, x)] - \mathbb{E}_{p_{t+1}}[\ell(u^*_t, x)].$$

Then, the regret $R_t$ is bounded in probability as follows:

$$\text{Prob} \left( R_t \leq \frac{4W_t}{(t + 1)^2} + TF_t + a\mu + 2L(u^*_t) \epsilon \right) \geq \rho(t)$$

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 $T$.

**The Switch function:** Consider $u \in \mathbb{R}, F^5 : u \to \max\{0, 1 - u\}$, which is differentiable almost everywhere. For a given $\mu > 0$, we compute

$$F^5_\mu(u) := \min_{z \in \mathbb{R}} \left\{ \max\{0, 1 - z\} + \frac{1}{2\mu} \| z - u \|^2 \right\}$$

$$= \min_{z \in \mathbb{R}} \left\{ \max\{1 - u, 0\} + \frac{1}{2\mu} \| z - u \|^2 \right\}$$

$$= \left\{ \begin{array}{ll} 1 - \frac{1}{\mu} u, & \text{if } u > 1 - \mu \\ 1 - u - \frac{1}{\mu} u, & \text{if } u \leq 1 - \mu \end{array} \right.$$ 

resulting in

$$F^5_\mu(u) := \left\{ \begin{array}{ll} 1 - \frac{1}{\mu} u, & \text{if } u > 1 - \mu \\ \frac{1}{2\mu} \| 1 - u \|^2, & \text{if } u \leq 1 - \mu \end{array} \right.$$ 

with the smoothing parameter $(1/2, 1)$.
i.e.,
\[
F_t = \max_{k \in T} \{ |G^*_k - G^*_t| \} + \bar{L}
\]
where \( G^*_k := G(\mu, u^*_t) \) 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 \( u_t \), then, we have
\[
\liminf_{t \to \infty} \Pr(\{ R_t \leq TF_t + a\mu \} \geq 1 - \beta
\]
with \( \beta \) a given, arbitrary confidence value.

Theorem VI.1 quantifies the dynamic regret of online decisions \( u \) w.r.t. solutions to (P1) in high probability. Notice that, the regret bound is dominated by terms: \( T F_t, a\mu \), and \( L(u_t^*)\epsilon \), which mainly depend on three factors: the data-driven parameters \( \epsilon, \eta \), and \( \mu \) of the solution system (5), the variation \( F_t \) over optimal objective values, and the parameters \( T, \beta, \gamma \), and \( \ell \) 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 \( \epsilon \); 2) a proper selection of the loss function \( \ell \) 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 objective 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.

**Algorithm 1: Opal(\mathcal{I})**

1. Select \( \{ f^{(i)} \}_i, \ell, \beta, \mathcal{U}, u_0, \mu, \), and \( t = 1; \)
2. repeat
3. Update dataset \( \mathcal{I} := \mathcal{I}_t; \)
4. Compute \( \alpha := \alpha_t \) as in [29];
5. Select \( \bar{\mathcal{P}}_{t+1} \in (2) \) and \( \bar{c} := (\ell(T, \beta, \alpha, \mathcal{U}) \) in (4);
6. Run dynamical system (5) for \( u := u_t; \)
7. Apply \( u \) 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]
\[
\begin{align*}
x_{t+1} &= x_t + h \cos(\theta_t) d_{1,t} + h w_{1,t} \\
y_{t+1} &= y_t + h \sin(\theta_t) d_{1,t} + h w_{2,t} \\
\theta_{t+1} &= \theta_t - h d_{2,t} + h w_{3,t}
\end{align*}
\]
where components of states \( x_t := (x_t, y_t, \theta_t) \in \mathbb{R}^2 \times \{-\pi, \pi\} \) represent vehicle position and orientation on the 2-D plane. We take the discretization parameter \( h = 0.01 \) and assume sub-Gaussian uncertainty \( u_t := (w_{1,t}, w_{2,t}, w_{3,t}) \in \mathbb{R}^3 \) to be a zero-mean, mixture of Gaussian and uniform distributions with \( \sigma = 0.5 \). The intermediate variable \( d_t := (d_{1,t}, d_{2,t}) \) depends on the wheel radius \( r = 0.15 \) m, the distance between wheels \( \bar{m} \), the controlled left-right wheel speed \( u_t := (v_{1,t}, v_{r,t}) \) and an unknown parameter \( \epsilon_t := (e_{1,t}, e_{2,t}) \), which depends on the wheel quality and road conditions. For simplicity, we assume that the planner adapts the system (6) with \( e_t \equiv (0,0) \) and \( u_t \equiv (0,0,0) \), and the vehicle can move over three types of road zones, the regular zone with \( e(1) := (0,0) \), the slippery zone with \( e(2) := (4,0) \), and the sandy zone with \( e(3) := (-1.2, -0.2) \), where locations of these zones are described in Fig. 2.

To adapt the proposed approach, we consider Problem (P) with the following loss function:
\[
\ell(u, x, y, \theta) = \frac{1}{2} \left\| u - u^{\text{ref}} \right\|^2 + \frac{1}{14\sqrt{2}} |x - x^{\text{ref}}|
\]
\[
+ \frac{1}{4\sqrt{2}} |y - y^{\text{ref}}| + \frac{289}{8} \left( \cos(\theta) - \cos(\theta^{\text{ref}}) \right)^2
\]
\[
+ \frac{289}{8} \left( \sin(\theta) - \sin(\theta^{\text{ref}}) \right)^2
\]
where \( (u^{\text{ref}}, x^{\text{ref}}, y^{\text{ref}}, \theta^{\text{ref}}) \) 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 (6), with \( u_t \equiv (0,0,0) \), and
\[
\begin{align*}
i &= 1, & e_1 &= 0, & e_2 &= 0 \\
i &= 2, & e_1 &= 10, & e_2 &= 0 \\
i &= 3, & e_1 &= 0, & e_2 &= 10.
\end{align*}
\]

Note that the ground truth parameter \( \alpha^* := (1,0,0) \) in the regular zone, \( \alpha^* := (0.6,0.4,0) \) in the slippery zone, and \( \alpha^* := (1.14, -0.12, -0.02) \) in the sandy zone. At each time \( t \), we have access to model sets \( \{ f^{(i)} \}_i \) and as well as the real-time dataset \( \mathcal{I}_t \) with size \( \mathcal{I}_0 = 100 \), which corresponds to the moving time window of order 0.1 s. 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 the characterization of the uncertain vehicle states, learning of the unknown road-condition parameter \( e \), and control toward 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 \( u^{\text{ref}} \) and the actual control \( u \) which reduces the weighted tracking error in road uncertainty.

