Safe Non-Stochastic Control of Linear Dynamical Systems

Hongyu Zhou, Vasileios Tzoumas

Abstract— We study the problem of *safe control of linear dynamical systems corrupted with non-stochastic noise*, and provide an algorithm that guarantees (i) zero constraint violation of convex time-varying constraints, and (ii) bounded dynamic regret, *i.e.*, bounded suboptimality against an optimal clairvoyant controller that knows the future noise a priori. The constraints bound the values of the state and of the control input such as to ensure collision avoidance and bounded control effort. We are motivated by the future of autonomy where robots will safely perform complex tasks despite real-world unpredictable disturbances such as wind and wake disturbances. To develop the algorithm, we capture our problem as a sequential game between a linear feedback controller and an adversary, assuming a known upper bound on the noise's magnitude. Particularly, at each step $t = 1, \ldots, T$, first the controller chooses a linear feedback control gain $K_t \in \mathcal{K}_t$, where $\mathcal{K}_t$ is constructed such that it guarantees that the safety constraints will be satisfied; then, the adversary reveals the current noise w_t and the controller suffers a loss $f_t(K_t)$ — *e.g.*, f_t represents the system's tracking error at t upon the realization of the noise. The controller aims to minimize its cumulative loss, despite knowing w_t only after K_t has been chosen. We validate our algorithm in simulated scenarios of safe control of linear dynamical systems in the presence of bounded noise.

I. INTRODUCTION

In the future, robots will be leveraging their on-board control capabilities to complete safety-critical tasks such as package delivery [1], target tracking [2], and disaster response [3]. To complete such complex tasks, the robots need to *reliably* overcome a series of key challenges:

a) Challenge I: Time-Varying Safety Constraints: The robots need to ensure their own safety and the safety of their surroundings. For example, robots often need to ensure that they follow prescribed collision-free trajectories or that their control effort is kept under prescribed levels. Such safety requirements take the form of *time-varying* state and control input constraints, and can make the planning of control inputs computationally hard [4], [5].

b) Challenge II: Unpredictable Noise: The robots' dynamics are often corrupted by unknown and non-stochastic noise, *i.e.*, noise that is not necessarily i.i.d. Gaussian and, broadly, stochastic. For example, aerial and marine vehicles often face non-stochastic winds and waves, respectively [6], [7]. But the current control algorithms primarily rely on stochastic noise (*e.g.*, Gaussian-structured), compromising thus the robots' ability to ensure safety [8], [9].

The above challenges motivate the development of safe control algorithms against unpredictable noise. State-of-the-art methods that aim to address this problem either rely on robust control [10]–[15] or on online learning for control [16]–[22]. But the robust methods are often conservative

and computationally heavy since they simulate the system dynamics over a lookahead horizon assuming a worst-case noise realization that matches a known upper bound on the magnitude of the noise. To reduce conservatism and increase efficiency, researchers have recently focused on online learning for control methods via Online Convex Optimization (OCO) [23], [24]. The online learning for control methods typically rely on the *Online Gradient Descent* (**OGD**) algorithm and its variants, offering *bounded regret* guarantees, *i.e.*, bounded suboptimality with respect to an optimal (possibly time-varying) clairvoyant controller [16]–[22]. However, the current online methods address only time-invariant safety constraints.

Contributions. Our goal is to achieve online control of linear dynamical systems subject to time-varying safety constraints, despite unpredictable noise. To this end, we formalize the problem of *Safe Non-Stochastic Control of Linear Dynamical Systems* (Safe-NSC). Safe-NSC can be interpreted as a sequential game between a linear feedback controller and an adversary. Particularly, at each step $t = 1, \ldots, T$, first the controller chooses a linear feedback control gain $K_t \in \mathcal{K}_t$, where $\mathcal{K}_t$ is constructed such that it guarantees that the safety constraints will be satisfied; then, the adversary reveals the current noise w_t and the controller suffers a loss $f_t(K_t)$ — *e.g.*, f_t represents the system's tracking error at t upon the realization of the noise.

Safe-NSC is challenging since the controller aims to guarantee bounded dynamic regret despite knowing w_t only after K_t has been chosen.

