

Safe Control with Minimal Regret

Andrea Martin

ANDREA.MARTIN@EPFL.CH

Institute of Mechanical Engineering, École Polytechnique Fédérale de Lausanne, Switzerland

Luca Furieri

LUCA.FURIERI@EPFL.CH

Institute of Mechanical Engineering, École Polytechnique Fédérale de Lausanne, Switzerland

Florian Dörfler

DORFLER@ETHZ.CH

Department of Information Technology and Electrical Engineering, ETH Zürich, Switzerland

John Lygeros

JLYGEROS@ETHZ.CH

Department of Information Technology and Electrical Engineering, ETH Zürich, Switzerland

Giancarlo Ferrari-Trecate

GIANCARLO.FERRARITRECATI@EPFL.CH

Institute of Mechanical Engineering, École Polytechnique Fédérale de Lausanne, Switzerland

Editors: R. Firoozi, N. Mehr, E. Yel, R. Antonova, J. Bohg, M. Schwager, M. Kochenderfer

Abstract

As we move towards safety-critical cyber-physical systems that operate in non-stationary and uncertain environments, it becomes crucial to close the gap between classical optimal control algorithms and adaptive learning-based methods. In this paper, we present an efficient optimization-based approach for computing a finite-horizon robustly safe control policy that minimizes dynamic regret, in the sense of the loss relative to the optimal sequence of control actions selected in hindsight by a clairvoyant controller. By leveraging the system level synthesis framework (SLS), our method extends recent results on regret minimization for the linear quadratic regulator to optimal control subject to hard safety constraints, and allows competing against a safety-aware clairvoyant policy with minor modifications. Numerical experiments confirm superior performance with respect to finite-horizon constrained \mathcal{H}_2 and \mathcal{H}_∞ control laws when the disturbance realizations poorly fit classical assumptions.

Keywords: safe adaptive control, dynamic regret, system level synthesis

1. Introduction

Classical \mathcal{H}_2 and \mathcal{H}_∞ control theories have studied how to optimally regulate the behavior of a linear dynamical system driven by a stochastic or worst-case disturbance process, respectively. In the \mathcal{H}_2 setting, control actions are chosen so to minimize the expected control cost incurred by the learner. Conversely, in the \mathcal{H}_∞ setting, the agent minimizes the worst-case control cost across all disturbance realizations (Zhou and Doyle, 1998). Both paradigms commit to a control strategy that is tailored to the presumed nature of the disturbance process, and that is blindly pursued regardless of the observed perturbation sequence. Hence, these control policies may suffer high cumulative costs if the disturbance realizations do not match the class of perturbations hypothesized a priori.

Classical literature on online learning has studied sequential decision-making algorithms that learn from experience, as repeated interactions with a *memoryless* environment reveal more aspects of the problem at hand. In this setting, agents make no assumptions about the benign or adversarial nature of the environment, and they dynamically adjust their strategy based on information deduced

from previous rounds to minimize regret (Shalev-Shwartz et al., 2011; Hazan, 2016). Informally, one can think of regret as measuring the loss suffered by a learner relative to the optimal policy in hindsight.

When the environment evolves dynamically, online learning comes with significant new challenges, since the losses agents observe depend not only on their current action, but also on their past decisions. Recent years have witnessed an increasing interest in applying modern statistical and algorithmic techniques to classical control problems. Initiated by Abbasi-Yadkori and Szepesvári (2011), several works have approached the problem of adaptively controlling a linear dynamical system perturbed by a stochastic (Dean et al., 2018; Cohen et al., 2019; Lale et al., 2020) or adversarial (Agarwal et al., 2019; Foster and Simchowitz, 2020; Hazan et al., 2020; Simchowitz et al., 2020) disturbance process from the perspective of *policy regret* minimization. In this framework, control algorithms are designed to compete with the best *static* policy selected in hindsight from a parametric class, and performance guarantees are expressed in terms of sublinear regret bounds against this idealized benchmark. Intuitively, attaining sublinear regret is desirable, as it implies that the average difference between the cost suffered by the learning algorithm and that of the best fixed a posteriori strategy converges to zero over time.

When the online learner interacts with a dynamic environment, policy regret minimization algorithms may yield loose performance certificates (Hazan and Seshadhri, 2009). Indeed, while these methods effectively approach the best fixed strategy in the chosen comparator class, this static benchmark could also incur a high control cost. For instance, no single state-feedback controller can perform well in a scenario where disturbances alternate between being drawn according to a well-behaved stochastic process and being chosen adversarially (Goel and Hassibi, 2020b).

This consideration motivates the design of algorithms that compete against the *clairvoyant* control law that selects the globally optimal dynamic sequence of control actions in hindsight (Goel and Hassibi, 2020a). This concept of *dynamic regret*, which was first introduced in Zinkevich (2003), has recently been studied in the full-information setting in Goel and Hassibi (2020b); Sabag et al. (2021), and in the measurement-feedback setting in Goel and Hassibi (2021). Assuming knowledge of the underlying dynamics and control costs, these works have not only provided tight dynamic regret bounds in terms of the energy of the disturbance sequence,¹ but have also computed a control law that exactly minimizes the worst-case loss relative to the clairvoyant optimal policy; specifically, regret-optimal controllers have been explicitly derived, both in state-space and input-output form, via elegant reductions to classical \mathcal{H}_∞ synthesis and Nehari noncausal approximation problems (Nehari, 1957).