**Example (Lane-Changing Behavior Adaptation):** In this scenario, we assume the initial system state \( x_0 = (10, 0, \pi/2) \). 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
Example of the (gray) planned trajectory and (black) controlled system trajectory in various road zones, with the system state $x = (x, y, \theta)$. 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.

To demonstrate the learning effect of the algorithm, we show in Fig. 4 components $\alpha_1$ and $\alpha_2$ of $\alpha := (\alpha_1, \alpha_2, \alpha_3)$, where the black lines indicate value of the ground truth $\alpha^*$ on the planned trajectory and the gray lines represent the learned, real-time estimate of $\alpha_1$ and $\alpha_2$ at the actual vehicle position. Notice that $\alpha^*$ 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 $\epsilon$ and $\gamma$.

As the proposed loss function $\ell$ 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 nonlinear and uncertain, evaluating the actual optimal objective value of Problem (P) is difficult. Therefore, it is 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 $\ell$ and as well as the realized objective value of Problem (P2), where the loss $\ell$ 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 $\ell$ is only locally convex in $x$. 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 $x_0 = (0, 30, 0)$. 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 $u \in \mathcal{U}$ of a unit wealth to three assets. At each time $t$, the agent receives random return rates $x_t = [x_{t1}, x_{t2}, x_{t3}]^\top$ of assets from some unknown and uncertain dynamics

$$ x_{t+1} = x_t + hA(t) + hw_t, \quad \text{with some } x_0 \in \mathbb{R}^3 $$

where $h = 10^{-3}$ is a stepsize, the vector $A(t)$ is randomly generated, unknown and piecewise constant, and the uncertainty vector $w_t$ 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 $w_t$ are zero and $x_3 \equiv 1$. Over time, an example of the realized loss $\ell$ and the achieved objective of (P2).

![Fig. 4.](image)

Fig. 4. (a) (Gray) Control signal provided by the planner and an example of the (black) control signal derived from the proposed approach. (b) Realized loss $\ell$ and the achieved objective of (P2).

![Fig. 5.](image)

Fig. 5. Component $\alpha_1$ and $\alpha_2$ of the real-time parameter $\alpha := (\alpha_1, \alpha_2, \alpha_3)$ in the learning procedure.

Fig. 6. Example of random returns $x = (x_1, x_2, x_3)$, 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. Component \( a_1 \) and \( a_2 \) of the real-time parameter \( \alpha := (a_1, a_2, a_3) \) in learning, where the values \( a_1^* \) and \( a_2^* \) are the online-inaccessible ground truth. Notice the responsive behavior of the proposed learning algorithm.

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

\[ x_{t+1} \text{ as in } (P1), \text{ we equivalently write it as in form } (P2'), \text{ where} \]

\[ G_\mu(t, u) = \frac{1}{T} \sum_{k \in T} F_\mu^S \left( \left\langle u, \frac{p_{k,t}}{r_0} \right\rangle + q_t \right) \]

with functions \( F_\mu^S \) and \( F_\mu^R \), and real-time data \( p_{k,t} \) and \( q_t \) determined as in Problem 2. We claim that \( G_\mu(t, u) \) has a time-dependent Lipschitz gradient constant in \( u \) given by \( \text{Lip}(G_\mu) = q_t/r_0 + 1/(r_0^2 T) \sum_{k \in T} \|p_{k,t}\|^2 \), and we use \( \varepsilon := 1/\text{Lip}(G_\mu) \) in the solution system (5) to compute the online decisions.

Fig. 9. Realized loss \( \ell \) and the achieved objective of (P2).

Fig. 9. Realized loss \( \ell \) and the achieved objective of (P2).

VII. CONCLUSION

In this article, we proposed a unified solution framework for online learning and optimization problems in the 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 nonsmooth 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 \( \ell, G, \) and \( G_\mu \) be those in Lemma V.2 on examples of (P2'). We now derive \( \nabla G_\mu := \nabla_u G_\mu(t, u) \) as follows.

**Problem 1 (Optimal control Under Uncertainty):**

\[
\nabla_u G_\mu(t, u) = \frac{1}{\mu} u + \frac{1}{T} \sum_{k \in T} \nabla_u F_\mu(p_{k,t}) + \frac{2}{T} \sum_{i=1}^p \nabla_u F_\mu(H_k^{(i)})
\]

where, for each \( k \in T \), the term \( \nabla_u F_\mu(p_{k,t}) \) is

\[
\frac{1}{\mu} \left( \sum_{i=1}^p \alpha_i f_2^{(i)}(t, \tilde{x}_t) \right)^\top p_{k,t}, \quad \text{if } \|p_{k,t}\| \leq \mu
\]

\[
\frac{1}{\|p_{k,t}\|} \sum_{i=1}^p \alpha_i f_2^{(i)}(t, \tilde{x}_t)^\top p_{k,t}, \quad \text{otherwise}
\]
and, for $k \in \mathcal{T}$, $i \in \{1, \ldots, p\}$, the term $\nabla u F_\mu(H_k^{(i)})$ is
\[
\begin{cases}
-\frac{1}{\mu} (f_\mu^{(i)}(t, x_i)) ^\top H_k^{(i)}, & \text{if } \|H_k^{(i)}\| \leq \mu \\
-\frac{1}{\mu} (f_\mu^{(i)}(t, x_i)) ^\top H_k^{(i)}, & \text{otherwise}.
\end{cases}
\]