We make the following contributions to solving Safe-NSC:

- *Algorithmic Contributions*: We introduce the algorithm *Safe Online Gradient Descent* (**Safe-OGD**), which generalizes the seminal *Online Gradient Descent* (**OGD**) [25] to the Safe-NSC setting to enable online non-stohastic control subject to time-varying constraints.
- *Technical Contributions*: We prove that the **Safe-OGD** controller has bounded dynamic regret against any safe linear feedback control policy (Theorem 1), given a known upper bound on the noise's magnitude. When the domain sets are time-invariant, we prove that the bound of **Safe-OGD** reduces to the bound in the standard (time-invariant) OCO setting [25] (Section V-B).[1]

Numerical Evaluations. We validate our algorithms in simulated scenarios. (Section VI and Appendix A). Specifically, we compare our algorithm with the safe H_2 and H_∞ in scenarios involving a quadrotor aiming to stay at a hovering position despite unpredictable disturbances (Section VI). We then compare our algorithm with state-of-the-art OCO with Memory (OCO-M) algorithm [17] in scenarios involving synthetic linear time-invariant systems (Appendix A). Our

Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109 USA; {zhouhy, vtzoumas}@umich.edu

[1]All the proofs are given in [26].

algorithm achieves in the simulations (i) comparable loss and better computation time performance than safe H_2 and H_∞, and (ii) better loss performance than OCO-M algorithm.

II. RELATED WORK

We next review (i) *Non-Stochastic Control*: regret optimal control and online learning for control; (ii) OCO with *time-invariant constraints* or *time-varying constraints*.

Regret-optimal control. Regret optimal control algorithms select control inputs upon simulating the future system dynamics across a lookahead horizon [10]–[15]. Specifically, these methods guarantee safety via solving a robust optimization problem that (ii-a) assumes a worst-case realization of noise, which can be pessimistic and time-consuming, thus compromising the quality of real-time control; and (ii-b) requires the a priori knowledge of a closed-form solution of the optimal controller over the lookahead horizon, which is not available in the safe non-stochastic control setting.

Online learning for control. Online learning algorithms select control inputs based on past information only [16]–[22]. By employing the OCO framework, they provide bounded regret guarantees against an optimal (potentially time-varying) clairvoyant controller even though the noise is unpredictable. However, they consider time-invariant state and control input constraints, in contrast to time-varying safety constraints in this paper. In more detail:

a) OCO with Time-Invariant Constraints: We focus our review on algorithms with bounded dynamic regret; for a broader review of OCO algorithms, we refer the reader to [24]. [27] prove that the optimal dynamic regret for OCO is $\Omega\left(\sqrt{T(1 + C_T)}\right)$, where $C_T \triangleq \sum_{t=2}^{T}\|\mathbf{v}_{t-1} - \mathbf{v}_t\|$ is the path length of the comparator sequence, and provide an algorithm matching this bound. The algorithm is based on *Online Gradient Descent* (**OGD**) [25], which is a projection-based algorithm: at each time step t, **OGD** chooses a decision $\mathbf{x}_t$ by first computing an intermediate decision $\mathbf{x}_t' = \mathbf{x}_{t-1} - \eta\nabla f_{t-1}(\mathbf{x}_{t-1})$ —given the previous decision $\mathbf{x}_{t-1}$, the gradient of the previously revealed loss $f_{t-1}(\mathbf{x}_{t-1})$, and a step size $\eta > 0$— and then projects $\mathbf{x}_t'$ back to the time-invariant domain set $\mathcal{X}$ to output the final decision $\mathbf{x}_t$.

b) OCO with Time-Varying Constraints: The problem of OCO *with Time-Varying Constraints* (OCO-TV) is defined as follows: At each time step t, first the optimizer chooses a decision $\mathbf{x}_t$ from a time-invariant domain set $\mathcal{X}$, and then the adversary reveals a convex loss function f_t as well as a vector-valued constraint $g_t(\mathbf{x}_t) \leq 0$. The optimizer in particular aims to minimize (i) the cumulative loss $\sum_{t=1}^{T} f_t(\mathbf{x}_t)$, and (ii) the cumulative constraint violation $\sum_{t=1}^{T} g_t(\mathbf{x}_t)$. In contrast thus to the setting in this paper, where the constraints must be satisfied at each time step, in the OCO-TV setting the optimizer may violate any of the constraints, aiming to only asymptotically guarantee in the best case no-regret constraint violation, *i.e.*, $\lim_{T\to\infty} \sum_{t=1}^{T} g_t(\mathbf{x}_t)/T = 0$ [28]–[33].

III. PROBLEM FORMULATION

We formulate the problem of *Safe Non-Stochastic Control of Linear Dynamical Systems* (Problem 1). To this end, we use the following notation and assumptions.

Linear Time-Varying System. We consider Linear Time-Varying (LTV) systems of the form

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

where $x_t \in \mathbb{R}^{d_x}$ is the state of the system, $u_t \in \mathbb{R}^{d_u}$ is the control input, and $w_t \in \mathbb{R}^{d_x}$ is the process noise.

Assumption 1 (Known System Matrices). *The system matrices, i.e., A_t and B_t, are known.*

Assumption 2 (Bounded System Matrices and Noise). *The system matrices and noise are bounded, i.e., $\|A_t\| \leq \kappa_A$, $\|B_t\| \leq \kappa_B$, and $w_t \in \mathcal{W} \triangleq \{w \mid \|w\| \leq W\}$, where κ_A, κ_B, and W are given positive numbers.*

Per Assumption 2, we assume no stochastic model for the process noise w_t. The noise may even be adversarial, subject to the bounds prescribed by W.