Dynamic regret minimization algorithms, which combine design criteria from online learning theory with continuous action spaces typical of control, have further been shown to nicely interpolate between the performance of classical \mathcal{H}_2 and \mathcal{H}_∞ controllers across both stochastic and adversarial environments (Goel and Hassibi, 2020b; Sabag et al., 2021). However, due to the lack of provable robustness guarantees, these control policies do not lend themselves to real-time applications that impose hard safety constraints on the physical variables of the system. Despite being critical to reliably deploy learning-based methods in modern engineering systems, the problem of designing safety-aware algorithms with non-asymptotic performance guarantees has been only recently approached in Nonhoff and Müller (2021), which assumes that the underlying linear dynamics are not affected by disturbances, and in Li et al. (2020, 2021), albeit from the standpoint

1. Dynamic regret bounds are often expressed in terms of some “regularity” of the perturbation sequence since dynamic regret scales linearly with time in the worst-case (Jadbabaie et al., 2015; Goel and Hassibi, 2020a; Zhao et al., 2020).

of policy regret minimization. To the best of our knowledge, the literature offers no characterization of a control policy that simultaneously attains minimum dynamic regret and guarantees compliance with hard safety constraints in face of the uncertain disturbance realizations.

In this paper, we present an efficient optimization-based approach for computing a finite-horizon safe regret-optimal control policy that exactly minimizes the loss relative to the clairvoyant optimal controller, while satisfying safety constraints defined over physical variables of the system. To do so, we leverage the recent system level parametrization (SLP) of linear dynamic controllers (Wang et al., 2019) as well as classical duality results from robust optimization. Compared to previous works (Goel and Hassibi, 2020b; Sabag et al., 2021), we directly treat the system *closed-loop responses* as design variables, and we reduce the safe controller synthesis task to solving a semidefinite program (SDP). The proposed method circumvents the need to repeatedly appeal to the whitening property of Kalman filters for calculating explicit causal matrix factorizations (Kailath et al., 2000), and it allows a one-shot computation of a safe regret-optimal controller, without requiring to iteratively solve suboptimal instances of the problem. As such, we believe the flexibility provided by this optimization perspective can pave the way towards analyzing more complex control problems through the lens of regret minimization. Our approach also permits to naturally capture the time-varying nature of the system dynamics, the control costs, and the safety constraints. We present numerical experiments to support using dynamic regret as control design criterion.

2. Problem Statement and Preliminaries

We consider known discrete-time linear time-varying dynamical systems with state-space equations

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

where $u_t \in \mathbb{R}^m$, $x_t \in \mathbb{R}^n$ and $w_t \in \mathbb{R}^n$ are the control input, the system state, and an exogenous disturbance, respectively. We do not make any assumptions about the statistical distribution of the disturbance process, which we allow to be of adversarial nature.

We study the evolution over a horizon $T \in \mathbb{N}$ of system (1), initialized at $x_0 \in \mathbb{R}^n$, when the control actions u_t are computed according to a time-varying linear feedback control policy

$$u_t = \sum_{k=0}^t K_{t,k} x_k, \quad \forall t \in \{0 \dots T-1\}. \quad (2)$$

For convenience, we compactly write signals and causal operators over T as

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{T-1} \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{T-1} \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} x_0 \\ w_0 \\ \vdots \\ w_{T-2} \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} K_{0,0} & 0_{m \times n} & \cdots & 0_{m \times n} \\ K_{1,0} & K_{1,1} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0_{m \times n} \\ K_{T-1,0} & K_{T-1,1} & \cdots & K_{T-1,T-1} \end{bmatrix},$$

and we denote the online cost incurred applying the control sequence \mathbf{u} in response to the disturbance realization \mathbf{w} by

$$\text{cost}(\mathbf{w}, \mathbf{u}) = \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix}^\top \begin{bmatrix} \mathcal{Q} & 0 \\ 0 & \mathcal{R} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix}, \quad (3)$$

where $\mathcal{Q} \succeq 0$ and $\mathcal{R} \succ 0$ represent the cost matrices coupling state and input signals at possibly different time instants, respectively.

Our goal is to synthesize a safe control policy in the form of (2) that exactly minimizes the loss relative to the clairvoyant policy, i.e., the offline control law that selects the globally optimal sequence of control actions with complete foreknowledge of the disturbance realizations. In other words, we wish to minimize the worst-case dynamic regret

$$\text{regret}(T, \mathbf{u}) = \max_{\|\mathbf{w}\|_2 \leq 1} \left(\text{cost}(\mathbf{w}, \mathbf{u}) - \min_{\mathbf{u}' \in \mathbb{R}^{mT}} \text{cost}(\mathbf{w}, \mathbf{u}') \right), \quad (4)$$

while complying with polytopic safety constraints defined over the system states and inputs as per

$$\mathcal{H} [\mathbf{x}^\top \quad \mathbf{u}^\top]^\top \leq \mathbf{h}, \quad \forall \mathbf{w} \in \mathbb{W} = \{\mathbf{w}' \in \mathbb{R}^{nT} : \mathcal{H}_w \mathbf{w}' \leq \mathbf{h}_w\}, \quad (5)$$

where \mathbb{W} is a compact polytope that contains an open neighborhood of the origin. For the rest of the paper, inequalities involving vectors are to be intended element-wise. As it is standard in the robust control literature (Rawlings et al., 2017), we formulate the following feasibility assumption:

Assumption 1 There exists a causal controller \mathbf{K} satisfying (5) for all perturbations $\mathbf{w} \in \mathbb{W}$.

By competing against the clairvoyant policy in (4), we shape the worst-case gain from the disturbance energy to the regret through system-dependent optimal performance weights. Ultimately, this leads to synthesizing control laws that are adaptive, in the sense that they strive to minimize the incurred control cost independently of how the disturbances are generated.

Remark 1 We rely on closed-loop control policies of the form (2) to dynamically adjust to the per-instance disturbance realizations. This is beneficial not only to track the performance of the clairvoyant controller, but also to fulfill the safety constraints (5). Selecting a priori the control actions to be applied, based on the open-loop prediction of the system evolution only, may lead to excessive conservatism, or even infeasibility (Bemporad, 1998).

2.1. System Level Synthesis

We briefly outline the necessary background on the SLS approach to optimal controller synthesis, and we refer to Wang et al. (2019) and Anderson et al. (2019) for a complete discussion. Akin to Youla-based and disturbance-feedback controllers (Youla et al., 1976; Goulart, 2007), the SLS approach shifts the synthesis problem from directly designing the controller to shaping the closed-loop maps from the exogenous disturbance to the state and input signals (Furieri et al., 2019; Zheng et al., 2020).