**Problem 2 (Online Resource Allocation):**

\[
\nabla u G_\mu(t, u) = \frac{1}{T} \sum_{k \in \mathcal{T}} \nabla u F_\mu(u, p_{k,t}) + q t \nabla u F_\mu(u)
\]

where
\[
\nabla u F_\mu(u) := \begin{cases}
\frac{\nabla u}{\mu}, & \text{if } \|u\| \leq \mu \\
\frac{1}{\mu} u, & \text{otherwise}
\end{cases}
\]

and, for each $k \in \mathcal{T}$, the gradient $\nabla u F_\mu(u, p_{k,t})$ is
\[
\begin{cases}
-p_{k,t}, & \text{if } (u, p_{k,t}) \leq 1 - \mu \\
-p_{k,t} / \mu, & \text{if } 1 - \mu \leq (u, p_{k,t}) < 1 \\
0, & \text{if } (u, p_{k,t}) \geq 1.
\end{cases}
\]
These explicit expressions provide ingredients for the solution system. With different selections of the norm, the expression varies accordingly.

**APPENDIX B
STABILITY ANALYSIS OF THE SOLUTION SYSTEM**

Here, we adapt the 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], [40], and [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 to (P2'). To simplify the discussion, the notation we used in this subsection is different from that in the main body of the article. Consider the online problem, analogous to (P2'), defined as follows:

\[
\min_{x \in \mathcal{X}} f_t(x), \quad t = 0, 1, 2, \ldots
\]

where $f_t(x)$ is locally Lipschitz in $x$ with the parameter $h(x)$ and, at each time $t$, the objective function $f_t$ is $m_t$-strongly convex and $L_t$-smooth, with $m_t \geq 0$ 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

\[
x_{t+1} = \Pi(y_t - \alpha_t \nabla f_t(y_t)),
\]

\[
y_{t+1} = x_{t+1} + \beta_{t+1} (x_{t+1} - x_t)
\]

with some $y_0 = x_0 \in \mathcal{X}$, where $\alpha_t \leq 1/L_t$ and $\beta_t$ is selected iteratively, following

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

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

\[
\Pi(x) = \arg\min_{z \in \mathbb{R}^n} \frac{1}{2} \|z - x\|^2 + \alpha_t \ell(z)
\]

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

\[
x - \Pi(x) \in \alpha_t \partial \ell(\Pi(x))
\]

where the r.h.s. is the subdifferential of $\ell$ at $\Pi(x)$. Equivalently, we write the above condition as

\[
\Pi(x) = x - \alpha_t \partial \ell(\Pi(x)).
\]

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

\[
x_{t+1} = y_t - \alpha_t \nabla f_t(y_t) - \alpha_t \partial \ell(w_t),
\]

\[
y_{t+1} = x_{t+1} + \beta_{t+1} (x_{t+1} - x_t)
\]

\[
w_t = x_{t+1}.
\]

Note that (10) is not an explicit online algorithm, as the state $x_{t+1}$ 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 $x^*_t$. To do this, let $z_t := (x_t - x^*_t, x_{t+1} - x^*_{t+1})$ denote the tracking error vector and represent (10) as the error dynamical system

\[
z_{t+1} = A_t z_t + B^w_t u_t + B^v_t v_t
\]

with $z_t = (x_t - x^*_t, x_{0} - x^*_0)$. Using the gradient input $u_t := \nabla f_t(y_t) + \partial \ell(w_t)$, the reference signal $v_t := (x^*_t - x_{t+1}^*, x_{t+1}^* - x_{t+1}^*)$, the matrices

\[
A_t = \begin{bmatrix} 1 + \beta_t & -\beta_t \\ 0 & 1 \end{bmatrix}, \quad B^w_t = \begin{bmatrix} -\alpha_t \\ 0 \end{bmatrix}, \quad B^v_t = \begin{bmatrix} \beta_t & -1 \\ 0 & 0 \end{bmatrix}
\]

and the auxiliary variables

\[
y_t - x^*_t = \begin{bmatrix} 1 + \beta_t & -\beta_t \\ 0 & 1 \end{bmatrix} z_t + \begin{bmatrix} \alpha_t \\ 0 \end{bmatrix} v_t
\]

\[
w_t - x^*_t = \begin{bmatrix} 1 & 0 \end{bmatrix} z_{t+1} + \begin{bmatrix} 0 & 1 \end{bmatrix} v_t.
\]

We provide the following stability analysis of the system.

**Theorem B.1 (Stability of (9)):** Consider the solution algorithm (9), or equivalently (10). Then,

1. For each $t \geq 1$, we have the following:

\[
f_t(x_t) - f_t(x_{t+1}) \geq \xi_t ^\top X_{1,t} \xi_t
\]

\[
f_t(x^*_t) - f_t(x_{t+1}) \geq \xi_t ^\top X_{2,t} \xi_t
\]

Here, $\xi_t := (z_t, u^t, v^t),$ and

\[
X_{1,t} := \begin{bmatrix} 1/2 \end{bmatrix}
\]

\[
\begin{bmatrix}
m\beta^2 & -m\beta^2 & -\beta & m\beta^2 & 0 \\
-m\beta^2 & m\beta^2 & \beta & -m\beta^2 & 0 \\
-\beta & \beta & \alpha(2 - L\alpha) & -\beta & 0 \\
m\beta^2 & -m\beta^2 & -\beta & m\beta^2 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

\[
X_{2,t} := \begin{bmatrix} 1/2 \end{bmatrix}
\]

\[
\begin{bmatrix}
m(1 + \beta)^2 & -\eta & -(1 + \beta) & \eta & 0 \\
-\eta & m\beta^2 & \beta & -m\beta^2 & 0 \\
-(1 + \beta) & \beta & \alpha(2 - L\alpha) & -\beta & 0 \\
\eta & -m\beta^2 & -\beta & m\beta^2 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

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

---

LI et al.: ONLINE OPTIMIZATION AND AMBIGUITY-BASED LEARNING OF DISTRIBUTIONALLY UNCERTAIN DYNAMIC SYSTEMS 6163

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
2) Given the horizon parameter $T_0 \in \mathbb{Z}_{\geq 0}$ with $T = \min\{t - 1, T_0\}$. Then, for any $t \geq 2$, the solution $x_t$ from (9) achieves

$$
f_t(x_t) - f_t(x_t^*) \leq \frac{4G_t}{(t + 2)^2} + TF_t + TK_t + \frac{4(t - T - 1 + \delta)^2}{(t + 2)^2}(f_{t-T}(x_{t-T}) - f_{t-T}(x_{t-T}^*))
$$

where the time-dependent parameters $G_t$, $F_t$, and $K_t$ are determined by $f_t$, $\alpha_t$, and $\beta_t$.