Safety Constraints. We consider the states and control inputs for all t must satisfy polytopic constraints of the form

$$x_t \in \mathcal{S}_t \triangleq \{x \mid L_{x,t}x \leq l_{x,t}\}, \ \forall\{w_\tau \in \mathcal{W}\}_{\tau=1}^{t-1},$$
$$u_t \in \mathcal{U}_t \triangleq \{u \mid L_{u,t}u \leq l_{u,t}\}, \quad (2)$$

for given $L_{x,t}$, $l_{x,t}$, $L_{u,t}$, and $l_{u,t}$.[2]

Linear-Feedback Control Policy. We consider a linear state feedback control policy $u_t = K_t x_t$ such that

$$\|K_t\| \leq \kappa, \ \|A_t - B_t K_t\| \leq 1 - \gamma, \quad (3)$$

for given $\kappa > 0$ and $\gamma \in (0, 1)$, where K_t will be optimized online. The constraint $\|K_t\| \leq \kappa$ ensures K_t is chosen from a compact decision set , and the constraint $\|A_t - B_t K_t\| \leq 1 - \gamma$ ensures the state is bounded for all t; both constraints enable bounding the dynamic regret of the proposed online optimization algorithm. To ensure that also the safety constraints in eq. (2) are satisfied, we impose additional constraints on K_t later in the paper (Lemma 2 presented in Section IV).

Remark 1 (Removal of the constraint $\|A_t - B_t K_t\| \leq 1 - \gamma$). *The constraint can be removed by employing a sequence K_t^s of (ϵ, γ) sequentially stabilizing controllers, i.e., setting $u_t = -K_t x_t - K_t^s x_t$, where K_t^s is sequentially stabilizing [20] and $\|K_t x_t\|$ is bounded.*

Loss Function. We consider loss functions (control costs) that satisfy the following assumption.

Assumption 3 (Convex and Bounded Loss Function with Bounded Gradient). *$c_t(x_{t+1}, u_t) : \mathbb{R}^{d_x} \times \mathbb{R}^{d_u} \to \mathbb{R}$ is convex in x_{t+1} and u_t. Further, when $\|x\| \leq D$, $\|u\| \leq D$ for some $D > 0$, then $|c_t(x, u)| \leq \beta D^2$ and $\|\nabla_x c_t(x, u)\| \leq GD, \|\nabla_u c_t(x, u)\| \leq GD$, for given β and G.*

An example of a loss function that satisfies Assumption 3 is the quadratic loss $c_t(x_{t+1}, u_t) = x_{t+1} Q x_{t+1}^\top + u_t R u_t^\top$.

Control Performance Metric. We design the control inputs u_t to ensure both safety and a control performance comparable to an optimal clairvoyant policy that selects u_t knowing the future noise realizations w_t a priori.

---

[2]Our results hold true also for any convex state and control input constraints. We focus on polytopic constraints for simplicity in the presentation.

Algorithm 1: Safe Online Gradient Descent (**Safe-OGD**) for Safe-NSC (Problem 1).

Input: Time horizon T; step size η.
Output: Control u_t at each time step $t = 1, \ldots, T$.

1: Initialize $K_1 \in \mathcal{K}_1$;
2: **for** each time step $t = 1, \ldots, T$ **do**
3:     Output $u_t = -K_t x_t$;
4:     Observe the state x_{t+1} and calculate the noise $w_t = x_{t+1} - A_t x_t - B_t u_t$;
5:     Suffer the loss $c_t(x_{t+1}, u_t)$;
6:     Express the loss function in K_t as $f_t(K_t) : \mathbb{R}^{d_u \times d_x} \to \mathbb{R}$;
7:     Obtain gradient $\nabla_K f_t(K_t)$;
8:     Obtain domain set $\mathcal{K}_{t+1}$;
9:     Update $K'_{t+1} = K_t - \eta \nabla_K f_t(K_t)$;
10:    Project $K_{t+1} = \Pi_{\mathcal{K}_{t+1}}(K'_{t+1})$;
11: **end for**

Definition 1 (Dynamic Policy Regret). *The dynamic policy regret is defined as follows:*

$$\text{Regret-NSC}_T^D = \sum_{t=1}^{T} c_t(x_{t+1}, u_t) - \sum_{t=1}^{T} c_t(x_{t+1}^*, u_t^*), \quad (4)$$

where (i) both sums in eq. (4) are evaluated with the same noise $\{w_1, \ldots, w_T\}$, which is the noise experienced by the system during its evolution per the control inputs $\{u_1, \ldots, u_T\}$, (ii) $u_t^* \triangleq -K_t^* x_t$ is the optimal linear feedback control input in hindsight, i.e., the optimal input given a priori knowledge of w_t, (iii) $x_{t+1}^* \triangleq A_t x_t + B_t u_t^* + w_t$ is the state reached by applying the optimal control inputs u_t^* from state x_t, and (iv) x_{t+1}^* and u_t^* satisfy constraints in eq. (2) for all t.