Let Z be the block-downshift operator, namely a matrix with identity matrices along its first block sub-diagonal and zeros elsewhere, and define $\mathcal{A} = \text{blkdiag}(A_0, A_1, \dots, A_{T-2}, 0_{n \times n})$ and $\mathcal{B} = \text{blkdiag}(B_0, B_1, \dots, B_{T-2}, 0_{n \times m})$. Then, the evolution of the input and state trajectories of system (1) over the control horizon can be compactly expressed as

$$\mathbf{x} = Z\mathcal{A}\mathbf{x} + Z\mathcal{B}\mathbf{u} + \mathbf{w}.$$

Similarly, the closed-loop system behavior under the feedback law $\mathbf{u} = \mathbf{K}\mathbf{x}$ can be described, as a function of the perturbation vector \mathbf{w} , through the non-convex relations

$$\begin{aligned} \mathbf{x} &= (I - Z(\mathcal{A} + \mathcal{B}\mathbf{K}))^{-1} \mathbf{w} = \Phi_x \mathbf{w}, \\ \mathbf{u} &= \mathbf{K}(I - Z(\mathcal{A} + \mathcal{B}\mathbf{K}))^{-1} \mathbf{w} = \Phi_u \mathbf{w}. \end{aligned} \quad (6)$$

Let $\Phi_x = (I - Z(\mathcal{A} + \mathcal{B}\mathbf{K}))^{-1}$ and $\Phi_u = \mathbf{K}(I - Z(\mathcal{A} + \mathcal{B}\mathbf{K}))^{-1}$ denote the system closed-loop responses induced by the controller \mathbf{K} in (6), and observe that these operators inherit a lower block-diagonal causal structure. One can show that there exists a controller \mathbf{K} such that $\mathbf{x} = \Phi_x \mathbf{w}$ and $\mathbf{u} = \Phi_u \mathbf{w}$ if and only if

$$(I - Z\mathcal{A})\Phi_x - Z\mathcal{B}\Phi_u = I; \quad (7)$$

we call pairs $\{\Phi_x, \Phi_u\}$ that satisfy (7) *achievable*.

Based on (7), many optimal control problems of practical interest can be equivalently posed as an optimization over the convex set of system responses $\{\Phi_x, \Phi_u\}$. For instance, one can express classical unconstrained \mathcal{H}_2 and \mathcal{H}_∞ control problems in terms of the achievable closed-loop responses by exploiting a priori assumptions on the disturbance process, i.e., $\mathbf{w} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \Sigma_w)$, where $\Sigma_w \succ 0$ denotes the covariance matrix of \mathbf{w} , in the \mathcal{H}_2 setting, or \mathbf{w} adversarially chosen in the \mathcal{H}_∞ setting. For the sake of comparison, we report these reformulations (Anderson et al., 2019) below on the left and right sides of the page, respectively:

$$\begin{aligned} \min_{\Phi_x, \Phi_u} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & \mathcal{R}^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \Sigma_w^{\frac{1}{2}} \right\|_F^2 & \quad \min_{\Phi_x, \Phi_u} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & \mathcal{R}^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{2 \rightarrow 2}^2 \\ \text{subject to } [I - Z\mathcal{A} \quad -Z\mathcal{B}] \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I, & \quad \text{subject to } [I - Z\mathcal{A} \quad -Z\mathcal{B}] \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I, \quad (8) \\ \Phi_x, \Phi_u \text{ with causal sparsities,} & \quad \Phi_x, \Phi_u \text{ with causal sparsities.} \quad (9) \end{aligned}$$

Here, $\|\cdot\|_F$ and $\|\cdot\|_{2 \rightarrow 2}$ denote the Frobenius and the induced 2-norm of a matrix, respectively. Thanks to convexity, the optimal system responses $\{\Phi_x^*, \Phi_u^*\}$ can be computed efficiently and the corresponding optimal control policy in the form of (2) can then be recovered by $\mathbf{K}^* = \Phi_u^* \Phi_x^{*-1}$.

3. Proposed Methodology

In this section, we first adapt useful results from Hassibi et al. (1999), and we then present a novel tractable method to design regret-optimal control policies that further comply with hard safety constraints over the system states and inputs. Our idea is to exploit the SLP of achievable closed-loop responses to characterize as solutions of convex optimization problems both the clairvoyant optimal policy, which solves the inner minimization in (4), and a safe regret-optimal causal controller.

3.1. The Clairvoyant Optimal Controller

In the full-information setting considered in this paper, there exists a unique noncausal control law that outperforms any other controller for every disturbance realization – the clairvoyant policy. To show this, let \mathcal{F} and \mathcal{G} denote the causal response operators comprising the Markov parameters that encode the linear dynamics (1) as $\mathbf{x} = \mathcal{F}\mathbf{u} + \mathcal{G}\mathbf{w}$. Then, the control cost (3) can be expressed as

$$\text{cost}(\mathbf{w}, \mathbf{u}) = (\mathcal{F}\mathbf{u} + \mathcal{G}\mathbf{w})^\top \mathcal{Q}(\mathcal{F}\mathbf{u} + \mathcal{G}\mathbf{w}) + \mathbf{u}^\top \mathcal{R}\mathbf{u} = \mathbf{u}^\top \mathcal{P}\mathbf{u} + 2\mathbf{u}^\top \mathcal{F}^\top \mathcal{Q}\mathcal{G}\mathbf{w} + \mathbf{w}^\top \mathcal{G}^\top \mathcal{Q}\mathcal{G}\mathbf{w},$$

where $\mathcal{P} = \mathcal{R} + \mathcal{F}^\top \mathcal{Q}\mathcal{F} \succ 0$. Observing that $\mathcal{Q}\mathcal{F}\mathcal{P}^{-1}\mathcal{F}^\top \mathcal{Q} + \mathcal{Q}(I + \mathcal{F}\mathcal{R}^{-1}\mathcal{F}^\top \mathcal{Q})^{-1} = \mathcal{Q}$ thanks to the Woodbury matrix identity, the incurred control cost can further be written as

