Beyond $\tilde{O}(\sqrt{T})$ Constraint Violation for Online Convex Optimization with Adversarial Constraints

Abhishek Sinha **Rahul Vaze**
School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai 400005, India
abhishek.sinha@tifr.res.in, rahul.vaze@gmail.com

Abstract

We study Online Convex Optimization with adversarial constraints (COCO). At each round a learner selects an action from a convex decision set and then an adversary reveals a convex cost and a convex constraint function. The goal of the learner is to select a sequence of actions to minimize both regret and the cumulative constraint violation (CCV) over a horizon of length T. The best-known policy for this problem achieves $O(\sqrt{T})$ regret and $\tilde{O}(\sqrt{T})$ CCV. In this paper, we improve this by trading off regret to achieve substantially smaller CCV. This trade-off is especially important in safety-critical applications, where satisfying the safety constraints is non-negotiable. Specifically, for any bounded convex cost and constraint functions, we propose an online policy that achieves $\tilde{O}(\sqrt{dT}+T^{\beta})$ regret and $\tilde{O}(dT^{1-\beta})$ CCV, where d is the dimension of the decision set and $\beta \in [0,1]$ is a tunable parameter. We begin with a special case, called the CONSTRAINED EXPERT problem, where the decision set is a probability simplex and the cost and constraint functions are linear. Leveraging a new adaptive small-loss regret bound, we propose a computationally efficient policy for the CONSTRAINED EXPERT problem, that attains $O(\sqrt{T\ln N}+T^{\beta})$ regret and $\tilde{O}(T^{1-\beta}\ln N)$ CCV for N number of experts. The original problem is then reduced to the CONSTRAINED EXPERT problem via a covering argument. Finally, with an additional M-smoothness assumption, we propose a computationally efficient first-order policy attaining $O(\sqrt{MT}+T^{\beta})$ regret and $\tilde{O}(MT^{1-\beta})$ CCV.

1 Introduction

Online Convex Optimization (OCO) is a standard framework for sequential decision-making under adversarial uncertainty [Hazan, 2022]. At each round $1 \le t \le T$, a learner selects an action x_t from a convex decision set $\mathcal{X}$ with finite diameter D. The environment then reveals a convex, Lipschitz continuous cost function $f_t : \mathcal{X} \mapsto \mathbb{R}$, and the learner incurs a cost of $f_t(x_t)$. The objective is to minimize the regret relative to the best fixed action in hindsight. For any comparator action $x^\star \in \mathcal{X}$, the regret is defined as:

$$\text{Regret}_T(x^\star) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(x^\star). \tag{1}$$

The worst-case regret is defined to be $\sup_{x^\star \in \mathcal{X}} \text{Regret}_T(x^\star)$. It is well-known that simple algorithms such as Online Mirror Descent (OMD) attain an $O(\sqrt{T})$ worst-case regret, which is also minimax optimal [Hazan, 2022].

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

Online Convex Optimization with adversarial constraints (COCO) generalizes the standard OCO framework and underpins a variety of emerging applications, including AI safety [Amodei et al., 2016, Sun et al., 2017], fair allocation [Sinha, 2024], online ad markets with budget constraints [Liakopoulos et al., 2019], and multi-task learning [Ruder, 2017, Dekel et al., 2006]. In COCO, on each round $1 \le t \le T$, the learner selects an action x_t from a convex decision set $\mathcal{X}$ with diameter D. The adversary then reveals two convex functions - a cost function $f_t : \mathcal{X} \mapsto \mathbb{R}$ and a constraint function $g_t : \mathcal{X} \mapsto \mathbb{R}$. For simplicity, we make the following mild assumption:

Assumption 1 (Bounded Cost and Constraints)**.** *The cost and constraint functions are bounded within their domain. In particular, via appropriate translation and scaling, we assume that* $0 \le f_t, g_t \le 1, \forall t$.

See Section 8.1 in the Appendix for a brief discussion on Assumption 1. The constraint function g_t corresponds to a constraint of the form $h_t(x) \le 0$ where we define $g_t(x) \equiv \max(0, h_t(x))$. Thus, $g_t(x_t)$ quantifies the penalty incurred by the learner for violating the hard constraint of $h_t(x) \le 0$ at round t. If an action $x^\star$ satisfies $g_t(x^\star) = 0, \forall t$, then it is feasible throughout the horizon. To measure long-term constraint violation of a policy, we define Cumulative Constraint Violation (CCV) as:

$$\text{CCV}_T = \sum_{t=1}^{T} g_t(x_t). \tag{2}$$

Since the constraint on each round is revealed *after* the learner selects action - and may be chosen adversarially - it is generally impossible for an online policy to satisfy the constraints on every round. Therefore, to ensure the problem is well-posed, one must impose some restrictions on the constraint functions [Mannor et al., 2009]. In the COCO literature, the following feasibility assumption is universally made [Sinha and Vaze, 2024, Guo et al., 2022, Neely and Yu, 2017, Yuan and Lamperski, 2018, Yi et al., 2021].

We define the feasible set $\mathcal{X}^\star$ to be the subset of actions that satisfy the constraints across *all rounds*:

$$\mathcal{X}^\star = \{x \in \mathcal{X} : g_t(x) = 0, \ \forall t \ge 1\}. \tag{3}$$

Assumption 2 (Feasibility)**.** *The feasible set is non-empty,* i.e., $\mathcal{X}^\star \neq \varnothing$.

The feasibility assumption is not essential for our results. In particular, our algorithmic and analytical techniques naturally extend to the more general setting where the feasible actions are allowed to violate the constraints up to a prescribed budget of $B_T \ge 0$ [Sarkar et al., 2025]. Please refer to Appendix 8.6 and Theorem 6 for this generalization.

In the COCO problem, the regret of any policy is computed relative to the best feasible action in hindsight, *i.e.,*

$$\text{Regret}_T = \sup_{x^\star \in \mathcal{X}^\star} \text{Regret}_T(x^\star). \tag{4}$$

Goal: The standard objective in the COCO problem is to design an online policy that achieves both small regret and small cumulative constraint violation (CCV). In this work, our primary focus is on designing a flexible framework that minimizes the CCV to the extent possible, while ensuring that the regret remains sublinear in the horizon length T. This trade-off is particularly important in safety-critical applications, such as autonomous driving, where reducing constraint violation (*e.g.,* safety breaches) takes precedence over minimizing regret (*e.g.,* fuel or battery optimization, commute time reduction).

Background and Our Contribution

Recently, Sinha and Vaze [2024] proposed a computationally efficient first-order policy for the COCO problem, which achieves $O(\sqrt{T})$ regret and $\tilde{O}(\sqrt{T})$ CCV. They also established the tightness of their results in the high-dimensional regime, where the dimension d of the decision set $\mathcal{X}$ is at least T. It is well-known that even in the standard OCO problem, where $g_t = 0, \forall t$, the regret is lower bounded by $\Omega(\sqrt{T})$, even for $d = 1$ [Hazan, 2022, Theorem 3.2]. Thus the unconditional $O(\sqrt{T})$ regret guarantee for COCO cannot be improved. However, the question of whether one can achieve a CCV substantially smaller than $\tilde{O}(\sqrt{T})$ under additional natural assumptions was left open.

Our main contribution in this paper is to affirmatively answer this question. In particular, we show that in the fixed-dimensional setting where $d \ll T$, it is possible to achieve significantly smaller

cumulative constraint violation (CCV) while appropriately trading off the regret. Furthermore, when the cost and constraint functions are smooth, we show that a computationally efficient online gradient descent-based policy can achieve improved guarantees. A summary of our results is provided in Table 1. From the table, we also note that with an appropriate choice of the parameters ($\beta = 1$), our proposed policy achieves $O(\ln T)$ CCV in the special case of ONLINE CONSTRAINT SATISFACTION (OCS) problem when all cost functions are zero and the only goal is to satisfy the constraints [Sinha and Vaze, 2024].

A key technical ingredient in our analysis is the use of *small-loss* regret bounds, also known as $L^\star$ bounds in the online learning literature [Cesa-Bianchi and Lugosi, 2006, Orabona, 2019]. These bounds yield guarantees that improve upon the minimax optimal $O(\sqrt{T})$ rate for standard regret minimization problem when the fixed comparator incurs a small cumulative loss. In Section 2, we extend these classical results by proposing a new adaptive policy for the EXPERT problem that achieves a small-loss regret bound in the general setting where the per-round loss vectors can be potentially *unbounded*. In Section 3, we consider a special case of COCO, called the CONSTRAINED EXPERT problem, where the decision set is an N-dimensional simplex and the cost and constraint functions are linear. In Section 4, we reduce the general COCO problem to the CONSTRAINED EXPERT problem via a covering argument. In Section 5, we give a computationally efficient first-order policy with improved bounds for smooth and convex functions. Due to space constraints, experimental results have been deferred to Section 8.7 in the Appendix.

