Proceedings of Machine Learning Research vol 134:1–32, 2021

Lazy OCO: Online Convex Optimization on a Switching Budget

Uri Sherman URISHERMAN@MAIL.TAU.AC.IL
Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel
Tomer Koren TKOREN@TAUEX.TAU.AC.IL
Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel
Google Research, Tel Aviv, Israel

Abstract

We study a variant of online convex optimization where the player is permitted to switch decisions at most S times in expectation throughout T rounds. Similar problems have been addressed in prior work for the discrete decision set setting, and more recently in the continuous setting but only with an adaptive adversary. In this work, we aim to fill the gap and present computationally efficient algorithms in the more prevalent oblivious setting, establishing a regret bound of $O(T/S)$ for general convex losses and $\widetilde{O}(T/S^2)$ for strongly convex losses. In addition, for stochastic i.i.d. losses, we present a simple algorithm that performs $\log T$ switches with only a multiplicative $\log T$ factor overhead in its regret in both the general and strongly convex settings. Finally, we complement our algorithms with lower bounds that match our upper bounds in some of the cases we consider.

1. Introduction

We study online convex optimization with limited switching. In the classical online convex optimization (OCO) problem, a player and an adversary engage in a T-round game, where in each round, the player chooses a decision $w_t \in W \subseteq \mathbb{R}^d$, and the adversary responds with a loss function $f_t : W \to \mathbb{R}$. The losses f_t are convex functions over W which is also convex and traditionally referred to as the decision set. Each round incurs a loss of $f_t(w_t)$ against the player, whose objective is to minimize her cumulative loss. The performance of the player is then measured by her regret, defined as the difference between her cumulative loss and that of the best fixed decision in hindsight;

$$\sum_{t=1}^{T} f_t(w_t) - \min_{w \in W} \sum_{t=1}^{T} f_t(w).$$

This theoretical framework has found diverse applications in recent years, many of which benefit from player strategies that switch decisions sparingly. In adaptive network routing (Awerbuch and Kleinberg, 2008) switching decisions amounts to changing packet routes, which should be kept to a minimum as it may lead to severe networking problems (see, e.g., Feamster et al., 2014). When investing in the stock market, transactions may be associated with fixed commission costs, and thus trading strategies that change stock positions infrequently are of value. As another example, Geulen et al. (2010) approach online buffering by devising a low switching variant of the well known Multiplicative Weights algorithm. In addition, recent applications of OCO in online reinforcement learning and control problems involve addressing the fact that changing policies introduces short-term penalties, and thus could benefit from keeping the number of policy switches to a minimum (e.g., Cohen et al., 2018, 2019; Agarwal et al., 2019a,b; Foster and Simchowitz, 2020).

This motivates the study of regret bounds achievable when we limit the number of decision switches the player is allowed to perform. When the limit is applied to the expected number of decision switches, we arrive at a variant of the standard model we shall refer to as *lazy OCO*. In particular, we say that an OCO algorithm is S-lazy if the expected number of switches it performs over T rounds is less than S. A closely related problem where the player is charged a fixed price $c > 0$ per switch has been the focus of several works in the past, though mainly in the context of experts or multi-armed bandit problems (Dekel et al., 2014; Geulen et al., 2010; Altschuler and Talwar, 2018). It is not hard to see this problem, which we refer to as switching-cost OCO, is effectively equivalent to lazy OCO. (We defer formal details to the full version of the paper (Sherman and Koren, 2021).)

Perhaps the most natural approach for this problem would be to divide the T rounds into S equally sized time-blocks, and treat the cumulative loss of each block as a single loss function. This effectively reduces the game to the standard unconstrained OCO setting with S rounds, and a Lipschitz constant larger by a factor of T/S. This method has been termed "blocking argument" and dates back at least to Merhav et al. (2002) who used it to obtain an $O(T/\sqrt{S})$ regret bound on the prediction with expert advice problem. It is not hard to see that this strategy also yields $O(T/\sqrt{S})$ regret in the general convex setting. Recently, Chen et al. (2019) prove that this is in fact optimal against an *adaptive* adversary with linear losses. However, in the oblivious adversary setting, stronger results may be achieved owed to the power of randomization. For example, several works (Kalai and Vempala, 2005; Geulen et al., 2010; Devroye et al., 2013) have obtained a stronger $O(T/S)$ bound for the experts problem by employing randomized player strategies. In the general convex setting results have been more scarce, though the same bound has been achieved by Anava et al. (2015), who adapt the method of Geulen et al. (2010) to the continuous online optimization setup.

In this work, we aim to further develop our understanding of lazy OCO by addressing a number of questions that have remained open. First, the algorithm in Anava et al. (2015) relies on complex procedures for sampling from log-concave distributions and thus does not admit a practical implementation; further, while their result is optimal for $S = \Omega(\sqrt{T})$ in terms of dependence on T, it is not clear a priori whether an $O(\sqrt{T})$ regret bound may be obtained with less switches. Moreover, to the best of our knowledge, no results have been established in the strongly convex lazy OCO setting, neither for adaptive adversaries nor for oblivious ones. This work aims to fill these gaps and obtain a more coherent understanding of the lazy OCO problem.

1.1. Our contributions

We make the following contributions:

- *Regret upper bounds.* We present a computationally efficient S-lazy algorithm, achieving an $O(\sqrt{T} + \sqrt{d}T/S)$ regret bound for general convex losses, and an $\widetilde{O}(dT/S^2)$ bound for strongly convex losses. Compared to the algorithm of Anava et al. (2015), our scheme is substantially simpler and more efficient, and also features a better dependence on the dimension ($d \to \sqrt{d}$) for general convex losses. More importantly, it extends to the strongly convex case where it obtains improved regret bounds, which does not seem to be the case for their algorithm.[1]

- *Regret lower bounds.* For the general convex case, we prove an $\Omega(T/S)$ lower bound for the regret of any S-lazy algorithm, matching our upper bound in this setting in terms of dependence

1. In fact, a closer look into the regret analysis of Anava et al. (2015) reveals that it does not at all exploit the convexity of the losses: by taking a discretization of the decision set and using, e.g., the Shrinking-Dartboard algorithm of Geulen et al. (2010), one would obtain essentially the same regret guarantee (albeit not in polynomial time).

on T and S. For strongly convex losses, we prove an $\Omega(T/S^4)$ lower bound for an oblivious adversary, and an $\Omega(T/S)$ lower bound for an adaptive adversary. The result in the adaptive setting matches up to a logarithmic factor the $O((T/S)\log S)$ upper bound obtained by a straightforward application of the blocking technique in this setting (see the full version of the paper (Sherman and Koren, 2021) for details).

- *Regret bounds for stochastic i.i.d. losses.* For the special case of a stochastic i.i.d. adversary, we present an algorithm that performs $O(\log T)$ switches while introducing only a multiplicative $O(\log T)$ factor overhead in the regret bound compared to unrestricted OCO.