Problem Definition. We formally define the problem of *Safe Non-Stochastic Control of Linear Dynamical Systems*:

Problem 1 (Safe Non-Stochastic Control of Linear Dynamical Systems (Safe-NSC)). *Assume the initial state of the system is safe, i.e., $x_0 \in \mathcal{S}_0$. At each $t = 1, \ldots, T$, first a control input $u_t \in \mathcal{U}_t$ is chosen; then, the noise $w_t \in \mathbb{R}^{d_x}$ is revealed, the system evolves to state $x_{t+1} \in \mathcal{S}_{t+1}$, and suffers a loss $c_t(x_{t+1}, u_t)$. The goal is to guarantee states and control inputs that satisfy the constraints in eq. (2) for all t and that minimize the dynamic policy regret.*

IV. **Safe-OGD** ALGORITHM

We present **Safe-OGD** (Algorithm 1) with bounded dynamic regret for Safe-NSC. **Safe-OGD** first initializes $K_1 \in \mathcal{K}_1$, where $\mathcal{K}_1$ is defined per Lemma 2 (line 1). At each iteration t, Algorithm 1 evolves to state x_{t+1} with the control inputs $u_t = -K_t x_t$ and obtain the noise w_t (lines 3-4). After that, the cost function is revealed and the algorithm suffers a loss of $c_t(x_{t+1}, u_t)$ (line 5). Then, **Safe-OGD** expresses $c_t(x_{t+1}, u_t)$ as a function of K_t, denoted as $f_t(K_t) \triangleq c_t((A_t - B_t K_t)x_t + w_t, -K_t x_t)$ —which is convex in K_t, given A_t, B_t, x_t, and w_t, per Lemma 1 below— and obtains the gradient $\nabla_K f_t(K_t)$ (lines 6-7). To

ensure safety, **Safe-OGD** constructs the domain set $\mathcal{K}_{t+1}$ per Lemma 2 (line 8), which requires one step ahead knowledge of $L_{x,t+1}, L_{u,t+1}, l_{x,t+1}, l_{u,t+1}$. Finally, **Safe-OGD** updates the control gain and projects it back to $\mathcal{K}_{t+1}$ (lines 9-10).

Lemma 1 (Convexity of Loss function in Control Gain). *The loss function $c_t(x_{t+1}, u_t) : \mathbb{R}^{d_x} \times \mathbb{R}^{d_u} \to \mathbb{R}$ is convex in K_t.*

Lemma 2 (Set of Control Gains that Guarantee Safety). *By choosing $K_t \in \mathcal{K}_t$, where*

$$\mathcal{K}_t \triangleq \{K \mid -L_{x,t} B_t K x_t \leq l_{x,t} - L_{x,t} A_t x_t - W\|L_{x,t}\|,$$
$$-L_{u,t} K x_t \leq l_{u,t}, \ \|K\| \leq \kappa, \ \|A_t - B_t K\| \leq 1 - \gamma\}, \quad (5)$$

then, $x_{t+1} \in \mathcal{S}_{t+1}$ and $u_t \in \mathcal{U}_t$ at each time step t.

Assumption 4 (Recursive Feasibility). *We assume that the domain set $\mathcal{K}_t$ is non-empty for all t, $t \in \{1, \ldots, T\}$.*[3]

V. DYNAMIC REGRET ANALYSIS

We present the dynamic regret bound for **Safe-OGD** against any comparator sequence (Theorem 1). The bound reduces to the bound of standard OCO when the optimization domain is time-invariant (Remark 2 in Section V-B). We use the notation:

- $\Pi_\mathcal{K}(\cdot)$ is a projection operation onto the set $\mathcal{K}$;
- $\bar{K}_{t+1} \triangleq \Pi_{\mathcal{K}_t}(K'_{t+1})$ is the decision would have been chosen at time step $t + 1$ if $\mathcal{K}_{t+1} = \mathcal{K}_t$;
- $\zeta_t \triangleq \|\bar{K}_{t+1} - K_{t+1}\|_F$ is the distance between $\bar{K}_{t+1}$ and K_{t+1}, which are the projection of K'_{t+1} onto sets $\mathcal{K}_t$ and $\mathcal{K}_{t+1}$, respectively. Thus, it quantifies how fast the safe domain set changes —ζ_t is 0 when $\mathcal{K}_t = \mathcal{K}_{t+1}$;
- $S_T \triangleq \sum_{t=1}^{T} \zeta_t$ is the cumulative variation of decisions due to time-varying domain sets —S_T becomes 0 when domain sets are time-invariant;
- $C_T \triangleq \sum_{t=2}^{T} \|K_{t-1}^\star - K_t^\star\|_F$ is the path length of the sequence of comparators. It quantifies how fast the optimal control gains change.

