2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

Regret Optimal Control for Uncertain Stochastic Systems

Andrea Martin, Luca Furieri, Florian Dörfler, John Lygeros, and Giancarlo Ferrari-Trecate

Abstract— **We consider control of uncertain linear time-varying stochastic systems from the perspective of regret minimization. Specifically, we focus on the problem of designing a feedback controller that minimizes the loss relative to a clairvoyant optimal policy that has foreknowledge of both the system dynamics and the exogenous disturbances. In this competitive framework, establishing robustness guarantees proves challenging as, differently from the case where the model is known, the clairvoyant optimal policy is not only inapplicable, but also impossible to compute without knowledge of the system parameters. To address this challenge, we embrace a scenario optimization approach, and we propose minimizing regret robustly over a finite set of randomly sampled system parameters. We prove that this policy optimization problem can be solved through semidefinite programming, and that the corresponding solution retains strong probabilistic out-of-sample regret guarantees in face of the uncertain dynamics. Our method naturally extends to include satisfaction of safety constraints with high probability. We validate our theoretical results and showcase the potential of our approach by means of numerical simulations.**

I. Introduction

Inspired by online optimization and learning methods, control of dynamical system has recently been studied through the lens of regret minimization [1]. This emerging paradigm aims at designing efficient control laws that minimize the worst-case loss relative to an optimal policy in hindsight. Algorithms with provable regret certificates hence offer attractive performance guarantees that – in contrast with the stochastic and worst-case assumptions typical of $\mathcal{H}_2$ and $\mathcal{H}_\infty$ controllers [2] – hold independently of how disturbances are generated.

Most prior work in this area employs gradient methods to deal with adversarially chosen cost functions and perturbations, and shows that the resulting control law achieves sublinear regret against expressive policy classes [1], [3]. A parallel line of research, initiated by [4], [5], studies the problem of competing against the optimal control actions selected by a clairvoyant (noncausal) policy, without imposing any parametric structure on this benchmark policy.

For the case of known cost functions, the formulation of [4], [5] has received increasing interest thanks to: optimality of the clairvoyant benchmark policy, possibility of computing the regret-minimizing controller, and remarkable performance reported in several applications, including longitudinal motion control of a helicopter and control of a wind energy conversion system [6]. In particular, among recent contributions, [7] and [8] proposed an efficient optimization-based synthesis framework to incorporate safety constraints, [9] established recursive feasibility and stability guarantees for receding horizon regret optimal control, [10] and [6] investigated the closely related metric of competitive ratio, [5] and [11] considered state estimation problems, and [12] studied connections with imitation learning.

Despite these advances, an important open challenge is how to track the performance of the clairvoyant optimal policy without knowledge of the underlying dynamics. In fact, as the systems under control become increasingly complex, assuming availability of precise mathematical models appears more and more unrealistic. Nevertheless, to the best of our knowledge, only [13] approached this problem, showing that several iterative control algorithms that combine system identification with gradient descent methods, e.g., [1], [3], also achieve, asymptotically, near-optimal competitive ratio relative to the clairvoyant optimal policy. However, this result only holds asymptotically and does not allow synthesizing control policies that, given a set of admissible plants, guarantee that the regret relative to the clairvoyant optimal policy is minimized robustly.

Towards addressing these issues, in this paper we present a solution to the robust regret minimization problem based on scenario optimization [14], [15], which is applicable to uncertain stochastic linear time-varying systems affected by a priori unknown but measurable disturbance processes.[1] A key challenge lies in handling the different impacts that parametric uncertainty has on the closed-loop behavior achieved by the clairvoyant benchmark policy, on the one hand, and by the causal controller to be designed on the other. In fact, simultaneously accounting for these effects has not yet been achieved following the analysis methods used in [17], [18] to derive suboptimality and sample complexity bounds for classical linear quadratic control problems.

For several control applications, including robotics, building energy management, and power grids, designing a single state feedback policy that attains robust performance across all admissible system dynamics can prove overly conservative. Instead, it is beneficial to optimize for a unique closed-loop behavior – while allowing the state feedback law that achieves it to vary – leveraging a posteriori measurements of exogenous perturbations such as external forces, solar ra-

A. Martin, L. Furieri, and G. Ferrari-Trecate are with the Institute of Mechanical Engineering, EPFL, Switzerland. E-mail addresses: {andrea.martin, luca.furieri, giancarlo.ferraritrecate}@epfl.ch.

F. Dörfler and J. Lygeros are with the Department of Information Technology and Electrical Engineering, ETH Zürich, Switzerland. E-mail addresses: {dorfler, jlygeros}@ethz.ch.

Research supported by the Swiss National Science Foundation (SNSF) under the NCCR Automation (grant agreement 51NF40_80545). Luca Furieri is also grateful to the SNSF for the Ambizione grant PZ00P2_208951.

[1] These include but are not limited to the class of linear parameter-varying systems – a middle ground between linear and nonlinear dynamics [16].

diation, and electricity demands for control implementation.

Motivated as above, we show how convex optimization and sampling techniques can be used to synthesize a disturbance feedback robust control policy with provable regret guarantees in spite of the uncertain dynamics. In particular, building upon [14], [15], we propose constructing a scenario problem by appropriately sampling over the space of uncertain parameters. We prove that the policy that minimizes regret robustly over the considered scenarios can be computed via semidefinite programming, and that this solution exhibits generalization capabilities – in the sense that the resulting regret bound holds true for all but a small fraction of uncertainty realizations whose probability is no larger than a prespecified tolerance level. Our approach naturally extends to include satisfaction of safety constraints with high probability. The advantages of our probabilistic design method are twofold. First, contrary to worst-case solutions, which are known to be computationally hard to evaluate, and coherently with the theory of scenario optimization, our approach uses a finite number of randomly sampled uncertainty realizations only, and thus calls for the solution of a convex program – albeit with a size that increases with the number of considered scenarios. Second, as opposed to probabilistic solutions based on scenario optimization with classical $\mathcal{H}_\infty$ objectives, our method leverages the cost of the optimal policy in hindsight to yield performance guarantees that are tailored to the specific uncertainty and disturbance realizations. In turn, as we validate by means of numerical simulations, this often allows us to reduce conservatism of $\mathcal{H}_\infty$ methods by establishing tighter upper bounds on the realized cost – which in turn translate into improved closed-loop performance across all system dynamics for several disturbance profiles of practical relevance.

