Parameter-Free Dynamic Regret for Unconstrained Linear Bandits

Alberto Rumi*
Intesa Sanpaolo AI Research

Andrew Jacobsen*
Università degli Studi di Milano,
Politecnico di Milano

Nicolò Cesa-Bianchi
Università degli Studi di Milano

Fabio Vitale
Intesa Sanpaolo Innovation Center

Abstract

We study dynamic regret minimization in unconstrained adversarial linear bandit problems. In this setting, a learner must minimize the cumulative loss relative to an arbitrary sequence of comparators $\boldsymbol{u}_1, \dots, \boldsymbol{u}_T$ in $\mathbb{R}^d$, but receives only *point-evaluation feedback* on each round. We provide a simple approach to combining the guarantees of several bandit algorithms, allowing us to optimally adapt to the number of switches $S_T = \sum_t \mathbb{I}\{\boldsymbol{u}_t \neq \boldsymbol{u}_{t-1}\}$ of an arbitrary comparator sequence. In particular, we provide the *first* algorithm for linear bandits achieving the optimal regret guarantee of order $\mathcal{O}\big(\sqrt{d(1+S_T)T}\big)$ up to poly-logarithmic terms *without prior knowledge of* S_T, thus resolving a long-standing open problem.

1 INTRODUCTION

Online learning is a framework that models sequenctial decision-making against adversarial environments (Shalev-Shwartz et al., 2012; Orabona, 2019). Formally, learning is modelled as a game played over a convex decision set $\mathcal{W} \subseteq \mathbb{R}^d$ for T rounds of interaction. On each round $t \in [T]$, the learner chooses $\boldsymbol{w}_t \in \mathcal{W}$, the adversary chooses chooses a loss function $f_t : \mathcal{W} \to \mathbb{R}$, and the learner incurs loss $f_t(\boldsymbol{w}_t)$. We focus in particular on the *unconstrained linear bandit* setting, wherein $\mathcal{W} = \mathbb{R}^d$, the losses are linear functions $f_t(\boldsymbol{w}) = \langle \boldsymbol{\ell}_t, \boldsymbol{w} \rangle$ for some $\boldsymbol{\ell}_t \in \mathbb{R}^d$, and on each round the learner observes only the *scalar* output of their decision, $\langle \boldsymbol{\ell}_t, \boldsymbol{w}_t \rangle$.

The classical way to evaluate the performance of online algorithms is through the notion of *regret*, which compares the learner's expected cumulative loss to that of the best fixed strategy chosen in hindsight. While this can be a natural benchmark in some settings, it implicitly assumes that the environment is *stationary*, i.e., that the optimal action does not change over time. In many practical scenarios, however, this can be unrealistic: user preferences, market conditions, or system dynamics can evolve unpredictably over time, leading to situations where minimizing regret against a fixed comparator becomes meaningless. Instead, in this paper we focus on the more general notion of *dynamic regret*, in which the learner's performance is evaluated relative to an arbirary comparator *sequence* $\boldsymbol{u}_1, \dots, \boldsymbol{u}_T$ in $\mathbb{R}^d$:

$$\mathbb{E}\left[R_T(\boldsymbol{u}_1, \dots, \boldsymbol{u}_T)\right] = \mathbb{E}\left[\sum_{t \in [T]} \langle \boldsymbol{\ell}_t, \boldsymbol{w}_t - \boldsymbol{u}_t \rangle\right] ,$$

The standard notion of (static) regret is recovered when $\boldsymbol{u}_1 = \cdots = \boldsymbol{u}_T$. In this work, we consider a partially-oblivious model in which the comparator sequence is chosen obliviously at the start of the game, while the loss sequence is free to be chosen adaptively based on the learner's past decisions.

Notice that the "complexity" of the comparator sequence contributes to characterizing the difficulty of a given problem. If the comparator is allowed to change arbitrarily on each round, there is no hope to achieve sublinear dynamic regret since the comparator sequence can ensure $\sum_{t=1}^T \ell_t(\boldsymbol{u}_t) = 0$ even against completely unpredictable losses. On the other hand, we know that in the special case $\boldsymbol{u}_1 = \cdots = \boldsymbol{u}_T$ we can ensure sublinear regret, since this is just the usual linear bandits setting. Thus, we are typically interested

*Equal contribution. Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

in algorithms which make dynamic regret guarantees that gracefully *adapt* to some measure of complexity or *variability* of the comparator sequence. The classic way to quantify the variability of the comparator sequence is via its *path-length*:

$$P_T = \sum_{t=2}^{T} \|\boldsymbol{u}_{t-1} - \boldsymbol{u}_t\| \ ,$$

which offers a measure of cumulative drift of the comparator sequence. A related measure of variability is the *switching metric*, which more coarsely measures the number of times that the comparator changes over time:

$$S_T = \sum_{t=2}^{T} \mathbb{I}[\boldsymbol{u}_t \neq \boldsymbol{u}_{t-1}] \ .$$

While dynamic or switching regret provides a more realistic benchmark for non-stationary environments, it introduces significant challenges. Notably, achieving regret bounds that scale optimally with the comparator variability in bandit settings has previously only been achieved using preliminary knowledge of the comparator variability or a known upper bound on it (Agarwal et al., 2017; Marinov & Zimmert, 2021; Luo et al., 2022), which is prior knowledge which practitioners rarely have in practice. Parameter-free methods which eliminate the need for prior knowledge of the comparator sequence are relatively well understood in the full-information setting (Zhang et al., 2018; Cutkosky, 2020; Jacobsen & Cutkosky, 2022, 2023; Zhang et al., 2023; Jacobsen & Cutkosky, 2024), but no prior works have successfully achieved parameter-free dynamic regret guarantees in the adversarial bandit setting. Under bandit feedback, many of the parameter-free techniques developed for the full information setting are either not directly applicable or result in a suboptimal dependence on the time horizon and path length.

1.1 OUR CONTRIBUTIONS

We obtain optimal dynamic regret guarantees of order $\widetilde{\mathcal{O}}\big(\sqrt{d(1+S_T)T}\big)$ against an adaptive adversary without prior knowledge of the number of changes S_T, resolving a long-standing open problem (Marinov & Zimmert, 2021; Luo et al., 2022; Auer et al., 2002). Key to obtaining this result is a technique for combining the guarantees of *comparator-adaptive* base algorithms inspired by a clever result in the full information setting (Cutkosky, 2019), which we adapt to bandit feedback via a sampling trick. This simple trick enables us to easily combine the outputs of several bandit algorithms to achieve the best of their respective dynamic regret guarantees, effectively enabling us to "tune" hyperparameters on-the-fly. Such hyperparameter tuning arguments have been attempted by several prior works using sophisticated mixture-of-experts style arguments (Agarwal et al., 2017; Marinov & Zimmert, 2021; Luo et al., 2022), but have only achieved the optimal $\sqrt{S_T}$ dependence by leveraging *a priori* knowledge of S_T. We expect that our approach might also find applications in other settings where bandit-over-bandit ensemble strategies have failed in the past.