Table 1 lists our contributions compared to the relevant state-of-the-art bounds. Our upper bounds are discussed in Sections 3 and 4, and the general convex lower bound in Section 5. For the lower bounds in the strongly convex setting, see the full version of the paper (Sherman and Koren, 2021).

1.2. Key ideas and techniques

Our starting point for designing lazy algorithms in the (adversarial, oblivious) OCO setup is the general idea present in Follow-the-Lazy-Leader (FLL) algorithm of Kalai and Vempala (2005), where perturbations are introduced for obfuscating small changes in the player's (unperturbed) decisions. Then, the perturbations may be correlated in such a way that preserves the marginal distribution of decisions, and at the same time have sufficient overlap in total variation, which allows for the player to avoid switching altogether across several consecutive rounds. However, unlike Kalai and Vempala (2005) who study the linear case, we are interested in the general convex and strongly convex settings, which pose a number of additional challenges.

First, the subset of perturbed objectives cannot be fixed in advance and must be determined per round in a dynamical fashion during execution of the algorithm. This is because the particular perturbation that leads to the same decision being used across rounds depends on the loss sequence in a way that mandates ad-hoc coupling between consecutive minimization objectives. Specifically, in our algorithm it depends on the gradient of the loss evaluated at the decision from the previous round. It is not hard to see that the linear case allows for a broader set of solutions to this task, as the gradient of a linear loss is the same regardless of the point at which it is evaluated.

In addition, in the general non-linear convex setting, the unperturbed minimizers need to be stabilized so that consecutive decision distributions overlap sufficiently in total variation. To that end, a regularization component is added to obtain the desired relation between regret and the number of decision switches. Thus, unlike Follow-the-Perturbed-Leader-type algorithms that introduce perturbations for promoting stability, we draw our stability properties from a regularization component while using perturbations only for inducing proximity in total variation, which in turn allows the algorithm to resample decisions less frequently.

Finally, our regret bound for strongly convex losses makes use of two additional ideas that were key in achieving the improved dependence on the switches parameter S. First, perhaps surprisingly, the perturbation variance has to be *increased* at a rate that is in accordance with the increasing curvature of the per round minimization objective (despite the fact that the unperturbed decision actually becomes more and more stable with time). The second and more crucial observation is that the regret penalty introduced by perturbations on top of the hypothetical "be-the-leader" strategy can be bounded much more efficiently for strongly convex losses: our analysis reveals that this penalty depends on the distance between the deterministic, unperturbed minimizer and the perturbed random

Table 1: S-lazy OCO bounds, omitting factors other than S and T. Our contributions are in boldface.

SETTING	ADVERSARY	LOWER BOUND	UPPER BOUND
Experts	Oblivious	T/S Geulen et al. (2010)	T/S [a] Kalai and Vempala (2005) [b]
OCO	Adaptive	$T/\sqrt{S}$ Chen et al. (2019)	$T/\sqrt{S}$ Chen et al. (2019)
	Oblivious	$\boldsymbol{T/S}$	$\boldsymbol{T/S}$ [a]
	i.i.d.	$\sqrt{T}$	$\boldsymbol{\sqrt{T}\log T}$ [c]
OCO Strongly Convex	Adaptive	$\boldsymbol{T/S}$	$(T/S)\log S$
	Oblivious	$\boldsymbol{T/S^4}$	$\boldsymbol{(T/S^2)\log T}$
	i.i.d.	$\log T$	$\boldsymbol{\log^2 T}$ [c]

[a] For $S = O(\sqrt{T})$.
[b] Also Geulen et al. (2010); Devroye et al. (2013); Altschuler and Talwar (2018).
[c] For $S \geq 1 + \log T$.

one; crucially, with strong convexity, this distance shrinks rapidly with the number of steps at a rate that compensates for the increased perturbation variance.

1.3. Additional related work

Prior work on low switching strategies in online learning has been mostly concerned with the switching-cost perspective. All bounds we present here pertain to algorithms with an expected number of switches bounded by S, so that they are easily comparable. For completeness, their equivalent original switching-cost forms can be found in the full version of the paper (Sherman and Koren, 2021).

Experts. The experts problem with switching costs has been extensively studied, giving rise to several algorithms such as Follow-the-Lazy-Leader (FLL) (Kalai and Vempala, 2005), Shrinking-Dartboard (Geulen et al., 2010) and Perturbation-Random-Walk (Devroye et al., 2013), all of which achieve $O(T/S)$ regret known to be optimal due to a matching lower bound of Geulen et al. (2010). Recently, Altschuler and Talwar (2018) study experts and multi-armed bandits in the setting where the player is given a *hard cap* on the number of switches she is allowed (see the full version of the paper (Sherman and Koren, 2021) for a discussion of this variant of the model); they develop a framework converting Follow-the-Perturbed-Leader (FPL) type algorithms that work in expectation to ones with high probability guarantees, and leverage this result to achieve an upper bound of $\widetilde{O}(T/S)$ for $S = O(\sqrt{T})$, shown to be tight.

Multi-armed bandits. Unlike experts, the multi-armed bandit problem has proved to exhibit a more significant dependence on the number of switches, setting apart switching-cost regret from the standard unconstrained setting. An $O(T/\sqrt{S})$ upper bound was obtained by a blocking argument (Arora et al., 2012) applied to the EXP3 algorithm (Auer et al., 2002). A matching $\widetilde{\Omega}(T/\sqrt{S})$ lower bound was proved by Dekel et al. (2014). In the case of stochastic i.i.d. losses, Cesa-Bianchi et al. (2013) present an $\widetilde{O}(\sqrt{T})$ algorithm for multi-armed bandits that performs $O(\log \log T)$ switches.

Online convex optimization. To the best of our knowledge, Anava et al. (2015) establish the first and only $O(T/S)$ upper bound in the general convex setting with an oblivious adversary, albeit with a computationally intensive algorithm whose running time is bounded by a high-degree polynomial in the dimension. More recently, Chen et al. (2019) study lazy OCO in the general convex setting with an adaptive adversary and prove a tight $\Theta(T/\sqrt{S})$ result. Our work is thus complementary to theirs as we study lazy OCO in the oblivious setting, where stronger upper bounds turn out to be possible. Also relevant to our work is the paper of Jaghargh et al. (2019), who propose a Poisson process based algorithm for both general and strongly convex losses, although their results are suboptimal compared to those presented here.

Movement costs. Also related to lazy OCO is the study of movement costs in online learning, where the player pays a switching cost proportional to the distance between consecutive decisions. This variant was studied in the context of multi-armed bandits (Koren et al., 2017a,b), and is at the core of the well known metrical-task-systems (MTS) framework in competitive analysis (Borodin et al., 1992; Borodin and El-Yaniv, 2005). In particular, the continuous variant of MTS has been the subject of several works both in the low dimensional setting (Bansal et al., 2015; Antoniadis and Schewior, 2017), and in the high dimensional setting where it has been recently termed *smoothed OCO* (Chen et al., 2018; Goel et al., 2019; Shi et al., 2020). MTS differs from lazy OCO in a number of important ways; we refer to Blum and Burch (2000); Buchbinder et al. (2012); Andrew et al. (2013) for an extensive discussion of the relations between competitive analysis and regret minimization.