A. Dynamic Regret Bound of **Safe-OGD**

We prove the following regret bound for **Safe-OGD**.

Theorem 1 (Dynamic Regret Bound of **Safe-OGD**). *Consider the Safe-NSC problem. **Safe-OGD** achieves against any sequence of comparators $(K_1^\star, \ldots, K_T^\star) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_T$,*

$$\text{Regret-NSC}_T^D \leq \frac{\eta T G_f^2}{2} + \frac{7 D_f^2}{4\eta} + \frac{D_f C_T}{\eta} + \frac{D_f S_T}{\eta}, \quad (6)$$

where $G_f \triangleq G D d_x d_u (\kappa_B + 1)$, $D_f \triangleq 2\kappa\sqrt{d}$, $D \triangleq \max\{\frac{W}{\gamma}, \frac{W\kappa}{\gamma}\}$, and $d \triangleq \min\{d_u, d_x\}$.

Specifically, for $\eta = \mathcal{O}\left(\frac{1}{\sqrt{T}}\right)$, we have

$$\text{Regret-NSC}_T^D \leq \mathcal{O}\left(\sqrt{T}\left(1 + C_T + S_T\right)\right). \quad (7)$$

The dependency on C_T results from the time-varying sequence of comparators. Specifically, any optimal dynamic regret bound for OCO is $\Omega\left(\sqrt{T(1 + C_T)}\right)$, and thus the bound necessarily depends on C_T in the worst case [27].

---

[3]The discussion on recursive feasibility is given in [26, Appendix C].

The dependency on S_T results from the domain sets being time-varying. S_T is zero when the domain sets are time-invariant (Remark 2). Thus, S_T can be sublinear in decision-making applications where any two consecutive safe sets differ a little (*e.g.*, in high-frequency control applications where the control input is updated every a few tenths of milliseconds, then, the safety set may change only a little between consecutive time steps).

B. Regret Bounds in the Time-Invariant Domain Case

When the domain set is time-invariant, the regret bounds in eq. (7) reduce to the results in the standard OCO setting

Remark 2 (Regret Bounds in the Time-Invariant Domain Case). *When the domain set is time-invariant, i.e., $\mathcal{K}_1 = \cdots = \mathcal{K}_T$, we have $S_T = 0$ by definition. Hence, the dynamic regret bound in eq. (7) reduces to $\mathcal{O}\left(\sqrt{T}\left(1 + C_T\right)\right)$, i.e., it becomes equal to the dynamic regret bound of **OGD** in the standard OCO setting [25].*

VI. NUMERICAL EVALUATIONS

We compare **Safe-OGD** with the safe H_2 and H_∞ controllers in simulated scenarios of safe control of a quadrotor aiming to stay at a hovering position. We implement H_2 and H_∞ controllers based on [34, eqs. (2.15) & (2.19)] and use [13, Theorem 3] to account for safety constraints. We implement H_2 and H_∞ with three different horizons, *i.e.*, $N = 1$, 5, 10. Supplementary numerical experiments, that compare **Safe-OGD** with OCO-M controllers [17], are presented in Appendix A. Our code is open-sourced at: https://github.com/UM-iRaL/Non-Stochastic-Control.

Tested Noise Types. We corrupt the system dynamics with diverse noise drawn for the Gaussian, Uniform, Gamma, Beta, Exponential, or Weibull distribution.

Simulation Setup. We consider a quadrotor model with state vector its position and velocity, and control input its roll, pitch, and total thrust. The quadrotor's goal is to stay at a predefined hovering position. To this end, we focus on its linearized dynamics, taking the form

$$x_{t+1} = Ax_t + Bu_t + w_t \tag{8}$$

where

$$A = \begin{bmatrix} 1 & 0 & 0 & 0.1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} -\frac{4.91}{100} & 0 & 0 \\ 0 & \frac{4.91}{100} & 0 \\ 0 & 0 & \frac{1}{200} \\ -\frac{98.1}{100} & 0 & 0 \\ 0 & \frac{98.1}{100} & 0 \\ 0 & 0 & \frac{1}{10} \end{bmatrix}.$$

We choose the safety constraints:

$$\begin{aligned} -\mathbf{1}_{6\times 1} \leq x_t &\leq \mathbf{1}_{6\times 1}, \\ [-\pi \ -\pi \ -20]^\top \leq u_t &\leq [\pi \ \pi \ 20]^\top, \end{aligned} \tag{9}$$

and we assume noise such that $\|w_t\| \leq 0.1$ for all t.

