On the Power of Optimism in Constrained Online Convex Optimization

Haobo Zhang, **Hengqun Guo** and **Xin Liu***
ShanghaiTech University, Shanghai, China
{zhanghb2023, guohq, liuxin7}@shanghaitech.edu.cn

Abstract

This paper studies the constrained online convex optimization problem (COCO) where the learner makes sequential decisions within a constrained set. We present Optimistic-COCO, an adaptive gradient-based algorithm that incorporates optimistic design with the Lyapunov optimization technique. The proposed algorithm achieves strong theoretical guarantees: 1) Optimistic-COCO provides a tight gradient-variation regret bound and constant constraint violation; 2) Optimistic-COCO is environment-agnostic, utilizing adaptive learning rates that rely solely on causal information. These results resolve an open question posed in prior work regarding whether an adaptive algorithm can achieve problem-dependent regret and constant constraint violation in COCO. We establish these robust guarantees through carefully designed adaptive parameters and a refined multi-step Lyapunov drift analysis. Experimental results further validate our theoretical findings, demonstrating the practical efficacy of the proposed algorithm.

1 Introduction

Constrained Online Convex Optimization (COCO) is a generalization of the online convex optimization framework (OCO) and has become increasingly popular as more online decision-making problems require balancing objectives with adherence to operational constraints. For example, in an advertisement platform [Goldfarb and Tucker, 2011], the advertisers aim to maximize the click/conversion rates within the weekly or monthly budgets; in the network traffic management [Mannor and Tsitsiklis, 2006], the operator aims to minimize the latency of network flows within the network's capacity limits; in the portfolio management [Li and Hoi, 2014], the investor seeks to maximize returns while adhering to risk constraints and maintaining a balanced allocation across various classes. In COCO, at each round t, the learner makes a decision x_t from a constrained set and then observes the full information of $f_t(\cdot)$. The goal is to minimize regret $\sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(x^*)$, while satisfying $g(x_t) \leq 0,\ \forall t \in [T]$, where x^* is the optimal solution to the offline problem

$$\min_{x\in\mathcal{X}_0} \sum_{t=1}^{T} f_t(x) \quad s.t.\ \ g(x) \leq 0.$$

A straightforward solution to COCO is a projection-based gradient descent method [Hazan and others, 2016], which projects the decision into the feasible region at each round defined by

$$\mathcal{X} = \{x \in \mathcal{X}_0 \mid g(x) \leq 0\}.$$

However, the projection operator is often computationally burdensome and might be impractical, especially when the constraint set is complicated (i.e., g is a complex function). To address the challenge, [Mahdavi *et al.*, 2012; Yu and Neely, 2020; Qiu *et al.*, 2023] studied slightly relaxed *long-term* constraints, where the constraints are allowed to be violated and as long as they are satisfied over the long term. In other words, the constraint violation over T rounds $\mathcal{V}(T) := \sum_{t=1}^{T} g(x_t)$ should be minimal. In this problem setting, [Mahdavi *et al.*, 2012] proposes a regularized primal-dual subgradient algorithm and achieves $\mathcal{O}(\sqrt{T})$ regret and $\mathcal{O}(T^{3/4})$ violation bound. The regret and violation performance is greatly improved to $\mathcal{O}(\sqrt{T})$ and $\mathcal{O}(1)$ in [Yu and Neely, 2020] with a novel "drift-plus-penalty" framework. However, these methods present their results with respect to T, which only leads to worst-case performance. But, intuitively, when the loss functions change slowly (or even remain static) in a benign environment, one would expect a better regret than the $O(\sqrt{T})$ worst-case regret.

To account for this, [Chiang *et al.*, 2012] introduced a type of gradient variation defined as

$$V_T^{\max} = \sum_{t=1}^{T} \max_{x\in\mathcal{X}_0} \|\nabla f_t(x) - \nabla f_{t-1}(x)\|^2,$$

which measures the dynamics of the environment. Using this definition, [Chiang *et al.*, 2012] achieves a problem-dependent regret of $\mathcal{O}(\sqrt{V_T^{\max}})$ in OCO through an optimistic-type two-step mirror descent algorithm. This approach captures the worst-case $\mathcal{O}(\sqrt{T})$ regret, but also improves the regret bound when the variations are small, offering better guarantees than the traditional $\mathcal{O}(\sqrt{T})$ bound.

*Corresponding Author

Algorithm	Regret	Violation	Env-Agnostic
[Mahdavi *et al.*, 2012]	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(T^{3/4})$	✗
[Yu *et al.*, 2017]	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	✗
[Yu and Neely, 2020]	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(1)$	✗
[Qiu *et al.*, 2023]	$\mathcal{O}(\sqrt{V_T^{\max}})$	$\mathcal{O}(1)$	✗
Optimistic-COCO	$\boldsymbol{\mathcal{O}(\sqrt{V_t})}$	$\boldsymbol{\mathcal{O}(1)}$	✓

Table 1: Comparison with related works. 'Env-Agnostic' indicates that the algorithm does not require prior knowledge of non-causal information such as T or V_T. We define the problem-dependent polynomial $V_t := \sum_{s=1}^t \|\nabla f_s(x_s) - \nabla f_{s-1}(x_s)\|^2$ which captures the dynamics of the problem. Note that, different from [Chiang *et al.*, 2012] and [Qiu *et al.*, 2023], we define the gradient variation V_t on the current decision x_t instead of the maximization over set $\mathcal{X}_0$ in $V_t^{\max}$. This definition is more relaxed since $\|\nabla f_t(x_t) - \nabla f_{t-1}(x_t)\|^2$ can be zero in some setting, while $\max_{x\in\mathcal{X}_0} \|\nabla f_t(x) - \nabla f_{t-1}(x)\|^2$ would be large. The gradient-variation bound was first achieved in COCO by [Qiu *et al.*, 2023], but their approach requires prior knowledge of the gradient variation $V_T^{\max}$. Optimistic-COCO introduces an adaptive design that provides guarantees of any time regret and violation without needing such prior information, operating effectively under an environment-agnostic setting.

More recently, [Qiu *et al.*, 2023] studied OCO with constraints and established both a problem-dependent regret and an $\mathcal{O}(1)$ violation bound, by assuming the non-causal information of $V_T^{\max}$.