II. Problem Statement and Preliminaries

A. Dynamics, control objective, and constraints

We consider an uncertain discrete-time linear time-varying dynamical system described by the state-space equation

$$x_{t+1} = A_t(\theta_t)x_t + B_t(\theta_t)u_t + E_t(\theta_t)w_t\,, \tag{1}$$

where $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^m$, $\theta_t \in \mathbb{R}^d$ and $w_t \in \mathbb{R}^p$ are the system state, the control input, a vector of uncertain parameters that characterize the family of admissible plants, and a measurable disturbance process, respectively. We focus on optimizing the closed-loop behavior of this uncertain system over a finite-time planning horizon of length $T \in \mathbb{N}$, and let $\boldsymbol{x} = (x_0, x_1, \dots, x_{T-1})$, $\boldsymbol{u} = (u_0, u_1, \dots, u_{T-1})$, $\boldsymbol{w} = (x_0, w_0, \dots, w_{T-2})$, and $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_{T-1})$ for compactness. On the one hand, we do not make any assumptions about the statistical properties of the exogenous disturbance process $\boldsymbol{w}$, that can also be adversarially selected. On the other hand, we assume that $\boldsymbol{\theta}$ is drawn according to a probability distribution $\mathbb{P}_{\boldsymbol{\theta}}$ with a possibly unknown and unbounded support set $\boldsymbol{\Theta}$. This probability measure may reflect a priori knowledge about the actual likelihood of each realization of the system parameters, or may simply encode the relative importance that we attribute to each uncertainty instance. In particular, we do not require $\mathbb{P}_{\boldsymbol{\theta}}$ to be known explicitly, but rely on a set $\mathcal{D} = \{\boldsymbol{\theta}^1, \dots, \boldsymbol{\theta}^N\}$ of $N \in \mathbb{N}$ independent samples only. Finally, we assume that the matrices $E_t(\theta_t)$ are full column rank for all $t \in \mathbb{I}_T = \{0, \dots, T-1\}$ and for all θ_t such that $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Motivated by the regret optimal control framework of [4], [5], we consider the problem of designing a causal decision policy $\boldsymbol{\pi} = (\pi_0, \dots, \pi_{T-1})$, with $u_t = \pi_t(x_0, \dots, x_t, w_0, \dots, w_{t-1})$, that closely tracks the performance of an ideal clairvoyant policy $\boldsymbol{\psi} = (\psi_0, \dots, \psi_{T-1})$. Importantly, we allow the noncausal benchmark policy $\boldsymbol{\psi}$ to select the control actions with foreknowledge of both the exogenous disturbance $\boldsymbol{w}$ and the system dynamics $\boldsymbol{\theta}$, i.e., $u_t = \psi_t(x_0, \dots, x_t, w_0, \dots, w_{T-2}, \theta_0, \dots, \theta_{T-1})$. More specifically, for any fixed $\boldsymbol{w}$ and $\boldsymbol{\theta}$, let

$$J(\boldsymbol{\pi}, \boldsymbol{w}, \boldsymbol{\theta}) = \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{u}^\top \boldsymbol{R}\boldsymbol{u}\,, \tag{2}$$

with $\boldsymbol{Q} \succeq 0$ and $\boldsymbol{R} \succ 0$, denote the control cost incurred by playing the policy $\boldsymbol{\pi}$, and define the per-instance regret of $\boldsymbol{\pi}$ relative to $\boldsymbol{\psi}$ as:

$$\mathtt{R}(\boldsymbol{\pi}, \boldsymbol{\psi}, \boldsymbol{w}, \boldsymbol{\theta}) = J(\boldsymbol{\pi}, \boldsymbol{w}, \boldsymbol{\theta}) - J(\boldsymbol{\psi}, \boldsymbol{w}, \boldsymbol{\theta})\,. \tag{3}$$

Building upon ideas proposed in [4], [5] for the case where the system dynamics (1) are perfectly known, we then formulate the robust regret minimization problem as follows:

$$\mathtt{R}^\star(\boldsymbol{\psi}) = \inf_{\boldsymbol{\pi}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \max_{\|\boldsymbol{w}\|_2 \leq 1} \mathtt{R}(\boldsymbol{\pi}, \boldsymbol{\psi}, \boldsymbol{w}, \boldsymbol{\theta})\,. \tag{4}$$

A solution $\boldsymbol{\pi}^\star$ to (4), if any, guarantees that its cost is always at most $\mathtt{R}^\star(\boldsymbol{\psi})$ higher than that of the ideal, yet inapplicable, benchmark policy $\boldsymbol{\psi}(\boldsymbol{w}, \boldsymbol{\theta})$ that minimizes (2) – no matter how $\boldsymbol{w}$ is generated and which $\boldsymbol{\theta}$ realize.

As modern engineering systems often feature safety-critical components, we include in the synthesis problem a robust constraint satisfaction requirement. In particular, we define a polytopic safe set in the space of state and input trajectories as follows:

$$\mathcal{S}(\boldsymbol{\theta}) = \{(\boldsymbol{x}, \boldsymbol{u}) : \boldsymbol{H}_x(\boldsymbol{\theta})\boldsymbol{x} + \boldsymbol{H}_u(\boldsymbol{\theta})\boldsymbol{u} \leq \boldsymbol{h}(\boldsymbol{\theta})\}\,. \tag{5}$$

Then, we consider the objective of solving (4) while ensuring that $(\boldsymbol{x}, \boldsymbol{u}) \in \mathcal{S}(\boldsymbol{\theta})$ robustly for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{w}$ belonging to a compact disturbance set $\mathcal{W}(\boldsymbol{\theta})$ defined as

$$\mathcal{W}(\boldsymbol{\theta}) = \{\boldsymbol{w} : \boldsymbol{w} = \boldsymbol{H}_w(\boldsymbol{\theta})\boldsymbol{d},\ \|\boldsymbol{d}\|_2 \leq 1\}\,. \tag{6}$$

In particular, we note that (6) reduces to the bounded energy constraint $\|\boldsymbol{w}\|_2 \leq 1$ used in (4) if $\boldsymbol{H}_w(\boldsymbol{\theta}) = \boldsymbol{I}$. Other values of $\boldsymbol{H}_w(\boldsymbol{\theta})$ instead allow considering different assumptions on $\boldsymbol{w}$ for what concerns safety and performance, providing extra design flexibility that one can exploit to strike a balance between these two critical – yet often competing – aspects.