$$\text{cost}(\mathbf{w}, \mathbf{u}) = [\mathcal{P}\mathbf{u} + \mathcal{F}^\top \mathcal{Q}\mathcal{G}\mathbf{w}]^\top \mathcal{P}^{-1}[\mathcal{P}\mathbf{u} + \mathcal{F}^\top \mathcal{Q}\mathcal{G}\mathbf{w}] + \mathbf{w}^\top \mathcal{G}^\top \mathcal{Q}(I + \mathcal{F}\mathcal{R}^{-1}\mathcal{F}^\top \mathcal{Q})^{-1}\mathcal{G}\mathbf{w},$$

so to highlight the presence of a first non-negative term and of a second addend that does not depend on \mathbf{u} . Solving for the cost-minimizing \mathbf{u} by setting the former term equal to zero, we obtain the following input-output description of the clairvoyant controller:

$$\mathbf{u}^* = \arg \min_{\mathbf{u}' \in \mathbb{R}^{mT}} \text{cost}(\mathbf{w}, \mathbf{u}') = -(\mathcal{R} + \mathcal{F}^\top \mathcal{Q} \mathcal{F})^{-1} \mathcal{F}^\top \mathcal{Q} \mathcal{G} \mathbf{w}. \quad (10)$$

Moreover, the control cost suffered by this offline optimal policy, as a function of the sampled disturbance sequence \mathbf{w} , is given by

$$\text{cost}(\mathbf{w}, \mathbf{u}^*) = \mathbf{w}^\top \mathcal{G}^\top \mathcal{Q} (I + \mathcal{F} \mathcal{R}^{-1} \mathcal{F}^\top \mathcal{Q})^{-1} \mathcal{G} \mathbf{w}. \quad (11)$$

Note that, despite not having specified a priori any parametric structure on such offline policy, the globally optimal dynamic sequence of control actions selected in hindsight by the clairvoyant policy can be expressed as a noncausal linear function of past and future disturbances (Hassibi et al., 1999).

A state-space description of the clairvoyant controller (10) has recently been derived in Goel and Hassibi (2020a) and in Foster and Simchowit (2020) via dynamic programming. Our first result is that the optimal clairvoyant controller (10) can also be computed by solving a convex SLS problem that imposes no causal constraints on the structure of $\{\Phi_x, \Phi_u\}$. As we establish in Section 3.3, the formulation we propose allows one to explicitly include safety requirements in the definition of the benchmark clairvoyant policy.

Lemma 2 *The closed-loop responses $\{\Phi_x^{nc}, \Phi_u^{nc}\}$ associated with the optimal noncausal controller \mathbf{u}^* defined in (10) can be computed as*

$$\{\Phi_x^{nc}, \Phi_u^{nc}\} = \arg \min_{\Phi_x, \Phi_u} \left\| \begin{bmatrix} \mathcal{Q}^{\frac{1}{2}} & 0 \\ 0 & \mathcal{R}^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \Sigma_w^{\frac{1}{2}} \right\|_F^2 \in \arg \min_{\Phi_x, \Phi_u} \left\| \begin{bmatrix} \mathcal{Q}^{\frac{1}{2}} & 0 \\ 0 & \mathcal{R}^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{2 \rightarrow 2}^2 \quad (12)$$

subject to (8), subject to (8).

Moreover, the control cost (11) incurred by the noncausal clairvoyant controller on a specific disturbance vector \mathbf{w} is given by

$$\text{cost}(\mathbf{w}, \mathbf{u}^*) = \left\| \begin{bmatrix} \mathcal{Q}^{\frac{1}{2}} & 0 \\ 0 & \mathcal{R}^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x^{nc} \\ \Phi_u^{nc} \end{bmatrix} \mathbf{w} \right\|_2^2. \quad (13)$$

Proof Observe that the objective functions of the optimization problems in (12) are equivalent to $\mathbb{E}_{\mathbf{w}}[\text{cost}(\mathbf{w}, \mathbf{u})]$ and $\max_{\|\mathbf{w}\|_2 \leq 1} \text{cost}(\mathbf{w}, \mathbf{u})$, respectively. Recall that the clairvoyant control law \mathbf{u}^* defined in (10) is the unique offline policy that minimizes $\text{cost}(\mathbf{w}, \mathbf{u})$ for every \mathbf{w} . Hence, for every control sequence \mathbf{u} , linearity of the expectation operator implies that

$$\mathbb{E}_{\mathbf{w}}[\text{cost}(\mathbf{w}, \mathbf{u})] - \mathbb{E}_{\mathbf{w}}[\text{cost}(\mathbf{w}, \mathbf{u}^*)] = \mathbb{E}_{\mathbf{w}}[\text{cost}(\mathbf{w}, \mathbf{u}) - \text{cost}(\mathbf{w}, \mathbf{u}^*)] \geq 0.$$

Similarly, $\max_{\|\mathbf{w}\|_2 \leq 1} \text{cost}(\mathbf{w}, \mathbf{u}) \geq \max_{\|\mathbf{w}\|_2 \leq 1} \text{cost}(\mathbf{w}, \mathbf{u}^*)$ since $\text{cost}(\mathbf{w}, \mathbf{u}) \geq \text{cost}(\mathbf{w}, \mathbf{u}^*)$ point-wise for every \mathbf{w} and every \mathbf{u} . As the clairvoyant policy \mathbf{u}^* minimizes both $\mathbb{E}_{\mathbf{w}}[\text{cost}(\mathbf{w}, \mathbf{u})]$ and $\max_{\|\mathbf{w}\|_2 \leq 1} \text{cost}(\mathbf{w}, \mathbf{u})$, we deduce that the corresponding closed-loop responses $\{\Phi_x^{nc}, \Phi_u^{nc}\}$ belong to the set of minimizers of both optimization problems in (12). Moreover, we note that $\left\| \mathcal{Q}^{\frac{1}{2}} \Phi_x \Sigma_w^{\frac{1}{2}} \right\|_F^2 + \left\| \mathcal{R}^{\frac{1}{2}} \Phi_u \Sigma_w^{\frac{1}{2}} \right\|_F^2$, with $\Phi_x = (I - Z\mathcal{A})^{-1}(I + Z\mathcal{B}\Phi_u)$ as per (7), is strictly convex