However, the existing approaches are not adaptive or environment-agnostic. They rely on prior knowledge of the total number of rounds, T, or more challenging, the total gradient variations of the loss functions, $V_T^{\max}$, to properly schedule the learning rate or other penalty parameters in the algorithm. This assumption is overly strict in practice because the learner can only observe the current loss function (or the casual cumulative gradient variation $V_t^{\max}$). Estimating the full variation of the problem at the beginning is extremely challenging, if not impossible. Therefore, an open question raised by [Qiu *et al.*, 2023] is both critical and challenging:

Is it possible to design an adaptive algorithm for COCO such that it achieves a problem-dependent regret and constant violation without any prior knowledge of the key environment parameters?

In this paper, we provide a positive answer to this question by introducing Optimistic-COCO, an efficient first-order primal-dual method that incorporates the optimistic gradient descent and Lyapunov optimization technique. Our contributions are summarized as follows:

- **Algorithm Design:** The design of Optimistic-COCO is inspired by optimistic gradient descent method and the Lyapunov optimization technique. For the primal domain, Optimistic-COCO achieves *fully adaptive* by leveraging historical information through optimistic gradient descent that captures the variation of the environment and well-designed adaptive parameters to control the loss-constraint tradeoff. For the dual domain, it designs an alternative virtual queue update to further track the constraint violation. Our algorithm does not require any prior knowledge of the total gradient variation and time horizon. It is adaptive to the non-casual variation and can effectively balance regret and constraint violation.
- **Theoretical guarantees:** Optimistic-COCO establishes anytime regret and violation bounds of $\mathcal{R}(t) = \mathcal{O}(\sqrt{V_t})$ and $\mathcal{V}(t) = \mathcal{O}(1)$, where the gradient violation is $V_t := \sum_{s=1}^t \|\nabla f_s(x_s) - \nabla f_{s-1}(x_s)\|^2$. This not only addresses the open question in [Qiu *et al.*, 2023] regarding whether an adaptive and non-casual algorithm can achieve a problem-dependent regret and constant violation but also improves the regret to a relaxed one $O(\sqrt{V_t})$ (against $O(\sqrt{V_T^{\max}})$). We derive these strong results through the carefully designed virtual queue, adaptive trade-off parameters, and a refined multi-step Lyapunov drift analysis of the virtual queue process. Specifically, we establish a relationship between the trade-off parameter ϕ_t and the gradient variation V_t and prove that it greatly enhances our bounds. A detailed comparison to the most related results is summarized in Table 1.
- **Experiments:** We evaluate Optimistic-COCO using synthetic experiments in both dynamic and slowly changing environments. Our experimental results demonstrate that Optimistic-COCO outperforms baseline algorithms in both settings (even the algorithm in [Qiu *et al.*, 2023] that uses the non-casual knowledge of the total gradient variation). Moreover, it is observed that in a slowly changing environment, Optimistic-COCO achieves a much better performance than other traditional worst-case algorithms. This proves our algorithm's effectiveness in different environments and demonstrates the theoretical results.

1.1 Related Work

Online convex optimization is a broad field that has been extensively studied with numerous variants [Chen *et al.*, 2017; Yuan and Lamperski, 2018; Mokhtari *et al.*, 2020; Yi *et al.*, 2021; Guo *et al.*, 2022; Sinha and Vaze, 2024; Liu *et al.*, 2024]. In this paper, we focus on the works most relevant to ours.

"Optimism" in OCO: The concept of optimism in online convex optimization has been widely studied in [Hazan and Kale, 2010; Chiang *et al.*, 2012; Rakhlin and Sridharan, 2013; Yang *et al.*, 2014] to establish the problem-dependent regret beyond the traditional worst-case guarantee. The type of gradient variation result is first derived in [Hazan and Kale, 2010] for online linear optimization with the optimistic-style algorithm. [Chiang *et al.*, 2012] studies linear and smooth loss functions to achieve $\mathcal{O}(\sqrt{V_T^{\max}})$ regret bound. [Rakhlin and Sridharan, 2013] generalizes the results in [Hazan and Kale, 2010] by introducing optimism in Follow the Regularized Leader and Online Mirror Descent. [Yang *et al.*, 2014] extends results in [Chiang *et al.*, 2012] to the non-smooth setting and provides a guarantee of $\mathcal{O}(\sqrt{V_T^{\max}})$ in the general non-Euclidean space. These results leverage the power of "optimism", though it remains uncertain whether this approach can be effectively applied to constrained settings to achieve strong performance.

COCO: Constrained online convex optimization was initialized by [Mahdavi *et al.*, 2012] where a regularized primal-dual algorithm is proposed to achieve a $\mathcal{O}(\sqrt{T})$ regret bound and a $\mathcal{O}(T^{3/4})$ violation bound. It can also achieve an $\mathcal{O}(T^{2/3})$ bound for both regret and violation when the constraint function $g(x)$ is linear. [Jenatton *et al.*, 2016] extends these results to the setting that the total round T is unknown, providing regret and violation bound of $\mathcal{O}\left(T^{\max\{c,1-c\}}\right)$ and $\mathcal{O}\left(T^{1-c/2}\right)$, respectively, with $c \in (0,1)$ as a trade-off parameter. But this result is much worse than ours. [Yu *et al.*, 2017] shows $\mathcal{O}(\sqrt{T})$ bound both in regret and constraint violation, but they study a different constraint setting, which is stochastic with the environment. When the Slater's condition holds, [Yu and Neely, 2020] proposes a "drift-plus-penalty" framework to achieve $\mathcal{O}(\sqrt{T})$ regret and $\mathcal{O}(1)$ violation bound. However, these results fail to capture the variation of the loss functions, making them unable to achieve improved performance even when the environment is benign. Motivated by the optimism in OCO, a primal-dual gradient descent-ascent algorithm is introduced in [Qiu *et al.*, 2023] that achieves problem-dependent $\mathcal{O}(\sqrt{V_T^{\max}})$ regret and $\mathcal{O}(1)$ violation bound. Recently, the paper [Lekeufack and Jordan, 2024] extends [Sinha and Vaze, 2024] to the setting of both adversarial and fixed constraints by incorporating an optimistic design on a surrogate Lagrange function. While they can achieve $\mathcal{O}(\sqrt{V_T})$ regret and $\mathcal{O}(\log T)$ violation bound in the fixed constraint setting, they focus on hard constraint violation and require prior knowledge of the total gradient variation V_T to schedule the learning rates in the algorithms carefully. As we discussed before, obtaining such non-casual information is usually infeasible or even impossible in the practical system.