B. Linear disturbance feedback policy

In general, it is well-known that optimizing over the function space of feedback policies is computationally intractable. Therefore, as common in the control literature [19], throughout this paper we restrict our attention to linear disturbance feedback policies of the form $\boldsymbol{u} = \boldsymbol{\Phi}_u\boldsymbol{w}$, with $\boldsymbol{\Phi}_u$

lower block-triangular to enforce causality. Note that linear policies attain minimum regret against the optimal sequence of control actions in hindsight if the system dynamics are known and the safety constraints are not active [4], [5].

Let us define through diagonal concatenation of matrices the operators $\boldsymbol{A}(\boldsymbol{\theta}) = \operatorname{blkdiag}(A_0(\theta_0), \dots, A_{T-1}(\theta_{T-1}))$, $\boldsymbol{B}(\boldsymbol{\theta}) = \operatorname{blkdiag}(B_0(\theta_0), \dots, B_{T-1}(\theta_{T-1}))$, and $\boldsymbol{E}(\boldsymbol{\theta}) = \operatorname{blkdiag}(I_n, E_0(\theta_0), \dots, E_{T-2}(\theta_{T-2}))$. Further, let $\boldsymbol{Z}$ denote the block-downshift operator, namely, a matrix with identity matrices along its first block sub-diagonal and zeros elsewhere. With this notation in place, we observe that the closed-loop state trajectory under the feedback law $\boldsymbol{u} = \boldsymbol{\Phi}_u \boldsymbol{w}$ can be expressed as a linear function of $\boldsymbol{w}$ as per:

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{Z}\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{x} + \boldsymbol{Z}\boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{u} + \boldsymbol{E}(\boldsymbol{\theta})\boldsymbol{w}\,, \qquad (7)\\ &= (\boldsymbol{I} - \boldsymbol{Z}\boldsymbol{A}(\boldsymbol{\theta}))^{-1}(\boldsymbol{Z}\boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{\Phi}_u + \boldsymbol{E}(\boldsymbol{\theta}))\boldsymbol{w} := \boldsymbol{\Phi}_x(\boldsymbol{\theta})\boldsymbol{w}\,. \end{aligned}$$

C. On the choice of the clairvoyant benchmark policy

We conclude our problem formulation by commenting on the choice of the clairvoyant benchmark policy $\boldsymbol{\psi}$. Extending ideas from [4], [5] to the case where the model is uncertain, a meaningful objective is that of competing against the best sequence of control actions in hindsight, without imposing any structure on $\boldsymbol{\psi}$. In this case, it can be shown by adapting the derivations of [2], [7] that:

$$\boldsymbol{\psi}(\boldsymbol{w}, \boldsymbol{\theta}) = -(\boldsymbol{R} + \boldsymbol{F}(\boldsymbol{\theta})^\top \boldsymbol{Q}\boldsymbol{F}(\boldsymbol{\theta}))^{-1}\boldsymbol{F}(\boldsymbol{\theta})^\top \boldsymbol{Q}\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{w}\,, \quad (8)$$

where $\boldsymbol{F}(\boldsymbol{\theta}) = (\boldsymbol{I} - \boldsymbol{Z}\boldsymbol{A}(\boldsymbol{\theta}))^{-1}\boldsymbol{Z}\boldsymbol{B}(\boldsymbol{\theta})$ and $\boldsymbol{G}(\boldsymbol{\theta}) = (\boldsymbol{I} - \boldsymbol{Z}\boldsymbol{A}(\boldsymbol{\theta}))^{-1}\boldsymbol{E}(\boldsymbol{\theta})$ are the causal response operators that encode the uncertain dynamics (1) as $\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{\theta})\boldsymbol{u} + \boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{w}$. Differently from the model-based setting considered in [4], [5], however, the (nonlinear) dependence of $\boldsymbol{\psi}$ on the uncertain system dynamics $\boldsymbol{\theta}$ makes it impossible to compute the actual benchmark policy – and hence also the policy that minimizes regret against it. To get around this problem without sacrificing the instance-wise optimality of $\boldsymbol{\psi}$ – as would result, for instance, by constructing a benchmark policy that achieves robust performance across all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ – in the next section we present a randomized approach based on the scenario optimization framework [14], [15].

III. Main Results

In this section, we show how a causal control policy with probabilistic certificates of regret and safety can be efficiently computed in spite of the uncertain dynamics. To do so, we first construct a scenario approximation of the robust regret minimization problem in (4) by restricting our focus to a finite number of uncertainty instances only. Then, inspired by [7], we prove that the policy that safely minimizes regret over the considered scenarios can be expressed as the solution of a semidefinite optimization problem. Finally, leveraging results from the theory of uncertain convex programs [14], [15], we derive strong guarantees on the probability of both out-of-sample regret bound and safety constraint violation. We collect all proofs in our technical report [20].

In what follows and by inspection of (7) and (8), we let $\boldsymbol{\Psi}_u(\boldsymbol{\theta}) = -(\boldsymbol{R} + \boldsymbol{F}(\boldsymbol{\theta})^\top \boldsymbol{Q}\boldsymbol{F}(\boldsymbol{\theta}))^{-1}\boldsymbol{F}(\boldsymbol{\theta})^\top \boldsymbol{Q}\boldsymbol{G}(\boldsymbol{\theta})$ and $\boldsymbol{\Psi}_x(\boldsymbol{\theta}) = \boldsymbol{F}(\boldsymbol{\theta})\boldsymbol{\Psi}_u(\boldsymbol{\theta}) + \boldsymbol{G}(\boldsymbol{\theta})$ denote the closed-loop system responses that map $\boldsymbol{w}$ to the control actions selected by $\boldsymbol{\psi}$ and to the corresponding state trajectory, respectively. Further, with slight abuse of notation, we will often use $\boldsymbol{\Phi}_u$ and $\boldsymbol{\Psi}_u$ instead of $\boldsymbol{\pi}$ and $\boldsymbol{\psi}$, respectively. We start by introducing the following epigraphic form of the robust safe regret minimization problem:

$$\begin{aligned} &\inf_{\boldsymbol{\Phi}_u, \gamma} \ \gamma && (9a)\\ &\text{subject to } \ \boldsymbol{\Phi}_x(\boldsymbol{\theta}) = \boldsymbol{F}(\boldsymbol{\theta})\boldsymbol{\Phi}_u + \boldsymbol{G}(\boldsymbol{\theta})\,, && (9b)\\ &\qquad \max_{\boldsymbol{w} \in \mathcal{W}(\boldsymbol{\theta})} \ \boldsymbol{H}_x(\boldsymbol{\theta})\boldsymbol{\Phi}_x(\boldsymbol{\theta})\boldsymbol{w} + \boldsymbol{H}_u(\boldsymbol{\theta})\boldsymbol{\Phi}_u\boldsymbol{w} \leq \boldsymbol{h}(\boldsymbol{\theta})\,, && (9c)\\ &\qquad \max_{\|\boldsymbol{w}\|_2 \leq 1} \ \mathtt{R}(\boldsymbol{\Phi}_u, \boldsymbol{\Psi}_u(\boldsymbol{\theta}), \boldsymbol{w}, \boldsymbol{\theta}) \leq \gamma\,, \ \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}\,; && (9d) \end{aligned}$$

we denote the optimal value of (9) by $\bar{\mathtt{R}}^\star(\boldsymbol{\Psi}_u(\boldsymbol{\theta}))$. Despite we narrowed attention to linear disturbance feedback policies, (9) remains intractable if $\boldsymbol{\Theta}$ has infinite cardinality. Besides, strong duality results do not apply in a straightforward way as $\boldsymbol{\Theta}$ is not assumed to be connected, let alone convex.

Motivated by the scenario optimization framework [14], [15], we therefore propose replacing the maximization over $\boldsymbol{\Theta}$ with a maximization over the finite set $\mathcal{D} = \{\boldsymbol{\theta}^1, \dots, \boldsymbol{\theta}^N\}$ of randomly sampled uncertainty realizations only. In this way, we approximate (9) with its scenario counterpart:

$$\begin{aligned} &\min_{\boldsymbol{\Phi}_u, \gamma} \ \gamma && (10a)\\ &\text{subject to } \ \boldsymbol{\Phi}_x(\boldsymbol{\theta}^k) = \boldsymbol{F}(\boldsymbol{\theta}^k)\boldsymbol{\Phi}_u + \boldsymbol{G}(\boldsymbol{\theta}^k)\,, && (10b)\\ &\qquad \max_{\boldsymbol{w} \in \mathcal{W}(\boldsymbol{\theta}^k)} \ \boldsymbol{H}_x^k\boldsymbol{\Phi}_x(\boldsymbol{\theta}^k)\boldsymbol{w} + \boldsymbol{H}_u^k\boldsymbol{\Phi}_u\boldsymbol{w} \leq \boldsymbol{h}^k\,, && (10c)\\ &\qquad \max_{\|\boldsymbol{w}\|_2 \leq 1} \ \mathtt{R}(\boldsymbol{\Phi}_u, \boldsymbol{\Psi}_u(\boldsymbol{\theta}^k), \boldsymbol{w}, \boldsymbol{\theta}^k) \leq \gamma\,, \ \forall \boldsymbol{\theta}^k \in \mathcal{D}\,, && (10d) \end{aligned}$$

with $\boldsymbol{H}_x^k = \boldsymbol{H}_x(\boldsymbol{\theta}^k)$, $\boldsymbol{H}_u^k = \boldsymbol{H}_u(\boldsymbol{\theta}^k)$, and $\boldsymbol{h}^k = \boldsymbol{h}(\boldsymbol{\theta}^k)$ for brevity. In particular, note that the infimum in (10a) is attained since only a finite number of uncertainty realizations $\boldsymbol{\theta}^k \in \mathcal{D}$ are considered, and since, for every $\boldsymbol{\theta}^k$, $\boldsymbol{R} \succ 0$ implies that the regret (3) is radially unbounded with respect to $\boldsymbol{\Phi}_u$. Building upon the reformulations proposed in [7], [8] for the case of known system dynamics, the next proposition shows that (10) can be solved by means of standard convex optimization techniques.

Proposition 1: The scenario optimization problem (10) is equivalent to the following semidefinite program:

$$\begin{aligned} &\min_{\boldsymbol{\Phi}_u, \gamma} \ \gamma && (11a)\\ &\text{subject to } \ (10b)\,, \ \forall \boldsymbol{\theta}^k \in \mathcal{D}\,, \ \forall i \in \{1, \dots, S\}\,,\\ &\qquad \left\|(\boldsymbol{H}_x^k\boldsymbol{\Phi}_x(\boldsymbol{\theta}^k) + \boldsymbol{H}_u^k\boldsymbol{\Phi}_u)_i \boldsymbol{H}_w^k\right\|_2 \leq \boldsymbol{h}^k\,, && (11b)\\ &\begin{bmatrix} \boldsymbol{I} & \begin{bmatrix} \boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{\Phi}_x(\boldsymbol{\theta}^k) \\ \boldsymbol{R}^{\frac{1}{2}}\boldsymbol{\Phi}_u \end{bmatrix} \\ \star & \gamma\boldsymbol{I} + \begin{bmatrix} \boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{\Psi}_x(\boldsymbol{\theta}^k) \\ \boldsymbol{R}^{\frac{1}{2}}\boldsymbol{\Psi}_u(\boldsymbol{\theta}^k) \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{\Psi}_x(\boldsymbol{\theta}^k) \\ \boldsymbol{R}^{\frac{1}{2}}\boldsymbol{\Psi}_u(\boldsymbol{\theta}^k) \end{bmatrix} \end{bmatrix} \succeq 0\,, && (11c) \end{aligned}$$

where $\boldsymbol{H}_w^k = \boldsymbol{H}_w(\boldsymbol{\theta}^k)$, S is the number of constraints in (5), and $\star$ denotes entries that can be inferred from symmetry.

We remark that the operators $\boldsymbol{\Psi}_x(\boldsymbol{\theta}^k)$ and $\boldsymbol{\Psi}_u(\boldsymbol{\theta}^k)$ in (11c) are the noncausal system responses associated with a benchmark policy that is optimal for the specific realization $\boldsymbol{\theta}^k$ of the uncertain system parameters. For each $\boldsymbol{\theta}^k \in \mathcal{D}$, enforcing (11c) hence requires to first evaluate the corresponding optimal closed-loop behavior in hindsight using (8). Establishing regret guarantees relative to the clairvoyant optimal policy $\boldsymbol{\Psi}_u(\boldsymbol{\theta}^k)$, which is impossible to compute without knowledge of $\boldsymbol{\theta}^k$, constitutes our main motivation towards adopting sampling techniques in a competitive setting, shedding light on an interesting application of scenario optimization beyond those in stochastic model predictive control [21].