Correlated sampling. The algorithms we present are based on a lazy sampling procedure for sampling from maximal couplings (see Section 2.3). This procedure bears similarity to the well-known correlated sampling problem (Broder, 1997; Kleinberg and Tardos, 2002), where two players are given two probability distributions and are required to produce samples with minimal disagreement probability. As the players are not allowed to communicate, this problem is crucially different than sampling from maximal couplings; see Bavarian et al. (2020) for a more elaborate discussion.

2. Preliminaries

We start by giving a precise definition of our model and describe techniques and basic tools we use.

2.1. Problem setup

We describe the setting of lazy OCO, within which we develop all results presented in the paper. In this setting, an oblivious adversary chooses convex loss functions $f_t : W \longrightarrow \mathbb{R}$ over a convex domain $W \subseteq \mathbb{R}^d$. The game proceeds for T rounds, where in round t the player chooses $w_t \in W$, suffers loss $f_t(w_t)$, and observes f_t as feedback. We denote by $\mathcal{R}_T$ the player's regret;

$$\mathcal{R}_T := \sum_{t=1}^{T} f_t(w_t) - \min_{w \in W} \sum_{t=1}^{T} f_t(w),$$

and by $\mathcal{S}_T$ the number of decision switches she performs; $\mathcal{S}_T := \sum_{t=1}^{T-1} \mathbb{1}\{w_{t+1} \neq w_t\}$. When it is not clear from context, we may write $\mathcal{R}_T(\mathcal{A})$ and $\mathcal{S}_T(\mathcal{A})$ to make explicit which player we are referring to. We are interested in the asymptotic behavior of the player's regret, under the restriction she is obligated to perform a limited number $S \in [T]$ of switches in expectation; $\mathbb{E}\,\mathcal{S}_T \leq S$. We say $\mathcal{A}$ is an S-lazy algorithm if it satisfies that for any loss sequence $\mathbb{E}\,\mathcal{S}_T \leq S$.

2.2. Basic definitions and tools

The *diameter* of a set $W \subseteq \mathbb{R}^d$ is defined as $\max_{x,y\in W} \|x-y\|$. We denote by $\Pi_W(x) := \arg\min_{w\in W} \|w-x\|^2$ the orthogonal projection of a point $x \in \mathbb{R}^d$ onto W, but usually omit the subscript W and write $\Pi(x)$ unless the context requires to be explicit. For two probability distributions p, q over a sample space $\mathcal{X}$, we denote by

$$D_{KL}(p \parallel q) := \mathop{\mathbb{E}}_{x\sim p}\left[\log \frac{p(x)}{q(x)}\right],$$

the *KL-divergence* between p and q, and write $\|p-q\|_{TV}$ to denote their *total variation distance*;

$$\|p-q\|_{TV} := \max_{B\subseteq\mathcal{X}} \left\{p(B) - q(B)\right\}.$$

Also, a two dimensional random variable $(\mathbf{X}, \mathbf{Y})$ is a *coupling* of p and q if its marginals satisfy $\mathbf{X} \sim p$ and $\mathbf{Y} \sim q$. Throughout the paper, we use freely the well known facts described next. The total variation distance is related to the L1 norm as follows;

$$\|p-q\|_{TV} = \frac{1}{2}\|p-q\|_1, \tag{1}$$

and to the KL-divergence by *Pinsker's Inequality* (e.g., Cover, 1999);

$$\|p-q\|_{TV} \leq \sqrt{\frac{1}{2} D_{KL}(p \parallel q)}. \tag{2}$$

Finally, we denote by $\mathcal{N}(\mu, \sigma^2)$ the Gaussian distribution with mean $\mu \in \mathbb{R}^d$ and variance $\sigma^2 I \in \mathbb{R}^{d\times d}$ where $\sigma > 0$ and I is the identity matrix. For $\mu_1, \mu_2 \in \mathbb{R}^d$ and any $\sigma > 0$ we have that

$$D_{KL}\Big(\mathcal{N}(\mu_1, \sigma^2) \parallel \mathcal{N}(\mu_2, \sigma^2)\Big) = \frac{\|\mu_1 - \mu_2\|^2}{2\sigma^2}. \tag{3}$$

2.3. Sampling from maximal couplings

Algorithm 1 presented below provides a mechanism to maximally couple consecutive decision distributions. A similar procedure for sampling from maximal couplings can be found in the literature in various places, see e.g., Jacob et al. (2020). The desired properties of the algorithm follow from the two lemmas stated next. For completeness, we provide their proofs in the full version of the paper (Sherman and Koren, 2021). Throughout the paper, within an algorithmic context, we use the calligraphic font ($\mathcal{P}, \mathcal{Q}$, etc.) to denote computational objects that provide $O(1)$ oracle access to evaluate the density and to sample from a probability distribution.

Lemma 1. *Running LazySample($x, \mathcal{Q}, \mathcal{P}$) with $x \sim \mathcal{Q}$, we have that $\mathcal{P}$ is sampled from with probability $\|\mathcal{Q} - \mathcal{P}\|_{TV}$, where randomness is over choice of x and execution of the algorithm.*

Lemma 2. *Assume we run LazySample($x, \mathcal{Q}, \mathcal{P}$) with $x \sim \mathcal{Q}$, then the algorithm generates a return value distributed according to $\mathcal{P}$ in expected $O(1)$ time.*

Algorithm 1 LazySample

input: x, $\mathcal{Q}$, $\mathcal{P}$
Sample $z \sim \mathrm{Unif}[0, \mathcal{Q}(x)]$
If $\mathcal{P}(x) > z$, **return** x
Otherwise, repeat;
    Sample $y \sim \mathcal{P}$, and $z' \sim \mathrm{Unif}[0, \mathcal{P}(y)]$
    If $z' > \mathcal{Q}(y)$, **return** y

3. Lazy OCO

In this section, we present and analyze our lazy OCO algorithm for convex losses in the oblivious adversarial setup. The algorithm has a regularization component and generates decisions that are minimizers of a perturbed cumulative loss on each round. As such, it can be viewed as a natural combination of the well known Follow-the-Perturbed-Leader (FPL) algorithm (Kalai and Vempala, 2005) and Follow-the-Regularized-Leader (FTRL) meta-algorithm, with regularization being intrinsic in the strongly convex case. The resulting algorithm, given in Algorithm 2, is thus named Follow-the-Perturbed-Regularized-Lazy-Leader (FTPRLL).