2 Constrained Online Convex Optimization

In this section, we formally define the COCO problem and introduce necessary assumptions. Let $\mathcal{X}_0 \in \mathbb{R}^d$ be a closed convex set. We define stacked vector $\boldsymbol{g}(x)$ as $[g_1(x), g_2(x), \cdots, g_m(x)]^\top$, where each $g_i(x)$ is a scalar-value constraint function evaluated at the input x and m is the number of constraints. Assume $f_t(x), \forall t \in [T]$ and $g_i(x), \forall i \in [m]$ are convex and continuous functions. When not specified, we consider $\|\cdot\|$ to be an L_2-norm.

We define the function $h(x)$ as α-strongly convex if $h(x)$ satisfies

$$h(x) \leq h(y) + \nabla h(x)^\top (x-y) - \frac{\alpha}{2}\|x-y\|^2.$$

for any $x, y \in \mathbb{R}^d$. The cumulative gradient variation at round t is defined as

$$V_t := \sum_{s=1}^{t} \|\nabla f_s(x_s) - \nabla f_{s-1}(x_s)\|^2. \tag{1}$$

For ease of exposition, we also denote $V_t = \sum_{s=1}^t \|\nabla f_s - \nabla f_{s-1}\|^2$.

In COCO, at each round t, the learner selects a decision x_t from $\mathcal{X}_0$ and then observes the full information about $f_t(x)$. To quantify the performance of generated sequence $\{x_1, x_2, \cdots, x_T\}$ by an algorithm, we compare them with a static baseline, which is the solution to the following offline COCO problem:

$$\min_{x\in\mathcal{X}_0} \sum_{t=1}^{T} f_t(x) \ \text{ s.t. } g_i(x) \leq 0, \ \forall i \in [m]. \tag{2}$$

Let x^* be the optimal solution to (2), define the anytime regret and violation at round t as follows:

$$\mathcal{R}(t) := \sum_{s=1}^{t} f_s(x_s) - \sum_{s=1}^{t} f_s(x^*) \tag{3}$$

$$\mathcal{V}(t) := \max_i \sum_{s=1}^{t} g_i(x_s). \tag{4}$$

Before introducing our algorithm and theoretical results, we present the following standard technical assumptions in COCO.

Assumption 1. *The decision set $\mathcal{X}_0$ is convex and bounded with diameter D such that $\|x - x'\| \leq D, \forall x, x' \in \mathcal{X}_0$.*

Assumption 2. *The loss function $f_t(x)$ satisfies $|f_t(x) - f_t(x')| \leq F\|x - x'\|$, $\|\nabla f_t(x) - \nabla f_t(x')\| \leq F\|x - x'\|$, $\forall x, x' \in \mathcal{X}_0, t \in [T]$.*

Assumption 3. *The constraint function $\boldsymbol{g}(x)$ satisfies $\forall x, x' \in \mathcal{X}_0$, $\|\boldsymbol{g}(x) - \boldsymbol{g}(x')\| \leq G\|x - x'\|$, $\|\boldsymbol{g}(x)\| \leq R$, and $\|\nabla g_i(x) - \nabla g_i(x')\| \leq G\|x - x'\|$, $\forall i \in [m]$.*

Assumptions 2 and 3 indicate the Lipschitz continuous of functions and function gradients of both loss and constraint functions. The bound of the constraint function follows directly from the continuity of $\boldsymbol{g}(x)$ and Assumption 1.

Assumption 4. *There exists a point $\hat{x} \in \mathcal{X}_0$ and a positive constant $\delta > 0$ such that $g_i(\hat{x}) \leq -\delta$, $\forall i \in [m]$.*

This assumption, known as Slater's condition [Boyd and Vandenberghe, 2004], indicates the existence of a strictly feasible solution to the constrained problem and is crucial for establishing a constant violation bound. However, our algorithm is fully adaptive and does not require prior knowledge of δ for implementation; it is introduced solely to facilitate the analysis. Finally, note that these assumptions are common in online learning and easy to satisfy in practice [Cesa-Bianchi and Lugosi, 2006; Hazan *et al.*, 2007].

3 Optimistic-COCO

In this section, we introduce Optimistic-COCO, an efficient optimistic gradient descent algorithm based on Lyapunov optimization techniques. To present our algorithm, we first define the approximated Lagrange function as

$$L_t(x_t, \boldsymbol{Q}_t) = f_t(x_t) + \phi_{t-1}\boldsymbol{Q}_t^\top \boldsymbol{g}(x_t),$$

where virtual queue vector $\boldsymbol{Q}_t = [Q_t^1, Q_t^2, \cdots, Q_t^m]^\top$ represents the dual variable that controls constraint violation of decisions. In Optimistic-COCO, at each round t, it first computes the optimistic gradient θ_t and optimizes a surrogate decision function w.r.t. x. After submitting decision x_{t+1} and observing the loss function, the algorithm calculates the causal gradient variation V_{t+1} to update the trade-off parameter ϕ_{t+1} and learning rate η_{t+1}. The virtual queue is then updated based on the current violation. We summarize Optimistic-COCO in Algorithm 1.