1.2 RELATED WORKS

Dynamic regret in online learning. The study of dynamic regret was initiated by Herbster & Warmuth (1998, 2001) in the online regression and classification settings. These results were first extended to the more general setting of online convex optimization (OCO) in the seminal work of Zinkevich (2003), where it was shown that online gradient descent achieves a bound of order $\mathcal{O}\big((1+P_T)\sqrt{T}\big)$. Yang et al. (2016) improved the rate to $\mathcal{O}(\sqrt{(1+P_T)T})$ by leveraging prior knowledge of P_T, and this bound was later shown to be minimax optimal by Zhang et al. (2018), who also provided the first algorithm achieving a matching upper bound without prior knowledge of P_T.

Since then, several works have achieved generalizations of the $\mathcal{O}(\sqrt{(1+P_T)T})$ guarantee by incorporating more adaptive, data-dependent quantities. These refinements typically replace the dependence on T dependence with terms like the cumulative sum of the squared gradient norms $\sum_t \|\nabla \ell_t(\boldsymbol{w}_t)\|^2$, or the temporal variation of the loss $\sum_t \sup_{\boldsymbol{x}} |\ell_t(\boldsymbol{x}) - \ell_{t-1}(\boldsymbol{x})|$ (Cutkosky, 2020; Campolongo & Orabona, 2021). Significant effort in recent years has also been dedicated to also removing boundedness assumptions on the domain $\mathbb{R}^d$ to adapt automatically to maximum comparator norm $\max_t \|\boldsymbol{u}_t\|$ (Jacobsen & Cutkosky, 2022, 2023; Zhang et al., 2023). Various improvements in adaptivity can also be obtained under additional assumptions on the losses such as smoothness (Mokhtari et al., 2016; Zhao et al., 2020; Jacobsen & Cutkosky, 2024; Zhao et al., 2024; Jacobsen et al., 2025).

Comparator-adaptive methods. Key to our results is the notion of *comparator adaptive* online learning, where the goal is to design algorithms that adapt to the complexity of the comparator sequence (e.g., its maximum norm or a measure of its variability) without requiring prior knowledge about it. In the static comparator setting, this idea has been extensively studied under full-information feedback, where the optimal guarantee $R_T(\boldsymbol{u}) = O(\|\boldsymbol{u}\|\sqrt{T})$, for any $\boldsymbol{u} \in \mathbb{R}^d$, can be obtained up to logarithmic terms (Mcmahan & Streeter, 2012; McMahan & Orabona, 2014; Orabona & Pál, 2016; Cutkosky & Orabona,

2018). Notably, Cutkosky (2019) showed that algorithms making comparator-adaptive guarantees can be easily combined, obtaining regret proportional to the best among them; this observation will be crucial to our approach in Section 2. In both linear and convex settings with bandit feedback, comparator-adaptive bounds were studied by van der Hoeven et al. (2020) where they consider static regret and propose a black-box reduction approach, taking inspiration from the full-information reduction of Cutkosky & Orabona (2018).

Dynamic regret with bandit feedback. In the bandit setting, the study of the closely-related notion of switching regret was initiated by Auer (2002), where a bound of $\mathcal{O}(\sqrt{(1+S_T)T})$ was obtained using prior knowledge of the number of switches S_T. The optimal regret bound for the non-stationary *stochastic* bandit setting without *a priori* variational knowledge was first obtained in Auer et al. (2018). Interestingly, Marinov & Zimmert (2021) showed the impossibility of obtaining a $\mathcal{O}(\sqrt{(1+S_T)T})$ bound against *adaptive* adversaries in settings with finite policy classes, leaving the question open for unconstrained and oblivious settings.

More broadly, ensemble and meta-algorithm strategies have been proposed to adapt to unknown environment parameters in bandits. However, naive bandit-over-bandit schemes tend to incur significant overhead; for instance, running EXP4 (Auer et al., 2002) as a meta-algorithm over EXP3 base learners would yield $\mathcal{O}\left(T^{2/3}\right)$ regret due to the extra exploration needed (Odalric & Munos, 2011; Cheung et al., 2019). While these methods are flexible, they often pay a price in terms of worse regret bounds or higher variance, particularly in dynamic or tuning-free scenarios. A notable example that avoids some of these issues is the CORRAL framework and its extensions (Agarwal et al., 2017; Luo et al., 2022; Marinov & Zimmert, 2021), which leverages a carefully-chosen sequence of regularizers at the meta-algorithm level which is used to cancel out some of the additional variance at the base-algorithm level. However, in the context of adapting to a dynamic comparator sequence, this approach still fails to obtain $\sqrt{(1+S_T)T}$ regret unless S_T is known *a priori*, since the hyperparameters of the meta-algorithm itself need to be set with knowledge of S_T.

Notation. Without further specifications, the notation $\|\cdot\|$ refers to the Euclidean norm. The dual norm of $\boldsymbol{w}$ is denoted by $\|\boldsymbol{w}\|_* = \sup_{\boldsymbol{\ell}:\|\boldsymbol{\ell}\|\le 1}\langle \boldsymbol{\ell}, \boldsymbol{w}\rangle$. Given a norm $\|\cdot\|$ and $\rho \ge 0$, we denote by $\mathcal{B}_\rho := \{\boldsymbol{x} : \|\boldsymbol{x}\| \le \rho\}$ the closed ball of radius ρ, or the unit ball when the radius is not specified. The notation $\mathcal{O}(\cdot)$ hides constant factors and $\widetilde{\mathcal{O}}(\cdot)$ hides constant and logarithmic factors.