Reference	Regret	CCV	Complexity	Assumptions
Jenatton et al. [2016]	$O(T^{\max(\beta,1-\beta)})$	$O(T^{1-\beta/2})$	Projection	FC
Neely and Yu [2017]	$O(\sqrt{T})$	$O(\sqrt{T}/\eta)$	Conv-OPT	Slater condition
Yuan and Lamperski [2018]	$O(T^{\max(\beta,1-\beta)})$	$O(T^{1-\beta/2})$	Projection	FC
Yu and Neely [2020]	$O(\sqrt{T})$	$O(1/\eta)$	Conv-OPT	Slater, FC
Yi et al. [2021]	$O(T^{\max(\beta,1-\beta)})$	$O(T^{(1-\beta)/2})$	Conv-OPT	FC
Guo et al. [2022]	$O(\sqrt{T})$	$O(T^{3/4})$	Conv-OPT	-
Yi et al. [2023]	$O(T^{\max(\beta,1-\beta)})$	$O(T^{1-\beta/2})$	Conv-OPT	-
Sinha and Vaze [2024]	$O(\sqrt{T})$	$\tilde{O}(\sqrt{T})$	Projection	-
Vaze and Sinha [2025]	$O(\sqrt{T})$	Instance dependent	Projection	-
This paper	$O(\sqrt{T\ln N} + T^{\beta})$	$\tilde{O}(T^{1-\beta}\ln N)$	$O(N)$	CONSTR. EXPERT
This paper	$\tilde{O}(\sqrt{dT} + T^{\beta})$	$\tilde{O}(dT^{1-\beta})$	$O(T^d)$	d-dim. decision set
This paper	$O(\sqrt{MT} + T^{\beta})$	$\tilde{O}(MT^{1-\beta})$	Projection	Smooth

Table 1: Summary of the key results on the COCO problem. In the above table, $\beta \in [0, 1]$ is a tunable parameter, $\eta > 0$ denotes the Slater's constant, N denotes the number of experts, d denotes the dimension of the decision set $\mathcal{X}$, M denotes the smoothness constant, and $\tilde{O}(\cdot)$ hides polylogarithmic factors in T. CONSTR. EXPERT refers to the CONSTRAINED EXPERT problem described in Section 3, Conv-OPT refers to solving a constrained convex optimization problem on each round, Projection refers to the Euclidean projection operation on the decision set $\mathcal{X}$, and FC refers to the Fixed Constraints setting.

Intuition for the results: We now give some intuition for why the Cumulative Constraint Violation (CCV) can be expected to be made smaller than the current-best bound of $\tilde{O}(\sqrt{T})$ by crucially utilizing small-loss regret bounds. Consider the Online Constraint Satisfaction (OCS) problem, introduced by Sinha and Vaze [2024], where all cost functions are identically equal to zero and the goal is to minimize the CCV only. Let the constraint functions (*i.e.,* $\{g_t\}_{t\geq 1}$'s), be non-negative, M-smooth, and convex.

For solving this problem, we run the Online Gradient Descent policy on the sequence of constraint functions with an adaptive step size schedule. Specifically, we choose the next action as $x_{t+1} = \text{PROJ}_{\mathcal{X}}(x_t - \eta_t \nabla_t)$, where $\nabla_t \equiv \nabla g_t(x_t)$ and the step sizes are adaptively chosen as $\eta_t = D/\sqrt{2\sum_{\tau=1}^{t}\|\nabla_\tau\|_2^2}, t \geq 1$. The following small-loss regret bound achieved by this policy for non-negative smooth functions is well-known [Orabona, 2019, Theorem 4.25] (please see the

statement of Theorem 4 in Section 5 for a quick reference):

$$\sum_{t=1}^{T} g_t(x_t) - \sum_{t=1}^{T} g_t(u) \leq 4D^2M + 4D\sqrt{M\sum_{t=1}^{T} g_t(u)}, \ \forall u \in \mathcal{X}. \tag{5}$$

Now, if we choose the comparator u to be a feasible action by setting $u = x^\star, x^\star \in \mathcal{X}^\star$, we have $g_t(x^\star) = 0, \forall t$. Thus the regret bound (5) implies $\text{CCV}_T = \sum_{t=1}^{T} g_t(x_t) \leq 4D^2M$ - which is a *constant* independent of T. This result is surprising as the $O(\sqrt{T})$ lower bound for regret holds even for linear functions [Hazan, 2022].

In this paper, we generalize this observation by exploring the trade-off between regret and CCV while taking into account both the cost and constraint functions. Our main result roughly says that any $(\text{Regret}_T, \text{CCV}_T)$ pair with $\text{Regret}_T \geq \tilde{\Theta}(\sqrt{T})$ and $\text{Regret}_T \times \text{CCV}_T = \tilde{O}(T)$ is achievable.

1.1 Prior work

Online convex optimization with constraints has been extensively studied under various modeling assumptions. Table 1 provides a summary of key results in the literature.

In the fixed-constraint setting, where $g_t = g$ for all t, Yi et al. [2021] proposed a policy that achieves $O(T^{\max(\beta,1-\beta)})$ regret and $O(T^{(1-\beta)/2})$ CCV. Under the stronger assumption of *Slater's condition*, which requires the existence of a uniformly strictly feasible action $x^\star \in \mathcal{X}$ such that $g_t(x^\star) \leq -\eta$ for some constant $\eta > 0$ and all t, Yu and Neely [2020] showed that the CCV can be reduced to $O(1/\eta)$ while maintaining $O(\sqrt{T})$ regret.

The problem becomes significantly more challenging in the presence of time-varying adversarial constraints. Under Slater's condition, Neely and Yu [2017] developed an algorithm achieving $O(\sqrt{T})$ regret and $O(\sqrt{T}/\eta)$ CCV. However, since this condition is quite strong, difficult to verify in practice, and leads to vacuous CCV bounds as $\eta \to 0^+$, recent works have focused on avoiding this assumption.

Guo et al. [2022] proposed an algorithm that needs to solve a separate offline convex optimization problem in each round, achieving $O(\sqrt{T})$ regret and $O(T^{3/4})$ CCV. Subsequently, Sinha and Vaze [2024] introduced a simpler gradient-based policy that improves the CCV bound to $\tilde{O}(\sqrt{T})$ while still maintaining $O(\sqrt{T})$ regret. See also Lekeufack and Jordan [2024], Lu et al. [2025], Supantha and Sinha [2025], Sarkar et al. [2025] for various extensions of their result. More recently, Vaze and Sinha [2025] proposed an algorithm that achieves $O(\sqrt{T})$ regret and constant CCV for some special classes of the constraint sets.

The algorithms proposed in this paper make use of Lipschitz-adaptive small-loss regret bounds. Small-loss regret bounds for the standard regret minimization problem with *bounded* Lipschitz constants, where the regret scales with the cumulative loss of the benchmark (instead of the time horizon T), have been well studied [Cesa-Bianchi et al., 1997, Auer et al., 2002, Hazan and Kale, 2010]. Mhammedi et al. [2019] propose Lipschitz-adaptive algorithms in the EXPERT setting, which rely on multiple restart phases and incur additional computational overhead. Similarly, Cesa-Bianchi et al. [2007] uses doubling trick to estimate the unknown parameters. Restarting learning algorithms during their course of execution is practically wasteful as it discards all past data prior to the restarts. To the best of our knowledge, continuously adaptive, scale-free variants of the small-loss regret bounds in the EXPERT setting have not been investigated before in the literature. In the following, we review the EXPERT problem and derive an adaptive small-loss regret bound.

2 Preliminaries: Adaptive Small-Loss Regret Bound for the EXPERT Problem

The *Prediction with Expert Advice* problem, also known as the EXPERT problem in the literature, refers to a repeated game where, in each round $t \geq 1$, the learner chooses a probability distribution p_t over a set of N experts (experts may be identified with the set of actions). After that, the adversary chooses a bounded loss vector $l_t \in [0,1]^N$, where $l_t(i)$ denotes the loss for the ith expert, $i \in [N]$. Consequently, the learner incurs an expected loss of $f_t(p_t) = \langle l_t, p_t \rangle$ in round t. The full loss vector l_t is revealed to the learner at the end of round t. The learner's objective is to choose a sequence of distributions $\{p_t\}_{t\geq 1}$ to minimize its regret over T rounds. The EXPERT problem is the full-information counterpart of the Multi-armed Bandit (MAB) problem [Bubeck et al., 2012].

The Exponential Weights algorithm, also known as Hedge, is a well-known solution to the EXPERT problem. The Hedge algorithm selects the ith expert on round t with probability

$$p_t(i) = \exp\big(-\eta \sum_{\tau=1}^{t-1} l_\tau(i)\big)/Z, \; i \in [N], \tag{6}$$

where $\eta > 0$ is a suitably chosen learning rate and Z is the normalizing constant. Hedge is known to achieve the following *small-loss* regret bound [Cesa-Bianchi and Lugosi, 2006, Corollary 2.4]:

$$\text{Regret}_T = \tilde{L}_T - L_T(i^\star) \leq \sqrt{2L_T(i^\star)\ln N} + \ln N, \tag{7}$$

where $\tilde{L}_T = \sum_{\tau=1}^{T} \langle l_\tau, p_t \rangle$ is the learner's cumulative loss, $L_T(i^\star) = \sum_{\tau=1}^{T} l_\tau(i^\star)$ is the cumulative loss of a comparator arm $i^\star \in [N]$. See also the Squint algorithm proposed by Koolen and Van Erven [2015]. The small-loss bound (7) offers a significant improvement over the standard $O(\sqrt{T})$ regret bound when the comparator's loss is small, *i.e.,* $L_T(i^\star) = o(T)$. While the bound (7) is well-known, it assumes that that all loss values are bounded [Cesa-Bianchi and Lugosi, 2006, Section 2.4]. However, for reasons that will become clear in the sequel, we cannot use (7) in our algorithm, as we will need to control the regret when the losses, defined appropriately by the algorithm, are unbounded.

To address this technical challenge, Theorem 1 presents a small-loss regret bound for the EXPERT problem that generalizes (7) by accommodating potentially unbounded losses. This regret bound is achieved by an adaptive variant of the Hedge policy that employs a self-confident variable learning rate. The full pseudocode for this policy is provided in Algorithm 4 in Appendix 8.2.