Algorithm 1 Optimistic-COCO

Initialization: $x_0,\ x_1 \in \mathcal{X}_0,\ \boldsymbol{Q}_0 = \boldsymbol{0},\ L_0(x) = L_1(x) = 0,\ \forall x \in \mathcal{X}_0$, adaptive learning rates $\eta_t = \Theta(\frac{1}{\sqrt{V_t}})$ and $\phi_t = \Theta(\sqrt{V_t})$.
for $t = 1, \cdots, T$ **do**
    Calculate the optimistic gradient:

$$\theta_t = 2\nabla_x L_t(x_t, \boldsymbol{Q}_t) - \nabla_x L_{t-1}(x_{t-1}, \boldsymbol{Q}_{t-1}). \quad (5)$$

    Take adaptive gradient decision:

$$x_{t+1} = \underset{x\in\mathcal{X}_0}{\arg\min}\ \langle \theta_t, x\rangle + \frac{1}{2\eta_t}\|x - x_t\|^2. \quad (6)$$

    Observe f_{t+1} **and calculate:** ϕ_{t+1} and η_{t+1}.
    Update virtual queue:

$$\boldsymbol{Q}_{t+1} = [\boldsymbol{Q}_t + \boldsymbol{g}(x_{t+1})]^+ . \quad (7)$$

end for

To further explain the underlying intuition in the algorithm, we recall the offline COCO problem in (2) and introduce a Lagrange function to it:

$$L_t(x, \boldsymbol{\lambda}) = f_t(x) + \boldsymbol{\lambda}^\top \boldsymbol{g}(x),$$

where $\boldsymbol{\lambda}$ is the Lagrange multiplier associated with constraint functions. Since f_t is observed after making the decision x_t in online optimization. An efficient and classical method is to approximate L_t with first-order estimation that:

$$L_{t-1}(x_{t-1}) + \langle \nabla L_{t-1}(x_{t-1}), x - x_{t-1}\rangle.$$

The problem can then be solved by applying gradient descent to both the primal and dual updates, which has been proven to ensure an $\mathcal{O}(\sqrt{T})$ regret guarantee, as shown in [Mahdavi *et al.*, 2012; Yu and Neely, 2020]. However, to achieve an improved problem-dependent guarantee, we require a more effective estimator capable of detecting environmental changes, along with a refined dual design to balance the tradeoff between loss and constraints.

In the primal domain, inspired by the concept of "optimism" in [Rakhlin and Sridharan, 2013], we design an optimistic gradient estimator:

$$\theta_t = 2\nabla_x L_t(x_t, \boldsymbol{Q}_t) - \nabla_x L_{t-1}(x_{t-1}, \boldsymbol{Q}_{t-1}).$$

Rather than relying solely on gradient information from the previous round, this estimator incorporates the momentum of gradients, accounting for function variation. With this updated gradient, we can compute the decision for the next time step by optimizing the surrogate function in (6). Note that this equation does not require computing the minimum of any function, but instead involves a gradient-based descent, which is computationally efficient.

In the decision in (6), we also introduce a regularization (or smoothing) term $\frac{1}{2\eta_t}\|x - x_t\|^2$ to stabilize the algorithm, where $\eta_t = \Theta(\frac{1}{\sqrt{V_t}})$ represents the learning rate. Intuitively, when the environment changes slowly, it incurs a smaller η_t, encouraging the algorithm to make more stable decisions that prioritize exploitation. Moreover, since we have no prior knowledge of the total number of rounds T or the total gradient variation V_T, the design of the adaptive learning rate should be carefully chosen. The specific value of η_t will be further discussed in the theorem statements in the following section. It is important to note that this design is non-trivial in constrained optimization, where the loss functions, constraints, and dual variables (or virtual queues) are all interdependent, making the analysis challenging.

For the dual domain, we employ the virtual queues $\boldsymbol{Q}$ to approximate the dual variables. Intuitively, $\boldsymbol{Q}$ captures the cumulative violation and encourages more conservative decisions when the queue length grows large. The traditional update rule of virtual queues also imposed "optimism" in the dual domain to achieve *constant* constraint violation, however, we design a novel update for the virtual queue in Optimistic-COCO which differs from previous approaches in two aspects:

- Our algorithm leverages the unique feature of alternative updating in networking systems such that we can leverage real queues as the signal to estimate dual variables directly. Specifically, at time slot t, we can observe the feedback from the current decision x_{t+1} and use it to update the queue length. In contrast, traditional "optimistic" optimization methods [Yu and Neely, 2020; Qiu *et al.*, 2023] update both the primal variable x_{t+1} and the dual variable $\boldsymbol{Q}_{t+1}$ simultaneously, relying only on the previous decision x_t.
- We do not introduce a trade-off parameter in the constraint function $\boldsymbol{g}(x)$ when updating the virtual queue, as seen in previous works. Instead, the adaptive trade-off parameter $\phi_t = \Theta(\sqrt{V_t})$, which balances regret and violation, is only applied to $\boldsymbol{g}(x)$ when making decisions. In the next section, we will show that simply adding the trade-off parameter to the virtual queue update, as in [Yu and Neely, 2020] and [Qiu *et al.*, 2023], undermines the algorithm's ability to achieve real adaptiveness.

Remark 1. *We note that the definition of gradient variation used in this paper* $\left(V_T = \sum_{t=1}^T \|\nabla f_t(x_t) - \nabla f_{t-1}(x_t)\|^2\right)$ *is more relaxed and practical. Compared to the previous definition* $\left(V_T^{max} = \sum_{t=1}^T \max_{x\in\mathcal{X}_0} \|\nabla f_t(x) - \nabla f_{t-1}(x)\|^2\right)$ *in [Chiang* et al.*, 2012] and [Qiu* et al.*, 2023], it is not necessary to solve a maximization problem in each round with the* *full knowledge of gradient functions. While the computation is simple when f is linear or quadratic, it might become even more time-consuming than the algorithm's runtime when the loss function f is complex. However, with our definition and analysis, the algorithm only needs to compute the gradient variation at the current decision x_t in each round, thereby avoiding the heavy computational overhead in the existing formulation.*

4 Main Results

In this section, we first present our main theoretical results on the bound of regret and constraint violation.

Theorem 1. *Let the parameters be*

$$\phi_t = \max\left\{\sqrt{V_t} + \|\boldsymbol{Q}_t\| + 2F + R, \phi_{t-1}\right\}$$
$$\frac{1}{2\eta_t} = \max\left\{4G\phi_t\|\boldsymbol{Q}_t\| + 4M^2\phi_t + 4F, \frac{1}{2\eta_{t-1}}\right\}$$

Under Assumptions 1-4, Optimistic-COCO algorithm achieves the following regret and violation for any time $t \in [T]$

$$\mathcal{R}(t) = \mathcal{O}(\sqrt{V_t}),$$
$$\mathcal{V}(t) = \mathcal{O}\left(1/\delta\right) = \mathcal{O}(1).$$

where δ is the Slater's constant.