2 COMBINING GUARANTEES VIA UNIFORM SAMPLING

Our approach is inspired by a framework for combining guarantees of online learning algorithms proposed by Cutkosky (2019) for the simpler Online Covex Optimization (OCO) setting. The framework applies to a class of algorithms which are *adaptive* to the norm of an arbitrary comparator to obtain $\mathbb{R}_T(\boldsymbol{u}) \le \widetilde{O}(\|\boldsymbol{u}\|\sqrt{T})$, uniformly over all $\boldsymbol{u} \in \mathbb{R}^d$ *simultaneously* (McMahan & Abernethy, 2013; McMahan & Orabona, 2014; Orabona & Pál, 2016; Cutkosky & Orabona, 2018). One of the key properties that characterizes these comparator-adaptive guarantees is the fact the regret at the origin is constant, i.e., $R_T(\mathbf{0}) = \mathcal{O}(1)$.

The key insight from Cutkosky (2019) is that if we have N such OCO algorithms $\mathcal{A}_1, \dots, \mathcal{A}_N$, we can achieve a regret competitive with the best guarantee among them, $R_T(\boldsymbol{u}) = \mathcal{O}(\min_i R_T^{\mathcal{A}_i}(\boldsymbol{u}))$, by simply adding the iterates together. To see why, let $\boldsymbol{w}_t^{(i)}$ denote the output of $\mathcal{A}_i$ on round t, and observe that if we play $\boldsymbol{w}_t = \sum_{i\in[N]} \boldsymbol{w}_t^{(i)}$, then for any $j \in [N]$ we have

$$
\begin{aligned}
R_T(\boldsymbol{u}) &:= \sum_{t=1}^{T} \langle \boldsymbol{\ell}_t, \boldsymbol{w}_t - \boldsymbol{u}\rangle \\
&= \sum_{t=1}^{T} \left\langle \boldsymbol{\ell}_t, \boldsymbol{w}_t^{(j)} - \boldsymbol{u}\right\rangle + \sum_{i\neq j}\sum_{t=1}^{T} \left\langle \boldsymbol{\ell}_t, \boldsymbol{w}_t^{(i)}\right\rangle \\
&= R_T^{\mathcal{A}_j}(\boldsymbol{u}) + \sum_{i\neq j} R_T^{\mathcal{A}_i}(\mathbf{0}) \\
&= \mathcal{O}\left(R_T^{\mathcal{A}_j}(\boldsymbol{u}) + N\right),
\end{aligned}
$$

where the last step holds because each algorithm guarantees $R_T^{\mathcal{A}_i}(\mathbf{0}) = \mathcal{O}(1)$. Moreover, since this holds for any $j \in [N]$, it must hold for the best among them.

While the iterate-adding approach is elegant in the full-information OCO setting, it is not directly applicable under bandit feedback. The fundamental issue is that when playing the sum $\boldsymbol{w}_t = \sum_{i\in[N]} \boldsymbol{w}_t^{(i)}$, the feedback observed is only the total feedback $\langle \boldsymbol{\ell}_t, \sum_i \boldsymbol{w}_t^{(i)}\rangle$, making it difficult to give proper feedback to any individual learner since the observed feedback includes includes contributions from all other learners' decisions. Instead, we make the following observation: if on each round we sample one of the algorithms uniformly at random and play *only its action* on round t, then *in expectation*, this is equivalent to playing $\sum_i \boldsymbol{w}_t^{(i)}/N$. As a result, we can still apply *nearly* the same iterate-adding argument outlined above, but we can now ac-

Algorithm 1 Uniform Sampling Interface

Input: Base algorithms $\left(\mathcal{A}_n\right)_{n=1}^N$
for $t = 1, \dots, T$ **do**
  Get $\boldsymbol{w}_t^{(i)}$ from $\mathcal{A}_i$ for all $i \in [N]$
  Sample $I_t \sim \text{Uniform}(N)$ and play $\boldsymbol{w}_t = \boldsymbol{w}_t^{(I_t)}$
  Receive feedback $\phi(\boldsymbol{w}_t, \ell_t)$
  Send $\widehat{\boldsymbol{\ell}}_t^{(i)} = \phi(\boldsymbol{w}_t, \ell_t)\mathbb{I}\{I_t = i\}$ to $\mathcal{A}_i$ for $i \in [N]$
end for

curately assign feedback since only one learner's action is played on each round. The main difference from the iterate-adding approach is that we need to rescale the comparator to account for the $1/N$ factor that shows up in $\sum_i \boldsymbol{w}_t^{(i)}/N$.

The procedure described above is summarized in Algorithm 1 for a generic feedback oracle ϕ, which returns a feedback signal $\phi(\boldsymbol{w}_t, \ell_t)$ given the decision $\boldsymbol{w}_t$. This captures first-order feedback when $\phi(\boldsymbol{w}_t, \ell_t) = \boldsymbol{\ell}_t$ and bandit feedback when $\phi(\boldsymbol{w}_t, \ell_t) = \ell_t(\boldsymbol{w}_t)$. To build intuition, we begin with a warm-up result that analyzes the performance of this strategy in the first-order feedback setting.

Proposition 2.1. *Let $\mathcal{A}_1, \dots, \mathcal{A}_N$ be online learning algorithms and let $\boldsymbol{w}_t^{(i)}$ denote the output of $\mathcal{A}_i$ on round t. Suppose that for all i, $\mathcal{A}_i$ guarantees $R_T^{\mathcal{A}_i}(\mathbf{0}) = \sum_{t=1}^T f_t(\boldsymbol{w}_t^{(i)}) - f_t(\mathbf{0}) \le G\epsilon$ for any sequence of G-Lipschitz linear loss functions $f_1, \dots, f_T$. Then for any sequence $\boldsymbol{u}_1, \dots, \boldsymbol{u}_T$ in $\mathbb{R}^d$, Algorithm 1 guarantees*

$$\mathbb{E}\left[R_T(\boldsymbol{u}_1, \dots, \boldsymbol{u}_T)\right] \le (N-1)\epsilon G + \min_{n \in [N]} \mathbb{E}\left[R_T^{\mathcal{A}_n}(N\boldsymbol{u}_1, \dots, N\boldsymbol{u}_T)\right]$$

where $R_T^{\mathcal{A}_n}(N\boldsymbol{u}_1, \dots, N\boldsymbol{u}_T)$ denotes the dynamic regret of $\mathcal{A}_n$ played against losses $\widehat{\boldsymbol{\ell}}_t^{(n)} = \boldsymbol{\ell}_t \mathbb{I}\{I_t = n\}$.