Theorem 1. *Consider the* EXPERT *problem with N experts. Let the vector l_t denote the losses of the experts at round $t \geq 1$, where $l_t(i) \geq 0, \forall i, t$. The losses need not be uniformly bounded above for all rounds. Let G_t be an upper bound to $\|l_t\|_\infty$ satisfying the following conditions*[1]*:*

1. *The sequence $\{G_t\}_{t\geq 1}$ is monotonically non-decreasing,* i.e., $G_t \geq G_{t-1}, \forall t \geq 1, G_0 = 1$.
2. *The growth of G_t in consecutive rounds is bounded:* $\max_{1\leq t\leq T} \frac{G_t}{G_{t-1}} \leq \gamma$ *for some known constant $\gamma \geq 1$.*

Then the following adaptive Hedge algorithm, which selects the ith expert with probability $p_t(i) \propto \exp(-\eta_t L_{t-1}(i)), \forall i$, with an adaptive learning rate $\eta_t = \frac{1}{\sqrt{G_{t-1}}}\sqrt{\frac{\ln N}{\tilde{L}_{t-1}+\gamma G_{t-1}}}$, achieves the following regret bound:

$$\tilde{L}_T - L_T(i^\star) \leq 2\gamma\sqrt{L_T(i^\star)G_T \ln N} + 7\gamma^2 G_T \ln N. \tag{8}$$

In the above, $\tilde{L}_t = \sum_{\tau=1}^{t} \langle p_\tau, l_\tau \rangle$ denotes the algorithm's cumulative loss up to round t and $L_T(i^\star) = \sum_{\tau=1}^{T} l_\tau(i^\star)$ is the cumulative loss of any comparator expert $i^\star \in [N]$ up to round T.

See Appendix 8.3 for a proof of Theorem 1.

Remarks: 1. The monotonicity assumption on the sequence $\{G_t\}_{t\geq 1}$ entails no loss of generality, since one can always define a new sequence of upper bounds as $G'_t := \max_{1\leq\tau\leq t} G_\tau, t \geq 1$, which is monotonic by construction.

2. The proof of Theorem 1 is non-trivial because the value of G_T is unknown in advance. As a result, we cannot simply rescale all losses by G_T and directly apply the small-loss bound in (7) to the normalized losses. Instead, we carefully design an adaptive learning rate schedule that accounts for the growth of the loss scale and the cumulative loss of the best expert over time.

3 The CONSTRAINED EXPERT problem: Simplex Decision Set, Linear Cost, and Linear Constraint Functions

In this section, we focus on an important special case of the COCO problem, called CONSTRAINED EXPERT problem, where the cost and constraint functions are linear and the decision set is the

[1] Naturally, the $\{G_t\}_{t\geq 1}$ sequence is causal *i.e.,* G_t can be computed only after observing the loss vector l_t on round t.

$N-1$-dimensional simplex Δ_N, *i.e.,*

$$\mathcal{X} = \Delta_N = \left\{ p \in \mathbb{R}^N : \sum_{i=1}^{N} p_i = 1, p_i \geq 0, \forall i \right\}.$$

In this problem, at each round t, the learner first computes a distribution $p_t \in \Delta_N$ over N experts, and then it samples an expert from the distribution p_t. Choosing expert i incurs a cost of $f_t(i)$ and a constraint violation of $g_t(i)$. Consequently, the learner incurs an expected cost of $f_t(p_t) \equiv \langle f_t, p_t \rangle$ and an expected constraint violation of $g_t(p_t) \equiv \langle g_t, p_t \rangle$ on round t[2]. Both the cost vector f_t and the constraint vector g_t are revealed to the learner at the end of the round. The feasibility assumption (Assumption 2), specialized to this setting, implies that there exists at least one expert $i^\star \in [N]$ such that $g_t(i^\star) = 0, \forall t$. The learner's objective is to generate a sequence of distributions $\{p_t\}_{t=1}^T$ that minimizes both the regret and the cumulative constraint violation (CCV).

3.1 An Online Policy for the CONSTRAINED EXPERT Problem

Let $\Phi : \mathbb{R} \mapsto \mathbb{R}$ be a non-decreasing convex Lyapunov function, to be specified later, and let $i^\star$ be an uniformly feasible expert with $g_t(i^\star) = 0, \forall t$. As stated earlier, the existence of $i^\star$ is guaranteed by Assumption 2. Let $Q(t)$ denote the cumulative constraint violation (CCV) up to round t, which evolves as follows:

$$Q(t) = Q(t-1) + \langle g_t, p_t \rangle. \tag{9}$$

We now use the regret decomposition framework introduced by Sinha and Vaze [2024] to design an online policy for the CONSTRAINED EXPERT problem. Using the convexity of the Lyapunov function $\Phi(\cdot)$, we have for any $1 \leq t \leq T$:

$$\begin{aligned} \Phi(Q(t)) - \Phi(Q(t-1)) &\leq \Phi'(Q(t))(Q(t) - Q(t-1)) \\ &\overset{(a)}{=} \Phi'(Q(t))\langle g_t, p_t \rangle \\ &\overset{(b)}{=} \Phi'(Q(t))(\langle g_t, p_t \rangle - g_t(i^\star)), \end{aligned} \tag{10}$$

where (a) follows from Eqn. (9) and (b) uses the fact that $g_t(i^\star) = 0, \forall t$. Adding the term $\langle f_t, p_t \rangle - f_t(i^\star)$ to both sides of the inequality, we obtain

$$\begin{aligned} &\Phi(Q(t)) - \Phi(Q(t-1)) + (\langle f_t, p_t \rangle - f_t(i^\star)) \\ &\leq \langle f_t + \Phi'(Q(t))g_t, p_t \rangle - (f_t(i^\star) + \Phi'(Q(t))g_t(i^\star)). \end{aligned} \tag{11}$$

Define the tth surrogate cost function $\hat{f}_t : \Delta_N \mapsto \mathbb{R}$ to be the following linear function:

$$\hat{f}_t(p) \coloneqq \langle f_t + \Phi'(Q(t))g_t, p \rangle, \quad t \geq 1. \tag{12}$$

Summing up Eqn. (11) for $1 \leq t \leq T$, we obtain the following regret decomposition inequality:

$$\Phi(Q(T)) - \Phi(Q(0)) + \text{Regret}_T(i^\star) \leq \text{Regret}'_T(i^\star). \tag{13}$$

In Eqn. (13), $\text{Regret}_T(i^\star) \equiv \sum_{t=1}^T \langle f_t, p_t - e_{i^\star} \rangle$ and $\text{Regret}'_T(i^\star) \equiv \sum_{t=1}^T \langle \hat{f}_t, p_t - e_{i^\star} \rangle$ correspond to the regret for the original cost functions $\{f_t\}_{t=1}^T$ and the surrogate cost functions $\{\hat{f}_t\}_{t=1}^T$, respectively, with respect to a feasible expert $i^\star \in [N]$.

Algorithm 1 presents our proposed online policy for the CONSTRAINED EXPERT problem. It employs the adaptive Hedge subroutine from Theorem 1 to select the sampling distributions $\{p_t\}_{t \geq 1}$ which minimize $\text{Regret}'_T(i^\star)$ appearing on the RHS of Eqn. (13). Observe that the surrogate cost function $\hat{f}_t$ involves the term $\Phi'(Q(t))$, which may grow indefinitely with t as $Q(t)$ grows. Hence, it is imperative to use the adaptive Hedge algorithm (Algorithm 4) which can handle unbounded loss, in contrast to the standard Hedge algorithm (6), which assumes bounded loss vectors. The following theorem gives an upper bound to the regret and CCV achieved by Algorithm 1.

Theorem 2. *Consider the Lyapunov function $\Phi(x) = e^{\lambda x}$, where $\lambda = T^{-(1-\beta)}/(2c \ln N), c = 10$, and $\beta \in [0, 1]$ is a tunable parameter. Then, under Assumptions 1 and 2, Algorithm 1 achieves the following guarantees for the* CONSTRAINED EXPERT *problem for any feasible expert $i^\star \in [N]$:*

$$\textit{Regret}_T(i^\star) = O(\sqrt{T \ln N} + T^\beta + \ln N), \ \textit{CCV}_T = O(T^{1-\beta} \ln N \ln T).$$

In particular, if $f_t = 0, \forall t$, upon setting $\beta = 1$, we obtain $\textit{CCV}_T = O(\ln N \ln T)$.

[2]To simplify the notations, we use the same symbol for denoting a linear function and its associated coefficient vector.

Algorithm 1 Algorithm for the CONSTRAINED EXPERT problem

Input: Number of experts N, Horizon length T, Cost vectors $\{f_t\}_{t=1}^T$ and Constraint vectors $\{g_t\}_{t=1}^T$, Tunable parameter $\beta \in [0,1]$.

Parameter settings: $\Phi(x) = e^{\lambda x}, \lambda = T^{-(1-\beta)}/(2c\ln N), c = 10$.

Initialization: Set $Q(0) \leftarrow 0$, Algorithm's loss $\tilde{L}_0 \leftarrow 0$, Experts' cumulative loss $L_0 \leftarrow \mathbf{0}, G_0 = 1 + \Phi'(Q(0))$.

for each $t = 1 : T$ **do**

Compute self-confident learning rate

$$\eta_t = \frac{1}{\sqrt{G_{t-1}}}\sqrt{\frac{\ln N}{\tilde{L}_{t-1} + \gamma G_{t-1}}}. \tag{14}$$

Play adaptive Hedge by choosing the ith expert with probability

$$p_t(i) = \exp(-\eta_t L_{t-1}(i)) / \sum_{j=1}^{N} \exp(-\eta_t L_{t-1}(j)), \ \forall i \in [N].$$

Observe the vectors f_t, g_t. Incur a cost of $\langle f_t, p_t \rangle$ and constraint violation of $\langle g_t, p_t \rangle$.

Update CCV and G_t:

$$Q(t) = Q(t-1) + \langle g_t, p_t \rangle, \ \ G_t = 1 + \Phi'(Q(t)).$$

Compute the surrogate cost vector: $\hat{f}_t = f_t + \Phi'(Q(t))g_t$. (coordinate-wise vector addition)

Update the cumulative loss of the algorithm and the cumulative loss of each expert

$$\tilde{L}_t = \tilde{L}_{t-1} + \langle \hat{f}_t, p_t \rangle, \ \ L_t(i) = L_{t-1}(i) + \hat{f}_t(i), \ \ i \in [N].$$

end for each

Experimental results, presented in Section 8.7 of the Appendix, qualitatively support the expected variations of Regret and CCV as a function of the parameter β.