The key idea is that stability introduced by regularization causes minimizers of the *unperturbed* objectives to move in small steps, thereby encouraging consecutive decisions—minimizers of the *perturbed* objectives—to overlap in total variation. This, combined with the lazy sampling sub-routine Algorithm 1, produces a low switching algorithm. Importantly, we note that while in FPL the perturbations are the source of stability, here regularization accounts for stability, and the perturbations serve to obfuscate the shifts between consecutive decisions.

Algorithm 2 Follow-The-Perturbed-Regularized-Lazy-Leader (FTPRLL)

input: perturbation parameters $\sigma_1, \ldots, \sigma_T$, regularizer $R : W \to \mathbb{R}$
Sample $p_1 \sim \mathcal{N}(0, \sigma^2)$
$w_1 \leftarrow \arg\min_{w \in W} \left\{ p_1^\mathsf{T} w + R(w) \right\}$
for $t = 1$ to T **do**
    Play w_t, Observe f_t
    $p_{t+1} \leftarrow \mathrm{LazySample}\big(p_t - \nabla f_t(w_t), \mathcal{N}(-\nabla f_t(w_t), \sigma_t^2), \mathcal{N}(0, \sigma_{t+1}^2)\big)$
    $w_{t+1} \leftarrow \arg\min_{w \in W} \left\{ \sum_{i=1}^t f_i(w) + p_{t+1}^\mathsf{T} w + R(w) \right\}$
end for

The expected number of switches performed by Algorithm 2 is governed solely by the gradient bound G and the perturbation variance sequence $\sigma_1^2, \ldots, \sigma_T^2$. The lemma we state below gives the bound in general form.

Lemma 3. *Running Algorithm 2 with $\sigma_1 \le \sigma_2 \le \ldots \le \sigma_T$ on any sequence of G-Lipschitz convex losses $f_1, \ldots, f_T$, it is guaranteed that*

$$\mathbb{E}\, \mathcal{S}_T \le \frac{G}{2} \sum_{t=1}^T \frac{1}{\sigma_t}.$$

The proof of Lemma 3 shows that the way Algorithm 2 correlates the perturbations results in the cumulative objectives ϕ_t frequently sharing minimizers across rounds. By Lemma 1, the lazy sampler

Algorithm 1 returns the particular perturbation that leads to the same minimizer with probability that is precisely the total variation between the two distributions it is given as arguments. Since these are Gaussians with means off by the current round's gradient, it follows their total variation and thus the switch probability is bounded by the gradient norm divided by the current round's perturbation standard deviation.

Proof of Lemma 3. Denote the optimization objective at time t by

$$\phi_t(w) := \sum_{i=1}^{t-1} f_i(w) + p_t^\top w + R(w), \tag{4}$$

and note that we have the following relation between consecutive objectives;

$$\phi_{t+1}(w) = \phi_t(w) + f_t(w) + (p_{t+1} - p_t)^\top w.$$

Observe that whenever $p_{t+1} = p_t - \nabla f_t(w_t)$, we have that $w_t = \arg\min_{w \in W} \phi_t(w)$ is also the (unique) minimizer of ϕ_{t+1} over W; Indeed, let $x \in W$, and verify optimality conditions by checking

$$\begin{aligned}\nabla\phi_{t+1}(w_t)^\top(x - w_t) &= \nabla\phi_t(w_t)^\top(x - w_t) + \nabla f_t(w_t)^\top(x - w_t) + (p_{t+1} - p_t)^\top(x - w_t) \\ &= \nabla\phi_t(w_t)^\top(x - w_t) \\ &\geq 0,\end{aligned}$$

where the last inequality follows by optimality conditions of w_t (which is the unique minimizer of ϕ_t due to strong convexity). This implies that $\Pr(w_{t+1} \neq w_t) = \Pr\left(p_{t+1} \neq p_t - \nabla f_t(w_t)\right)$. Given the above, we wish to upper bound the expected number of times $p_{t+1} \neq p_t - \nabla f_t(w_t)$ occurs. To that end, observe that by Lemma 2 we have $p_t \sim \mathcal{N}(0, \sigma_t^2)$, hence $p_t - \nabla f_t(w_t) \sim \mathcal{N}(-\nabla f_t(w_t), \sigma_t^2)$. Therefore, applying Lemma 1 we obtain

$$\Pr\left(p_{t+1} \neq p_t - \nabla f_t(w_t)\right) = \left\|\mathcal{N}(-\nabla f_t(w_t), \sigma_t^2) - \mathcal{N}(0, \sigma_{t+1}^2)\right\|_{TV} \leq \frac{\|\nabla f_t(w_t)\|}{2\sigma_t}.$$

where in the above derivation we substitute for the Gaussian KL Eq. (3) and apply Pinsker's Inequality Eq. (2). Now

$$\mathbb{E}\,\mathcal{S}_T = \sum_{t=1}^{T-1} \Pr(w_{t+1} \neq w_t) \leq \sum_{t=1}^{T-1} \frac{\|\nabla f_t(w_t)\|}{2\sigma_t} \leq \frac{G}{2} \sum_{t=1}^{T} \frac{1}{\sigma_t},$$

as desired. ∎

3.1. The general convex case

We start with the general convex case where the regret analysis is simpler. Here we introduce stability into the algorithm by means of L2 regularization, with $R(w) = \frac{1}{2\eta}\|w - w_0\|^2$ for some $w_0 \in W$. Below, we state and prove our theorem giving the guarantees of Algorithm 2 when tuned for general convex losses.

Theorem 4. *When the loss functions $f_1, \ldots, f_T$ are G-Lipschitz and convex, running Algorithm 2 with $\sigma_t = \sigma$ for all t and $R(w) = \frac{1}{2\eta} \|w - w_0\|^2$, $w_0 \in W$, we obtain*

$$\mathbb{E}\, \mathcal{R}_T \leq 2\eta G^2 T + \frac{D^2}{2\eta} + \sigma\sqrt{d}D \quad \text{and} \quad \mathbb{E}\, \mathcal{S}_T \leq \frac{GT}{2\sigma},$$

where D is the L2 diameter of W. In particular, setting $\sigma = GT/2S$ and $\eta = D/2G\sqrt{T}$ we obtain $\mathbb{E}\, \mathcal{S}_T \leq S$ and $\mathbb{E}\, \mathcal{R}_T = O(\sqrt{T} + \sqrt{d}T/S)$.