Proof. For all $t \ge 1$, let $\mathcal{F}_{t-1}$ be the σ-algebra generated by the history up to the start of round t. Let $\widehat{\boldsymbol{\ell}}_t^{(n)} = \mathbb{I}\{I_t = n\}\boldsymbol{\ell}_t$, and note that $\boldsymbol{w}_t = \sum_i \mathbb{I}\{I_t = i\}\boldsymbol{w}_t^{(i)}$. Observe that for any $n \in [N]$ we can bound the expected dynamic regret by

$$\begin{aligned}
\mathbb{E}\Big[R_T(\boldsymbol{u}_1, \dots, \boldsymbol{u}_T)\Big] &= \mathbb{E}\left[\sum_{t=1}^T \langle \boldsymbol{\ell}_t, \boldsymbol{w}_t - \boldsymbol{u}_t \rangle\right] \\
&= \mathbb{E}\left[\sum_{t=1}^T \left\langle \boldsymbol{\ell}_t, \sum_{i=1}^N \mathbb{I}\{I_t = i\}\boldsymbol{w}_t^{(i)} \right\rangle\right] \\
&\quad - \mathbb{E}\left[\sum_{t=1}^T \langle \boldsymbol{\ell}_t, \boldsymbol{u}_t \rangle \underbrace{\mathbb{E}\Big[N\,\mathbb{I}\{I_t = n\} \,\Big|\, \mathcal{F}_{t-1}\Big]}_{=1}\right] \\
&= \mathbb{E}\left[\sum_{i=1}^N \sum_{t=1}^T \left\langle \widehat{\boldsymbol{\ell}}_t^{(i)}, \boldsymbol{w}_t^{(i)} \right\rangle\right] - \mathbb{E}\left[\sum_{t=1}^T \left\langle \widehat{\boldsymbol{\ell}}_t^{(n)}, N\boldsymbol{u}_t \right\rangle\right] \\
&= \mathbb{E}\left[\sum_{t=1}^T \left\langle \widehat{\boldsymbol{\ell}}_t^{(n)}, \boldsymbol{w}_t^{(n)} - N\boldsymbol{u}_t \right\rangle + \sum_{i \ne n} \underbrace{\sum_{t=1}^T \left\langle \widehat{\boldsymbol{\ell}}_t^{(i)}, \boldsymbol{w}_t^{(i)} \right\rangle}_{=R_T^{\mathcal{A}_i}(\mathbf{0})}\right] \\
&\le \mathbb{E}\left[R_T^{\mathcal{A}_n}(N\boldsymbol{u}_1, \dots, N\boldsymbol{u}_T)\right] + (N-1)G\epsilon\,.
\end{aligned}$$

The proof is concluded by observing n was arbitrary, so we may choose the best $n \in [N]$. □

This theorem illustrates how uniform sampling acts as a rudimentary coordination mechanism for comparator-adaptive algorithms. Because each learner individually adapts to an arbitrary comparator, our combined strategy, on average, can be compared against the performance of any single algorithm in the collection. As we will see in the following sections, this mechanism enables us to "tune" hyperparameters on-the-fly, allowing us to adapt to unknown problem parameters such as the switching number S_T without any prior knowledge.

It is important to note that, unlike the iterate-adding approach of Cutkosky (2019), this method for combining guarantees can ends up increasing the comparator norm by a factor of N. Therefore, in the context of OCO, the uniform sampling approach is a generally a worse option than the iterate-adding approach. Indeed, this trick is primarily of interest in settings where the iterate-adding adding approach would complicate feedback, such as partial-feedback settings where the iterate adding approach makes individual loss estimation difficult.

3 LINEAR BANDITS

Recall from the previous section that the key property that we need to apply the uniform sampling strategy (Algorithm 1) is that the base algorithms $\mathcal{A}_i$ guarantee a comparator-adaptive property, $R_T(\mathbf{0}) = \mathcal{O}(1)$. To obtain guarantees of this form in the bandit setting, we use a scale/direction decomposition introduced by van der Hoeven et al. (2020), which adapts a well-known reduction from the full-information setting to bandit feedback (Cutkosky & Orabona, 2018). The idea is to decompose the action $\boldsymbol{w}_t$ into a *scale* $v_t \in \mathbb{R}$ and a *direction* $\boldsymbol{\beta}_t \in \mathcal{B} = \left\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| \le 1\right\}$, which will be learned by separate online learning algorithms.

Algorithm 2 Scale / Direction Decomposition (van der Hoeven et al., 2020)

Input: Scale learner $\mathcal{A}_\mathcal{V}$, direction learner $\mathcal{A}_\mathcal{B}$
for $t = 1, \dots, T$ **do**
  Get scale prediction v_t from $\mathcal{A}_\mathcal{V}$ and direction prediction $\boldsymbol{\beta}_t$ from $\mathcal{A}_\mathcal{B}$
  Play $\boldsymbol{w}_t = v_t \boldsymbol{\beta}_t$
  Observe $\langle \boldsymbol{w}_t, \boldsymbol{\ell}_t \rangle$ and compute $g_t = \langle \boldsymbol{w}_t, \boldsymbol{\ell}_t \rangle / v_t$
  Send $\langle \boldsymbol{\ell}_t, \boldsymbol{\beta}_t \rangle = \langle \boldsymbol{\ell}_t, \boldsymbol{w}_t \rangle / v_t$ to $\mathcal{A}_\mathcal{B}$ as the feedback on round t
  Send $v \mapsto \langle \boldsymbol{\ell}_t, \boldsymbol{\beta}_t \rangle v$ to $\mathcal{A}_\mathcal{V}$ as the tth loss
end for

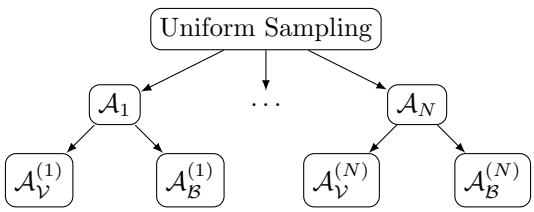

Figure 1: Illustration of how the Uniform Sampling interface interacts with each base algorithm $\mathcal{A}_i$. Each base algorithm internally applies the direction and scale decomposition, using its own hyperparameters.