**Theorem B.1:** 1) By the $m$-strong convexity and $L$-smoothness of $f$, we have

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

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

(1a) Consider (12) with $(x, y) \equiv (x_t, y_t)$. We leverage $y_t = x_t + \beta(x_t - x_{t-1})$ and the distributive law\footnote{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 $a = \nabla f(y_t)$, $b = \partial\ell(w_t)$, $c = x_{t-1} - x_t$, $d = x_{t-1}^* - x_t^*$, $\alpha = (a + b)^\top(c - d)$ and $m = \frac{4}{(t + 2)^2}$, then we get}

$$
f(x_t) - f(y_t) \geq \beta\nabla f(y_t)^\top(x_{t-1} - x_t) + \frac{m\beta^2}{2}\|x_{t-1} - x_t\|^2
$$

$$
= \beta(\nabla f(y_t) + \partial\ell(w_t))^\top(x_{t-1} - x_t - x_{t-1}^* + x_t^*)
$$

$$
+ \frac{m\beta^2}{2}\|x_{t-1} - x_t - x_{t-1}^* + x_t^*\|^2
$$

$$
= \beta(\nabla f(y_t) + \partial\ell(w_t))^\top(x_{t-1}^* - x_t^*)
$$

$$
- \beta\partial\ell(w_t)^\top(x_t - x_{t-1})
$$

$$
+ \frac{m\beta^2}{2}\|x_{t-1}^* - x_t^*\|^2.
$$

We reorganize the right-hand-side (r.h.s.) into the matrix form as

$$
\frac{1}{2}\delta_t^\top \begin{pmatrix}
\frac{m\beta^2}{2} & -m\beta^2 & -\beta & m\beta^2 \\
-m\beta^2 & m\beta^2 & \beta & -m\beta^2 \\
\beta & -m\beta^2 & 0 & -\beta \\
m\beta^2 & -m\beta^2 & -\beta & m\beta^2
\end{pmatrix} \eta
$$

$$
\delta_t - \beta\partial\ell(w_t)^\top(x_{t-1} - x_t)
$$

with $\eta = (x_{t-1} - x_t^*, x_{t-1}^* - x_t^*, \nabla f(y_t) + \partial\ell(w_t), x_{t-1}^* - x_t^*)$.

(1b) Consider (13) with $(x, y) \equiv (x_t, y_t)$. We leverage $x_{t+1} = y_t - \alpha\nabla f(y_t) - \alpha\partial\ell(w_t)$ and the distributive law, resulting in

$$
f(y_t) - f(x_{t+1}) \geq \alpha\nabla f(y_t)^\top(\nabla f(y_t) + \partial\ell(w_t))
$$

$$
- \frac{L\alpha^2}{2}\|\nabla f(y_t) + \partial\ell(w_t)\|^2
$$

$$
= \frac{\alpha(2 - L\alpha)}{2}\|\nabla f(y_t) + \partial\ell(w_t)\|^2.
$$

Now, we sum the terms involving $\partial\ell(w_t)$ in the r.h.s. of inequalities in (1a) and (1b), leverage (10), and then apply the convexity of $\ell$, $x_t \in \mathcal{X}$, and $w_t = x_{t+1} \in \mathcal{X}$, to obtain the following:

$$
- \beta\partial\ell(w_t)^\top(x_{t-1} - x_t) - \alpha\partial\ell(w_t)^\top(\nabla f(y_t) + \partial\ell(w_t))
$$

$$
= -\partial\ell(w_t)^\top(x_t - x_{t-1}) \geq \ell(w_t) - \ell(x_t^*) = 0
$$

which results in $f(x_t) - f(x_{t+1}) \geq \xi_t^\top X_{t+1} \xi_t$.

Note that we have identified $(f, m, L, \alpha, \beta)$ with $(f_t, m_t, L_t, \alpha_t, \beta_t)$, and note that $\nabla f_t(x_t^*) + \partial\ell(x_t^*) = 0$.

(1c) Similarly, consider (12) with $(x, y) \equiv (x_t^*, y_t)$. From $y_t = x_t + \beta(x_t - x_{t-1})$ and the distributive law

$$
f(x_t^*) - f(y_t) \geq \nabla f(y_t)^\top(x_t^* - y_t) + \frac{m}{2}\|x_t^* - y_t\|^2
$$

$$
= (\nabla f(y_t) + \partial\ell(w_t))^\top(x_t^* - y_t) + \beta(x_t - x_{t-1}) + \beta(x_t - x_{t-1}^*)
$$

$$
+ \frac{m\beta^2}{2}\|x_t^* - x_{t-1}^*\|^2
$$

$$
= \frac{m\beta^2}{2}\|x_t^* - x_{t-1}^*\|^2
$$

$$
= \frac{1}{2}\delta_t^\top \begin{pmatrix}
m(1 + \beta)^2 & -\eta & (1 + \beta) & \eta \\
-\eta & m\beta^2 & 0 & -m\beta^2 \\
(1 + \beta) & 0 & -\beta & m\beta^2 \\
-\eta & -m\beta^2 & -\beta & m\beta^2
\end{pmatrix} \eta
$$

$$
- \partial\ell(w_t)^\top(x_t^* - y_t)
$$

with $\eta = (m(1 + \beta)^2 - (1 + \beta)\eta, (1 + \beta)\eta, \beta, -m\beta^2)^\top$. We add this inequality to that in (1b) and leverage

$$
- \partial\ell(w_t)^\top(x_t^* - y_t) - \alpha\partial\ell(w_t)^\top(\nabla f(y_t) + \partial\ell(w_t))
$$

$$
= -\partial\ell(w_t)^\top(x_t^* - w_t) \geq \ell(w_t) - \ell(x_t^*) = 0
$$

resulting in $f(x_t^*) - f(x_{t+1}) \geq \xi_t^\top X_{t+1} \xi_t$.