Remark 1: As the number of uncertainty samples in $\mathcal{D}$ increases, solving (11) through semidefinite programming may represent a major computational bottleneck. In Section IV, we numerically show how imposing a Toeplitz block structure on $\boldsymbol{\Phi}_u$ can substantially reduce this computational burden, at the price of an only slight increase in conservativeness in our regret bound. We refer the accompanying technical report [20] for more detailed discussion on scalability.

Let $\boldsymbol{\Phi}_u^\star(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \mathcal{D})$ and $\bar{\mathtt{R}}_N^\star(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \mathcal{D})$ denote the optimal policy and the optimal value of (10), respectively. Since only a finite subset of the constraints of (9) are considered in (10), we have that $\bar{\mathtt{R}}_N^\star \leq \bar{\mathtt{R}}^\star$, that is, $\bar{\mathtt{R}}_N^\star$ is an optimistic lower bound on the true minimax regret $\bar{\mathtt{R}}^\star$. Conversely, thanks to Proposition 1 and exploiting key results in scenario optimization, we now show that the solution of (10) is approximately feasible for (9) – in the sense that the measure of the set of original constraints that it violates rapidly approaches zero as N increases. Before formalizing this generalization property in the theorem below, we observe that multiple optimal policies for (11) may exist, since the function $\lambda_{\max}(\cdot)$ is not strongly convex. In this case, uniqueness of $\boldsymbol{\Phi}_u^\star(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \mathcal{D})$ can be enforced by designing a convex tie-break rule, e.g., a lexicographic criterion [22]. Conversely, if the safety constraints (10c) are overly restrictive, the scenario problem (11) may become infeasible; if this were the case, however, the original problem (9) would also certainly be infeasible, and one would need to consider broader classes of policies, or to relax the safety requirements, e.g., by introducing slack variables in (10c).

Theorem 1: Fix any violation and confidence levels, say ϵ and β, in the open interval $(0,1)$, and let δ and $\mathbb{P}_{\boldsymbol{\theta}}^N$ denote the number of optimization variables in (10) and the N-fold product distribution $\mathbb{P}_{\boldsymbol{\theta}} \times \cdots \times \mathbb{P}_{\boldsymbol{\theta}}$ with N terms, respectively. If the scenario optimization problem (10) is feasible and $N > \delta$ satisfies $\sum_{j=0}^{\delta-1} \binom{N}{j} \epsilon^j (1-\epsilon)^{N-j} \leq \beta$, then, with probability of at least $1-\beta$ given a dataset $\mathcal{D} \sim \mathbb{P}_{\boldsymbol{\theta}}^N$, it holds that:

$$\mathbb{P}_{\boldsymbol{\theta}}\Big(\max_{\|\boldsymbol{w}\|_2 \leq 1} \mathtt{R}(\boldsymbol{\Phi}_u^\star, \boldsymbol{\Psi}_u(\boldsymbol{\theta}), \boldsymbol{w}, \boldsymbol{\theta}) \leq \bar{\mathtt{R}}_N^\star\,, \\ \text{and } (\boldsymbol{x}, \boldsymbol{u}) \in \mathcal{S}(\boldsymbol{\theta})\,,\ \forall \boldsymbol{w} \in \mathcal{W}(\boldsymbol{\theta})\Big) \geq 1 - \epsilon\,. \quad (12)$$

Theorem 1 presents an explicit sample complexity bound that, given a priori specified ϵ and β, ensures that the safety and regret guarantees extend to all but at most a fraction ϵ of unseen dynamics $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ with probability $1-\beta$. As well-known in the literature on scenario optimization, the minimum number of scenarios $N(\epsilon, \beta)$ required to fulfill the conditions of Theorem 1 grows linearly with ϵ^{-1}, yet at most logarithmically with β^{-1}. Hence, even if a very small β is selected – so that (12) holds with practical certainty – the number of scenarios to be sampled remains manageable, see also [21]. Further, we note that the condition on N given in Theorem 1 is tight for fully-supported problems [15]; a simpler, albeit not tight, sufficient condition on N is [14]:

$$N \geq 2\epsilon^{-1}(\delta + \log(\beta^{-1}))\,. \quad (13)$$

A. Comparison with worst-case oriented synthesis

Our main motivation towards embracing a scenario perspective is that randomized approaches allow us to explicitly compute $\psi(\boldsymbol{w}, \boldsymbol{\theta})$ by replacing the uncertain system dynamics with their sampled counterparts. Regret bounds relative to the instance-wise optimal benchmark $\psi(\boldsymbol{w}, \boldsymbol{\theta})$ are attractive, as they yield upper bounds on the closed-loop cost that adapt to the realized dynamics $\boldsymbol{\theta}$ and perturbation $\boldsymbol{w}$. To illustrate this point more thoroughly, let us consider an alternative design based on a classical worst-case $\mathcal{H}_\infty$ objective:

$$\begin{aligned} \{\boldsymbol{\Phi}_{u,\mathtt{H}}^\star,\ \bar{\mathtt{H}}_N^\star\} = \ & \underset{\boldsymbol{\Phi}_u, \gamma}{\arg\min}\ \gamma \\ & \text{subject to (10b), (10c)}\,, \\ & \max_{\|\boldsymbol{w}\|_2 \leq 1} J(\boldsymbol{\Phi}_u, \boldsymbol{w}, \boldsymbol{\theta}^k) \leq \gamma\,,\ \forall \boldsymbol{\theta}^k \in \mathcal{D}\,. \end{aligned} \quad (14)$$

Leaving safety concerns aside to ease the discussion, the control policies $\boldsymbol{\Phi}_u^\star$ and $\boldsymbol{\Phi}_{u,\mathtt{H}}^\star$ offer the following probabilistic performance guarantees:

$$J(\boldsymbol{\Phi}_u^\star, \boldsymbol{w}, \boldsymbol{\theta}) - J(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \boldsymbol{w}, \boldsymbol{\theta}) \leq \bar{\mathtt{R}}_N^\star\,, \quad (15)$$

$$J(\boldsymbol{\Phi}_{u,\mathtt{H}}^\star, \boldsymbol{w}, \boldsymbol{\theta}) \leq \bar{\mathtt{H}}_N^\star\,, \quad (16)$$

for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and any $\boldsymbol{w}$ with $\|\boldsymbol{w}\|_2 \leq 1$. In particular, while the $\mathcal{H}_\infty$ solution provides a single pessimistic upper bound on the closed-loop cost as per (16), our regret optimal policy gives a non-uniform certificate shaped by $J(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \boldsymbol{w}, \boldsymbol{\theta})$ as per (15). Moreover, our upper bound on $J(\boldsymbol{\Phi}_u^\star, \boldsymbol{w}, \boldsymbol{\theta})$ is tighter than that on $J(\boldsymbol{\Phi}_{u,\mathtt{H}}^\star, \boldsymbol{w}, \boldsymbol{\theta})$ whenever

$$J(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \boldsymbol{w}, \boldsymbol{\theta}) \leq \bar{\mathtt{H}}_N^\star - \bar{\mathtt{R}}_N^\star\,. \quad (17)$$