in Φ_u thanks to $\mathcal{Q} \succeq 0$, $\mathcal{R} \succ 0$ and $\Sigma_w \succ 0$. Hence, the pair $\{\Phi_x^{nc}, \Phi_u^{nc}\}$ constitutes the unique global minimizer of the optimization problem on the left-hand side of (12). Lastly, (13) follows by substituting $\mathbf{x}^{nc} = \Phi_x^{nc} \mathbf{w}$ and $\mathbf{u}^{nc} = \Phi_u^{nc} \mathbf{w}$ in (3). \blacksquare

Note that the optimization problems in (12) mirror the classical \mathcal{H}_2 and \mathcal{H}_∞ control formulations presented in Section 2.1. Indeed, the only difference is the absence of the sparsity constraints (9) on the noncausal closed-loop responses $\{\Phi_x^{nc}, \Phi_u^{nc}\}$ corresponding with the offline optimal policy.

Linearity of the clairvoyant policy (10) with respect to the disturbance realizations is central to the regret-optimal synthesis approach proposed in Goel and Hassibi (2020b) in the absence of safety requirements. However, finding the proper change of variables that allows to reduce a regret-suboptimal control problem to a suboptimal \mathcal{H}_∞ problem might be nontrivial, as it requires to analytically characterize a causal factorization of $\gamma^2 I + \mathcal{G}^\top \mathcal{Q} (I + \mathcal{F} \mathcal{R}^{-1} \mathcal{F}^\top \mathcal{Q})^{-1} \mathcal{G}$, where γ denotes the performance level to be tuned iteratively. Indeed, the method of Goel and Hassibi (2020b) involves repeated applications of the whitening property of Kalman filters (Kailath et al., 2000).

3.2. The Safe Regret-Optimal Controller

We now show that linearity of the optimal offline policy (12) is also key to computing a safe regret-optimal controller with efficient numerical programming techniques. Formulating the nested minimization of (4) as a convex optimization problem constitutes our main result.

Theorem 3 *Let Assumption 1 hold. Consider the evolution of the linear time-varying system (1) over a horizon of length $T \in \mathbb{N}$, and the constrained regret-optimal control problem:*

$$\{\Phi_x^{sr}, \Phi_u^{sr}\} \in \arg \min_{\Phi_x, \Phi_u} \text{regret}(T, \mathbf{u}) \quad (14a)$$

subject to (8) – (9),

$$\mathcal{H} \begin{bmatrix} \Phi_x^\top & \Phi_u^\top \end{bmatrix}^\top \mathbf{w} \leq \mathbf{h}, \quad \forall \mathbf{w} : \mathcal{H}_w \mathbf{w} \leq \mathbf{h}_w. \quad (14b)$$

Then, (14) is equivalently formulated as the following convex optimization problem:

$$\{\Phi_x^{sr}, \Phi_u^{sr}\} \in \arg \min_{\Phi_x, \Phi_u, \mathbf{Z}, \lambda} \lambda \quad (15a)$$

subject to (8) – (9), $\lambda > 0$,

$$\mathbf{Z}^\top \mathbf{h}_w \leq \mathbf{h}, \quad \mathcal{H} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = \mathbf{Z}^\top \mathcal{H}_w, \quad \mathbf{Z}_{ij} \geq 0, \quad (15b)$$

$$\begin{bmatrix} I & \begin{bmatrix} \mathcal{Q}^{\frac{1}{2}} & 0 \\ 0 & \mathcal{R}^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \\ \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix}^\top \begin{bmatrix} \mathcal{Q}^{\frac{1}{2}} & 0 \\ 0 & \mathcal{R}^{\frac{1}{2}} \end{bmatrix} & \lambda I + \begin{bmatrix} \Phi_x^{nc} \\ \Phi_u^{nc} \end{bmatrix}^\top \begin{bmatrix} \mathcal{Q} & 0 \\ 0 & \mathcal{R} \end{bmatrix} \begin{bmatrix} \Phi_x^{nc} \\ \Phi_u^{nc} \end{bmatrix} \end{bmatrix} \succeq 0. \quad (15c)$$

Proof In light of (13), (14) can be equivalently expressed as

$$\min_{\Phi_x, \Phi_u} \max_{\|\mathbf{w}\|_2 \leq 1} \mathbf{w}^\top \underbrace{\left(\begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix}^\top \begin{bmatrix} \mathcal{Q} & 0 \\ 0 & \mathcal{R} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} - \begin{bmatrix} \Phi_x^{nc} \\ \Phi_u^{nc} \end{bmatrix}^\top \begin{bmatrix} \mathcal{Q} & 0 \\ 0 & \mathcal{R} \end{bmatrix} \begin{bmatrix} \Phi_x^{nc} \\ \Phi_u^{nc} \end{bmatrix} \right)}_{\Delta(\Phi_x, \Phi_u)} \mathbf{w} \quad (16)$$

subject to (8) – (9), (14b).