2) Let us define the time-varying function

$$
V_t(z_t) := \begin{bmatrix} z_t \\ x_t^* - x_{t-1}^* \end{bmatrix}^\top H_t \begin{bmatrix} z_t \\ x_t^* - x_{t-1}^* \end{bmatrix}
$$

where we take

$$
H_t := \frac{1}{2\alpha(t - 1)} \begin{pmatrix}
\delta_{t-1} \\ 1 - \delta_{t-1}
\end{pmatrix} \begin{pmatrix}
\delta_{t-1} & 1 - \delta_{t-1}, \delta_{t-1}
\end{pmatrix}
$$

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

$$
V_{t+1}(z_{t+1}) - \frac{\alpha_t - 1}{\alpha_t} V_t(z_t) = \xi_t^\top J_t \xi_t
$$

where $\xi_t := (z_t, u_t, v_t)$, which are those defined in (11), resulting in $\xi_t := (x_t^* - x_{t-1}^*, x_{t-1} - x_{t-1}^*, \nabla f_t(y_t) + \partial\ell(w_t), x_t^* - x_{t-1}^*, x_t^* - x_{t-1}^*)$ and

$$
J_t = \frac{1}{2\alpha_t}.
Let us compute, unnumbered equation shown at the bottom of this page, and then achieve
\[ \xi_t^T (J_t - M_t) \xi_t = \left[ \begin{array}{c} z_t \\ x_t^* - x_{t-1}^* \end{array} \right]^T N_{1,t} \left[ \begin{array}{c} z_t \\ x_t^* - x_{t-1}^* \end{array} \right] + \left[ \begin{array}{c} z_t \\ x_t^* - x_{t-1}^* \end{array} \right]^T N_{2,t} \left[ \begin{array}{c} z_t \\ x_t^* - x_{t-1}^* \end{array} \right] - \alpha_t (1 - L_t \alpha_t) u_t^T u_t \]

with, for each \( t \geq 1 \),
\[ N_{1,t} := \frac{1}{2} \left( \begin{array}{cccc} -m_t(\delta_{t-1}^2 - 1) & m_t \beta_t \delta_{t-1} & -m_t \beta_t \delta_{t-1} & 0 \\ m_t \beta_t \delta_{t-1} & -m_t \beta_t \delta_{t-1} & m_t \beta_t \delta_{t-1} & 0 \\ -m_t \beta_t \delta_{t-1} & m_t \beta_t \delta_{t-1} & -m_t \beta_t \delta_{t-1} & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \]
\[ \succeq \frac{m_t}{2} \left( \begin{array}{cccc} -(\delta_{t-1}^2 - 1) & \beta_t \delta_{t-1} & 0 & 0 \\ \beta_t \delta_{t-1} & -(\beta_t^2 \delta_{t-1}^2 - 1) & 0 & 0 \\ 0 & 0 & -\beta_t \delta_{t-1} & 0 \\ 0 & 0 & 0 & -\beta_t \delta_{t-1} \end{array} \right) \geq 0 \]
and, using the fact that \( \delta_t > (t+1)/2, \forall t \geq 0 \), we have
\[ N_{2,t} := \frac{1}{2} \left( \begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\beta_t \delta_{t-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \leq 0 \]
Then, if we select \( \alpha_t \leq 1/L_t \), it results in
\[ \xi_t^T (J_t - M_t) \xi_t \leq 0. \]
We rewrite it as
\[ V_{t+1}(z_{t+1}) - \frac{\alpha_t}{\alpha_t} V_t(z_t) \leq \xi_t^T M_t \xi_t \]
\[ \leq \delta_{t-1}^2 (f_t(x_t) - f_t(x_{t+1})) + \delta_t (f_t(x_t^*) - f_t(x_{t+1})) \]
\[ = -\delta_{t-1}^2 (f_t(x_{t+1}) - f_t(x_t^*)) + \delta_t^2 (f_t(x_t) - f_t(x_t^*)). \]
As \( f_t \) being locally Lipschitz in \( t \), there exists a nonnegative function \( h(x) \) such that
\[ f_t(x_{t+1}) - f_t(x_t^*) \leq h(x_{t+1}) \]
resulting in
\[ V_{t+1}(z_{t+1}) - \frac{\alpha_t}{\alpha_t} V_t(z_t) \leq -\delta_{t-1}^2 (f_t(x_{t+1}) - f_t(x_{t+1}^*)) + \delta_{t-1}^2 (f_t(x_t) - f_t(x_t^*)) \]
\[ -\delta_t^2 (f_t(x_{t+1}^*) - f_t(x_t^*)) + \delta_t^2 h(x_{t+1}) \forall t. \]
Summing up the above set of inequalities over the moving horizon window \( t \in T = \{t-1, \ldots, t-T\} \), where \( T = \min\{t-1, T_0\} \) with some \( T_0 \in \mathbb{Z}_{>0} \), we obtain
\[ V_t(z_t) + \sum_{k \in T} \left( 1 - \frac{\alpha_{k-1}}{\alpha_k} \right) V_k(z_k) - V_{t-T}(z_{t-T}) \]
\[ \leq -\delta_{t-1}^2 (f_t(x_t) - f_t(x_t^*)) \]
\[ + \delta_{t-T-1}^2 (f_{t-T}(x_{t-T}) - f_{t-T}(x_{t-T}^*)) \]
\[ - \sum_{k \in T} \delta_k^2 (f_k(x_{k+1}^*) - f_k(x_k^*)) + \sum_{k \in T} \delta_k^2 h(x_{k+1}). \]
Let us denote by \( G_t, K_t, \) and \( F_t \), respectively, the horizon accumulated potential, the bound of the locally Lipschitz function \( h_1 \), and the variation bound of the optimal objective values. That is,
\[ G_t := V_{t-T}(z_{t-T}) - V_t(z_t) - \sum_{k \in T} \left( 1 - \frac{\alpha_{k-1}}{\alpha_k} \right) V_k(z_k) \]
\[ K_t := \max_{k \in T} \{h(x_{k+1})\} \]
\[ F_t := \max_{k \in T} \{f_k(x_{k+1}^*) - f_k(x_k^*)\}. \]
Then, using the fact that \( 1) \delta_{t-1} \geq (t+2)/2, \forall t \geq 0; 2) \delta_{t-1} \leq t - T + \delta_t \) with \( \delta_t = (1 + \sqrt{5})/2; \) and \( 3) \delta_t \) is monotonically increasing, we have
\[ f_t(x_t) - f_t(x_t^*) \leq 4G_t \]
\[ + \frac{4(t-T-1 + \delta_0)}{(t+2)^2} (f_{t-T}(x_{t-T}) - f_{t-T}(x_{t-T}^*)) \]
Note that, when \( t \leq T_0 + 1 \), we have \( T = t - 1 \). This gives
\[ f_t(x_t) - f_t(x_t^*) \leq 4G_t \]
\[ + \frac{4\delta_0^2}{(t+2)^2} (f_1(x_1) - f_1(x_1^*)). \]