This result implies that Optimistic-COCO can achieve problem-dependent regret while maintaining constant constraint violation. Here, V_t represents the gradient variation that captures the problem's dynamics. When the environment is benign and the loss function remains fixed, this regret bound indicates no gap between the offline optimal solution and the online decision (since $\|\nabla f_t(x_t) - \nabla f_{t-1}(x_t)\| = 0$), whereas the traditional $\mathcal{O}(\sqrt{T})$ regret still incurs a sub-linear gap. Furthermore, since the gradient of loss functions is bounded according to Assumption 2, our results also generalize the previous $\mathcal{O}(\sqrt{T})$ bound for worst-case scenarios, where constant variation persists across all rounds. Therefore, our problem-dependent variation bound reflects the dynamic and complexity of the problem, providing a stronger guarantee for the algorithm.

In Optimistic-COCO, a non-increasing learning rate is employed to guide decision-making. The algorithm is fully adaptive and does not rely on prior knowledge of the environment, such as V_T or T, unlike previous approaches [Yu and Neely, 2020; Qiu *et al.*, 2023]. Instead, the learning rate increases progressively based on the accumulated gradient variation over time. Initially, when the gradient variation is small, the algorithm favors more aggressive exploration. As the experiment progresses, the algorithm transitions towards more stable behavior. Additionally, when the loss function is fixed, the algorithm maintains a constant learning rate by the definition of η_t, which is consistent with the optimization of a known convex problem. This constant learning rate also prevents excessively large updates at the start of the iterations. Furthermore, the trade-off parameter is adaptive, and its strategic placement is critical to achieving the algorithm's adaptivity and ensuring that constant constraint violation is maintained.

Proof of Theorem 1

We give some proof sketches of our theoretical results in this section and leave the full proof in the Appendix. To prove our main theorem, we first introduce the following critical lemma that provides a unified bound on both regret and drift.

Lemma 1. *Under the Optimistic-COCO algorithm, we have for any $x \in \mathcal{X}_0$ such that*

$$f_{t+1}(x_{t+1}) - f_{t+1}(x) + \frac{\phi_t}{2}\|\boldsymbol{Q}_{t+1}\|^2 - \frac{\phi_t}{2}\|\boldsymbol{Q}_t\|^2$$
$$\leq \phi_t \boldsymbol{Q}_{t+1}^\top \boldsymbol{g}(x) - \frac{\phi_t}{2}\|\boldsymbol{Q}_{t+1} - \boldsymbol{Q}_t\|^2 + \langle \epsilon_t, x_{t+1} - x\rangle$$
$$+ \frac{1}{2\eta_t}\|x - x_t\|^2 - \frac{1}{2\eta_t}\|x - x_{t+1}\|^2 - \frac{1}{2\eta_t}\|x_{t+1} - x_t\|^2.$$

The proof of Theorem 1 follows this key lemma which provides a bound on the one-step regret and Lyapunov drift as a whole. Note this lemma holds for all $x \in \mathcal{X}_0$ including the optimal decision x^*, which facilitates the subsequent proofs, so we highlight some key proof steps here.

Proof. For the Optimistic-COCO decision, we define

$$h(x) := \langle \theta_t, x\rangle + \frac{1}{2\eta_t}\|x - x_t\|^2, \;\; \forall x \in \mathcal{X}_0.$$

It is not hard to check $h(x)$ is $1/\eta_t-$strongly convex due to the quadratic term. Recall that x_{t+1} is the optimal solution to $h(x)$ in Optimistic-COCO, we have for any $x \in \mathcal{X}_0$

$$\langle \theta_t, x_{t+1}\rangle + \frac{1}{2\eta_t}\|x_{t+1} - x_t\|^2 \qquad (8)$$
$$\leq \langle \theta_t, x\rangle + \frac{1}{2\eta_t}\|x - x_t\|^2 - \frac{1}{2\eta_t}\|x - x_{t+1}\|^2.$$

which is one of the properties of the strongly convex function. Further define an error term $\epsilon_t = \nabla L_{t+1}(x_{t+1}, \boldsymbol{Q}_{t+1}) - \theta_t$ and reformulate (8) as

$$\langle \nabla_x L_{t+1}(x_{t+1}, \boldsymbol{Q}_{t+1}), x_{t+1} - x\rangle + \frac{1}{2\eta_t}\|x_{t+1} - x_t\|^2$$
$$\leq \langle \epsilon_t, x_{t+1} - x\rangle + \frac{1}{2\eta_t}\|x - x_t\|^2 - \frac{1}{2\eta_t}\|x - x_{t+1}\|^2.$$

Combine it with the convexity of $L_{t+1}(x, \boldsymbol{Q})$ w.r.t. x, we have the inequality

$$f_{t+1}(x_{t+1}) + \phi_t \boldsymbol{Q}_{t+1}^\top \boldsymbol{g}(x_{t+1}) - f_{t+1}(x) - \phi_t \boldsymbol{Q}_{t+1}^\top \boldsymbol{g}(x)$$
$$\leq \langle \epsilon_t, x_{t+1} - x\rangle + \frac{1}{2\eta_t}\|x - x_t\|^2 - \frac{1}{2\eta_t}\|x - x_{t+1}\|^2$$
$$- \frac{1}{2\eta_t}\|x_{t+1} - x_t\|^2. \qquad (9)$$

To cancel the term $\boldsymbol{Q}_{t+1}^\top \boldsymbol{g}(x_{t+1})$, define a function of the queue dynamic

$$h(\boldsymbol{Q}) := -\boldsymbol{Q}^\top \boldsymbol{g}(x_{t+1}) + \frac{1}{2}\|\boldsymbol{Q} - \boldsymbol{Q}_t\|^2, \;\; \boldsymbol{Q} \geq \boldsymbol{0}.$$

Since $\boldsymbol{Q}_{t+1}$ is the optimal solution to $h(\boldsymbol{Q})$ in Optimistic-COCO, we have for any $\boldsymbol{Q} \geq \boldsymbol{0}$

$$\begin{aligned} &-\boldsymbol{Q}_{t+1}^{\top}\boldsymbol{g}(x_{t+1}) + \frac{1}{2}\|\boldsymbol{Q}_{t+1} - \boldsymbol{Q}_t\|^2 \\ \leq &-\boldsymbol{Q}^{\top}\boldsymbol{g}(x_{t+1}) + \frac{1}{2}\|\boldsymbol{Q} - \boldsymbol{Q}_t\|^2 - \frac{1}{2}\|\boldsymbol{Q} - \boldsymbol{Q}_{t+1}\|^2. \end{aligned}$$