As we will numerically show in the next section, (17) not only holds consistently over several classes of disturbances, but this tighter guarantee in terms of upper bounds often translates into improved performance, that is, $J(\boldsymbol{\Phi}_u^\star, \boldsymbol{w}, \boldsymbol{\theta}) \leq J(\boldsymbol{\Phi}_{u,\mathtt{H}}^\star, \boldsymbol{w}, \boldsymbol{\theta})$, no matter which $\boldsymbol{\theta}$ realizes. In this sense, regret minimization can alleviate the conservatism of (14).

IV. Numerical Results

In this section, we first validate numerically the probabilistic regret guarantee we have established in Theorem 1, and we then show how this guarantee allows improving the overall closed-loop performance in face of the uncertain system dynamics. For our experiments, we consider a discrete-time stochastic mass-spring-damper system described by the uncertain linear dynamics:

$$x_{t+1} = \begin{bmatrix} 1 & T_s \\ -\frac{(k+\delta_k)T_s}{m} & 1 - \frac{(c+\delta_c)T_s}{m} \end{bmatrix} x_t + \begin{bmatrix} 0 \\ \frac{T_s}{m} \end{bmatrix} u_t + w_t\,,$$

with mass $m = 1\,\mathrm{kg}$, nominal spring and damping constants $k = 1\,\mathrm{N\,m^{-1}}$ and $c = 1\,\mathrm{N\,m^{-1}\,s}$, respectively, and sampling time $T_s = 1\,\mathrm{s}$. This simple model is often used to describe the behavior of several physical systems, including vibrating structures, suspension systems, and mechanical oscillators; the uncertain parameters $\theta = \begin{bmatrix}\delta_k & \delta_c\end{bmatrix}^\top$ can thus model deviations from the nominal parameters arising in the mass production process of these devices. We assume that θ is constant over the control horizon $T = 20$, and that it is uniformly distributed, i.e., $\delta_k \sim \mathcal{U}_{[-0.2,0.2]}$ and $\delta_c \sim \mathcal{U}_{[-0.2,0.2]}$. We define the control cost (2) by letting $\boldsymbol{Q} = \boldsymbol{I}_{20} \otimes \boldsymbol{I}_2$ and $\boldsymbol{R} = \boldsymbol{I}_{20}$, where $\otimes$ denotes the Kronecker product. For simplicity and to focus on the advantages brought about by regret minimization, we assume that no safety constraints are imposed on the system.

To corroborate our main theoretical result in Theorem 1, we repeatedly solve (11), each time considering a dataset $\mathcal{D}_i$ with an increasing number N_i of training scenarios. In particular, for each $\boldsymbol{\theta}^k \in \mathcal{D}_i$, we use (8) to compute $\boldsymbol{\Psi}_u(\boldsymbol{\theta}^k)$ as the closed-loop map associated with the unconstrained optimal policy in hindsight; according to (7), we obtain the corresponding $\boldsymbol{\Psi}_x(\boldsymbol{\theta}^k)$ by $\boldsymbol{F}(\boldsymbol{\theta}^k)\boldsymbol{\Psi}_u(\boldsymbol{\theta}^k) + \boldsymbol{G}(\boldsymbol{\theta}^k)$. Then, given a set of 10000 independently sampled uncertainty instances for validation, we estimate the probability in (12) by recording how often the optimal policy $\boldsymbol{\Phi}^\star_u(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \mathcal{D}_i)$ fails to comply with the associated regret bound $\bar{\mathtt{R}}^\star_{N_i}(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \mathcal{D}_i)$. To showcase the effect of time-invariant controller structure discussed in Remark 1, we repeat these experiments while including in (11) the additional constraint that the solution $\widehat{\boldsymbol{\Phi}}^\star_u(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \mathcal{D}_i)$ has constant block diagonal terms. We denote the regret bound associated to $\widehat{\boldsymbol{\Phi}}^\star_u(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \mathcal{D}_i)$ by $\hat{\mathtt{R}}^\star_{N_i}(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \mathcal{D}_i) \geq \bar{\mathtt{R}}^\star_{N_i}(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \mathcal{D}_i)$. In Figure 1, we plot the evolution of the empirical violation probabilities $V(\boldsymbol{\Phi}^\star_u, \boldsymbol{\Psi}_u(\boldsymbol{\theta}), \mathcal{D}_i) := \bar{V}_N$ and $V(\widehat{\boldsymbol{\Phi}}^\star_u, \boldsymbol{\Psi}_u(\boldsymbol{\theta}), \mathcal{D}_i) := \widehat{V}_N$ associated with $\boldsymbol{\Phi}^\star_u$ and $\widehat{\boldsymbol{\Phi}}^\star_u$, respectively, as a function of the dataset size.[2] For completeness, we also display the (non-tight) theoretical upper bounds on the violation probability ϵ given by (13) for $\beta = 0.1$. In Figure 2, we compare the regret certificates $\bar{\mathtt{R}}^\star_N$ and $\hat{\mathtt{R}}^\star_N$ provided by the control policies $\boldsymbol{\Phi}^\star_u$ and $\widehat{\boldsymbol{\Phi}}^\star_u$, respectively, as well as the computation times $\bar{\tau}_N$ and $\hat{\tau}_N$ required to evaluate them via semidefinite optimization.[3] Besides validating our theoretical results, these figures allow us to draw the following observations. First, the approximate solution $\widehat{\boldsymbol{\Phi}}^\star_u$ with constant block diagonal terms guarantees regret at most 9% higher than $\boldsymbol{\Phi}^\star_u$ with high probability, yet its evaluation requires a computation time $\hat{\tau}_N$ that is lower than $\bar{\tau}_N$ by an entire order of magnitude. Second, consistently with the intuition that simpler models are less prone to overfit, we observe that $\widehat{\boldsymbol{\Phi}}^\star_u$ achieves better generalization than $\boldsymbol{\Phi}^\star_u$, as the out-of-sample empirical violation probability $\widehat{V}_N$ is consistently smaller than $\bar{V}_N$. Third, the quantities $\widehat{V}_N$ and $\hat{\mathtt{R}}^\star_N$ rapidly converge to their corresponding limit values as N increases, suggesting that the minimax solution to (9) could be practically approximated by sampling a limited number of uncertainty instances only. Motivated by these considerations and with the aim of further reducing the computational complexity of our scheme, we plan to study the possible application of wait-and-judge [22] and constraint removal [23] approaches in future work.