We first proceed to derive a tractable formulation for the objective function in (16). Since the offline optimal controller attains minimum control cost on every perturbation sequence, we have that $\Delta(\Phi_x, \Phi_u) \succeq 0$ for all $\{\Phi_x, \Phi_u\}$. Consequently, there exists a matrix $\delta(\Phi_x, \Phi_u) \succeq 0$ such that $\Delta = \delta^\top \delta$,² where we have omitted the dependence of Δ and δ on $\{\Phi_x, \Phi_u\}$ to ease readability. Then, leveraging well-known properties of induced matrix norms, we obtain

$$\max_{\|\mathbf{w}\|_2 \leq 1} \mathbf{w}^\top \Delta \mathbf{w} = \max_{\|\mathbf{w}\|_2 \leq 1} \|\delta \mathbf{w}\|_2^2 = \|\delta\|_{2 \rightarrow 2}^2 = \sigma_{\max}^2(\delta) = \lambda_{\max}(\Delta),$$

where $\sigma_{\max}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the largest singular value and the largest eigenvalue of a matrix, respectively. Hence, minimizing dynamic regret corresponds to minimizing the maximum eigenvalue of Δ , which depends quadratically on $\{\Phi_x, \Phi_u\}$. Building upon classical results on semidefinite programming for eigenvalue minimization (see, e.g., Section 2.2 in Boyd et al. (1994)), $\lambda_{\max}(\Delta)$ is equivalently computed as $\min_{\lambda} \lambda$ subject to $\lambda I - \Delta \succeq 0$, $\lambda > 0$. Recalling the definition of Δ in (16) and exploiting the Schur complement, we obtain the desired expression (15c).

For the safety constraints, we apply dualization to eliminate the universal quantifier from (14b). In particular, we recognize that each row of $\max_{\mathbf{w} \in \mathbb{W}} \mathcal{H} \begin{bmatrix} \Phi_x^\top & \Phi_u^\top \end{bmatrix}^\top \mathbf{w}$ is a linear optimization problem (Goulart, 2007). Exploiting duality, we reformulate the constraints as

$$\begin{aligned} \max_{\mathbf{w} \in \mathbb{W}} \left(\mathcal{H} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right)_i \mathbf{w} &= \min_{\mathbf{z}_i} \mathbf{h}_w^\top \mathbf{z}_i, \\ \text{subject to } \mathbf{z}_i &\geq 0, \mathcal{H}_w^\top \mathbf{z}_i = \left(\mathcal{H} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right)_i^\top, \end{aligned}$$

where the vector \mathbf{z}_i represents the dual vector associated with the i -th row of the maximization. Finally, by combining the dual variables that arise from each row into the matrix \mathbf{Z} , we obtain the linear constraints (15b). \blacksquare

We note that the method in Goel and Hassibi (2020b) allows computing the unconstrained regret-optimal controller by solving a series of Riccati recursions. Instead, our approach to the more general constrained case relies on solving (15) through semidefinite programming, which poses challenges to scalability, in general. We refer the interested reader to Ahmadi and Majumdar (2019); Zheng et al. (2017) and the references therein for state-of-the-art techniques that exploit diagonal dominance and chordal sparsity to improve scalability. Last, notice that a safe regret-optimal policy in the form of (2) is recovered by $\mathbf{K}^{sr} = \Phi_u^{sr} \Phi_x^{sr-1}$.

3.3. Competing Against Safety-Aware Control Benchmarks

The safe regret-optimal control law characterized in Theorem 3 competes against the unconstrained clairvoyant policy (12). Alternatively, one could aim to minimize the loss relative to an optimal safety-aware noncausal policy that selects the control actions with complete knowledge of past and future disturbances, but has to obey the same safety constraints as the online controller. We note that a similar idea has recently been studied in Li et al. (2021), which aims to bound the regret with respect to the best safe state-feedback causal controller in hindsight. Instead, our proposal is to

2. Such a factorization can be computed from the eigendecomposition $\Delta = V \Lambda V^{-1}$. Since Δ is positive semidefinite, its eigenvalues are non-negative and its eigenvectors are orthogonal, hence $\Delta = V \Lambda V^\top = V \Lambda^{\frac{1}{2}} (V \Lambda^{\frac{1}{2}})^\top = \delta^\top \delta$.

synthesize safe control laws that minimize regret against safe clairvoyant policies through convex programming techniques. Specifically, we show that the optimization perspective of Lemma 2 and Theorem 3 can also be used for defining safety-aware control benchmarks and, subsequently, safe regret-optimal policies that compete against them.

Corollary 4 *Let Assumption 1 hold. The closed-loop responses $\{\Phi_x^{snc}, \Phi_u^{snc}\}$ associated with a safe clairvoyant policy that is optimal either in the \mathcal{H}_2 or in the \mathcal{H}_∞ sense are computed by solving*

$$\min_{\Phi_x, \Phi_u, \mathbf{Z}} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & \mathcal{R}^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \Sigma_w^{\frac{1}{2}} \right\|_F^2 \quad \text{and} \quad \min_{\Phi_x, \Phi_u, \mathbf{Z}} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & \mathcal{R}^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{2 \rightarrow 2}^2 \quad (17)$$

subject to (8), (15b), subject to (8), (15b),

respectively. Moreover, a safe regret-optimal policy that competes against a safe clairvoyant controller is computed by solving (15), provided that $\{\Phi_x^{nc}, \Phi_u^{nc}\}$ is replaced by $\{\Phi_x^{snc}, \Phi_u^{snc}\}$ in (15c).

Note that Assumption 1 guarantees that both optimization problems in (17) are feasible, since their solution spaces include that of (15). We recall that in the unconstrained setting both cost formulations in (17) were minimized by the same noncausal input sequence (12) as per Lemma 2. Conversely, in the constrained case there may not exist a safe clairvoyant control law that outperforms all other safe policies on every disturbance instance. Hence, the minimizers of the two optimization problems in (17) may be different, and their optimality should only be understood in the \mathcal{H}_2 or \mathcal{H}_∞ sense. This observation is consistent with the partial-information setup studied in (Goel and Hassibi, 2021).

The results of Theorem 3 and Corollary 4 naturally extend to more complex settings. For instance, the proposed optimization standpoint allows one to constrain the system closed-loop responses to lie in any arbitrary set \mathcal{S} , provided that it admits a convex representation. In particular, constraints of the form $\{\Phi_x, \Phi_u\} \in \mathcal{S}$ could model additional performance requirements, structural constraints that arise from the distributed nature of the system under control, or sparsity surrogate requirements.