Setting $\boldsymbol{Q} = \boldsymbol{0}$ yields

$$\boldsymbol{Q}_{t+1}^{\top}\boldsymbol{g}(x_{t+1}) \geq \frac{1}{2}\|\boldsymbol{Q}_{t+1}\|^2 - \frac{1}{2}\|\boldsymbol{Q}_t\|^2 + \frac{1}{2}\|\boldsymbol{Q}_{t+1} - \boldsymbol{Q}_t\|^2.$$

Plugging it in inequality (9) and re-arranging the terms completes the proof. □

Constraint Violation

Following the virtual queue update rule in Optimistic-COCO, we directly have

$$\mathcal{V}(t) = \max_i \sum_{s=1}^{t} g_i(x_s) \leq \|\boldsymbol{Q}_t\|, \ \forall\, t \in [T]. \tag{10}$$

which implies that the constraint violation of Optimistic-COCO is bounded by the value of virtual queue and the problem reduces to determining the bound of the virtual queue. We remark that, in previous works [Yu and Neely, 2020; Qiu *et al.*, 2023], their algorithms typically introduced a trade-off parameter ϕ into the constraint function when updating the virtual queue $\boldsymbol{Q}_t$, leading to a violation bound of $\|\boldsymbol{Q}_t\|/\phi$. This method is valid when the total gradient variation V_T and time horizon T are known to set ϕ. However, when these parameters are unknown, incorporating an adaptive parameter ϕ_t into the constraint function invalidates inequality (10), making it impossible to use virtual queue to establish the violation bound. Therefore, in our algorithm, we do not impose ϕ_t on the constraint function during the virtual queue update and employ a "Lyapunov drift" analysis to Lemma 1 to derive the bound on $\|\boldsymbol{Q}_t\|$.

Lyapunov drift analysis is used to investigate the stability properties of control policies where a stable policy results in bounded queue lengths. In our analysis, when Slater's condition holds as stated in Assumption 4, we study the upper bound of queue length through a refined multiple-step Lyapunov drift analysis.

Lemma 2. *Set $x = \hat{x}$ satisfy Slater's condition in Lemma 1, i.e., $g_i(\hat{x}) \leq -\delta$ $(\delta > 0)$, $\forall i \in [m]$. Under Assumptions 1-4, there exists a positive constant C and an integer K such that the following multi-step Lyapunov drift holds:*

$$\begin{aligned} &\frac{1}{2}\|\boldsymbol{Q}_{t+K}\|^2 - \frac{1}{2}\|\boldsymbol{Q}_t\|^2 \\ \leq &(-K\delta + G + D^2)\|\boldsymbol{Q}_t\| + K^2\delta/2 + KC. \end{aligned}$$

The derivation of the Lyapunov drift over K steps is based on the one-step drift. Specifically, in the proof for a single step, we bound the error term $\langle \epsilon_t, x_{t+1} - x\rangle$ using both constant terms and quadratic terms like $\|x_{t+1} - x\|^2$, with the objective of applying a telescoping summation across K steps. Another key point in obtaining the (K)-step Lyapunov drift lies in determining the values of ϕ_t and η_t, which control the behavior of the drift.

This lemma implies that if the norm of the virtual queue satisfies $\|\boldsymbol{Q}_t\| \geq \frac{K^2\delta + KC}{K\delta - G - D^2}$, the drift becomes negative, leading to a decrease in the virtual queue norm over the next K steps, i.e., $\|\boldsymbol{Q}_{t+K}\|^2 \leq \|\boldsymbol{Q}_t\|^2$. Intuitively, when the current violation is large at step t, the algorithm will reduce the violation after K steps. This is the central mechanism that ensures our algorithm achieves constant constraint violation.

Lemma 3. *Under Assumptions 1-4, Optimistic-COCO ensures that*

$$\|\boldsymbol{Q}_t\| = \mathcal{O}\left(\frac{1}{\delta}\right) = \mathcal{O}(1)$$

This lemma is derived from Lemma 2, the update rule of Optimistic-COCO, and the idea of contradiction. Then, the proof of constraint violation in Theorem 1 is directly derived by combining inequality (10) and this lemma.

Regret Analysis

Let $x = x^*$ in Lemma 1 and note $\boldsymbol{Q}_{t+1}^{\top}\boldsymbol{g}(x^*) \leq 0$ because $\boldsymbol{Q}_{t+1} \geq \boldsymbol{0}$ from the update rule of virtual queue and x^* is a feasible solution to COCO such that $\boldsymbol{g}(x^*) \leq \boldsymbol{0}$. Then we have the following lemma.

Lemma 4. *Based on Lemma 1, set $\alpha_t = 8\eta_t$, we can bound the one-step regret as*

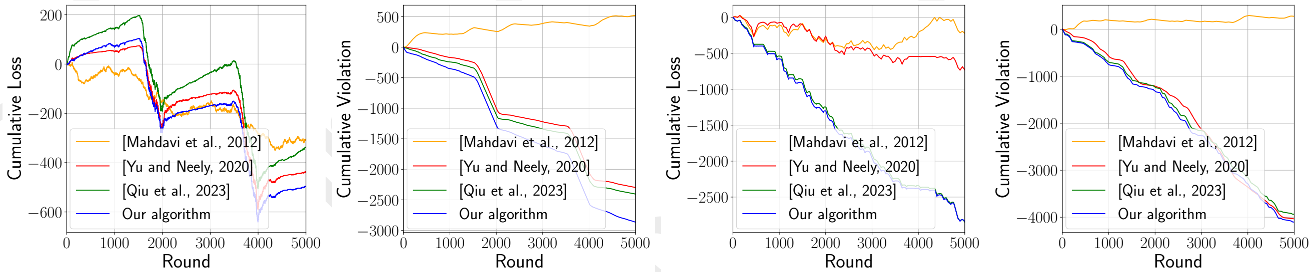
$$\begin{aligned} &f_{t+1}(x_{t+1}) - f_{t+1}(x^*) \qquad (11) \\ \leq &\frac{\phi_t}{2}\|\boldsymbol{Q}_t\|^2 - \frac{\phi_t}{2}\|\boldsymbol{Q}_{t+1}\|^2 + \frac{1}{2\eta_t}\|x^* - x_t\|^2 - \frac{1}{2\eta_t}\|x^* - \\ &x_{t+1}\|^2 + \frac{\phi_t}{2}\|\boldsymbol{Q}_t - \boldsymbol{Q}_{t-1}\|^2 - \frac{\phi_t}{2}\|\boldsymbol{Q}_{t+1} - \boldsymbol{Q}_t\|^2 + \frac{1}{2\alpha_t} \\ &\|x_t - x_{t-1}\|^2 - \frac{1}{2\alpha_t}\|x_{t+1} - x_t\|^2 + \frac{\alpha_t}{2}\|\nabla f_t - \nabla f_{t-1}\|^2 \\ &+ \sqrt{m}GD\left(\phi_{t-1} - \phi_{t-2}\right)\|\boldsymbol{Q}_{t-1}\| \end{aligned}$$