3.2 Proof of Theorem 2

Define the sequence $G_t \equiv 1 + \Phi'(Q(t)), t \geq 0$, where the Lyapunov function $\Phi(\cdot)$ is chosen as specified in the statement of Theorem 2. From the definition of the surrogate costs (Eqn. (12)), it follows that G_t is an upper bound to the maximum component of the surrogate cost vector:

$$\|\hat{f}_t\|_\infty \leq \|f_t\|_\infty + \Phi'(Q(t))\|g_t\|_\infty \leq 1 + \Phi'(Q(t)) = G_t, \forall t.$$

Since $Q(t)$ is non-decreasing in t and the function $\Phi(\cdot)$ is convex, it follows that the sequence $\{G_t\}_{t\geq 1}$ is also non-decreasing. Furthermore, from Lemma 2 in the Appendix, we have that $\max_{1\leq t\leq T} \frac{G_t}{G_{t-1}} \leq 1.08$. Thus all conditions in Theorem 1 are fulfilled with $\gamma = 1.08$, and we can use the small-loss regret bound (8) of the adaptive Hedge algorithm for the surrogate cost sequence $\{\hat{f}_t\}_{t=1}^T$. Let $i^\star \in [N]$ be a uniformly feasible expert which incurs zero constraint violation on every round, *i.e.,* $g_t(i^\star) = 0, \forall t$. The cumulative surrogate cost incurred by expert $i^\star$ can be upper bounded as:

$$L_T(i^\star) = \sum_{t=1}^{T} \left(f_t(i^\star) + \Phi'(Q(t))g_t(i^\star) \right) \leq T,$$

where we have used the fact that $\|f_t\|_\infty \leq 1$ and $g_t(i^\star) = 0, \forall t \geq 1$. Hence, using the small-loss regret bound (8), the regret for the surrogate cost functions with respect to any feasible arm $i^\star$ can be upper bounded as:

$$\text{Regret}'_T(i^\star) \leq c\sqrt{T(1 + \Phi'(Q(T))\ln N} + c(1 + \Phi'(Q(T)))\ln N, \tag{15}$$

where we have used the fact that $\max(2\gamma, 7\gamma^2) \leq 10 \equiv c$ (say). Using the inequality $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}, x \geq 0, y \geq 0$, and substituting the upper bound from Eqn. (15) into the regret decomposition

inequality (13), we obtain

$$\Phi(Q(T)) + \text{Regret}_T(i^\star) \quad \leq \quad \Phi(Q(0)) + c\Bigg(\sqrt{T\ln N} + \sqrt{T\Phi'(Q(T))\ln N} + \Phi'(Q(T))\ln N + \ln N\Bigg). \tag{16}$$

1. Bounding the CCV: Using the fact that $\text{Regret}_T(i^\star) \geq -L_T(i^\star) \geq -T$, inequality (16) yields

$$e^{\lambda Q(T)} \leq T + 1 + c\sqrt{T\ln N} + c\sqrt{\lambda T e^{\lambda Q(T)}\ln N} + (\lambda c \ln N)e^{\lambda Q(T)} + c\ln N. \tag{17}$$

Since $T \geq 1$, our choice of the parameter $\lambda = T^{-(1-\beta)}/(2c\ln N)$ ensures that $\lambda c \ln N \leq 1/2$. Using this inequality to bound the coefficient of the penultimate term on the RHS of Eqn. (17) and transposing, we obtain

$$e^{\lambda Q(T)} \leq 8\max\Bigg(T+1, c\sqrt{T\ln N}, c\sqrt{\lambda T\ln N}e^{\lambda Q(T)/2}, c\ln N\Bigg).$$

Since the maximum value on the RHS is achieved by at least one term on the right, comparing the LHS with each term on the RHS separately and simplifying, we obtain the following bound for CCV:

$$Q(T) \leq \lambda^{-1}(c_1 + c_2\ln T + c_3\ln\ln N), \tag{18}$$

where c_1, c_2, c_3 are universal constants.

2. Bounding the Regret: Transposing the term $\Phi(Q(T)) = e^{\lambda Q(T)}$ to the RHS of Eqn. (16) and using the fact that $\lambda c\ln N \leq 1/2$ for our chosen parameter λ, we obtain

$$\text{Regret}_T(i^\star) \leq 1 + c\sqrt{T\ln N} + c\sqrt{\lambda T\ln N}e^{\lambda Q(T)/2} - \frac{1}{2}e^{\lambda Q(T)} + c\ln N.$$

Using the fact that $ax - bx^2 \leq \frac{a^2}{4b}, \forall b > 0$, the above inequality implies the following regret bound:

$$\text{Regret}_T(i^\star) \leq 1 + c\sqrt{T\ln N} + \frac{c^2}{2}\lambda T\ln N + c\ln N. \tag{19}$$

Substituting $\lambda = \frac{T^{-(1-\beta)}}{2c\ln N}$ into Eqns. (19) and Eqn. (18), we obtain the following regret and CCV bounds:

$$\text{Regret}_T(i^\star) = O(\sqrt{T\ln N} + T^\beta + \ln N),\ \text{CCV}_T = O(T^{1-\beta}\ln N\ln T).$$

□

4 Convex Cost and Constraint functions

In this section, we generalize our previous results to the general convex setting. In particular, we make the following assumption.

Assumption 3 (Convexity and Lipschitzness)**.** *All cost and constraint functions are convex and G-Lipschitz. The decision set $\mathcal{X}$ is a bounded subset of the d-dimensional Euclidean space $\mathbb{R}^d$.*

We begin by recalling the notion of a δ-cover from Wainwright [2019]. See Figure 1 in the Appendix for a schematic.

Definition 1 (Covering number)**.** *A δ-cover of a set $\mathbb{T}$ with respect to a metric ρ is a set $\{\theta^1, \theta^2, \ldots, \theta^N\} \subseteq \mathbb{T}$ such that for each $\theta \in \mathbb{T}$, there exists some $i \in [N]$ such that $\rho(\theta, \theta^i) \leq \delta$. The δ-covering number $N(\delta; \mathbb{T}, \rho)$ is the cardinality of the smallest δ-cover.*

Construction: Let $\mathcal{N}_\delta = \{x^1, x^2, \ldots, x^{N_\delta}\}$ be the smallest δ-cover of the decision set $\mathcal{X}$ with $\delta = 1/T$. Since $\mathcal{X}$ is contained within a d-dimensional ball of diameter D, its covering number is bounded by $N_\delta \leq (1 + \frac{2D}{\delta})^d$ [Wainwright, 2019, Lemma 5.7]. We construct an instance of the CONSTRAINED EXPERT problem with N_δ experts where the ith expert corresponds to the point $x^i, i \in [N_\delta]$. The cost and constraint violation for the experts are defined as:

$$f_t^{\texttt{CE}}(i) = f_t(x^i),\ \ g_t^{\texttt{CE}}(i) = (g_t(x^i) - G\delta)^+,\ i \in [N_\delta]. \tag{20}$$

Intuitively, the learner reduces the complex continuous decision space to a finite set of representative points, allowing us to apply the CONSTRAINED EXPERT algorithm.

Feasibility: To show the feasibility of the above CONSTRAINED EXPERT problem, consider any feasible action in the decision set $x^\star \in \mathcal{X}$ that satisfies $g_t(x^\star) = 0, \forall t$. Let $x^{i^\star}$ be the nearest point in the δ-cover, such that $\|x^{i^\star} - x^\star\| \le \delta$. Using the Lipschitzness of the constraint function, we have

$$g_t^{\texttt{CE}}(i^\star) \equiv g_t(x^{i^\star}) \le g_t(x^\star) + G\|x^{i^\star} - x^\star\| \le 0 + G\delta.$$

Hence, from Eqn. (20), it follows that expert $i^\star$ is feasible as $g_t^{\texttt{CE}}(i^\star) = 0, \forall t$.

Algorithm: Our policy is summarized in Algorithm 2 where, on every round, we run Algorithm 1 on the CONSTRAINED EXPERT instance defined above. We then use the output distribution from Algorithm 1 to select the next action as the corresponding convex combination of the points in $\mathcal{N}_\delta$.

Algorithm 2 COCO Algorithm for Convex Cost and Constraints

1: **Input:** A minimal δ-cover of $\mathcal{X}$: $\mathcal{N}_\delta = \{x^1, x^2, \ldots, x^{N_\delta}\}$, with $\delta = 1/T$. Parameter $\beta \in [0, 1]$.
2: **for each** $t = 1 : T$ **do**
3: Run Algorithm 1 on the CONSTRAINED EXPERT problem with N_δ experts where the cost and constraint vectors are given by Eqn. (20). Obtain the distribution p_t over $\mathcal{N}_\delta$.
4: Play $x_t = \sum_i p_t(i) x^i$
5: **end for**

Analysis: By Eqn. (20) and Jensen's inequality, the cost and constraint violation on any round can be upper bounded by the cost and constraint violation of the CONSTRAINED EXPERT instance as:

$$f_t(x_t) = f_t(\sum_i p_t(i)x^i) \le \sum_i p_t(i) f_t(x^i) = \langle f_t^{\texttt{CE}}, p_t\rangle,\ g_t(x_t) = g_t(\sum_i p_t(i)x^i) \le \langle g_t^{\texttt{CE}}, p_t\rangle + G\delta.$$

In addition, using the Lipschitzness of f_t, we have $f_t(x^{i^\star}) - f_t(x^\star) \le G\delta, \forall t$. Hence, the regret and CCV of Algorithm 2 differs from that of the CONSTRAINED EXPERT instance by at most $G\delta T \le G$, which is a constant. Finally, using the fact that $\ln N_\delta = O(d \ln T)$, we invoke Theorem 2 to obtain the following bounds for Algorithm 2. The results are summarized in Theorem 3.

$$\text{Regret}_T(x^\star) = O(\sqrt{dT\ln(T)} + T^\beta + d\ln T),\ \text{CCV}_T = O(dT^{1-\beta}(\ln T)^2). \tag{21}$$

Theorem 3. *Under Assumptions 1, 2, and 3, Algorithm 2 achieves $\tilde{O}(\sqrt{dT} + T^\beta)$ regret and $\tilde{O}(dT^{1-\beta})$ CCV for any $\beta \in [0, 1]$.*

5 Convex and Smooth Cost and Constraint Functions

While the previous reduction-based approach, given by Algorithm 2, is interesting, it can be computationally prohibitive when the dimension d is large. To address this problem, we now propose an efficient gradient-based policy for the class of non-negative, smooth, and convex cost and constraint functions. We will make use of the following small-loss regret bound achieved by the Online Gradient Descent (OGD) policy for this class of functions.