We consider that the loss functions take the form of $c_t(x_{t+1}, u_t) = x_{t+1}^\top x_{t+1} + u_t^\top u_t$.

We simulate the setting for $T = 500$ time steps.

Remark 3 (Time-Varying Domain Set). *The domain set $\mathcal{K}_t$ for the quadrotor system is time-varying even though the safety constraints in eq. (9) are time-invariant, since $\mathcal{K}_t$ depends on the time-varying state x_t over T in eq. (5).*

Summary of Results. The simulation results are presented in Table I (cumulative loss performance) and Table II (running time). All methods ensure the safety constraints in eq. (9) are satisfied. Algorithm 1 demonstrates better performance in comparison to H_2 and H_∞ with $N = 1$ and $N = 5$ in terms of cumulative loss across the tested types of noise. H_2 and H_∞ with $N = 10$ incur lower cumulative loss than Algorithm 1. However, as shown in Table II, Algorithm 1 is computationally more efficient. Specifically, Algorithm 1 is 9 and 114 times faster than H_2 and H_∞ with $N = 10$ on average, respectively.

VII. CONCLUSION

We studied the problem of *Safe Non-Stochastic Control of Linear Dynamical Systems* (Problem 1), and provided the **Safe-OGD** algorithm that guarantees (i) zero constraint violation of convex time-varying constraints, and (ii) bounded dynamic regret against any linear time-varying control policy with safety guarantees (Theorem 1). We demonstrated that the dynamic regret bound of **Safe-OGD** reduces to that in the standard OCO setting [25] when the optimization domain is time-invariant (Remark 2).

We evaluated our algorithm in simulated scenarios of safe control of a quadrotor aiming to maintain a hovering position in the presence of unpredictable disturbances. We observed that the **Safe-OGD**-based controller achieved comparable cumulative loss and better computational time compared to safe H_2 and H_∞ controllers [13], [34].

Future Work. We will investigate the optimality of the regret bound of the **Safe-OGD** algorithm. We will also investigate conditions for the recursive feasibility of time-varying domain set $\mathcal{K}_t$. Further, we will apply the algorithm to real-world robotic systems (quadrotors) to demonstrate resilient online control against unpredictable wind. To this end, we will extend the algorithms to nonlinear systems.

REFERENCES

[1] E. Ackerman, "Amazon promises package delivery by drone: Is it for real?" *IEEE Spectrum, Web*, 2013.

[2] J. Chen, T. Liu, and S. Shen, "Tracking a moving target in cluttered environments using a quadrotor," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 446–453.

[3] A. Rivera, A. Villalobos, J. C. N. Monje, J. A. G. Mariñas, and C. M. Oppus, "Post-disaster rescue facility: Human detection and geolocation using aerial drones," in *IEEE 10 Conference*, 2016, pp. 384–386.

[4] J. B. Rawlings, D. Q. Mayne, and M. Diehl, *Model predictive control: Theory, computation, and design*. Nob Hill Publishing, 2017, vol. 2.

[5] F. Borrelli, A. Bemporad, and M. Morari, *Predictive control for linear and hybrid systems*. Cambridge University Press, 2017.

[6] O. Faltinsen, *Sea loads on ships and offshore structures*. Cambridge University Press, 1993, vol. 1.

[7] T. P. Sapsis, "Statistics of extreme events in fluid flows and waves," *Annual Reviews*, 2021.

[8] K. J. Åström, *Introduction to stochastic control theory*. Courier Corporation, 2012.

[9] F. Berkenkamp, "Safe exploration in reinforcement learning: Theory and applications in robotics," Ph.D. dissertation, ETH Zurich, 2019.

[10] G. Goel and B. Hassibi, "Regret-optimal control in dynamic environments," *arXiv preprint:2010.10473*, 2020.

TABLE I: Comparison of **Safe-OGD** with the safe H_2 and H_∞ controllers in terms of cumulative loss over.

Noise Distribution	Ours	$N = 1$		$N = 5$		$N = 10$	
		H_2	H_∞	H_2	H_∞	H_2	H_∞
Gaussian	44.05	61.81	93.44	47.96	52.03	30.66	48.69
Uniform	151.49	724.98	1859.61	331.32	323.42	100.21	53.86
Gamma	159.21	811.09	2082.12	372.52	364.26	112.90	60.77
Beta	186.98	836.41	2152.63	386.30	375.73	116.70	62.40
Exponential	126.69	552.73	1421.90	259.82	250.76	79.25	44.35
Weibull	195.71	873.09	2246.31	405.70	392.94	122.63	65.86
Average	142.50	643.35	1642.67	300.60	293.19	93.72	55.99
Standard Deviation	53.92	307.00	814.06	134.16	128.60	34.53	8.43

TABLE II: Comparison of **Safe-OGD** Algorithm 1 with the safe H_2 and H_∞ controllers in terms of computation time in seconds.