In particular, observe that for $M = \max_t \|\boldsymbol{u}_t\|$ we have

$$\begin{aligned}
R_T(\boldsymbol{u}_1 \dots \boldsymbol{u}_T) &= \sum_{t=1}^T \langle \boldsymbol{\ell}_t, \boldsymbol{w}_t - \boldsymbol{u}_t \rangle \\
&= \sum_{t=1}^T \langle \boldsymbol{\ell}_t, \boldsymbol{\beta}_t \rangle v_t - \langle \boldsymbol{\ell}_t, \boldsymbol{u}_t \rangle \\
&= \underbrace{\sum_{t=1}^T \langle \boldsymbol{\ell}_t, \boldsymbol{\beta}_t \rangle (v_t - M)}_{=: R_T^{\mathcal{A}_\mathcal{V}}(M)} + M \underbrace{\sum_{t=1}^T \langle \boldsymbol{\ell}_t, \boldsymbol{\beta}_t - \boldsymbol{u}_t / M \rangle}_{=: R_T^{\mathcal{A}_\mathcal{B}}(\boldsymbol{u}_1/M, \dots, \boldsymbol{u}_T/M)}
\end{aligned}$$

where we've added and subtracted $M \langle \boldsymbol{\ell}_t, \boldsymbol{\beta}_t \rangle$ in the last line. Therefore, in order to ensure the required comparator-adaptive property $R_T(\mathbf{0}) = \mathcal{O}(1)$ it suffices to provide a scale learner algorithm $\mathcal{A}_\mathcal{V}$ which guarantees $R_T^{\mathcal{A}_\mathcal{V}}(0) = \mathcal{O}(1)$ against the losses $v \mapsto \langle \boldsymbol{\ell}_t, \boldsymbol{\beta}_t \rangle v$. This is a 1-dimensional *static regret* problem for Online Linear Optimization with full information feedback, so we can apply any of the existing comparator-adaptive algorithms to ensure this property (Orabona & Pál, 2016; Cutkosky & Orabona, 2018; Mhammedi & Koolen, 2020; Jacobsen & Cutkosky, 2022). The following proposition then shows that when applied with the uniform sampling mechanism from the previous section, we are able to obtain a dynamic regret guarantee which matches that of the best algorithm in the set.

Proposition 3.1. *Pick N base algorithms implementing Algorithm 2, where each scale learner $\mathcal{A}_\mathcal{V}^{(n)}$ is over $\mathbb{R}$ and each direction learner $\mathcal{A}_\mathcal{B}^{(n)}$ is over $\mathcal{B} = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| \le 1\}$. Suppose that for any sequence of G-Lipschitz linear losses $f_1, \dots, f_T$ over $\mathbb{R}$ and for any $\epsilon > 0$, the regret of $\mathcal{A}_\mathcal{V}^{(n)}$ satisfies $R_T^{\mathcal{A}_\mathcal{V}^{(n)}}(0) = \sum_{t=1}^T f_t(v_t^{(n)}) - f_t(0) \le G\epsilon$. Then, for any $n \in [N]$ and any sequence $\boldsymbol{u}_1, \dots, \boldsymbol{u}_T$ in $\mathbb{R}^d$, Algorithm 1 guarantees*

$$\begin{aligned}
\mathbb{E}\left[R_T(\boldsymbol{u}_1, \dots, \boldsymbol{u}_T)\right] \le (N-1)G\epsilon + \mathbb{E}\left[R_T^{\mathcal{A}_\mathcal{V}^{(n)}}(MN)\right] \\
+ MN\mathbb{E}\left[R_T^{\mathcal{A}_\mathcal{B}^{(n)}}\Big(\frac{\boldsymbol{u}_1}{M}, \dots, \frac{\boldsymbol{u}_T}{M}\Big)\right],
\end{aligned}$$

where $M = \max_t \|\boldsymbol{u}_t\|$.

Proof. The result is immediate by applying Proposition 2.1 followed by the scale / direction decomposition. For any $n \in [N]$, we have

$$\begin{aligned}
\mathbb{E}\left[R_T(\boldsymbol{u}_1, \dots, \boldsymbol{u}_T)\right] &= \mathbb{E}\left[\sum_{t=1}^T \langle \boldsymbol{\ell}_t, \boldsymbol{w}_t - \boldsymbol{u}_t \rangle\right] \\
&\le \mathbb{E}\left[\sum_{t=1}^T \left\langle \widehat{\boldsymbol{\ell}}_t^{(n)}, \boldsymbol{\beta}_t^{(n)} v_t^{(n)} - N\boldsymbol{u}_t \right\rangle\right] + (N-1)G\epsilon \\
&= \mathbb{E}\Big[\sum_{t=1}^T \left\langle \widehat{\boldsymbol{\ell}}_t^{(n)}, \boldsymbol{\beta}_t^{(n)} \right\rangle v_t^{(n)} - \left\langle \widehat{\boldsymbol{\ell}}_t^{(n)}, \boldsymbol{\beta}_t^{(n)} \right\rangle NM\Big] \\
&\quad + \mathbb{E}\Big[\sum_{t=1}^T \left\langle \widehat{\boldsymbol{\ell}}_t^{(n)}, \boldsymbol{\beta}_t^{(n)} \right\rangle NM - \left\langle \widehat{\boldsymbol{\ell}}_t^{(n)}, N\boldsymbol{u}_t \right\rangle\Big] \\
&\quad + (N-1)G\epsilon \\
&= (N-1)G\epsilon + \mathbb{E}\left[R_T^{\mathcal{A}_\mathcal{V}^{(n)}}(MN)\right] \\
&\quad + MN\mathbb{E}\left[R_T^{\mathcal{A}_\mathcal{B}^{(n)}}\Big(\frac{\boldsymbol{u}_1}{M}, \dots, \frac{\boldsymbol{u}_T}{M}\Big)\right] . \qquad \square
\end{aligned}$$

As observed above, the key insight of van der Hoeven et al. (2020) is that the feedback received by the scale learner is actually full-information feedback; indeed, a scale learner's loss function $v \mapsto \langle \boldsymbol{\ell}_t, \boldsymbol{\beta}_t \rangle v$ can be precisely recovered from $\langle \boldsymbol{\ell}_t, \boldsymbol{w}_t \rangle = \langle \boldsymbol{\ell}_t, \boldsymbol{\beta}_t \rangle v_t$ by dividing it by the scale v_t, so no tricky loss estimation is required for the scale learner. Moreover, the scale learner faces a *static* regret problem, so overall to meet the condition of Proposition 3.1, it will suffice to apply any comparator-adaptive OCO algorithm for static regret as the scale learner, leading to a guarantee of the form $R_T^{\mathcal{A}_\mathcal{V}^{(n)}}(NM) = \widetilde{\mathcal{O}}(NM\sqrt{T})$. In what follows we choose Jacobsen & Cutkosky (2022, Algorithm 4) for concreteness, but really any algorithm guaranteeing $R_T(0) = \mathcal{O}(1)$ would suffice.