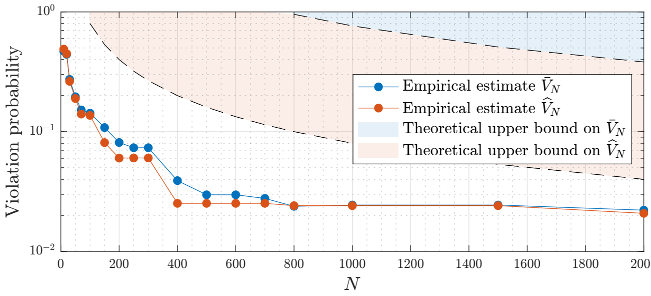


Fig. 1. Comparison between empirical regret violation probability and theoretical upper bound as a function of the number of sampled scenarios.

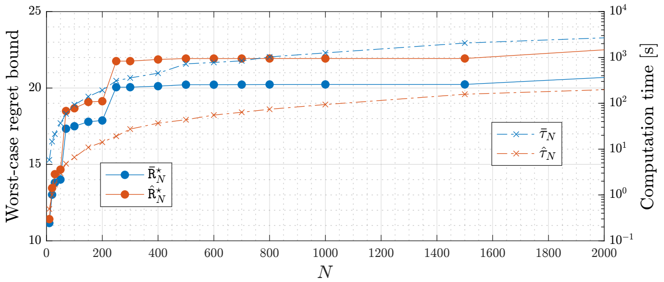


Fig. 2. Evolution of the probabilistic worst-case regret bounds (denoted by $\bar{\mathtt{R}}^\star_N$ and $\hat{\mathtt{R}}^\star_N$ on the left y-axis) and of the computation times (denoted by $\bar{\tau}_N$ and $\hat{\tau}_N$ on the right y-axis) for the exact and approximate solutions of (11), respectively, as a function of the number of considered scenarios.

Next, to illustrate the potential of our method, we compare the performance of the policies $\boldsymbol{\pi}_{\mathtt{R}}$ and $\boldsymbol{\pi}_{\mathtt{H}}$ computed solving (11) and (14), respectively, using $N = 5000$ random samples of δ_k and δ_c. For several classes of disturbances $\boldsymbol{w}$ often encountered in practice, we evaluate the control costs $J(\boldsymbol{\pi}_{\mathtt{R}}, \boldsymbol{w}, \boldsymbol{\theta})$, $J(\boldsymbol{\pi}_{\mathtt{H}}, \boldsymbol{w}, \boldsymbol{\theta})$ and $J(\boldsymbol{\psi}, \boldsymbol{w}, \boldsymbol{\theta})$ for 20 different values of $\boldsymbol{\theta}$. In Figure 3a, we plot $J(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \boldsymbol{w}, \boldsymbol{\theta})$ and compare it with $\bar{\mathtt{H}}^\star_N - \bar{\mathtt{R}}^\star_N$ to verify, according to (17), when (15) yields tighter upper bounds than (16) on the realized performance. In Figure 3b, we instead display the percentage increase in the cost due to using $\boldsymbol{\pi}_{\mathtt{H}}$ instead of $\boldsymbol{\pi}_{\mathtt{R}}$, that is,[4]

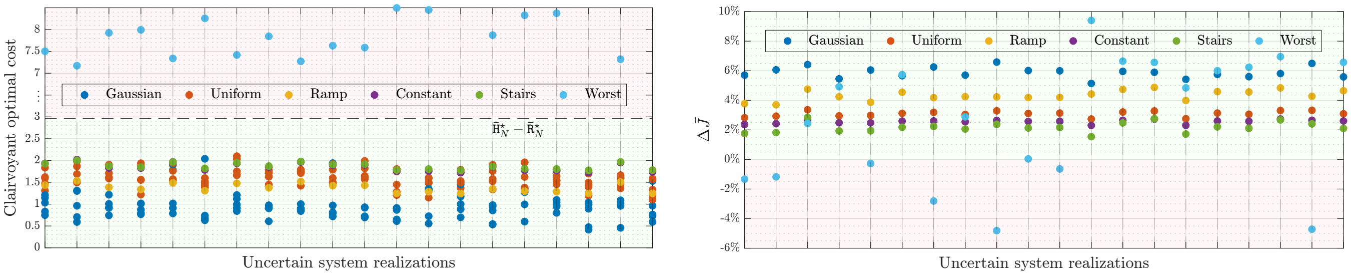
$$\Delta \bar{J}(\boldsymbol{w}, \boldsymbol{\theta}) = \frac{J(\boldsymbol{\pi}_{\mathtt{H}}, \boldsymbol{w}, \boldsymbol{\theta}) - J(\boldsymbol{\pi}_{\mathtt{R}}, \boldsymbol{w}, \boldsymbol{\theta})}{J(\boldsymbol{\pi}_{\mathtt{R}}, \boldsymbol{w}, \boldsymbol{\theta})} := \Delta \bar{J}\,.$$

As already observed in previous work for perfectly known systems [4], [5], [7], Figure 3 shows that regret minimization constitutes a viable control design strategy for improving the closed-loop performance when the disturbances do not match classical design assumptions – in terms of both lower upper bounds (Figure 3a) and lower realized costs

[2] The source code that reproduces our numerical examples is available at https://github.com/DecodEPFL/ScenarioSafeMinRegret.

[3] All optimization problems have been solved using MOSEK on a standard laptop computer with a 2.3 GHz Intel Core i9 CPU.

[4] For stochastic disturbances, results are averaged over 10^4 realizations.



(a) Control cost $J(\boldsymbol{\Psi}_u(\boldsymbol{\theta}), \boldsymbol{w}, \boldsymbol{\theta})$ incurred by the clairvoyant optimal policy.

(b) Average percentage increase in the cost incurred by $\boldsymbol{\pi}_{\mathrm{H}}$ relative to $\boldsymbol{\pi}_{\mathrm{R}}$.