Note that most of the terms in this lemma can be bounded by telescope summation when proving the regret bound because $\phi_{t+1} \geq \phi_t$ and $\frac{1}{2\eta_{t+1}} \geq \frac{1}{2\eta_t}$ from the definition of ϕ_t and η_t. And it can be seen that the terms $(\phi_{t-1} - \phi_{t-2})\|\boldsymbol{Q}_{t-1}\|$ and $\frac{\alpha_t}{2}\|\nabla f_t - \nabla f_{t-1}\|^2$ are part of the costs that our algorithm incurs to achieve adaptiveness. If we have prior knowledge of V_T, the bound is obvious. While in our proof, $\sum_{s=1}^{t}(\phi_{s-1} - \phi_{s-2})\|\boldsymbol{Q}_{s-1}\|$ can be well bounded by Lemma 3 and the term $\sum_{s=1}^{t-1}\frac{\alpha_s}{2}\|\nabla f_s - \nabla f_{s-1}\|^2$ can be bound by the following lemma.

Lemma 5. *Recall the definition of η_t in Theorem 1 and $\alpha_t = 8\eta_t$, we have the following inequality,*

$$\sum_{s=1}^{t-1}\frac{\alpha_s}{2}\|\nabla f_s - \nabla f_{s-1}\|^2 \leq \frac{1}{M^2}\sqrt{V_t}.$$

Now we are ready to present the final bound of regret. Taking summation of the inequality (11) over time step t, we get



(a) Cumulative Loss (b) Cumulative Violation (c) Cumulative Loss (d) Cumulative Violation

Figure 1: Experiment results in two settings. (a) and (b) are conducted in a highly dynamic environment. (c) and (d) are results in a slowly changing environment. Each value is plotted by averaging over 20 independent runs.

that, for all $t \in [T]$,

$$\begin{aligned}\mathcal{R}(t) &= \sum_{s=1}^{t-1} f_{s+1}(x_{s+1}) - f_{s+1}(x^*) \\ &\le C_1\phi_t\|\boldsymbol{Q}_t\| + C_2\sqrt{V_t} + C_3 \\ &= \mathcal{O}(\phi_t\|\boldsymbol{Q}_t\| + \sqrt{V_t}) \\ &= \mathcal{O}(\sqrt{V_t})\end{aligned}$$

where we use the definition of ϕ_t and $\|\boldsymbol{Q}_t\| = \mathcal{O}(1)$ from Lemma 3. C_1, C_2, C_3 are constants and the full proof can be found in the Appendix.

5 Experiment

In this section, we present a series of synthetic experiments designed to evaluate the performance of the Optimistic-COCO algorithm under both dynamic and slowly-changing environments. Our goal is to assess how well our algorithm adapts to varying conditions and to compare its performance against several well-established algorithms, specifically, those proposed in [Mahdavi *et al.*, 2012; Yu and Neely, 2020; Qiu *et al.*, 2023]. Similar to the setups used in [Yu and Neely, 2020; Yi *et al.*, 2021], we consider a scenario where the loss functions are linear, defined as $f_t(x) = \langle c_t, x\rangle$, where c_t is a time-varying coefficient. The constraint function is expressed as $Ax \le b$, with the decision variable x belonging to $\mathbb{R}^2$. The constraint matrix $A \in \mathbb{R}^{3\times 2}$ and vector $b \in \mathbb{R}^2$ are fixed and generated uniformly within the intervals $[0.1, 0.5]$ for A and $[0, 0.2]$ for b, respectively. We set the total number of rounds to $T = 5000$ and select the feasible set $\mathcal{X}_0 = [-1, 1]^2$. For each algorithm, we track and plot the cumulative loss and violation over time.

Dynamic environment: In the dynamic setting, we define $c_t = c_1(t) + c_2(t) + c_3(t)$, where $c_1(t)$ is uniformly drawn from $[-t^{1/10}, t^{1/10}]$, $c_2(t)$ is uniformly drawn from $[-1, 0]$ for $t \in [1, 1500] \cup [2000, 3500] \cup [4000, 5000]$ and from $[0, 1]$ otherwise, and $c_3(t) = (-1)^{\mu(t)}$, where $\mu(t)$ is a random permutation of the vector $[1, T]$. Figures 1(a) and 1(b) demonstrate that Optimistic-COCO outperforms the baseline algorithms in both regret and violation. From Figure 1(a), it can be seen that all algorithms achieve similar performance level of regret, consistent with the $\mathcal{O}(\sqrt{T})$ guarantee in the worst-case scenario under a highly dynamic environment. And algorithms with $\mathcal{O}(1)$ violation guarantee show better performance in Figure 1(b).