Theorem 4 (Orabona [2019], Theorem 4.25)**.** *Let $\mathcal{X}$ be a closed non-empty convex decision set with diameter D. Let $l_1, l_2, \ldots, l_T$ be an arbitrary sequence of non-negative convex and M-smooth functions. Let ∇_t be the gradient of l_t at $x_t, t \ge 1$. Pick any $x_1 \in \mathcal{X}$, set the step sizes adaptively as $\eta_t = D/\sqrt{2\sum_{\tau=1}^t \|\nabla_\tau\|_2^2}, t \ge 1$, and consider the Online Gradient Descent (OGD) policy with adaptive step sizes which selects the next action as: $x_{t+1} = \text{PROJ}_{\mathcal{X}}(x_t - \eta_t \nabla_t)$, where $\text{PROJ}_{\mathcal{X}}(\cdot)$ denotes the Euclidean projection operator on to the decision set $\mathcal{X}$. Then we have:*

$$\textit{Regret}_T(u) = \sum_{t=1}^T l_t(x_t) - \sum_{t=1}^T l_t(u) \le 4D\sqrt{M\sum_{t=1}^T l_t(u)} + 4D^2M,\ \ u \in \mathcal{X}. \tag{22}$$

In this section, we make the following assumption.

Assumption 4 (Convexity and Smoothness)**.** *All cost and constraint functions are convex and M-smooth.*

Algorithm and Analysis: Let $\Phi : \mathbb{R} \mapsto \mathbb{R}$ be a non-decreasing convex Lyapunov function, $x^\star \in \mathcal{X}$ be a feasible action, and $Q(t)$ be the CCV up to round t. Following identical arguments as in the CONSTRAINED EXPERT problem in Section 3, we define the surrogate cost function

$$\hat{f}_t(x) \coloneqq f_t(x) + \Phi'(Q(t))g_t(x),\ x \in \mathcal{X}. \tag{23}$$

From Assumption 4, it follows that all surrogate cost functions are non-negative, convex, and $M_T \equiv M(1 + \Phi'(Q(T))$-smooth. Our proposed algorithm is described in Algorithm 3 where we

Algorithm 3 COCO Algorithm for Smooth and Convex Cost and Constraints

1: **Initialization:** Choose $x_1 \in \mathcal{X}$ arbitrarily
2: **for each** $t = 1 : T$ **do**
3: Compute gradient $\nabla_t = \nabla \hat{f}_t(x_t)$ from Eqn. (23)
4: Set the step size $\eta_t = D/\sqrt{2\sum_{\tau=1}^{t} \|\nabla_\tau\|_2^2}$.
5: Choose next action using OGD: $x_{t+1} = \text{PROJ}_{\mathcal{X}}(x_t - \eta_t \nabla_t)$
6: **end for**

run the standard OGD policy on the surrogate cost functions $\{l_t \equiv \hat{f}_t\}_{t\geq 1}$, with adaptive step sizes given by Theorem 4. It is crucial to note that the step sizes $\{\eta_t\}_{t\geq 1}$, which are derived from the past gradients, are oblivious to the smoothness parameter M_T, which is unknown *a priori*. Using Eqn. (22) from Theorem 4, the regret for the surrogate costs for any feasible action $x^\star$ can be bounded as:

$$\text{Regret}'_T(x^\star) \leq 4D\sqrt{MT(1 + \Phi'(Q(T))} + 4(1 + \Phi'(Q(T))D^2 M. \tag{24}$$

We have used the fact that the cumulative surrogate cost is upper bounded by $\sum_{t=1}^{T} \hat{f}_t(x^\star) \leq T$ as $g_t(x^\star) = 0, \forall t \geq 1$. The regret bound (24) becomes algebraically identical to Eqn. (15) under the substitutions $c \leftarrow 4, \ln N \leftarrow D^2 M$. Thus, reusing the same analysis and parameter choices (including the Lyapunov function) from the proof of Theorem 2, we arrive at the following result.

Theorem 5. *Let $\beta \in [0, 1]$ be a tunable parameter, and define the Lyapunov function as $\Phi(x) = e^{\lambda x}$ with $\lambda = T^{-(1-\beta)}/(8D^2 M)$. Then, under Assumptions 1, 2, and 4, Algorithm 3 achieves the following guarantee for any feasible action $x^\star \in \mathcal{X}^\star$:*

$$\textit{Regret}_T(x^\star) = O(D\sqrt{MT} + T^\beta + D^2 M),\ \textit{CCV}_T = O(T^{1-\beta} M \ln T).$$

In particular, if $f_t = 0, \forall t$, upon setting $\beta = 1$, we obtain $\textit{CCV}_T = O(M \ln T)$.

See Section 8.6 in the Appendix for an extension of the above result that relaxes the feasibility assumption (Assumption 2) by allowing the benchmark $x^\star$ to violate the constraints within a prescribed long term budget of B_T.

6 Conclusion

In this paper we propose online policies for the COCO problem that achieve improved cumulative constraint violation (CCV) by carefully trading it off with regret. These results are particularly important in applications where violating the constraints are costly, such as autonomous driving or budget-constrained advertising. An important direction for future work is to design computationally efficient algorithms that achieve sharper CCV guarantees in the fixed-dimensional setting without the smoothness assumption. Additionally, it would be interesting to establish matching lower bounds.

7 Acknowledgement

This work was supported by the Department of Atomic Energy, Government of India, under project no. RTI4001 and by a Google India faculty research award. The authors gratefully acknowledge comments from the anonymous reviewers, which substantially improved the quality of the presentation.

References

Elad Hazan. *Introduction to online convex optimization*. MIT Press, 2022.

Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*, 2016.

Wen Sun, Debadeepta Dey, and Ashish Kapoor. Safety-aware algorithms for adversarial contextual bandit. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3280–3288. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/sun17a.html.

Abhishek Sinha. BanditQ:Fair Bandits with Guaranteed Rewards. In Negar Kiyavash and Joris M. Mooij, editors, *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, volume 244 of *Proceedings of Machine Learning Research*, pages 3227–3244. PMLR, 15–19 Jul 2024. URL https://proceedings.mlr.press/v244/sinha24a.html.

Nikolaos Liakopoulos, Apostolos Destounis, Georgios Paschos, Thrasyvoulos Spyropoulos, and Panayotis Mertikopoulos. Cautious regret minimization: Online optimization with long-term budget constraints. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3944–3952. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/liakopoulos19a.html.

Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.

Ofer Dekel, Philip M Long, and Yoram Singer. Online multitask learning. In *International Conference on Computational Learning Theory*, pages 453–467. Springer, 2006.

Shie Mannor, John N Tsitsiklis, and Jia Yuan Yu. Online learning with sample path constraints. *Journal of Machine Learning Research*, 10(3), 2009.

Abhishek Sinha and Rahul Vaze. Optimal algorithms for online convex optimization with adversarial constraints. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=TxffvJMnBy.

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.

Michael J Neely and Hao Yu. Online convex optimization with time-varying constraints. *arXiv preprint arXiv:1702.04783*, 2017.

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

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.

Dhruv Sarkar, Samrat Mukhopadhyay, and Abhishek Sinha. Online learning for approximately-convex functions with long-term adversarial constraints. *arXiv preprint arXiv:2508.16992*, 2025.

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

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

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.

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

Xinlei Yi, Xiuxian Li, Tao Yang, Lihua Xie, Yiguang Hong, Tianyou Chai, and Karl H Johansson. Distributed online convex optimization with adversarial constraints: Reduced cumulative constraint violation bounds under slater's condition. *arXiv preprint arXiv:2306.00149*, 2023.

Rahul Vaze and Abhishek Sinha. $o(\sqrt{T})$ static regret and instance dependent constraint violation for constrained online convex optimization. *arXiv preprint arXiv:2502.05019*, 2025.

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

Yiyang Lu, Mohammad Pedramfar, and Vaneet Aggarwal. Order-optimal projection-free algorithm for adversarially constrained online convex optimization. *arXiv preprint arXiv:2502.16744*, 2025.

Subhamon Supantha and Abhishek Sinha. Universal dynamic regret and constraint violation bounds for constrained online convex optimization. *arXiv preprint arXiv:2510.01867*, 2025.

Nicolo Cesa-Bianchi, Yoav Freund, David Haussler, David P Helmbold, Robert E Schapire, and Manfred K Warmuth. How to use expert advice. *Journal of the ACM (JACM)*, 44(3):427–485, 1997.

Peter Auer, Nicolo Cesa-Bianchi, and Claudio Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64(1):48–75, 2002.

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

Zakaria Mhammedi, Wouter M Koolen, and Tim Van Erven. Lipschitz adaptivity with multiple learning rates in online learning. In *Conference on Learning Theory*, pages 2490–2511. PMLR, 2019.

Nicolo Cesa-Bianchi, Yishay Mansour, and Gilles Stoltz. Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66(2):321–352, 2007.

Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

Wouter M Koolen and Tim Van Erven. Second-order quantile methods for experts and combinatorial games. In *Conference on Learning Theory*, pages 1155–1175. PMLR, 2015.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

Abhishek Sinha. Source code for "Beyond $\tilde{O}(\sqrt{T})$ Constraint Violation for Online Convex Optimization with Adversarial Constraints"; A. Sinha, R. Vaze. https://github.com/abhishek-sinha-tifr/COCO-improved-CCV, 2025.