Proof. Denote the minimization objective at time t by ϕ_t (defined in Eq. (4)), and set

$$y_{t+1} := \arg\min_{w \in W} \{\phi_t(w) + f_t(w)\}\,. \tag{5}$$

We have that

$$\begin{aligned}\mathbb{E}\, \mathcal{R}_T &= \mathbb{E}\Big[\sum_{t=1}^{T} f_t(w_t) - f_t(w^*)\Big] \\ &= \mathbb{E}\Big[\sum_{t=1}^{T} f_t(w_t) - f_t(y_{t+1})\Big] + \mathbb{E}\Big[\sum_{t=1}^{T} f_t(y_{t+1}) - f_t(w^*)\Big],\end{aligned} \tag{6}$$

where the expectation is taken over randomness of the algorithm, originating from the LazySampler (Algorithm 1) invocations and the initial Gaussian perturbation p_1. To bound the first term in Eq. (6), consider any perturbation p_t, and note that

$$\phi_t(w) + f_t(w) = \sum_{i=1}^{t} f_i(w) + p_t^\top w + \frac{1}{2\eta}\|w - w_0\|^2\,,$$

thus $\phi_t + f_t$ is $1/\eta$-strongly-convex. In addition, $\phi_t + f_t$ is minimized over W by y_{t+1}, and ϕ_t is minimized by w_t over W. Therefore by a standard bound on the stability of minimizers of strongly convex objectives (see Lemma 20 in the full version of the paper (Sherman and Koren, 2021)) we obtain $\|y_{t+1} - w_t\| \leq 2\eta G$, and then

$$\mathbb{E}\left[\sum_{t=1}^{T} f_t(w_t) - f_t(y_{t+1})\right] \leq \mathbb{E}\left[\sum_{t=1}^{T} G\,\|w_t - y_{t+1}\|\right] \leq 2\eta G^2 T.$$

The bound on the second term of Eq. (6) is given in the next lemma.

Lemma 5. *The hypothetical leaders regret is bounded as follows;*

$$\mathbb{E}\Big[\sum_{t=1}^{T} f_t(y_{t+1}) - f_t(w^*)\Big] \leq \frac{D^2}{2\eta} + \sigma\sqrt{d}D.$$

This concludes the proof of our regret bound. To establish the switches bound we invoke Lemma 3, and obtain $\mathbb{E}\, \mathcal{S}_T \leq \frac{G}{2}\sum_{t=1}^{T}\frac{1}{\sigma} = \frac{GT}{2\sigma}$. Finally, plugging $\sigma = GT/2S$ and $\eta = D/2G\sqrt{T}$ into the bounds we have established, the result follows. ■

The proof of Lemma 5 is a straightforward adaptation of the analysis for the linear case laid out in Kalai and Vempala (2005), and makes use of the well known Follow-the-leader Be-the-leader Lemma, stated next for completeness.

Lemma 6 (FTL-BTL, Kalai and Vempala, 2005). *Let h_t be any sequence of losses, and set $w_t^* := \arg\min_{w\in W} \sum_{s=1}^t h_t(w)$. Then*

$$\sum_{t=1}^T h_t(w_t^*) \le \sum_{t=1}^T h_t(w_T^*).$$

Proof of Lemma 5. Fix a perturbation sequence $p_1, \ldots, p_T$, and additionally define $p_0 = 0$. Consider the auxiliary loss sequence $\tilde{f}_0(w) = R(w)$, and for $t \ge 1$, $\tilde{f}_t(w) := f_t(w) + (p_t - p_{t-1})^\mathsf{T} w$. From Eq. (5) it follows that

$$y_{t+1} = \operatorname*{arg\,min}_{w\in W} \left\{ \sum_{i=0}^t \tilde{f}_i(w) \right\},$$

hence the BTL Lemma (Lemma 6) we obtain (for any w^*); $\sum_{t=0}^T \tilde{f}_t(y_{t+1}) \le \sum_{t=0}^T \tilde{f}_t(w^*)$. Substituting for the definition of $\tilde{f}_t$ and rearranging we get

$$\begin{aligned}\sum_{t=1}^T f_t(y_{t+1}) - f_t(w^*) &\le R(w^*) - R(y_1) + \sum_{t=1}^T (p_t - p_{t-1})^\mathsf{T}(w^* - y_{t+1}) \\ &\le \frac{D^2}{2\eta} + \sum_{t=1}^T (p_t - p_{t-1})^\mathsf{T}(w^* - y_{t+1}).\end{aligned}$$

Now, consider any perturbations distribution $\mathcal{Q}$, such that the marginals of the p_t's under $\mathcal{Q}$ are the same as the marginals of the p_t's under our actual lazy algorithm which we denote by $\mathcal{A}$. Recall that y_{t+1} defined by Eq. (5) depends only on randomness introduced by p_t. This implies that the y_t's are distributed the same under both $\mathcal{A}$ and $\mathcal{Q}$ as long as the marginals of the perturbations match. Hence

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathcal{A}}\left[\sum_{t=1}^T f_t(y_{t+1}) - f_t(w^*)\right] &= \sum_{t=1}^T \mathop{\mathbb{E}}_{\mathcal{A}}\left[f_t(y_{t+1}) - f_t(w^*)\right] \\ &= \sum_{t=1}^T \mathop{\mathbb{E}}_{\mathcal{Q}}\left[f_t(y_{t+1}) - f_t(w^*)\right] = \mathop{\mathbb{E}}_{\mathcal{Q}}\left[\sum_{t=1}^T f_t(y_{t+1}) - f_t(w^*)\right].\end{aligned}$$

By Lemma 2, for all t it holds that $p_t \sim \mathcal{N}(0, \sigma^2)$ when generated by our algorithm $\mathcal{A}$. Therefore, choosing $\mathcal{Q}$ by letting $p_1 \sim \mathcal{N}(0, \sigma^2)$, and setting $p_t = p_1$ for all $t \ge 2$, we achieve the same marginals as those induced by $\mathcal{A}$. This implies

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathcal{A}}\left[\sum_{t=1}^T f_t(y_{t+1}) - f_t(w^*)\right] &\le \frac{D^2}{2\eta} + \mathop{\mathbb{E}}_{\mathcal{Q}}\left[\sum_{t=1}^T (p_t - p_{t-1})^\mathsf{T}(w^* - y_{t+1})\right] \\ &\le \frac{D^2}{2\eta} + \mathop{\mathbb{E}}_{\mathcal{Q}}\left[\sum_{t=1}^T \|p_t - p_{t-1}\| \, \|w^* - y_{t+1}\|\right]\end{aligned}$$

$$\leq \frac{D^2}{2\eta} + D \mathop{\mathbb{E}}_{\mathfrak{Q}} \left[\sum_{t=1}^{T} \|p_t - p_{t-1}\| \right]$$
$$= \frac{D^2}{2\eta} + \sigma\sqrt{d}D,$$

as desired. ■

3.2. The strongly convex case

In this section, we state and prove Theorem 7 providing the guarantees of Algorithm 2 for the strongly convex setting. The performance here hinges on increasing the perturbations variance at a certain rate, accounting for the increasing curvature in the per round minimized objective. This, along with a careful analysis of the perturbed leaders regret, is key to achieving the quadratic gain in the guarantee.