**APPENDIX C**

**PROOFS OF LEMMAS AND THEOREMS**

**Proof of Lemma IV.1:** By the definition of the ambiguity set, we have that, for any distribution \( \mathcal{Q} \in \mathcal{P}_{t+1}(\alpha, u) \)
\[ d_W(Q, \hat{P}_{t+1}|u) \leq \hat{\varepsilon} \]
which, by the Kantorovich–Rubinstein theorem, is equivalent to
\[ \int_Z h(x)|Q(dx) - \int_Z h(x)|\hat{P}_{t+1}|u(dx) \leq \hat{\varepsilon} \quad \forall h \in \mathcal{L} \]
where \( \mathcal{L} \) is the set of functions with Lipschitz constant 1 and \( Z \) is the support of the random variable \( x \). For a given \( u \), let us

\[ M_t := \delta_{t-1}^2 X_{1,t} + \delta_{t-1} X_{2,t} = \frac{1}{2} \left( \begin{array}{cccc} m_t(\delta_{t-1}^2 - 1) & -m_t \beta_t \delta_{t-1} & -\delta_t \delta_{t-1} & m_t \beta_t \delta_{t-1} \\ -m_t \beta_t \delta_{t-1} & m_t \beta_t \delta_{t-1} & \beta_t \delta_{t-1} & -m_t \beta_t \delta_{t-1} \\ -\delta_t \delta_{t-1} & \beta_t \delta_{t-1} & m_t \beta_t \delta_{t-1} & 0 \\ m_t \beta_t \delta_{t-1} & -m_t \beta_t \delta_{t-1} & -\delta_t \delta_{t-1} & 0 \end{array} \right) \]
select $h$ to be
\[ h(x) := \frac{\ell(u, x)}{L(u)} \]
where $L$ is the positive Lipschitz function as in Assumption IV.1. Substituting $h$ to the above inequality, we have
\[ \int_Z \ell(u, x) dQ(dx) - \int_Z \ell(u, x) \hat{P}_{t+1}^\mu(dx) \leq \epsilon L(u) \]
or equivalently
\[ \mathbb{E}_Q[\ell(u, x)] \leq \mathbb{E}_{\hat{P}_{t+1}^\mu}(\ell(u, x)) + \epsilon L(u). \]
As the inequality holds for every $Q \in \mathcal{P}_{t+1}$, therefore
\[ \sup_{Q \in \mathcal{P}_{t+1}(\alpha, u)} \mathbb{E}_Q[\ell(u, x)] \leq \mathbb{E}_{\hat{P}_{t+1}^\mu}(\ell(u, x)) + \epsilon(t, T, \beta, \alpha, u) L(u). \]
\[
\text{Proof of Theorem IV.1:} \text{ We show this by constructing a distribution in the ambiguity set. By Assumption IV.2 on convex and gradient-accessible functions, there exists an index } j \in \mathcal{T} \text{ such that the derivative } \nabla_{x} \ell(u, x) \text{ at } (u, \bar{x}(j)), \bar{x}(j) := \sum_{i=1}^{p} \alpha_i s_{i}(\alpha, u), \text{ satisfies} \]
\[ \|\nabla_{x} \ell(u, \bar{x}(j))\| = L(u). \]
Now using this index $j$, we construct a parameterized distribution as follows:
\[ Q(\Delta x) = \frac{1}{T} \sum_{k \in \mathcal{T}, k \neq j} \delta \left\{ \sum_{i=1}^{p} \alpha_i s_{i}(\alpha, u) \right\} + \frac{1}{T} \delta[\bar{x}(j) + \Delta x] \]
where $\Delta x \in \mathbb{R}^n$ with $\|\Delta x\| \leq T \epsilon$. By the definition of the ambiguity set and, since the support of the distribution $P$ is $\mathcal{E}_{t+1} = \mathbb{R}^n$, we have $Q(\Delta x) \in \mathcal{P}_{t+1}(\alpha, u)$. Next, we quantify the lower bound of the following term:
\[ \mathbb{E}_Q(\Delta x) \mathbb{E}_Q[\ell(u, x)] - \mathbb{E}_{\hat{P}_{t+1}^\mu}(\ell(u, x)) \]
\[ = \frac{1}{T} \left( \ell(u, \bar{x}(j) + \Delta x) - \ell(u, \bar{x}(j)) \right). \]
By Assumption IV.2 on the convexity of $\ell$ on $x$, we have
\[ \ell(u, \bar{x}(j) + \Delta x) - \ell(u, \bar{x}(j)) \geq \nabla_{x} \ell(u, \bar{x}(j))^{\top} \Delta x. \]
Then, by selecting
\[ \Delta x := T \epsilon \nabla_{x} \ell(u, \bar{x}(j)) \]
we have
\[ \nabla_{x} \ell(u, \bar{x}(j))^{\top} \Delta x = T \epsilon L(u). \]
These results bound in
\[ \mathbb{E}_Q(\Delta x) \mathbb{E}_Q[\ell(u, x)] - \mathbb{E}_{\hat{P}_{t+1}^\mu}(\ell(u, x)) \geq \epsilon L(u). \]
As $Q(\Delta x) \in \mathcal{P}_{t+1}(\alpha, u)$, therefore
\[ \sup_{Q \in \mathcal{P}_{t+1}(\alpha, u)} \mathbb{E}_Q[\ell(u, x)] \geq \mathbb{E}_{\hat{P}_{t+1}^\mu}(\ell(u, x)) + \epsilon L(u). \]