Fig. 3. Closed-loop comparison between $\boldsymbol{\pi}_{\mathrm{H}}$ and our $\boldsymbol{\pi}_{\mathrm{R}}$: a priori performance guarantees and realized control cost for different disturbance profiles and different realizations of the uncertain system dynamics. Points in the green shaded area denote instances where the proposed regret minimization approach yields an advantage in terms of lower upper bound (Figure 3a) and realized performance (Figure 3b). We refer to our source code for a precise definition of the considered disturbance profiles.

(Figure 3b). Most importantly, our results show that regret optimal policies continue to offer these performance advantages consistently in face of the uncertain dynamics. Interestingly, we further observe that the policy $\boldsymbol{\pi}_{\mathrm{R}}$ often outperforms $\boldsymbol{\pi}_{\mathrm{H}}$ even for the worst-case disturbance $\boldsymbol{w}$. While this may seem counterintuitive, we note that $\boldsymbol{\pi}_{\mathrm{H}}$ ensures minimum cost on a single pair of worst-case disturbances and parameters $(\boldsymbol{w}_{\text{worst}}, \boldsymbol{\theta}_{\text{worst}})$ only. Conversely, for randomly sampled instances of the uncertain parameters $\boldsymbol{\theta} \neq \boldsymbol{\theta}_{\text{worst}}$, the policy $\boldsymbol{\pi}_{\mathrm{H}}$ retains no optimality guarantee on the cost that it incurs under the most averse perturbation $\boldsymbol{w}$ for that $\boldsymbol{\theta}$.

V. Conclusion

We have presented a novel method for convex synthesis of robust control policies with provable regret and safety guarantees in face of the uncertain stochastic dynamics. As the clairvoyant optimal policy we compete against is unknown in this setting, we have proposed sampling the space of parameters that characterize the system dynamics. Leveraging results from the theory of scenario optimization, we have shown that the policy that minimizes regret robustly over these randomly drawn uncertainty instances retains strong probabilistic out-of-samples guarantees. Finally, we have presented numerical experiments to corroborate our theoretical results, and to highlight the potential of regret minimization in adapting to heterogeneous dynamics and disturbance sequences. Interesting directions for future research encompass studying infinite-horizon control problems, addressing computational complexity challenges for real-time implementation, and extending the theory of this emerging competitive framework to systems with nonlinear dynamics.

References

[1] E. Hazan and K. Singh, "Introduction to online nonstochastic control," *arXiv preprint arXiv:2211.09619*, 2022.

[2] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite-quadratic estimation and control: a unified approach to $\mathcal{H}_2$ and $\mathcal{H}_\infty$ theories*. SIAM, 1999.

[3] N. Agarwal, B. Bullins, E. Hazan, S. Kakade, and K. Singh, "Online control with adversarial disturbances," in *International Conference on Machine Learning*. PMLR, 2019, pp. 111–119.

[4] O. Sabag, G. Goel, S. Lale, and B. Hassibi, "Regret-optimal controller for the full-information problem," in *2021 American Control Conference (ACC)*. IEEE, 2021, pp. 4777–4782.

[5] G. Goel and B. Hassibi, "Regret-optimal estimation and control," *IEEE Transactions on Automatic Control*, vol. 68, no. 5, pp. 3041–3053, 2023.

[6] O. Sabag, S. Lale, and B. Hassibi, "Optimal competitive-ratio control," *arXiv preprint arXiv:2206.01782*, 2022.

[7] A. Martin, L. Furieri, F. Dörfler, J. Lygeros, and G. Ferrari-Trecate, "Safe control with minimal regret," in *Learning for Dynamics and Control Conference*. PMLR, 2022, pp. 726–738.

[8] A. Didier, J. Sieber, and M. N. Zeilinger, "A system level approach to regret optimal control," *IEEE Control Systems Letters*, vol. 6, pp. 2792–2797, 2022.

[9] A. Martin, L. Furieri, F. Dörfler, J. Lygeros, and G. Ferrari-Trecate, "On the guarantees of minimizing regret in receding horizon," *arXiv preprint arXiv:2306.14561*, 2023.

[10] G. Goel and B. Hassibi, "Competitive control," *IEEE Transactions on Automatic Control*, vol. 68, no. 9, pp. 5162–5173, 2023.

[11] J.-S. Brouillon, F. Dörfler, and G. F. Trecate, "Minimal regret state estimation of time-varying systems," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 2595–2600, 2023.

[12] A. Martin, L. Furieri, F. Dörfler, J. Lygeros, and G. Ferrari-Trecate, "Follow the clairvoyant: an imitation learning approach to optimal control," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 2589–2594, 2023.

[13] G. Goel, N. Agarwal, K. Singh, and E. Hazan, "Best of both worlds in online control: Competitive ratio and policy regret," in *Learning for Dynamics and Control Conference*. PMLR, 2023, pp. 1345–1356.

[14] G. C. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE Transactions on automatic control*, vol. 51, no. 5, pp. 742–753, 2006.

[15] M. C. Campi and S. Garatti, "The exact feasibility of randomized solutions of uncertain convex programs," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1211–1230, 2008.

[16] R. Tóth, *Modeling and identification of linear parameter-varying systems*. Springer, 2010, vol. 403.

[17] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of the linear quadratic regulator," *Foundations of Computational Mathematics*, vol. 20, no. 4, pp. 633–679, 2020.

[18] L. Furieri, B. Guo, A. Martin, and G. Ferrari-Trecate, "Near-optimal design of safe output-feedback controllers from noisy data," *IEEE Transactions on Automatic Control*, vol. 68, no. 5, pp. 2699–2714, 2023.

[19] P. J. Goulart, E. C. Kerrigan, and J. M. Maciejowski, "Optimization over state feedback policies for robust control with constraints," *Automatica*, vol. 42, no. 4, pp. 523–533, 2006.

[20] A. Martin, L. Furieri, F. Dörfler, J. Lygeros, and G. Ferrari-Trecate, "Regret optimal control for uncertain stochastic systems," *arXiv preprint arXiv:2304.14835*, 2023.

[21] G. C. Calafiore and L. Fagiano, "Robust model predictive control via scenario optimization," *IEEE Transactions on Automatic Control*, vol. 58, no. 1, pp. 219–224, 2012.

[22] M. C. Campi and S. Garatti, "Wait-and-judge scenario optimization," *Mathematical Programming*, vol. 167, pp. 155–189, 2018.

[23] ——, "A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality," *Journal of optimization theory and applications*, vol. 148, no. 2, pp. 257–280, 2011.