Theorem 7. *When the loss functions $f_1, \ldots, f_T$ are G-Lipschitz and λ-strongly-convex, running Algorithm 2 with $\sigma_t = \sqrt{t}\sigma$ for all t and $R(w) \equiv 0$, we obtain*

$$\mathbb{E}\,\mathcal{R}_T \leq \frac{2G^2 + 2d\sigma^2}{\lambda}(1 + \log T) \quad \textit{and} \quad \mathbb{E}\,\mathcal{S}_T \leq \frac{G\sqrt{T}}{\sigma}.$$

In particular, setting $\sigma = G\sqrt{T}/S$ we obtain $\mathbb{E}\,\mathcal{S}_T \leq S$ and $\mathbb{E}\,\mathcal{R}_T = \widetilde{O}(dT/S^2)$.

Proof. Let $y_{t+1} = \arg\min_{w \in W} \{\phi_t(w) + f_t(w)\}$ (where ϕ_t is defined in Eq. (4)), and by following an argument similar to the general convex case we obtain

$$\mathbb{E}\,\mathcal{R}_T \leq \frac{2G^2}{\lambda}(1 + \log T) + \mathbb{E}\Big[\sum_{t=1}^{T} f_t(y_{t+1}) - f_t(w^*) \Big], \tag{7}$$

with the only difference being that $\phi_t + f_t$ is now $t\lambda$-strongly-convex. The bound on the second term in the above equation hinges on the increasing perturbation variance, which leads to the desired improvement when combined with strong convexity. Proceeding, we follow the same argument given in the proof of Lemma 5 to obtain

$$\mathbb{E}\Big[\sum_{t=1}^{T} f_t(y_{t+1}) - f_t(w^*) \Big] \leq \mathop{\mathbb{E}}_{\mathfrak{Q}} \left[\sum_{t=1}^{T} (p_t - p_{t-1})^\mathsf{T} (w^* - y_{t+1}) \right],$$

where this time we define $\mathfrak{Q}$ by $p_1 \sim \mathcal{N}(0, \sigma^2)$, and $p_t = \sqrt{t}p_1$ for all $t \geq 2$. Indeed, by Lemma 2 it holds that under our actual algorithm $p_t \sim \mathcal{N}(0, t\sigma^2)$, and therefore the marginals match those induced by $\mathfrak{Q}$. Next, we exploit the fact that p_t are zero mean in order to get rid of the non-random part of $w^* - y_{t+1}$. To that end, set $x_{t+1} := \arg\min_{w \in W} \left\{\sum_{i=1}^{t} f_i(w)\right\}$, and note x_{t+1} is deterministic. Therefore,

$$\mathbb{E}\left[\sum_{t=1}^{T} f_t(y_{t+1}) - f_t(w^*) \right] \leq \mathop{\mathbb{E}}_{\mathfrak{Q}} \left[\sum_{t=1}^{T} (p_t - p_{t-1})^\mathsf{T} (w^* - y_{t+1}) \right]$$
$$= \sum_{t=1}^{T} \mathop{\mathbb{E}}_{\mathfrak{Q}} [p_t - p_{t-1}]^\mathsf{T} (w^* - x_{t+1}) + \mathop{\mathbb{E}}_{\mathfrak{Q}} \left[\sum_{t=1}^{T} (p_t - p_{t-1})^\mathsf{T} (x_{t+1} - y_{t+1}) \right]$$

$$= \mathop{\mathbb{E}}_{\Omega}\left[\sum_{t=1}^{T}(p_t - p_{t-1})^{\mathsf{T}}(x_{t+1} - y_{t+1})\right].$$

Now, note that x_{t+1} and y_{t+1} minimize the same $t\lambda$-strongly-convex objective up to the additional perturbation vector p_t, therefore by Lemma 20; $\|x_{t+1} - y_{t+1}\| \leq \frac{2\|p_t\|}{t\lambda}$. Combining all of the above we obtain

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T} f_t(y_{t+1}) - f_t(w^*)\right] &\leq \mathop{\mathbb{E}}_{\Omega}\left[\sum_{t=1}^{T}(p_t - p_{t-1})^{\mathsf{T}}(x_{t+1} - y_{t+1})\right] \\
&\leq \mathop{\mathbb{E}}_{\Omega}\left[\sum_{t=1}^{T}\|p_t - p_{t-1}\|\,\|x_{t+1} - y_{t+1}\|\right] \\
&\leq \mathop{\mathbb{E}}_{\Omega}\left[\sum_{t=1}^{T}\|p_t - p_{t-1}\|\,\frac{2\,\|p_t\|}{t\lambda}\right] \\
&= \frac{2}{\lambda}\mathop{\mathbb{E}}_{\Omega}\left[\|p_1\|^2\right]\sum_{t=1}^{T}(\sqrt{t} - \sqrt{t-1})\frac{1}{\sqrt{t}} \\
&\leq \frac{2d\sigma^2}{\lambda}(1 + \log T),
\end{aligned}$$

which concludes the proof of the regret bound. For the switches guarantee, by Lemma 3 we have that

$$\mathbb{E}\,\mathcal{S}_T \leq \frac{G}{2\sigma}\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \leq \frac{G\sqrt{T}}{\sigma},$$

and plugging $\sigma = G\sqrt{T}/S$ in the bounds we have established completes the proof up to a trivial computation. ■

4. Lazy Stochastic OCO

In this section, we present a simple algorithm for the special case where the losses are drawn i.i.d. from some distribution of convex losses $\mathcal{F}$. The standard objective to be minimized here is the *pseudo* regret, defined by

$$\overline{\mathcal{R}}_T := \mathbb{E}\left[\sum_{t=1}^{T} f_t(w_t) - f_t(w^*)\right],$$

where $w^* = \arg\min_{w \in W} \mathbb{E}\left[f_1(w)\right]$ and the expectation is over the loss distribution $\mathcal{F}$, from which $f_1, \ldots, f_T$ are sampled i.i.d. Importantly, the minimizer of the expected loss defined above stays fixed for the duration of the game. This is in stark contrast to the situation of the general adversarial setting, and enables significantly better bounds achieved by non uniform blocking as outlined by Algorithm 3. Next, we state and prove Theorem 8 which summarizes the guarantees of Algorithm 3.