Slowly-changing environment: We also conduct experiments in a slowly-changing environment to justify the problem-dependent theoretical guarantees. The loss functions remain fixed within each 100-round segment. At the start of each segment, the linear factor c is uniformly drawn from $[0, 1]$, and $c_t = c$ remains constant for the following 100 rounds. In this setting, the regret bound of $\mathcal{O}(\sqrt{V_t})$ is much smaller than $\mathcal{O}(\sqrt{T})$ since the loss functions do not vary at each round. Figures 1(c) and 1(d) show that Optimistic-COCO consistently outperforms all baseline algorithms and significantly surpasses the algorithms in [Mahdavi *et al.*, 2012; Yu and Neely, 2020], which aligns with the theoretical gap between $\mathcal{O}(\sqrt{V_T})$ and $\mathcal{O}(\sqrt{T})$. Figure 1(d) demonstrates that algorithms with $\mathcal{O}(1)$ guarantee still maintain strong violation performance, with Optimistic-COCO performing slightly better

It's important to highlight that our algorithm is fully adaptive, whereas we provided non-causal information about T and V_T to algorithms in [Mahdavi *et al.*, 2012; Yu and Neely, 2020; Qiu *et al.*, 2023] to configure the learning rate. Despite this advantage, our algorithm achieves the best performance in both environments without requiring any prior knowledge, thereby validating our theoretical guarantees.

6 Conclusion

In this paper, we investigate constrained online convex optimization (COCO) and introduce Optimistic-COCO, a gradient-based algorithm that integrates optimistic designs with the "drift-plus-penalty" framework. We prove that Optimistic-COCO achieves a gradient-variation bound of $\mathcal{O}(\sqrt{V_t})$ for regret and constant bound for violation, thereby encompassing and improving upon state-of-the-art results. Our algorithm is fully adaptive, requiring no prior knowledge of the time horizon T or the total gradient variation V_T, effectively addressing the open problem highlighted in [Qiu *et al.*, 2023]. These advances are made possible through a refined design of the scaled virtual queue and an optimistic regret analysis. Our experimental results further corroborate the theoretical findings.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 62302305 and the Core Facility Platform of Computer Science and Communication, SIST, ShanghaiTech University.

References

[Boyd and Vandenberghe, 2004] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[Cesa-Bianchi and Lugosi, 2006] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[Chen *et al.*, 2017] Tianyi Chen, Qing Ling, and Georgios B Giannakis. An online convex optimization approach to proactive network resource allocation. *IEEE Transactions on Signal Processing*, 65(24):6350–6364, 2017.

[Chiang *et al.*, 2012] Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Conference on Learning Theory*, pages 6–1. JMLR Workshop and Conference Proceedings, 2012.

[Goldfarb and Tucker, 2011] Avi Goldfarb and Catherine Tucker. Online display advertising: Targeting and obtrusiveness. *Marketing Science*, 30(3):389–404, 2011.

[Guo *et al.*, 2022] Hengquan Guo, Xin Liu, Honghao Wei, and Lei Ying. Online convex optimization with hard constraints: Towards the best of two worlds and beyond. *Advances in Neural Information Processing Systems*, 35:36426–36439, 2022.

[Hazan and Kale, 2010] Elad Hazan and Satyen Kale. Extracting certainty from uncertainty: Regret bounded by variation in costs. *Machine learning*, 80:165–188, 2010.

[Hazan and others, 2016] Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

[Hazan *et al.*, 2007] Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2):169–192, 2007.

[Jenatton *et al.*, 2016] Rodolphe Jenatton, Jim Huang, and Cédric Archambeau. Adaptive algorithms for online convex optimization with long-term constraints. In *International Conference on Machine Learning*, pages 402–411. PMLR, 2016.

[Lekeufack and Jordan, 2024] Jordan Lekeufack and Michael I Jordan. An optimistic algorithm for online convex optimization with adversarial constraints. *arXiv preprint arXiv:2412.08060*, 2024.

[Li and Hoi, 2014] Bin Li and Steven CH Hoi. Online portfolio selection: A survey. *ACM Computing Surveys (CSUR)*, 46(3):1–36, 2014.

[Liu *et al.*, 2024] Xin Liu, Honghao Wei, and Lei Ying. Optimistic joint flow control and link scheduling with unknown utility functions. In *Proceedings of the Twenty-fifth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, pages 271–280, 2024.

[Mahdavi *et al.*, 2012] Mehrdad Mahdavi, Rong Jin, and Tianbao Yang. Trading regret for efficiency: online convex optimization with long term constraints. *The Journal of Machine Learning Research*, 13(1):2503–2528, 2012.

[Mannor and Tsitsiklis, 2006] Shie Mannor and John N Tsitsiklis. Online learning with constraints. In *International Conference on Computational Learning Theory*, pages 529–543. Springer, 2006.

[Mokhtari *et al.*, 2020] Aryan Mokhtari, Asuman E Ozdaglar, and Sarath Pattathil. Convergence rate of o(1/k) for optimistic gradient and extragradient methods in smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 30(4):3230–3251, 2020.

[Qiu *et al.*, 2023] Shuang Qiu, Xiaohan Wei, and Mladen Kolar. Gradient-variation bound for online convex optimization with constraints. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 9534–9542, 2023.

[Rakhlin and Sridharan, 2013] Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory*, pages 993–1019. PMLR, 2013.

[Sinha and Vaze, 2024] Abhishek Sinha and Rahul Vaze. Optimal algorithms for online convex optimization with adversarial constraints, 2024.

[Yang *et al.*, 2014] Tianbao Yang, Mehrdad Mahdavi, Rong Jin, and Shenghuo Zhu. Regret bounded by gradual variation for online convex optimization. *Machine learning*, 95:183–223, 2014.

[Yi *et al.*, 2021] Xinlei Yi, Xiuxian Li, Tao Yang, Lihua Xie, Tianyou Chai, and Karl Johansson. Regret and cumulative constraint violation analysis for online convex optimization with long term constraints. In *International conference on machine learning*, pages 11998–12008. PMLR, 2021.

[Yu and Neely, 2020] Hao Yu and Michael J Neely. A low complexity algorithm with $o(\sqrt{T})$ regret and o (1) constraint violations for online convex optimization with long term constraints. *Journal of Machine Learning Research*, 21(1):1–24, 2020.

[Yu *et al.*, 2017] Hao Yu, Michael Neely, and Xiaohan Wei. Online convex optimization with stochastic constraints. *Advances in Neural Information Processing Systems*, 30, 2017.

[Yuan and Lamperski, 2018] Jianjun Yuan and Andrew Lamperski. Online convex optimization for cumulative constraints. *Advances in Neural Information Processing Systems*, 31, 2018.