All that remains is to provide a bandit algorithm which can guarantee the optimal $\sqrt{dS_T T}$ switching regret on the unit-ball when tuned optimally. For this, we can apply Luo et al. (2022, Algorithm 2), which guarantees for any sequence $\widetilde{\boldsymbol{u}}_1, \ldots, \widetilde{\boldsymbol{u}}_T$ in $\mathcal{B} = \left\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| \leq 1\right\}$ that

$$\mathbb{E}\left[R_T^{\mathcal{A}_\mathcal{B}}(\widetilde{\boldsymbol{u}}_1, \ldots, \widetilde{\boldsymbol{u}}_T)\right] = \widetilde{\mathcal{O}}\left(\sqrt{d(1+S_T)T}\right),$$

so long as the hyperparameters are set optimally in terms of $S_T = \sum_{t=2}^T \mathbb{I}\{\widetilde{\boldsymbol{u}}_t - \widetilde{\boldsymbol{u}}_{t-1}\}$, T, and d. In particular, their approach combines instances of linear bandit algorithms using a carefully-designed multi-scale meta-algorithm, and obtaining the desired $\widetilde{\mathcal{O}}(\sqrt{d(1+S_T)T})$ bound requires that the hyperparameters *of the meta-algorithm* be set using knowledge of S_T, which is of course unknown in practice. However, using our Proposition 3.1 we can effectively *guess* S_T up to a constant factor by running an instance of their algorithm parameterized using $\widehat{S}_n$ for each $\widehat{S}_n$ in a geometrically-spaced grid of candidates

$$\mathcal{S} = \{2^n \wedge 2T, n = 0, 1, \ldots\} \cup \{0\}. \tag{1}$$

Note that even when $\widehat{S}_i$ is misspecified, the scale/direction decomposition above still ensures that the overall algorithm has the comparator-adaptive property $R_T(\mathbf{0}) = \mathcal{O}(1)$ due to the multiplicative dependence between the regret of the unit-ball learner and $M = \max_t \|\boldsymbol{u}_t\|$ in Proposition 3.1. Moreover, since $S_T \in [0, T-1]$ for any sequence of comparators, we know that there is always an $\widehat{S}_n \in \mathcal{S}$ satisfying $\widehat{S}_n \leq S_T \leq 2\widehat{S}_n$ while also maintaining $N = \mathcal{O}(\log(T))$. Therefore, we are able to leverage these facts to arrive at the following parameter-free dynamic regret guarantee.

Theorem 3.2. *Let $\mathcal{S}$ be defined as in Equation (1). For all $i \in \{0, \ldots, |\mathcal{S}|\}$, let $\mathcal{A}_i$ be an instance of Algorithm 2 such that the direction learner $\mathcal{A}_\mathcal{B}^{(i)}$ be characterized by Luo et al. (2022, Theorem 1) with $S = \widehat{S}_i \in \mathcal{S}$, and let $\mathcal{A}_\mathcal{V}^{(i)}$ be an instance of Jacobsen & Cutkosky (2022, Algorithm 4). Then for any sequence $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T$ in $\mathbb{R}^d$, Algorithm 1 guarantees*

$$\mathbb{E}\left[R_T(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T)\right] = \widetilde{\mathcal{O}}\left(MG\sqrt{N(1+S_T)T}\right).$$

where $S_T = \sum_{t=2}^T \mathbb{I}\{\boldsymbol{u}_t \neq \boldsymbol{u}_{t-1}\}$, $M = \max_t \|\boldsymbol{u}_t\|$, and $N = \lceil \log_2(T) + 1 \rceil$.

Proof. By Proposition 3.1, for any $n \in [|\mathcal{S}|]$, we have

$$\mathbb{E}\left[R_T(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T)\right] \leq \mathbb{E}\Big[(N-1)G\epsilon + R_T^{\mathcal{A}_\mathcal{V}^{(i)}}(MN) + MNR_T^{\mathcal{A}_\mathcal{B}^{(i)}}\left(\frac{\boldsymbol{u}_1}{M}, \ldots, \frac{\boldsymbol{u}_T}{M}\right)\Big].$$

where $M = \max_t \|\boldsymbol{u}_t\|$. The regret guarantee (Jacobsen & Cutkosky, 2022, Theorem 1) of any of the $\mathcal{A}_\mathcal{V}^{(i)}$ yields

$$\mathbb{E}\left[R_T^{\mathcal{A}_\mathcal{V}^{(i)}}(NM)\right] = \widetilde{O}\left(MN\mathbb{E}\left[\sqrt{\sum_t \left\langle \widehat{\boldsymbol{\ell}}_t^{(i)}, \boldsymbol{\beta}_t^{(i_t)}\right\rangle^2}\right]\right) = \widetilde{O}\left(MG\sqrt{NT}\right),$$

where we've used obliviousness of the comparator sequence and the facts that $\left\|\boldsymbol{\beta}_t^{(i_t)}\right\| \leq 1$ for all t and $\widehat{\boldsymbol{\ell}}_t^{(i)} = \boldsymbol{\ell}_t \mathbb{I}\{I_t = i\}$, so that $\mathbb{E}\left[\sqrt{\sum_{t=1}^T \|\widehat{\boldsymbol{\ell}}_t^{(i)}\|_*^2}\right] \leq \sqrt{\sum_{t=1}^T \mathbb{E}\left[\|\boldsymbol{\ell}_t\|_*^2 \mathbb{I}\{i_t = i\}\right]} \leq G\sqrt{T/N}$ via Jensen's inequality. Moreover, letting $\widetilde{\boldsymbol{u}}_t = \frac{\boldsymbol{u}_t}{M}$, there is a $\widehat{S}_i \in \mathcal{S}$ such that $\widehat{S}_i \leq S_T \leq 2\widehat{S}_i$, so applying Luo et al. (2022, Algorithm 2) with hyperparameters set according to Luo et al. (2022, Theorem 1) with $S = \widehat{S}_i$ yields

$$\mathbb{E}\left[R_T^{\mathcal{A}_\mathcal{B}^{(i)}}(\widetilde{\boldsymbol{u}}_1, \ldots, \widetilde{\boldsymbol{u}}_T)\right] = \widetilde{\mathcal{O}}\left(G\sqrt{d(1+S_T)T}\right),$$

so, combining with the previous display and the regret decomposition from Proposition 3.1

$$\mathbb{E}\left[R_T(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T)\right] = \widetilde{O}\Big(N\epsilon G + GM\sqrt{NT} + MNG\sqrt{d(1+S_T)T}\Big) = \widetilde{O}\left(GM\sqrt{d(1+S_T)T}\right),$$

where we've observed that $N = \lceil \log(T) + 1 \rceil$ and hidden poly-logarithmic factors. □