Haipeng Luo. Lecture 4. *Introduction to Online Learning*, 2017. https://haipeng-luo.net/courses/CSCI699/index.html.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We provide complete theorem statements along with detailed proofs for all claims presented in the paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Our results critically rely on the feasibility assumption (Assumption 2), which is standard in the COCO literature. While this assumption is widely adopted, it remains an interesting direction for future work to explore how and to what extent it can be relaxed. Additionally, as discussed in the conclusion, deriving matching lower bounds remains an open and important problem.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: We clearly state the assumptions under which our theoretical results hold. Complete proofs of all theoretical claims have been provided either in the main paper or in the Appendix.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The complete experimental set up with all relevant parameter values has been included in Section 8.7 of the Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The code has been made publicly available [Sinha, 2025].

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The complete experimental set up with all relevant parameter values has been included in Section 8.7 of the Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: Since our paper proves bounds on the worst-case performance metric, error-bars are not essential for the experimental results.

   Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Relevant details have been provided in Section 8.7 of the Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The paper conform, in every respect, with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: This is a theoretical paper and the authors do not see any immediate direct societal consequences of this paper.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.

8 Appendix

8.1 On the Boundedness Assumption

Assumption 1 implies that there exist constants K_f, K_g such that

$$|f_t(x)| \leq K_f, \;\; |g_t(x)| \leq K_g, \forall x \in \mathcal{X}, \forall t \geq 1.$$

Since the constraint function g_t is non-negative, the above implies

$$|f_t(x)| \leq K_f, \;\; 0 \leq g_t(x) \leq K_g, \forall x \in \mathcal{X}, \forall t \geq 1.$$

Now consider the following translated and scaled version of the cost and constraint functions:

$$\tilde{f}_t(x) \coloneqq \frac{f_t(x)}{2K_f} + \frac{1}{2}, \;\; \tilde{g}_t(x) \coloneqq \frac{g_t(x)}{K_g}, x \in \mathcal{X}, t \geq 1.$$

It is easy to verify that $0 \leq \tilde{f}_t, \tilde{g}_t \leq 1, \forall t \geq 1$. Hence, we can work with these modified cost and constraint functions.

P.S. With the feasibility assumption (Assumption 2), we can even obtain an explicit expression for K_g. Let $x^\star$ be a feasible action. Using the G-Lipschitzness of the function g_t, we have

$$g_t(x) \leq g_t(x^\star) + G\|x_t - x^\star\| \leq 0 + GD.$$

Thus we can take $K_g \equiv GD$. Consequently, we only need to assume the cost functions to be bounded.

8.2 Pseudocode for the Adaptive Hedge Policy

Algorithm 4 Adaptive Hedge Algorithm for the EXPERT problem with Unbounded Losses

Input: Number of experts N, Horizon length T, Non-negative and potentially unbounded loss vectors $\{l_t\}_{t=1}^T, \gamma \geq 1$.
Initialization: Algorithm's loss $\tilde{L}_0 = 0$, Losses of the experts $L_0 = \mathbf{0}$, $G_0 = 1$.
for each $t = 1 : T$ **do**
    Compute self-confident learning rate

$$\eta_t = \frac{1}{\sqrt{G_{t-1}}}\sqrt{\frac{\ln N}{\tilde{L}_{t-1} + \gamma G_{t-1}}}. \tag{25}$$

    Play adaptive Hedge by choosing the ith expert with probability

$$p_t(i) = \exp(-\eta_t L_{t-1}(i)) / \sum_{j=1}^{N} \exp(-\eta_t L_{t-1}(j)), \;\; \forall i \in [N].$$

    The loss vector l_t is revealed to the learner
    Update the algorithm's and experts' losses

$$\tilde{L}_t \leftarrow \tilde{L}_{t-1} + \langle l_t, p_t \rangle, \;\; L_t \leftarrow L_{t-1} + l_t$$

    G_t is an upper bound to $\|l_t\|_\infty$ such that $1 \leq G_{t+1}/G_t \leq \gamma$
end for each

8.3 Proof of Theorem 1

Our proof adapts the arguments given in Luo [2017] while accounting for unbounded losses. Let $L_t(i)$ denote the cumulative loss of the ith expert up to round t, *i.e.,* $L_t(i) = \sum_{\tau=1}^t l_\tau(i), i \in [N]$.

Define the potential function $\Phi_t(\eta) = \frac{1}{\eta}\log\left(\frac{1}{N}\sum_{i=1}^N \exp(-\eta L_t(i))\right)$. We have

$$\begin{aligned}
\Phi_t(\eta_t) - \Phi_{t-1}(\eta_t) &= \frac{1}{\eta_t}\ln\left(\frac{\sum_{i=1}^N \exp(-\eta_t L_t(i))}{\sum_{i=1}^N \exp(-\eta_t L_{t-1}(i))}\right) \\
&= \frac{1}{\eta_t}\ln\left(\sum_{i=1}^N p_t(i)\exp(-\eta_t l_t(i))\right) \\
&\overset{(a)}{\leq} \frac{1}{\eta_t}\ln\left(\sum_{i=1}^N p_t(i)\big(1-\eta_t l_t(i) + \frac{\eta_t^2 l_t^2(i)}{2}\big)\right) \\
&= \frac{1}{\eta_t}\ln\left(1-\eta_t\langle p_t, l_t\rangle + \frac{\eta_t^2}{2}\sum_{i=1}^N p_t(i) l_t^2(i)\right) \\
&\overset{(b)}{\leq} -\langle p_t, l_t\rangle + \frac{\eta_t}{2}\sum_{i=1}^N p_t(i) l_t^2(i).
\end{aligned}$$

where in inequality (a), we have used the fact that $e^{-y} \leq 1 - y + \frac{y^2}{2}, \forall y \geq 0$, and in (b), we have used the fact that $1 + y \leq e^y, \forall y \in \mathbb{R}$. Thus we have that

$$\langle p_t, \ell_t\rangle \leq \Phi_{t-1}(\eta_t) - \Phi_t(\eta_t) + \frac{\eta_t}{2}\sum_{i=1}^N p_t(i)\ell_t^2(i).$$

Summing the above inequality for $1 \leq t \leq T$ yields

$$\begin{aligned}
&\tilde{L}_T \\
=\ & \sum_{t=1}^T \langle p_t, l_t\rangle \\
\leq\ & \Phi_0(\eta_1) - \Phi_T(\eta_{T+1}) + \sum_{t=1}^T \frac{\eta_t}{2}\sum_{i=1}^N p_t(i)\ell_t^2(i) + \sum_{t=1}^T(\Phi_t(\eta_{t+1}) - \Phi_t(\eta_t)) \\
\overset{(a)}{\leq}\ & \frac{\ln N}{\eta_{T+1}} - \frac{1}{\eta_{T+1}}\ln\left(\exp(-\eta_{T+1}L_T(i^\star))\right) + \sum_{t=1}^T \frac{\eta_t}{2} G_t \sum_{i=1}^N p_t(i)\ell_t(i) + \sum_{t=1}^T(\Phi_t(\eta_{t+1}) - \Phi_t(\eta_t)) \\
\overset{(b)}{\leq}\ & \sqrt{(\tilde{L}_T + \gamma G_T)G_T \ln N} + L_T(i^\star) + \underbrace{\gamma\sqrt{G_T \ln N}\sum_{t=1}^T \frac{\langle p_t, \ell_t\rangle}{2\sqrt{\tilde{L}_{t-1} + \gamma G_{t-1}}}}_{(A)} + \underbrace{\sum_{t=1}^T(\Phi_t(\eta_{t+1}) - \Phi_t(\eta_t))}_{(B)},
\end{aligned} \tag{26}$$

where, in (a) we have used the fact that $\|l_t\|_\infty \leq G_t$, and in (b) we have used the expression (25) for the learning rate η_t along with the fact that $G_t \leq \gamma G_{t-1}$ and $G_t \leq G_T, \forall t$.

To bound term (A) in Eqn. (26), observe that

$$\tilde{L}_t = \tilde{L}_{t-1} + \langle p_t, l_t\rangle \leq \tilde{L}_{t-1} + G_t \leq \tilde{L}_{t-1} + \gamma G_{t-1}.$$

Hence,

$$\sum_{t=1}^T \frac{\langle p_t, \ell_t\rangle}{\sqrt{\tilde{L}_{t-1} + \gamma G_{t-1}}} \leq \sum_{t=1}^T \frac{\tilde{L}_t - \tilde{L}_{t-1}}{\sqrt{\tilde{L}_t}} \overset{(c)}{\leq} \int_{\tilde{L}_0}^{\tilde{L}_T} \frac{dx}{\sqrt{x}} = 2\sqrt{\tilde{L}_T},$$

where in inequality (c), we have used the monotonicity of the sequence $\{\tilde{L}_t\}_{t\geq 1}$.