4. Numerical Results

We test the performance of the proposed safe regret-optimal control law \mathcal{SR}_{nc} against classical constrained \mathcal{H}_2 and \mathcal{H}_∞ controllers, which can be computed by solving (17) subject to additional sparsity constraints (9). For our experiments, we consider the evolution of system (1), starting from the unknown initial condition $x_0 = 0$, with

$$A_t = \rho \begin{bmatrix} 0.7 & 0.2 & 0 \\ 0.3 & 0.7 & -0.1 \\ 0 & -0.2 & 0.8 \end{bmatrix}, \quad B_t = \begin{bmatrix} 1 & 0.2 \\ 2 & 0.3 \\ 1.5 & 0.5 \end{bmatrix}, \quad \forall t \in \{0 \dots T-1\},$$

where ρ is the spectral radius of the system, and $T = 30$. Letting $\rho = 0.7$, we first compute $\{\Phi_x^{nc}, \Phi_u^{nc}\}$ as the solution to (12), choosing $Q = I_{30} \otimes I_3$ and $\mathcal{R} = I_{30} \otimes I_2$ in (3), where \otimes denotes the Kronecker product. Subsequently, we solve the SDP in (15) to synthesize a control policy that minimizes dynamic regret while complying, for all possible initial conditions $-1 \leq x_0 \leq 1$ and perturbations $-1 \leq w_t \leq 1$, with safety constraints $-3 \leq x_t \leq 3$, $-2 \leq u_t \leq 2$, $\forall t \in$

$\{0 \dots T - 1\}$. We simulate the evolution of the closed-loop system perturbed by 1000 different realizations of a disturbance sequence uniformly distributed between -1 and 1 , i.e., $w_t \stackrel{\text{iid}}{\sim} \mathcal{U}_{[-1,1]}$ ³, and we verify that the input and state trajectories are safe in all rounds, as expected.

Then, for the two cases, $\rho = 0.7$ (open-loop stable system) and $\rho = 1.05$ (open-loop unstable system), we compare the average control cost suffered by these safe control laws when the disturbances are drawn according to a variety of stochastic and deterministic profiles. For the case $\rho = 1.05$, we relax the safety constraints and require that $-10 \leq x_t \leq 10$, $-10 \leq u_t \leq 10$, $\forall t \in \{0 \dots T - 1\}$ to achieve feasibility. We collect our results in Table 1.³ As expected, when the true perturbation sequence follows a Gaussian distribution, i.e., $w_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, I_3)$, \mathcal{SH}_2 achieves the best performance, closely followed by our safe regret-optimal control law \mathcal{SR}_{nc} when $\rho = 0.7$, as observed in Goel and Hassibi (2020b) for the unconstrained case. Similarly, when the disturbance realizations are chosen adversarially, \mathcal{SR}_{nc} nearly tracks the performance of \mathcal{SH}_∞ , which attains the lowest control cost. Instead, our \mathcal{SR}_{nc} consistently outperforms both \mathcal{SH}_2 and \mathcal{SH}_∞ in almost all other scenarios. We conjecture that this improvement is linked with the optimality of the control benchmark in (16), which is tailored to the system at hand. Furthermore, we observe that the performance increase may be very significant. This is made clear, for instance, in our experiments with $\rho = 0.7$, where \mathcal{SH}_2 and \mathcal{SH}_∞ always incur a loss that is at least 40% higher.

Table 1: Average control cost increase relative to the control policy denoted with **1**.

Open-loop stable system: $\rho = 0.7$				Open-loop unstable system: $\rho = 1.05$			
w	\mathcal{SH}_2	\mathcal{SH}_∞	\mathcal{SR}_{nc}	w	\mathcal{SH}_2	\mathcal{SH}_∞	\mathcal{SR}_{nc}
$\mathcal{N}(0, 1)$	1	+21.14%	+ 10.89%	$\mathcal{N}(0, 1)$	1	>+100%	+ 51.99%
$\mathcal{U}_{[0.5,1]}$	+63.42%	>+100%	1	$\mathcal{U}_{[0.5,1]}$	+36.60%	+9.91%	1
$\mathcal{U}_{[0,1]}$	+40.69%	>+100%	1	$\mathcal{U}_{[0,1]}$	+5.60%	+20.14%	1
1	+67.74%	>+100%	1	1	+44.57%	+7.25%	1
sin	+58.12%	>+100%	1	sin	+39.10%	+12.89%	1
sawtooth	+46.27%	>+100%	1	sawtooth	+26.53%	+13.52%	1
step	+66.49%	>+100%	1	step	+15.66%	1	1
stairs	+45.27%	>+100%	1	stairs	+15.37%	+0.46%	1
worst	+18.45%	1	+7.74%	worst	>+100%	1	+26.51%

5. Conclusion

We have presented a novel method for convex synthesis of regret-optimal control policies that comply with hard safety requirements. To do so, we have first characterized the clairvoyant policy by extending the SLS framework, and we have then nested the corresponding solution into a constrained regret-optimal SLS program. Numerical results show that control laws that safely minimize regret can adapt to heterogeneous disturbance sequences, effectively interpolating between the performance of, or even prevailing over, constrained \mathcal{H}_2 and \mathcal{H}_∞ controllers. Future work encompasses extensions to the infinite-horizon case, as well as distributed control and model-free scenarios.

3. The code that reproduces our numerical examples is available at <https://github.com/DecodEPFL/SafeMinRegret>. Please refer to the simulation code for a precise definition of the disturbance profiles that appear in Table 1.

Acknowledgments

Research supported by the Swiss National Science Foundation under the NCCR Automation (grant agreement 51NF40_180545).