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
\[ \inf_{u \in U} \mathbb{E}_{\hat{P}_{t+1}^\mu}(\ell(u, x)) + \epsilon(t, T, \beta, \alpha, u) L(u) \]
which is the Problem (P2).
\[
\text{Proof of Lemma IV.2:} \text{ This is the direct application of the definition of the local Lipschitz condition.} \]
\[
\text{Proof of Lemma VI.1:} \text{ First, we have} \]
\[ F_{\mu}(u) \leq F(u) + \frac{1}{2\mu} \|u - u\|^2 = F(u) \quad \forall u \in \mathcal{U}. \]
Then, we compute
\[ F(u) - F_{\mu}(u) = \sup_{z \in \mathcal{U}} \left\{ F(u) - F(z) - \frac{1}{2\mu} \|z - u\|^2 \right\} \]
\[ \leq \sup_{z \in \mathcal{U}} \left\{ g(u)^{\top}(u - z) - \frac{1}{2\mu} \|z - u\|^2 \right\} \]
\[ \leq \sup_{z} \left\{ g(u)^{\top}(u - z) - \frac{1}{2\mu} \|z - u\|^2 \right\} \]
\[ \leq \frac{\mu}{2} g(u)^{\top} g(u) \leq \frac{1}{\mu} \]
where the equality comes from the definition of $F_{\mu}(u)$, the first inequality leverages the convexity of $F$, the second one relaxes the constraint set, the third one applies the achieved optimizer $z^* = u - \mu g(u)$, 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 $F_{\mu}$ is convex and continuously differentiable, where its gradient $\nabla F_{\mu}$ is Lipschitz continuous with constant $1/\mu$. In addition, the minimizer $z^*(u)$ of $F_{\mu}$ is achievable and unique, resulting in an explicit gradient expression of $F_{\mu}$ as follows:
\[ \nabla F_{\mu}(u) = \frac{1}{\mu}(u - z^*(u)). \]
In addition, we claim that, if $F$ is $M$-strongly convex, $F_{\mu}$ is $M/(1 + \mu M)$-strongly convex, following [44, Th. 2.2]. Finally, we equivalently write the minimization problem as follows:
\[ \min_{u \in \mathcal{U}} F_{\mu}(u) = \min_{u \in \mathcal{U}} \min_{z \in \mathcal{U}} \left\{ F(z) + \frac{1}{2\mu} \|z - u\|^2 \right\} \]
\[ = \min_{u \in \mathcal{U}} \min_{z \in \mathcal{U}} \left\{ F(z) + \frac{1}{2\mu} \|z - u\|^2 \right\} \]
\[ = \min_{z \in \mathcal{U}} F(z) \]
where the first line applies the achievability of the minimizer of the problem that defines $F_{\mu}$, the second switches the minimization operators, and the third applies the fact that $u = z$ solves the inner problem. This concludes that any $u$ that minimizes $F_{\mu}$ also minimizes $F$, and vice versa.

\[
\text{Proof of Theorem VI.1:} \text{ Let us consider the solution system (5). At each time } t, \text{ we let select } \epsilon := \epsilon_t = 1/\text{Lip}(G_{\mu}), \text{ or equivalently, } \mu/b \text{ with } b = \max_{k \in T} b_k. \text{ Let } \eta_t \text{ satisfy} \]
\[ \delta_{-1} = 1, \delta_{t+1} := \frac{1 + \sqrt{1 + 4\delta_t^2}}{2}, \eta_t := \delta_{-1} - 1/\delta_t. \]
Then, by Theorem B.1 with $t \geq 2$, the following holds:
\[ G_{\mu}(t, u_t) - G_{\mu}(t, u^*_t) \leq \frac{4W_t}{(t + 2)^2} + TF_t \]
where $u^*_t$ is a solution to (P3), $T = \min\{t - 1, T_0\}$ 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 $T$, and $F_I$ is the variation bound of the optimal objective values in $T$. Specifically, we have

\[ F_I = \max_{k \in T} \left\{ |G_k(u_{k+1}, u_k^*) - G_k(k, u_k^*)| \right\} + \tilde{L} \]

with $\tilde{L}$ the variation bound of $G_k(u_{k+1})$ w.r.t. time $t$. Let us consider the storage function $V_t(z_t) := z_t^\top H_t z_t$, where $z_t := (u_t - u_t^*, u_{t-1} - u_{t-1}^*, u_t^* - u_{t-1}^*)$ and

\[ H_t := \frac{1}{2\epsilon_{t-1}} \begin{bmatrix} \delta_{t-1} & 1 - \delta_{t-1} \\ 1 - \delta_{t-1} & \delta_{t-1} \end{bmatrix} \]

where $\delta_{t-1} = 1 - \frac{\epsilon_{t-1}}{\epsilon_k}$.