Furthermore, it can be readily verified that the potential function $\Phi_t(\eta)$ is non-decreasing in η as $\Phi_t'(\eta) \geq 0$ [Luo, 2017]. Since the learning rate η_t is non-increasing in t, we conclude that each term in term (B) in Eqn. (26) is non-positive. Hence, using the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, from Eqn. (26), we have

$$\tilde{L}_T \leq 2\gamma\sqrt{\tilde{L}_T G_T \ln N} + L_T(i^\star) + G_T\sqrt{\gamma \ln N}.$$

In the above, we have used the fact that $\gamma \geq 1$. Solving the above quadratic inequality using Lemma 1 below, we conclude that the adaptive Hedge policy enjoys the following small-loss regret bound for the EXPERT problem:

$$\tilde{L}_T - L_T(i^\star) \leq 2\gamma\sqrt{L_T(i^\star)G_T \ln N} + 7\gamma^2 G_T \ln N.$$

8.4 Proof of Auxiliary Lemmas

Lemma 1 (A quadratic inequality)**.** *Consider the following quadratic inequality* $x^2 \le ax + b$*, where* $a \ge 0, b \ge 0$*. Then we have* $x^2 \le a^2 + b + a\sqrt{b}$*.*

Proof. Solving the given inequality using the usual quadratic formula and the fact that $\sqrt{x+y} \le \sqrt{x} + \sqrt{y},\ x, y \ge 0,$, we have

$$x \le \frac{a + \sqrt{a^2 + 4b}}{2} \le \frac{a + a + 2\sqrt{b}}{2} = a + \sqrt{b}.$$

Hence,

$$x^2 \le a(a + \sqrt{b}) + b = a^2 + b + a\sqrt{b}.$$

□

Lemma 2 (Bounding γ)**.** *Let* $\Phi(x) = e^{\lambda x}, \lambda = T^{-(1-\beta)}/(2c \ln N)$*, where* $c = 10$*. Define* $G_t = 1 + \Phi'(Q(t)), t \ge 0$ *and* $\gamma \equiv \max_{1 \le t \le T} \frac{G_t}{G_{t-1}}$*. Then we have* $\gamma \le 1.08$*.*

Proof. Since $0 \le g_t \le 1$, from Eqn. (9), we have

$$Q(t) = Q(t-1) + \langle g_t, p_t \rangle \le Q(t-1) + 1. \tag{27}$$

Thus

$$\frac{G_t}{G_{t-1}} = \frac{1 + \lambda e^{\lambda Q(t)}}{1 + \lambda e^{\lambda Q(t-1)}} \overset{(a)}{\le} \frac{1 + \lambda e^{\lambda} e^{\lambda Q(t-1)}}{1 + \lambda e^{\lambda Q(t-1)}} \le e^{\lambda}. \tag{28}$$

where in (a) we have used inequality (27). Since $N \ge 2$ and $T \ge 1$, we have $\lambda \le (2c \ln 2)^{-1}$. Hence, Eqn. (28) implies that

$$\gamma \le e^{1/(20 \ln 2)} \le 1.08.$$

□

8.5 Construction of a Minimal δ-cover of the Decision Set $\mathcal{X}$

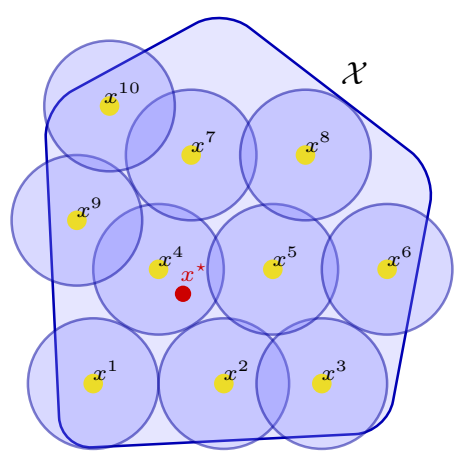

Figure 1: Schematic depicting the greedy construction of a minimal δ-cover of the decision set $\mathcal{X}$. If a point $x \in \mathcal{X}$ is not covered yet, we construct a ball of radius δ centred at x and include the point x in the δ-cover. The process continues till the entire set $\mathcal{X}$ is covered.

8.6 Relaxing the Feasibility Assumption

In this Section, we relax the feasibility Assumption 2 by allowing the benchmark actions to violate the constraints up to a given budget $B_T \geq 0$. In particular, we consider the following feasible set of actions which are budget feasible in the long term:

$$\mathcal{X}^\star_{B_T} = \{x \in \mathcal{X} : \sum_{t=1}^{T} g_t(x) \leq B_T\}. \tag{29}$$

We now replace Assumption 2 with the following assumption.

Assumption 5 (Long-term Feasibility)**.** *The feasible set is non-empty,* i.e., $\mathcal{X}^\star_{B_T} \neq \varnothing$.

Note that $\mathcal{X}^\star_{B_T} = \mathcal{X}^\star$ when $B_T = 0$. Since $0 \leq g_t \leq 1$ (Assumption 1), we clearly have $\mathcal{X}^\star_T = \mathcal{X}$, where $\mathcal{X}$ is the entire decision set. Thus, without any loss of generality, we can assume that $B_T \leq T$.

8.6.1 The Generalized Regret Decomposition Inequality

We now generalize the regret decomposition inequality (13) by taking into account the long-term feasibility constraint (29). Let $\Phi : \mathbb{R} \mapsto \mathbb{R}$ be any non-decreasing convex Lyapunov function defined as in Section 5 and let $x^\star \in \mathcal{X}^\star_{B_T}$ be any long-term feasible action satisfying the budget constraint (29). Recall that CCV evolves as $Q(t) = Q(t-1) + g_t(x_t)$. Hence, using the convexity of $\Phi(\cdot)$, we obtain

$$\begin{aligned}
&\Phi(Q(t)) - \Phi(Q(t-1)) + \big(f_t(x_t) - f_t(x^\star)\big) \\
&\leq \big(f_t(x_t) + \Phi'(Q(t))g_t(x_t)\big) - \big(f_t(x^\star) + \Phi'(Q(t))g_t(x^\star)\big) + \Phi'(Q(t))g_t(x^\star).
\end{aligned}$$

where, unlike in Section 5, the violation $g_t(x^\star)$ in this case could be strictly positive. Summing up the above inequalities, we conclude

$$\begin{aligned}
\Phi(Q(T)) - \Phi(Q(0)) + \text{Regret}_T(x^\star) &\overset{(a)}{\leq} \text{Regret}'_T(x^\star) + \Phi'(Q(T)) \sum_{t=1}^{T} g_t(x^\star) \\
&\overset{(b)}{\leq} \text{Regret}'_T(x^\star) + \Phi'(Q(T))B_T,
\end{aligned} \tag{30}$$

where, as before, $\text{Regret}_T(x^\star)$ and $\text{Regret}'_T(x^\star)$ denote the regrets for learning the original cost functions $\{f_t\}_{t\geq 1}$ and the surrogate cost functions $\{\hat{f}_t\}_{t\geq 1}$ respectively w.r.t. the long-term feasible benchmark $x^\star$ (see Eqn. (23) for the definition of the surrogate cost functions). In the above inequality, the upper bound in step (a) follows from the monotonicity of the CCV $(Q(t))_{t\geq 1}$ and the convexity of the Lyapunov function $\Phi(\cdot)$, while step (b) uses the budget constraint for $x^\star \in \mathcal{X}^\star_{B_T}$.

In the following, we consider the case of convex and smooth cost and constraint functions under the relaxed feasibility assumption (see Section 5). The analysis for the CONSTRAINED EXPERT problem is identical.

8.6.2 Convex and Smooth Cost and Constraints with Long-term Budget Constraints

Inequality (30) is of the same form as the regret decomposition inequality (13) under Assumption 2, albeit with an extra additive term $\Phi'(Q(T))B_T$ appearing on the right-hand side. By using the same exponential Lyapunov function as before (with the parameter λ now adapted to the budget B_T), we can analogously solve the above functional inequality and derive the corresponding regret and CCV bounds under long-term violation budget constraints.

The cumulative cost of the surrogate functions under any long-term feasible action $x^\star \in \mathcal{X}^\star_{B_T}$ can be upper bounded as

$$\sum_{t=1}^{T} \hat{f}_t(x^\star) = \sum_{t=1}^{T} f_t(x^\star) + \sum_{t=1}^{T} \Phi'(Q(t))g_t(x^\star) \overset{(a)}{\leq} T + \Phi'(Q(T)) \sum_{t=1}^{T} g_t(x^\star) \overset{(b)}{\leq} T + \Phi'(Q(T))B_T, \tag{31}$$

where in inequality (a), we have again made use of the convexity of the Lyapunov function and the non-decreasing property of the CCV, and in (b), we have used the budget constraint for $x^\star$. Hence,

using Eqn. (22) from Theorem 4, the regret for the surrogate costs under the action of Algorithm 3 can be upper bounded as:

$$\begin{aligned}\text{Regret}'_T(x^\star) &\leq 4D\sqrt{M(T+\Phi'(Q(T)B_T)(1+\Phi'(Q(T))}+4(1+\Phi'(Q(T))D^2M \\ &\leq c_1\sqrt{T}+c_2\sqrt{T\Phi'(Q(T))}+c_3\Phi'(Q(T))\sqrt{B_T}, \end{aligned} \tag{32}$$

where c_1, c_2, c_3 are generic constants which depend only on the problem-specific parameters M and D. If necessary, the reader can easily figure out the explicit values of these constants in each line of the derivation below.

As in Section 3.1, we now set $\Phi(\cdot)$ to be the exponential Lyapunov function, *i.e.,* $\Phi(x) := \exp(\lambda x)$, for a suitable value of the parameter λ which will be fixed later. With this choice, the regret decomposition inequality (30) yields

$$e^{\lambda Q(T)} + \text{Regret}_T(x^\star) \leq c_1\sqrt{T} + c_2\sqrt{\lambda T}e^{\lambda Q(T)/2} + c_3\lambda B_T e^{\lambda Q(T)}.$$

We now choose some λ such that $\lambda \leq (2c_3B_T)^{-1}$. Hence, the above equation yields

$$\frac{1}{2}e^{\lambda Q(T)} + \text{Regret}_T(x^\star) \leq c_1\sqrt{T} + c_2\sqrt{\lambda T}e^{\lambda Q(T)/2}. \tag{33}$$

The Regret and CCV bounds are obtained by solving the above inequality.