Noise Distribution	Ours	$N = 1$		$N = 5$		$N = 10$	
		H_2	H_∞	H_2	H_∞	H_2	H_∞
Average	0.1484	0.3712	0.6429	0.6033	1.3693	1.3854	17.0248
Standard Deviation	0.0342	0.0143	0.0116	0.0282	0.2741	0.0673	0.3691

[11] O. Sabag, G. Goel, S. Lale, and B. Hassibi, "Regret-optimal full-information control," *arXiv preprint:2105.01244*, 2021.
[12] G. Goel and B. Hassibi, "Regret-optimal measurement-feedback control," in *Learning for Dynamics and Control*, 2021, pp. 1270–1280.
[13] A. Martin, L. Furieri, F. Dörfler, J. Lygeros, and G. Ferrari-Trecate, "Safe control with minimal regret," in *Learning for Dynamics and Control Conference (L4DC)*, 2022, pp. 726–738.
[14] A. Didier, J. Sieber, and M. N. Zeilinger, "A system level approach to regret optimal control," *IEEE Control Systems Letters (L-CSS)*, 2022.
[15] H. Zhou and V. Tzoumas, "Safe perception-based control with minimal worst-case dynamic regret," *arXiv preprint arXiv:2208.08929*, 2022.
[16] E. Hazan, S. Kakade, and K. Singh, "The nonstochastic control problem," in *Algorithmic Learning Theory (ALT)*, 2020, pp. 408–421.
[17] N. Agarwal, B. Bullins, E. Hazan, S. Kakade, and K. Singh, "Online control with adversarial disturbances," in *International Conference on Machine Learning (ICML)*, 2019, pp. 111–119.
[18] Y. Li, S. Das, and N. Li, "Online optimal control with affine constraints," in *AAAI Conference on Artificial Intelligence (AAAI)*, vol. 35, no. 10, 2021, pp. 8527–8537.
[19] M. Simchowitz, K. Singh, and E. Hazan, "Improper learning for non-stochastic control," in *Conference on Learning Theory (COLT)*, 2020, pp. 3320–3436.
[20] P. Gradu, E. Hazan, and E. Minasyan, "Adaptive regret for control of time-varying dynamics," *arXiv preprint:2007.04393*, 2020.
[21] P. Zhao, Y.-H. Yan, Y.-X. Wang, and Z.-H. Zhou, "Non-stationary online learning with memory and non-stochastic control," *arXiv preprint arXiv:2102.03758*, 2021.
[22] H. Zhou, Z. Xu, and V. Tzoumas, "Efficient online learning with memory via frank-wolfe optimization: Algorithms with bounded dynamic regret and applications to control," *arXiv preprint arXiv:2301.00497*, 2023.
[23] S. Shalev-Shwartz *et al.*, "Online learning and online convex optimization," *Foundations and Trends® in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2012.
[24] E. Hazan *et al.*, "Introduction to online convex optimization," *Foundations and Trends in Optimization*, vol. 2, no. 3-4, pp. 157–325, 2016.
[25] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Interna. Conf. on Machine Learning (ICML)*, 2003, pp. 928–936.
[26] H. Zhou and V. Tzoumas, "Safe non-stochastic control of linear dynamical systems," *arXiv preprint arXiv:2308.12395*, 2023.

[27] L. Zhang, S. Lu, and Z.-H. Zhou, "Adaptive online learning in dynamic environments," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 31, 2018.
[28] X. Cao and K. R. Liu, "Online convex optimization with time-varying constraints and bandit feedback," *IEEE Transactions on automatic control*, vol. 64, no. 7, pp. 2665–2680, 2018.
[29] T. Chen and G. B. Giannakis, "Bandit convex optimization for scalable and dynamic iot management," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 1276–1286, 2018.
[30] X. Yi, X. Li, T. Yang, L. Xie, T. Chai, and K. H. Johansson, "Distributed bandit online convex optimization with time-varying coupled inequality constraints," *IEEE Transactions on Automatic Control*, vol. 66, no. 10, pp. 4620–4635, 2020.
[31] M. Mahdavi, R. Jin, and T. Yang, "Trading regret for efficiency: online convex optimization with long term constraints," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 2503–2528, 2012.
[32] G. Carnevale, A. Camisa, and G. Notarstefano, "Distributed online aggregative optimization for dynamic multi-robot coordination," *IEEE Transactions on Automatic Control*, 2022.
[33] S. Paternain and A. Ribeiro, "Online learning of feasible strategies in unknown environments," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2807–2822, 2016.
[34] J. Anderson, J. C. Doyle, S. H. Low, and N. Matni, "System level synthesis," *Annual Reviews in Control*, vol. 47, pp. 364–393, 2019.
[35] P. Zhao, Y.-X. Wang, and Z.-H. Zhou, "Non-stationary online learning with memory and non-stochastic control," in *International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR, 2022, pp. 2101–2133.