Then, we have

\[ W_t = V_{t-T}(z_{t-T}) - V_t(z_t) - \sum_{k \in T} \left( 1 - \frac{\epsilon_{k-1}}{\epsilon_k} \right) V_k(z_k) + (t - T + 1 + \delta_0)^2(f_{t-T}(x_{t-T}) - f_{t-T}(x_{t-T}^*)) \]

where the first two terms are the energy decrease in the horizon $T$; 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 $T$. Note how the selection of $\epsilon_t$ and $T$ affect $W_t$ (or $G_t$ in Theorem B.1). In the most conservative scenario, we select $\epsilon_t := \min\{\epsilon_{t-1}, \mu/b_t\}$ and $T_0 = \infty$, which results in a constant upper bound of $W_t$ as follows:

\[ W_t \leq V_1(z_1) + \delta_0^2(f_1(x_1) - f_1(x_1^*)) \]

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 $\epsilon_t := \min\{\epsilon_{t-1}, \mu/b_t\}$ but a finite $T_0$. Then, as $t$ is sufficiently large, we have

\[ W_t \leq V_{t-T}(z_{t-T}) + (t^2(f_{t-T}(x_{t-T}) - f_{t-T}(x_{t-T}^*))) \]

where, in this case, the bound (14) essentially depends on $F_t$ and $f_{t-T}(x_{t-T}) - f_{t-T}(x_{t-T}^*)$.

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

\[ G(t, u_t) - \alpha u_t \leq G(t, u_t^*) \]

and by Lemma V.1, we have that $u_t^*$ is a minimizer of (P2) if and only if it is that of (P2), and

\[ G(t, u_t^*) = G(t, u_t^*) \]

This results in

\[ G(t, u_t) - G(t, u_t^*) \leq \frac{4W_t}{(t + 2)^2} + TF_t + \alpha u_t \]

with an equivalent expression of $F_t$ as

\[ F_t = \max_{k \in T} \left\{ |G_k^* - G_k^*| \right\} + \tilde{L} \]

where $G_k^* := G(k, u_k^*)$ is the optimal objective value of (P2) or, later we see, equivalent to that of (P1).

Next, by Table IV.1 on the equivalence of (P1) and (P2), $u_t^*$ is a minimizer of (P2) if and only if it is also that of (P1), and

\[ G(t, u_t^*) \geq \sup_{Q \in P_{t+1} \alpha, u_t^*} \mathbb{E}_{Q} \left[ f(u_t^*, x) \right] \]

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

\[ \text{Prob} \left( \mathbb{P}_{t+1}| Q \in \mathbb{P}_{t+1} \right) \geq \rho(t) \]

or equivalently,

\[ \text{Prob} \left( \mathbb{E}_{\mathbb{P}_{t+1}} \left[ f(u_t^*, x) \right] \leq G(t, u_t^*) \right) \geq \rho(t) \]

with $\rho(t)$ as in Theorem III.1. Then, by (17), we know that $\mathbb{P}_{t+1}| Q \in \mathbb{P}_{t+1}$ if and only if $d_W(\mathbb{P}_{t+1}| Q, \tilde{P}_{t+1}| Q) \leq \bar{c}$, where $\bar{c}$ is selected as in Theorem III.1. Further, since $d_W$ is a metric, for any $Q \in \mathbb{P}_{t+1}$, we claim

\[ d_W(Q, \tilde{P}_{t+1}| Q) \leq d_W(Q, \tilde{P}_{t+1}| Q) + d_W(\tilde{P}_{t+1}| Q) \]

where $\bar{c} = \epsilon T + 2\epsilon T$.

By Assumption IV.1 and the same proof procedure of Lemma IV.1 on the above inequality, we have, for every $u_t^*$, the following:

\[ \sup_{Q \in P_{t+1}(\alpha, u_t^*)} \mathbb{E}_{Q} \left[ f(u_t^*, x) \right] \leq \mathbb{E}_{\mathbb{P}_{t+1}} \left[ f(u_t^*, x) \right] + 2L \bar{c} \]

By taking $u_t := u_t^*$ and using (16), we have

\[ G(t, u_t^*) \leq \mathbb{E}_{\mathbb{P}_{t+1}} \left[ f(u_t^*, x) \right] + 2L \bar{c} \]

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

\[ \mathbb{E}_{\mathbb{P}_{t+1}} \left[ f(u_t^*, x) \right] \]

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

\[ \lim_{t \to \infty} \text{Prob} \left( R_t \leq TF_t + \alpha u_t \right) \geq 1 - \beta. \]

References

[1] S. Shalev-Shwartz, “Online learning and online convex optimization,” Foundations Trends Mach. Learn., 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 Proc. Int. Conf. Adv. Neural Inf. Process. Syst., 2010, pp. 1984–1992.
[3] A. Simonetto, E. Dall’Anese, S. Paternain, G. Leus, and G. B. Giannakis, “Time-varying convex optimization: Time-structured algorithms and applications,” Proc. IEEE, vol. 108, pp. 2032–2048, Nov. 2020.
[4] M. Zinkevich, “Online convex programming and generalized infinitesimal gradient ascent,” in Proc. Int. Conf. Mach. Learn., 2003, pp. 928–936.
[5] E. Hazan, “Introduction to online convex optimization,” Foundations Trends Optim., vol. 2, no. 3/4, pp. 157–325, 2016.
[6] A. Mohktari, S. Shahrampour, A. Jadbabaie, and A. Ribeiro, “Online optimization in dynamic environments: Improved regret rates for strongly convex problems,” in Proc. IEEE Conf. Decis. Control, 2016, pp. 7195–7201.
[7] A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan, “Online optimization: Competing with dynamic comparators,” in Proc. 18th Int. Conf. Artif. Intell. Statis., pp. 398–406, 2015.
[8] A. Rakhlin and K. Sridharan, “Online learning with predictable sequences,” J. Mach. Learn. Res., vol. 16, pp. 155–186, 2015.
[9] E. Hall and R. Willert, “Online convex optimization in dynamic environments,” IEEE J. Sel. Topics Signal Process., vol. 9, no. 4, pp. 647–662, Jun. 2015.
[10] N. Chen, A. Agarwal, A. Wierman, S. Barman, and L. Andrew, “Online convex optimization using predictions,” in Proc. ACM Int. Conf. Meas. Model. Comput. Syst., 2015, pp. 191–204.