Bounding the CCV: Using the fact that $\text{Regret}_T(x^\star) \geq -T$, Eqn. (33) yields

$$e^{\lambda Q(T)} \leq c_1T + c_2\sqrt{\lambda T}e^{\lambda Q(T)/2} \leq 2\max(c_1T, c_2\sqrt{\lambda T}e^{\lambda Q(T)/2}).$$

This implies the following bounds for $Q(T)$:

$$Q(T) \leq \lambda^{-1}(c_1 + \log\lambda + \log T) = O(\lambda^{-1}\log T).$$

Bounding the Regret: Starting from inequality (33) once again, we have the following upper bound on regret

$$\text{Regret}_T(x^\star) \leq c_1\sqrt{T} + c_2\sqrt{\lambda T}e^{\lambda Q(T)/2} - \frac{1}{2}e^{\lambda Q(T)}.$$

Using the fact that $ax - bx^2 \leq \frac{a^2}{4b}, \forall b > 0$, and taking $x = e^{\lambda Q(T)/2}$, the regret can be further upper bounded as follows:

$$\text{Regret}_T(x^\star) \leq c_1\sqrt{T} + \frac{c_2^2}{2}\lambda T = O(\max(\sqrt{T}, \lambda T)).$$

Finally, choosing $\lambda = \min(\frac{1}{2c_3B_T}, T^{-(1-\beta)})$ for some $0 \leq \beta \leq 1$, we obtain the following trade-off

$$Q(T) = \tilde{O}(\max(B_T, T^{1-\beta})),\ \text{Regret}_T(x^\star) = O(\max(\sqrt{T}, T^\beta)). \tag{34}$$

As an example, if $B_T = O(T^{1/3})$, then by choosing $\beta = 2/3$, we obtain $\text{Regret}_T(x^\star) = O(T^{2/3})$ and $\text{CCV}_T = \tilde{O}(T^{1/3})$.

The above results are summarized in the following Theorem.

Theorem 6. *Let $B_T \geq 0$ be the prescribed long-term constraint violation budget. Consider the Lyapunov function $\Phi(x) = e^{\lambda x}$ with $\lambda = \min(\frac{1}{cB_T}, T^{-(1-\beta)})$ where $\beta \in [0,1]$ is a tunable parameter and c is a constant which depends on the problem parameters (D and M) as discussed above. Then, under Assumptions 1, 4, and 5, Algorithm 3 achieves the following guarantee for any long-term feasible benchmark action $x^\star \in \mathcal{X}^\star_{B_T}$:*

$$\textit{Regret}_T(x^\star) = O(\max(\sqrt{T}, T^\beta)),\ \tilde{O}(\max(B_T, T^{1-\beta})).$$

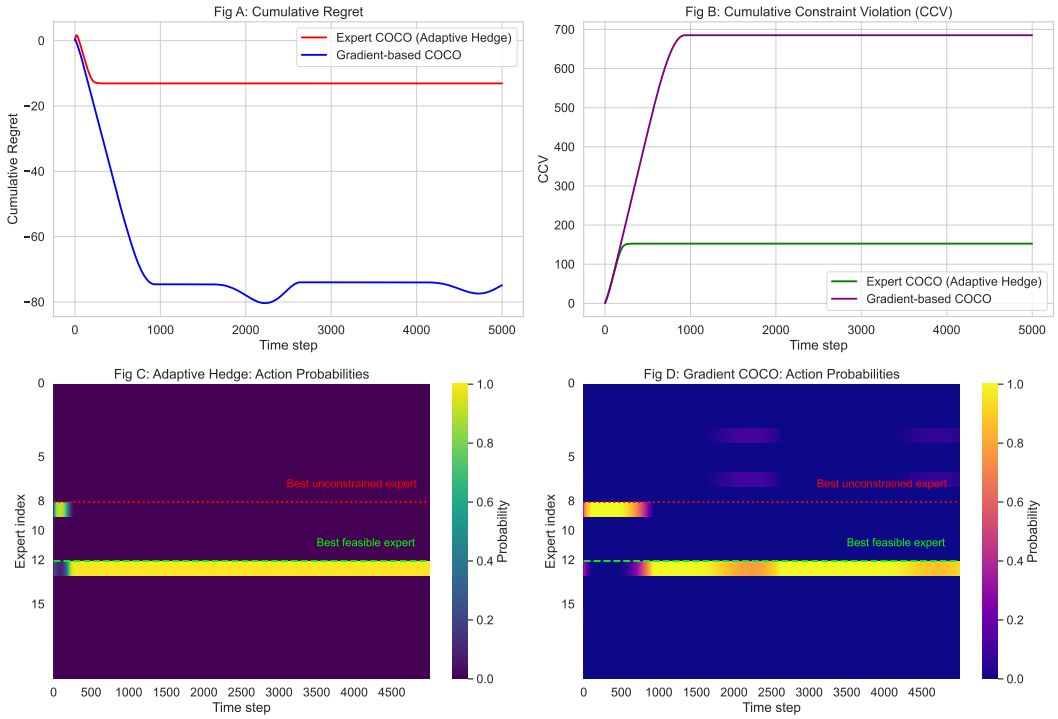

Figure 2: Comparison of the proposed adaptive Hedge-based policy (Algorithm 1) with $\beta = 0.75$ and the OGD-based policy from Sinha and Vaze [2024, Algorithm 1]. (A) Regret vs. time. (B) Cumulative Constraint Violation (CCV) vs. time. (C) Selection frequency of experts under our adaptive Hedge-based policy. (D) Selection frequency of experts under the OGD-based policy. The proposed policy quickly identifies and sticks to the best feasible expert, leading to significantly lower CCV, whereas the OGD-based policy initially selects infeasible experts and takes longer to converge.

8.7 Experiments

Problem instance: We consider the CONSTRAINED EXPERT problem on a synthetic dataset with $N = 20$ experts over a horizon of length $T = 5000$. Two experts are designated as special: the best feasible expert, denoted by $\mathtt{E}^\star$, and the best unconstrained expert, denoted by $\mathtt{UE}^\star$. The expert $\mathtt{E}^\star$ is feasible, with i.i.d. costs drawn from a distribution with mean $\bar{f}_{\mathtt{E}^\star} = 0.21$, and zero constraint violation on all rounds, i.e., $\bar{g}_{\mathtt{E}^\star} = 0.0$. In contrast, $\mathtt{UE}^\star$ is infeasible, with i.i.d. costs and constraint violations having a smaller mean $\bar{f}_{\mathtt{UE}^\star} = 0.11$ and higher average constraint violation $\bar{g}_{\mathtt{UE}^\star} = 0.91$, respectively. The remaining experts incur i.i.d. random costs with mean $\bar{f} = 0.41$ plus a zero-mean periodic component over time, and their constraint violations are i.i.d. with mean $\bar{g} = 0.6$. Additionally, two more experts, $\mathtt{DE}_1$ and $\mathtt{DE}_2$, distinct from both $\mathtt{E}^\star$ and $\mathtt{UE}^\star$, are made feasible by setting their constraint violations to zero on all rounds. Table 2 summarizes the parameters used in the experiments. The experiments have been run on a quad-core CPU with 8 GB RAM. The source code has been made publicly available [Sinha, 2025].

Expert(s)	Index	Average cost	Average constraint violation
$\mathtt{E}^\star$	$\#12$	$\bar{f}_{\mathtt{E}^\star} = 0.21$	$\bar{g}_{\mathtt{E}^\star} = 0.0$
$\mathtt{UE}^\star$	$\#8$	$\bar{f}_{\mathtt{E}^\star} = 0.11$	$\bar{g}_{\mathtt{E}^\star} = 0.91$
$\mathtt{DE}_1, \mathtt{DE}_2$	$\#3, \#6$	$\bar{f}_{\mathtt{E}^\star} = 0.41$	$\bar{g}_{\mathtt{E}^\star} = 0.0$
`The rest`	$[20] \setminus \{3, 6, 8, 12\}$	$\bar{f}_{\mathtt{E}^\star} = 0.41$	$\bar{g}_{\mathtt{E}^\star} = 0.6$

Table 2: Parameter settings for generating cost and constraint vectors for $N = 20$ experts.

Summary of the results. Parts (A) and (B) of Figure 2 compare the performance of our proposed adaptive Hedge-based policy (Algorithm 1) with $\beta = 0.75$ with that of the Online Gradient Descent

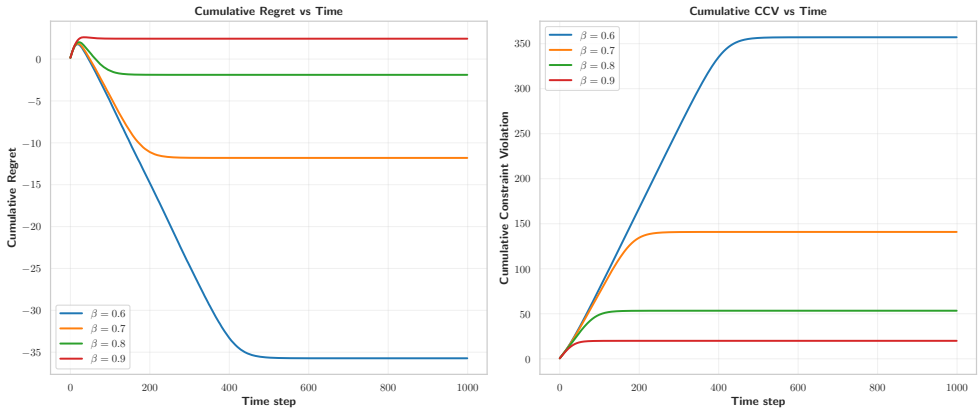

Figure 3: Performance comparison for different values of β

(OGD)-based policy proposed by Sinha and Vaze [2024, Algorithm 1]. From the plots, it is evident that the proposed algorithm incurs significantly lower cumulative constraint violation (CCV) while maintaining a sub-linear regret.

To gain deeper insight into the working of the algorithms, parts (C) and (D) of Figure 2 show the relative frequency with which each expert is selected by our algorithm and the OGD-based policy, respectively. These plots clearly demonstrate that the adaptive Hedge-based policy quickly identifies the best feasible expert and predominantly selects it thereafter. In contrast, the OGD-based policy initially incurs a substantial amount of constraint violation by frequently selecting infeasible experts. Only after a considerable number of rounds does it converge to the best feasible expert and begin exploiting it consistently.

Figure 3 illustrates the trade-off between regret and cumulative constraint violation (CCV) for different values of the tuning parameter β in the proposed policy. As β increases from 0.6 to 0.9, the algorithm incurs lower CCV at the expense of higher regret. This behavior aligns with the theoretical performance guarantees established in Theorem 2.