Theorem 8. *Assume $\mathcal{F}$ is a distribution of G-Lipschitz convex losses over a domain W of diameter D. Then running Algorithm 3 with step size $\eta_t = D/G\sqrt{t}$ guarantees $\mathcal{S}_T \leq 1 + \log T$ and*

$$\overline{\mathcal{R}}_T \leq 2DG\sqrt{T}(1 + \log T).$$

Algorithm 3 Lazy SGD

input: learning rates $\eta_1, \ldots, \eta_T > 0$
$k \leftarrow 0$; arbitrary $x_1 \in W$
for $t = 1$ to T **do**
  if $t = 2^k$ **then**
    $k \leftarrow k + 1$
    $\tilde{w}_k \leftarrow \Pi\left(\frac{1}{t}\sum_{s=1}^{t} x_t\right)$
  end if
  Play $w_t = \tilde{w}_k$; Observe f_t
  $x_{t+1} \leftarrow x_t - \eta_t \nabla f_t(x_t)$
end for

If we further assume losses sampled from $\mathfrak{F}$ are λ-strongly-convex, then running Algorithm 3 with step size $\eta_t = 1/\lambda t$ guarantees $\mathcal{S}_T \leq 1 + \log T$ and

$$\overline{\mathcal{R}}_T \leq \frac{G^2}{\lambda}(1 + \log T)^2.$$

Proof. We prove for the general convex case; the arguments for the strongly convex case are similar and thus omitted. Observe that the iterates x_t maintained by the algorithm are just decision variables of standard OGD with decreasing step $\eta_t = D/G\sqrt{t}$. By well known arguments (see e.g., Hazan, 2019) these obtain an any time $t \in [T]$ guarantee of

$$\sum_{s=1}^{t} f_s(x_s) - f_s(w) \leq 2DG\sqrt{t},$$

for any $w \in W$. Therefore, we have for any t,

$$\mathbb{E}\left[f_t(\tilde{w}_k) - f_t(w^*)\right] \leq \frac{1}{2^k}\sum_{s=1}^{2^k} \mathbb{E}\left[f_t(x_s) - f_t(w^*)\right] = \frac{1}{2^k}\mathbb{E}\left[\sum_{s=1}^{2^k} f_s(x_s) - f_s(w^*)\right] \leq \frac{2DG\sqrt{2^k}}{2^k}.$$

Now, set $T_k := \min\{2^k, T + 1\}$ and we obtain

$$\mathbb{E}\left[\sum_{t=1}^{T} f_t(w_t) - f_t(w^*)\right] = \sum_{k=0}^{\lfloor \log T \rfloor}\sum_{t=T_k}^{T_{k+1}-1} \mathbb{E}\left[f_t(\tilde{w}_k) - f_t(w^*)\right] \leq 2DG\sqrt{T}(1 + \log T),$$

which concludes the proof. ■

5. Lower bound for general convex losses

In this section, we establish an $\Omega\left(T/S\right)$ lower bound on the expected regret of any S-lazy algorithm in the general convex setting. We denote by Ber_p the Bernoulli distribution over $\{-1, 1\}$ that takes the value 1 w.p. $p \in (0, 1)$, and by Ber_p^j the joint distribution of j independent samples from Ber_p. First consider the standard unconstrained setup in the scalar case $W = [-1, 1]$. For $p, q \in (0, 1)$ sufficiently close, an adversary that plays $f_t(w) = b_t w$, with $b_t \sim \mathrm{Ber}_p$ is indistinguishable from one that draws $b_t \sim \mathrm{Ber}_q$, and an $\Omega(\sqrt{T})$ bound may be established. In the lazy OCO setting, the switching limit gives room for the adversary to repeat losses, effectively decreasing the amount of samples revealed and thereby allowing for a larger deviation between the loss distributions while maintaining their indistinguishability. Our next theorem provides a formal construction of this nature. We give a proof sketch here and defer the complete version to the full version of the paper (Sherman and Koren, 2021).

Theorem 9. *For any $S \in \mathbb{N}$, there exists a stochastic sequence of 1-Lipschitz convex losses over $W = [-1, 1]$, such that the expected regret of any S-lazy algorithm is $\Omega(T/S)$.*

Proof (sketch). Fix $T \in \mathbb{N}$, and let $\mathcal{A}$ be an arbitrary S-lazy algorithm. Let $p \in (0, 1)$, and we define the $\mathcal{F}(p, S)$ stochastic loss sequence as follows. Split the T rounds into $J := C^2S^2$ sections with $\tau := T/J$ consecutive rounds in each, where $C \in \mathbb{R}$ is a universal constant that will be determined later on. At the onset of each section $j \in [J]$, draw a single sample $b_j \sim \mathrm{Ber}_p$ and play $f_t(w) = b_j w$ for all rounds t that belong to section j. Consider the minimizer of the expected cumulative loss of $\mathcal{F}(p, S)$;

$$w^* := \arg\min_{w\in W} \mathop{\mathbb{E}}_{p}\left[\sum_{t=1}^{T} f_t(w)\right].$$

Clearly, w^* can perform no better than the realized minimizer in hindsight, and therefore it suffices to prove a lower bound with respect to it. Now, set $\epsilon := \frac{1}{8CS}$ and consider the two adversaries defined by $p_+ := (1 + \epsilon)/2$ and $p_- := (1 - \epsilon)/2$, along with their corresponding minimizers w^*_+ and w^*_-. Employing standard information theoretic arguments, we have that on any round where the player's decision w_t and the loss f_t are independent, the regret incurred against at least one of these adversaries will be $\Omega(\epsilon)$;

$$\mathop{\mathbb{E}}_{p_+}\left[f_t(w_t) - f_t(w^*_+)\right] \geq \frac{\epsilon}{4} \quad \textit{or} \quad \mathop{\mathbb{E}}_{p_-}\left[f_t(w_t) - f_t(w^*_-)\right] \geq \frac{\epsilon}{4}.$$

Importantly, since the player is allowed far fewer switches (S) than there are sections ($J = C^2S^2$), it follows that for most sections the player's decision is indeed independent of the loss. Proceeding, we consider the decomposition of the player's regret into two terms $\mathcal{R}_T = \mathcal{R}_{pos} + \mathcal{R}_{neg}$. The positive regret term $\mathcal{R}_{pos}$ includes all rounds belonging to sections where the player did not switch. These are precisely the rounds on which her decision is independent of the loss, and therefore by Lemma 16 a regret penalty of $\epsilon/4$ is suffered owed to at least one of the adversaries. Recall $\mathcal{S}_T$ denotes the (random variable) number of switches performed, and observe

$$\mathbb{E}\,\mathcal{R}_{pos} \geq \mathbb{E}\left[\frac{(J - \mathcal{S}_T)\tau\frac{\epsilon}{4}}{2}\right] \geq \frac{(J - S)\tau\frac{\epsilon}{4}}{2} = \frac{T}{64CS} - \frac{T}{64C^3S^2}.$$

The negative regret term $\mathcal{R}_{neg}$ on the other hand, includes all other rounds belonging to sections on which at least one decision switch was performed. The per round loss of $\mathcal{A}$ on these sections is trivially bounded by -1, therefore $\mathbb{E}\,\mathbb{R}_{neg} \geq -\mathbb{E}\,\mathcal{S}_T * \tau \geq -S * \tau$. Concluding, we obtain

$$\mathbb{E}\,\mathcal{R}_T = \mathbb{E}\left[\mathcal{R}_{neg} + \mathcal{R}_{pos}\right] \geq \frac{T}{CS}\left(\frac{1}{64} - \frac{1}{64C^2S} - \frac{1}{C}\right),$$

and the result follows by a choice of $C = 128$. ■

Acknowledgements

This work was partially supported by the Israeli Science Foundation (ISF) grant no. 2549/19, by the Len Blavatnik and the Blavatnik Family foundation, and by the Yandex Initiative in Machine Learning at Tel Aviv University.