We note that despite the simplicity of the argument, this result is in fact the first $\widetilde{\mathcal{O}}(\sqrt{(1+S_T)T})$ dynamic regret bound obtained without prior knowledge of S_T in non-stochastic linear bandit settings. One might worry that the result above might contradict existing impossibility results for $\sqrt{(1+S_T)T}$ dynamic regret against adaptive adversaries. Indeed, Marinov & Zimmert (2021) show that $S_T\sqrt{T}$ dynamic regret can be forced under very general conditions when bandit feedback is involved. Our result is avoiding this lower bound in two ways. First, the construction of Marinov & Zimmert (2021) uses *both* an adaptive comparator and an adaptive loss sequence, while our approach explicitly leverages obliviousness of the comparator sequence. Second, the lower bound construction in that work is for finite policy classes, which does not capture the unconstrained linear bandit problem. Intuitively, the reason why this distinction is important is that in

finite policy class settings one can construct instances where choosing the comparator action is strictly better than any other action, leading to a positive gap between the loss of the learner and comparator. As such, lower bounds can be constructed by quantifying how much exploration is needed to identify the optimal action, since any rounds of exploration will lead to a non-negative addition to the total regret. However, in an unconstrained setting, such gaps generally do not hold uniformly, and it is difficult in general to bound how much negative regret the learner may accumulate by making risky bets. For instance, it is possible for the learner to generate arbitrarily large negative regret on any round in which $\boldsymbol{w}_t$ is in the same half-plane as $-\boldsymbol{\ell}_t$ by taking $\|\boldsymbol{w}_t\|$ to be large.

3.1 ON THE ROLE OF OBLIVIOUSNESS

While our approach in the previous section successfully achieves a parameter-free $\widetilde{\mathcal{O}}(\sqrt{S_T T})$ dynamic regret bound against an adaptive loss sequence, the argument crucially relies on obliviousness *of the comparator sequence*. Most notably, Proposition 3.1 decomposes the expected dynamic regret in terms of the expected dynamic regret of an algorithm applied on the unit-ball, $M\mathbb{E}[R_T^{\mathcal{A}_{\mathcal{B}}}(\boldsymbol{u}_1/M, \dots, \boldsymbol{u}_T/M)]$, but this required moving the expectation passed $M = \max_t \|\boldsymbol{u}_t\|$. For this, M should at least be chosen obliviously, particularly so when we only have an *expected* regret guarantee for the direction learner.

More generally, the essential difficulty is that any time one hopes to leverage a bound on $\mathbb{E}[\|\widehat{\boldsymbol{\ell}}_t\|_*^2]$, a comparator-adaptive guarantee will require obliviousness of the comparator sequence to bypass the $M\sqrt{1+S_T}$ part of the bound. Indeed, suppose we could directly reduce the problem to OLO on the loss estimates $\widehat{\boldsymbol{\ell}}_t$ to obtain

$$\mathbb{E}\left[R_T(\boldsymbol{u}_1, \dots, \boldsymbol{u}_T)\right] \lessapprox \mathbb{E}\left[M\sqrt{(1+S_T)\sum_{t=1}^T \|\widehat{\boldsymbol{\ell}}_t\|_*^2}\right].$$

From here, we might hope to apply Jensen's inequality to arrive at

$$\mathbb{E}\left[R_T(\boldsymbol{u}_1, \dots, \boldsymbol{u}_T)\right] \lessapprox M\sqrt{(1+S_T)\sum_{t=1}^T \mathbb{E}\left[\|\widehat{\boldsymbol{\ell}}_t\|_*^2\right]},$$

but this is only valid if $M\sqrt{1+S_T}$ is deterministic. This issue could potentially be overcome by leveraging a loss estimate satisfying an *almost-sure bound* such as $\|\widehat{\boldsymbol{\ell}}_t\| = \widetilde{\mathcal{O}}(d\,\|\boldsymbol{\ell}_t\|)$, which would obviate the need for an oblivious comparator assumption entirely. We conjecture that this is indeed possible in unconstrained settings, though requires a different construction than our approach in the previous section. We leave this as an important direction for follow-up work.

Alternatively, one might hope to avoid these considerations by applying Cauchy-Schwarz inequality to get $\sqrt{\mathbb{E}\left[M^2(1+S_T)\right]\mathbb{E}[\sum_{t=1}^T \|\widehat{\boldsymbol{\ell}}_t\|_*^2]}$. However, this bound fails to be meaningful in general because it decouples the comparator from the loss sequence. Indeed, even with perfect loss estimates $\widehat{\boldsymbol{\ell}}_t = \boldsymbol{\ell}_t$ it is easy to define events such that the decoupled bound becomes vacuous even for $S_T = 0$ while $\sqrt{\mathbb{E}\left[M^2(1+S_T)\sum_{t=1}^T \|\widehat{\boldsymbol{\ell}}_t\|_*^2\right]}$ remains sublinear.

4 DISCUSSION

In this paper, we propose a remarkably simple approach to tuning the hyperparameters of unconstrained bandit algorithms on-the-fly, enabling adaptivity to problem parameters that cannot be directly observed or estimated, such as the switching number of the comparator sequence. This leads to the first optimal parameter-free dynamic regret bound for unconstrained linear bandits in adversarial settings. We expect that this technique can be applied in other contexts involving hard-to-estimate quantities, though it it crucially relies on the scale/direction decomposition discussed in Section 3. Extending this approach to more general constrained action sets remains nontrivial, as irregular geometries can significantly complicate the decoupling of scale and direction.

A natural question is whether it is possible to recover guarantees that scale with the standard path length P_T in linear bandits, instead of the more coarse switching guarantees obtained in this work. One possible approach might be to directly optimize in the action space—e.g., via mirror descent—and then construct a distribution p_t such that $\mathbb{E}_{p_t}[\boldsymbol{w}] = \boldsymbol{w}_t$, playing $\boldsymbol{w} \sim p_t$. However, the variance control of such sampling methods remain challenging in the unconstrained setting. Moreover, as mentioned in Section 3.1, we suspect that it is possible in the unconstrained linear bandit setting to remove the oblivious comparator assumption as well, resulting in parameter-free dynamic regret guarantees against a fully-adaptive adversary, though we leave this as a direction for future investigation.