APPENDIX

A. Supplementary Numerical Experiments

In this experiment, we compare our algorithm with state-of-the-art OCO-M controller [17]. We showcase that online optimization with memory does not necessarily result in superior performance.

Compared Algorithms. We compare the **Safe-OGD**-based controller with the memory-based **DAC** [17] controller.

TABLE III: Comparison of the **Safe-OGD** and **DAC** [17] controllers with two step sizes in terms of cumulative loss for 1000 time steps —the blue numbers correspond to the best performance and the red numbers correspond to the worse.

Noise Distribution	Sinusoidal Weights (eq. (12))				Step Weights (eq. (13))			
	Ours		**DAC**		Ours		**DAC**	
	η_1	η_2	η_1	η_2	η_1	η_2	η_1	η_2
Gaussian	1769	1732	1838	1561	952	913	991	860
Uniform	2839	2822	2649	2428	1555	1538	1508	1352
Gamma	845	690	30323	8193	591	423	29252	4746
Beta	3518	3494	3045	2628	1921	1899	1795	1489
Exponential	1359	1252	54470	20273	866	726	44821	9122
Weibull	1732	1540	73100	7271	1332	1118	72005	4776
Average	2010	1922	27571	7059	1203	1103	25062	3724
Standard Deviation	988	1042	30623	7032	491	541	29282	3166

Simulation Setup. We follow a setup similar to [35]. We consider linear systems of the form

$$x_{t+1} = Ax_t + Bu_t + w_t, \tag{10}$$

where (i) $x_t \in \mathbb{R}^2$, (ii) $u_t \in \mathbb{R}$, and (iii) w_t and the elements of A and B are sampled from various distributions, *i.e.*, Gaussian, Uniform, Gamma, Beta, Exponential, or Weibull distributions. We consider linear time-invariant systems and impose constraints only on the control input. This induces a time-<u>in</u>variant domain set of optimization, as required by the **DAC** controller [17]. Specifically, we use the control constraint $L_u u \leq l_u$, *i.e.*, $-L_u K x_t \leq l_u$. If we upper bound x_t with upper bound D' achieved by the OCO-M controller [17, Lemma 5.5], then the optimization domain in Lemma 2 becomes time-<u>in</u>variant, specifically,

$$\mathcal{K} \triangleq \{K \mid -L_u D' K \leq l_u, \|K\| \leq \kappa, \|A - BK\| \leq 1 - \gamma\}. \tag{11}$$

We compare the **Safe-OGD** and **DAC** controllers across two different step sizes η_1 and η_2 to investigate how the step sizes affect their performance. The **DAC** controller has a memory length of 10.

The loss function has the form $c_t(x_t, u_t) = q_t x_t^\top x_t + r_t u_t^\top u_t$, where q_t, $r_t \in \mathbb{R}$ are time-varying weights. Particularly, we consider the following two cases:

1) Sinusoidal weights defined as

$$q_t = \sin(t/10\pi), \ r_t = \sin(t/20\pi). \tag{12}$$

2) Step weights defined as

$$(q_t, r_t) = \begin{cases} \left(\frac{\log(2)}{2}, 1\right), & t \leq T/5, \\ (1, 1), & T/5 < t \leq 2T/5, \\ \left(\frac{\log(2)}{2}, \frac{\log(2)}{2}\right), & 2T/5 < t \leq 3T/5, \\ \left(1, \frac{\log(2)}{2}\right), & 3T/5 < t \leq 4T/5, \\ \left(\frac{\log(2)}{2}, 1\right), & 4T/5 < t \leq T. \end{cases} \tag{13}$$

Results. The results are summarized in Table III, showing that **Safe-OGD** outperforms **DAC** in terms of the average and standard deviation of cumulative loss. In more detail, **Safe-OGD** has comparable performance to **DAC** under Gaussian, Uniform, and Beta distributions, and is better under Gamma, Exponential, and Weibull distributions. We hypothesize that the reason for the latter is that the **DAC** controller minimizes a truncated unary loss, instead of the actual loss. In addition, the performance of **DAC** heavily relies on step size tuning, *e.g.*, under Gamma and Weibull distributions, as demonstrated by the large difference in cumulative loss across η_1 and η_2. By contrast, the cumulative loss of **Safe-OGD** varies less as we change the step size.