References

Naman Agarwal, Brian Bullins, Elad Hazan, Sham Kakade, and Karan Singh. Online control with adversarial disturbances. In *International Conference on Machine Learning*, pages 111–119. PMLR, 2019a.

Naman Agarwal, Elad Hazan, and Karan Singh. Logarithmic regret for online control. In *Advances in Neural Information Processing Systems*, pages 10175–10184, 2019b.

Jason Altschuler and Kunal Talwar. Online learning over a finite action set with limited switching. In *Conference On Learning Theory*, pages 1569–1573. PMLR, 2018.

Oren Anava, Elad Hazan, and Shie Mannor. Online learning for adversaries with memory: price of past mistakes. In *Advances in Neural Information Processing Systems*, pages 784–792, 2015.

Lachlan Andrew, Siddharth Barman, Katrina Ligett, Minghong Lin, Adam Meyerson, Alan Roytman, and Adam Wierman. A tale of two metrics: Simultaneous bounds on competitiveness and regret. In *Conference on Learning Theory*, pages 741–763. PMLR, 2013.

Antonios Antoniadis and Kevin Schewior. A tight lower bound for online convex optimization with switching costs. In *International Workshop on Approximation and Online Algorithms*, pages 164–175. Springer, 2017.

Raman Arora, Ofer Dekel, and Ambuj Tewari. Online bandit learning against an adaptive adversary: from regret to policy regret. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, pages 1747–1754, 2012.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.

Baruch Awerbuch and Robert Kleinberg. Online linear optimization and adaptive routing. *Journal of Computer and System Sciences*, 74(1):97–114, 2008.

Nikhil Bansal, Anupam Gupta, Ravishankar Krishnaswamy, Kirk Pruhs, Kevin Schewior, and Cliff Stein. A 2-competitive algorithm for online convex optimization with switching costs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2015.

Mohammad Bavarian, Badih Ghazi, Elad Haramaty, Pritish Kamath, Ronald L Rivest, and Madhu Sudan. Optimality of correlated sampling strategies. *Theory of Computing*, 16(1):1–18, 2020.

Avrim Blum and Carl Burch. On-line learning and the metrical task system problem. *Machine Learning*, 39(1):35–58, 2000.

Allan Borodin and Ran El-Yaniv. *Online computation and competitive analysis*. cambridge university press, 2005.

Allan Borodin, Nathan Linial, and Michael E Saks. An optimal on-line algorithm for metrical task system. *Journal of the ACM (JACM)*, 39(4):745–763, 1992.

Andrei Z Broder. On the resemblance and containment of documents. In *Proceedings. Compression and Complexity of SEQUENCES 1997 (Cat. No. 97TB100171)*, pages 21–29. IEEE, 1997.

Niv Buchbinder, Shahar Chen, Joshep Seffi Naor, and Ohad Shamir. Unified algorithms for online learning and competitive analysis. In *Conference on Learning Theory*, pages 5–1. JMLR Workshop and Conference Proceedings, 2012.

Nicolò Cesa-Bianchi, Ofer Dekel, and Ohad Shamir. Online learning with switching costs and other adaptive adversaries. In *Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 1*, pages 1160–1168, 2013.

Lin Chen, Qian Yu, Hannah Lawrence, and Amin Karbasi. Minimax regret of switching-constrained online convex optimization: No phase transition. *arXiv preprint arXiv:1910.10873*, 2019.

Niangjun Chen, Gautam Goel, and Adam Wierman. Smoothed online convex optimization in high dimensions via online balanced descent. In *Conference On Learning Theory*, pages 1574–1594. PMLR, 2018.

Alon Cohen, Avinatan Hasidim, Tomer Koren, Nevena Lazic, Yishay Mansour, and Kunal Talwar. Online linear quadratic control. In *International Conference on Machine Learning*, pages 1029–1038. PMLR, 2018.

Alon Cohen, Tomer Koren, and Yishay Mansour. Learning linear-quadratic regulators efficiently with only $\sqrt{T}$ regret. In *International Conference on Machine Learning*, pages 1300–1309. PMLR, 2019.

Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.

Ofer Dekel, Jian Ding, Tomer Koren, and Yuval Peres. Bandits with switching costs: $T^{2/3}$ regret. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 459–467, 2014.

Luc Devroye, Gábor Lugosi, and Gergely Neu. Prediction by random-walk perturbation. In *Conference on Learning Theory*, pages 460–473, 2013.

Nick Feamster, Jennifer Rexford, and Ellen Zegura. The road to sdn: an intellectual history of programmable networks. *ACM SIGCOMM Computer Communication Review*, 44(2):87–98, 2014.

Dylan Foster and Max Simchowitz. Logarithmic regret for adversarial online control. In *International Conference on Machine Learning*, pages 3211–3221. PMLR, 2020.

Sascha Geulen, Berthold Vöcking, and Melanie Winkler. Regret minimization for online buffering problems using the weighted majority algorithm. In *COLT*, pages 132–143. Citeseer, 2010.

Gautam Goel, Yiheng Lin, Haoyuan Sun, and Adam Wierman. Beyond online balanced descent: An optimal algorithm for smoothed online optimization. *Advances in Neural Information Processing Systems*, 32:1875–1885, 2019.

Elad Hazan. Introduction to online convex optimization. *arXiv preprint arXiv:1909.05207*, 2019.

Pierre E Jacob, John O'Leary, and Yves F Atchadé. Unbiased markov chain monte carlo methods with couplings. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(3): 543–600, 2020.

Mohammad Reza Karimi Jaghargh, Andreas Krause, Silvio Lattanzi, and Sergei Vassilvtiskii. Consistent online optimization: Convex and submodular. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2241–2250, 2019.

Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

Jon Kleinberg and Eva Tardos. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and markov random fields. *Journal of the ACM (JACM)*, 49(5): 616–639, 2002.

Tomer Koren, Roi Livni, and Yishay Mansour. Bandits with movement costs and adaptive pricing. In *Conference on Learning Theory*, pages 1242–1268. PMLR, 2017a.

Tomer Koren, Roi Livni, and Yishay Mansour. Multi-armed bandits with metric movement costs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 4122–4131, 2017b.

Neri Merhav, Erik Ordentlich, Gadiel Seroussi, and Marcelo J Weinberger. On sequential strategies for loss functions with memory. *IEEE Transactions on Information Theory*, 48(7):1947–1958, 2002.

Uri Sherman and Tomer Koren. Lazy oco: Online convex optimization on a switching budget. *arXiv preprint arXiv:2102.03803*, 2021.

Guanya Shi, Yiheng Lin, Soon-Jo Chung, Yisong Yue, and Adam Wierman. Online optimization with memory and competitive control. *arXiv e-prints*, pages arXiv–2002, 2020.

Aleksandrs Slivkins. Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272*, 2019.