Lastly, to obtain the desired $\mathcal{O}(\sqrt{d(1+S_T)T})$ bound, we had to resort to running a collection of $\mathcal{O}(\log(T))$ instances of the multi-scale corralling algorithm of Luo et al. (2022), which itself ends up requiring $\mathcal{O}(T\log T)$ per-round computation. We leave as an open question whether the optimal S_T dependence can still be obtained using the usual $\mathcal{O}(d\log T)$ per-round computation obtained in the full-information setting.

Acknowledgements

NCB and AJ acknowledge the financial support from the EU Horizon CL4-2022-HUMAN-02 research and innovation action under grant agreement 101120237, project ELIAS (European Lighthouse of AI for Sustainability) and from the EU Horizon CL4-2021-HUMAN-01 research and innovation action under grant agreement 101070617, project ELSA (European Lighthouse on Secure and Safe AI).

References

Agarwal, A., Luo, H., Neyshabur, B., and Schapire, R. E. Corralling a band of bandit algorithms. In *Conference on Learning Theory*, pp. 12–38. PMLR, 2017.

Auer, P. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.

Auer, P., Gajane, P., and Ortner, R. Adaptively tracking the best arm with an unknown number of distribution changes. In *European Workshop on Reinforcement Learning*, volume 14, pp. 375, 2018.

Campolongo, N. and Orabona, F. A closer look at temporal variability in dynamic online learning, 2021.

Cheung, W. C., Simchi-Levi, D., and Zhu, R. Learning to optimize under non-stationarity. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1079–1087. PMLR, 2019.

Cutkosky, A. Combining online learning guarantees. In Beygelzimer, A. and Hsu, D. (eds.), *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 895–913, Phoenix, USA, 2019. PMLR.

Cutkosky, A. Parameter-free, dynamic, and strongly-adaptive online learning. In III, H. D. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 2250–2259, Virtual, 2020. PMLR.

Cutkosky, A. and Orabona, F. Black-box reductions for parameter-free online learning in banach spaces. In *Conference On Learning Theory*, pp. 1493–1529. PMLR, 2018.

Herbster, M. and Warmuth, M. K. Tracking the best regressor. In *Proceedings of the eleventh annual conference on Computational learning theory*, pp. 24–31, 1998.

Herbster, M. and Warmuth, M. K. Tracking the best linear predictor. *Journal of Machine Learning Research*, 1(281-309):10–1162, 2001.

Jacobsen, A. and Cutkosky, A. Parameter-free mirror descent. In *Conference on Learning Theory*, pp. 4160–4211. PMLR, 2022.

Jacobsen, A. and Cutkosky, A. Unconstrained online learning with unbounded losses. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 14590–14630. PMLR, 2023.

Jacobsen, A. and Cutkosky, A. Online linear regression in dynamic environments via discounting. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 21083–21120. PMLR, 2024.

Jacobsen, A., Rudi, A., Orabona, F., and Cesa-Bianchi, N. Dynamic regret reduces to kernelized static regret. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Luo, H., Zhang, M., Zhao, P., and Zhou, Z.-H. Corralling a larger band of bandits: A case study on switching regret for linear bandits. In *Conference on Learning Theory*, pp. 3635–3684. PMLR, 2022.

Marinov, T. V. and Zimmert, J. The pareto frontier of model selection for general contextual bandits, 2021.

McMahan, B. and Abernethy, J. Minimax optimal algorithms for unconstrained linear optimization. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

Mcmahan, B. and Streeter, M. No-regret algorithms for unconstrained online convex optimization. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

McMahan, H. B. and Orabona, F. Unconstrained online linear learning in hilbert spaces: Minimax algorithms and normal approximations. In Balcan, M. F., Feldman, V., and Szepesvári, C. (eds.), *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pp. 1020–1039, Barcelona, Spain, 13–15 Jun 2014. PMLR.

Mhammedi, Z. and Koolen, W. M. Lipschitz and comparator-norm adaptivity in online learning. In

Abernethy, J. and Agarwal, S. (eds.), *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 2858–2887. PMLR, 09–12 Jul 2020.

Mokhtari, A., Shahrampour, S., Jadbabaie, A., and Ribeiro, A. Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 7195–7201. IEEE, 2016.

Odalric, M. and Munos, R. Adaptive bandits: Towards the best history-dependent strategy. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 570–578. JMLR Workshop and Conference Proceedings, 2011.

Orabona, F. A modern introduction to online learning. *CoRR*, abs/1912.13213, 2019.

Orabona, F. and Pál, D. Coin betting and parameter-free online learning. *Advances in Neural Information Processing Systems*, 29, 2016.

Shalev-Shwartz, S. et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

van der Hoeven, D., Cutkosky, A., and Luo, H. Comparator-adaptive convex bandits. *CoRR*, abs/2007.08448, 2020.

Yang, T., Zhang, L., Jin, R., and Yi, J. Tracking slowly moving clairvoyant: Optimal dynamic regret of online learning with true and noisy gradient. In *International Conference on Machine Learning*, pp. 449–457. PMLR, 2016.

Zhang, L., Lu, S., and Zhou, Z.-H. Adaptive online learning in dynamic environments. *Advances in neural information processing systems*, 31, 2018.

Zhang, Z., Cutkosky, A., and Paschalidis, Y. Unconstrained dynamic regret via sparse coding. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S. (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 74636–74670. Curran Associates, Inc., 2023.

Zhao, P., Zhang, Y.-J., Zhang, L., and Zhou, Z.-H. Dynamic regret of convex and smooth functions. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 12510–12520. Curran Associates, Inc., 2020.

Zhao, P., Zhang, Y.-J., Zhang, L., and Zhou, Z.-H. Adaptivity and non-stationarity: Problem-dependent dynamic regret for online convex optimization. *Journal of Machine Learning Research*, 25(98):1–52, 2024.

Zinkevich, M. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pp. 928–936, 2003.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable]

   We provide a clear and formal description of the mathematical setting, assumptions, and all algorithms presented. We include the analysis of the algorithms' through regret bounds and the computational complexity is trivially bounded by the number of base algorithms and their complexity.

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

   All assumptions are explicitly stated in the problem setup and within the theoretical claims. Complete proofs for all results are provided either inline or in the appendix

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Not Applicable]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]