References

- Yasin Abbasi-Yadkori and Csaba Szepesvári. Regret bounds for the adaptive control of linear quadratic systems. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 1–26. JMLR Workshop and Conference Proceedings, 2011.
- Naman Agarwal, Brian Bullins, Elad Hazan, Sham Kakade, and Karan Singh. Online control with adversarial disturbances. In *International Conference on Machine Learning*, pages 111–119. PMLR, 2019.
- Amir Ali Ahmadi and Anirudha Majumdar. DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization. *SIAM Journal on Applied Algebra and Geometry*, 3(2):193–230, 2019.
- James Anderson, John C Doyle, Steven H Low, and Nikolai Matni. System level synthesis. *Annual Reviews in Control*, 47:364–393, 2019.
- Alberto Bemporad. Reducing conservativeness in predictive control of constrained systems with disturbances. In *Proceedings of the 37th IEEE Conference on Decision and Control*, volume 2, pages 1384–1389. IEEE, 1998.
- Stephen Boyd, Laurent El Ghaoui, Eric Feron, and Venkataramanan Balakrishnan. *Linear matrix inequalities in system and control theory*. SIAM, 1994.
- Alon Cohen, Tomer Koren, and Yishay Mansour. Learning linear-quadratic regulators efficiently with only \sqrt{T} regret. In *International Conference on Machine Learning*, pages 1300–1309. PMLR, 2019.
- Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. Regret bounds for robust adaptive control of the linear quadratic regulator. In *Advances in Neural Information Processing Systems*, pages 4188–4197, 2018.
- Dylan Foster and Max Simchowitz. Logarithmic regret for adversarial online control. In *International Conference on Machine Learning*, pages 3211–3221. PMLR, 2020.
- Luca Furieri, Yang Zheng, Antonis Papachristodoulou, and Maryam Kamgarpour. An input–output parametrization of stabilizing controllers: Amidst youla and system level synthesis. *IEEE Control Systems Letters*, 3(4):1014–1019, 2019.
- Gautam Goel and Babak Hassibi. The power of linear controllers in LQR control. *arXiv preprint arXiv:2002.02574*, 2020a.
- Gautam Goel and Babak Hassibi. Regret-optimal control in dynamic environments. *arXiv preprint arXiv:2010.10473*, 2020b.

- Gautam Goel and Babak Hassibi. Regret-optimal measurement-feedback control. In *Learning for Dynamics and Control*, pages 1270–1280. PMLR, 2021.
- Paul James Goulart. *Affine feedback policies for robust control with constraints*. PhD thesis, University of Cambridge, 2007.
- Babak Hassibi, Ali H Sayed, and Thomas Kailath. *Indefinite-quadratic estimation and control: a unified approach to \mathcal{H}_2 and \mathcal{H}_∞ theories*. SIAM, 1999.
- Elad Hazan. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.
- Elad Hazan and Comandur Seshadhri. Efficient learning algorithms for changing environments. In *Proceedings of the 26th annual international conference on machine learning*, pages 393–400, 2009.
- Elad Hazan, Sham Kakade, and Karan Singh. The nonstochastic control problem. In *Algorithmic Learning Theory*, pages 408–421. PMLR, 2020.
- Ali Jadbabaie, Alexander Rakhlin, Shahin Shahrampour, and Karthik Sridharan. Online optimization: Competing with dynamic comparators. In *Artificial Intelligence and Statistics*, pages 398–406. PMLR, 2015.
- Thomas Kailath, Ali H Sayed, and Babak Hassibi. *Linear estimation*. Prentice Hall, 2000.
- Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. Logarithmic regret bound in partially observable linear dynamical systems. *Advances in Neural Information Processing Systems*, 33:20876–20888, 2020.
- Yingying Li, Subhro Das, and Na Li. Online optimal control with affine constraints. *arXiv preprint arXiv:2010.04891*, 2020.
- Yingying Li, Subhro Das, Jeff Shamma, and Na Li. Safe adaptive learning-based control for constrained linear quadratic regulators with regret guarantees. *arXiv preprint arXiv:2111.00411*, 2021.
- Zeev Nehari. On bounded bilinear forms. *Annals of Mathematics*, pages 153–162, 1957.
- Marko Nonhoff and Matthias A Müller. An online convex optimization algorithm for controlling linear systems with state and input constraints. In *2021 American Control Conference (ACC)*, pages 2523–2528. IEEE, 2021.
- James Blake Rawlings, David Q Mayne, and Moritz Diehl. *Model predictive control: theory, computation, and design*, volume 2. Nob Hill Publishing Madison, 2017.
- Oron Sabag, Gautam Goel, Sahin Lale, and Babak Hassibi. Regret-optimal full-information control. *arXiv preprint arXiv:2105.01244*, 2021.
- Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and trends in Machine Learning*, 4(2):107–194, 2011.

- Max Simchowitz, Karan Singh, and Elad Hazan. Improper learning for non-stochastic control. In *Conference on Learning Theory*, pages 3320–3436. PMLR, 2020.
- Yuh-Shyang Wang, Nikolai Matni, and John C Doyle. A system-level approach to controller synthesis. *IEEE Transactions on Automatic Control*, 64(10):4079–4093, 2019.
- Dante Youla, Hamid Jabr, and Jr Bongiorno. Modern Wiener-Hopf design of optimal controllers—part II: The multivariable case. *IEEE Transactions on Automatic Control*, 21(3):319–338, 1976.
- Peng Zhao, Yu-Jie Zhang, Lijun Zhang, and Zhi-Hua Zhou. Dynamic regret of convex and smooth functions. *Advances in Neural Information Processing Systems*, 33:12510–12520, 2020.
- Yang Zheng, Giovanni Fantuzzi, Antonis Papachristodoulou, Paul Goulart, and Andrew Wynn. Fast ADMM for semidefinite programs with chordal sparsity. In *2017 American Control Conference (ACC)*, pages 3335–3340. IEEE, 2017.
- Yang Zheng, Luca Furieri, Antonis Papachristodoulou, Na Li, and Maryam Kamgarpour. On the equivalence of Youla, system-level, and input–output parameterizations. *IEEE Transactions on Automatic Control*, 66(1):413–420, 2020.
- Kemin Zhou and John Comstock Doyle. *Essentials of robust control*, volume 104. Prentice hall Upper Saddle River, NJ, 1998.
- Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pages 928–936, 2003.