# Between Stochastic and Adversarial Online Convex Optimization: Improved Regret Bounds via Smoothness

**Sarah Sachs**
University of Amsterdam
Korteweg-de Vries Institute for Mathematics
s.c.sachs@uva.nl

**Hédi Hadiji**
University of Amsterdam
Korteweg-de Vries Institute for Mathematics
hedi.hadiji@gmail.com

**Tim van Erven**
University of Amsterdam
Korteweg-de Vries Institute for Mathematics
tim@timvanerven.nl

**Cristóbal Guzmán**
Pontificia Universidad Católica de Chile
Institute for Mathematical and Computational Eng.
Facultad de Matemáticas and Escuela de Ingeniería
crguzmanp@mat.uc.cl

## Abstract

Stochastic and adversarial data are two widely studied settings in online learning. But many optimization tasks are neither i.i.d. nor fully adversarial, which makes it of fundamental interest to get a better theoretical understanding of the world between these extremes. In this work, we establish novel regret bounds for online convex optimization in a setting that interpolates between stochastic i.i.d. and fully adversarial losses. By exploiting smoothness of the expected losses, these bounds replace dependence on the maximum gradient length by the variance of the gradients, which was previously known only for linear losses. In addition, they weaken the i.i.d. assumption by allowing, for example, adversarially poisoned rounds, which were previously considered in the expert and bandit setting. Our results extend this to the online convex optimization framework. In the fully i.i.d. case, our bounds match the rates one would expect from results in stochastic acceleration, and in the fully adversarial case, they gracefully deteriorate to match the minimax regret. We further provide lower bounds showing that our regret upper bounds are tight for all intermediate regimes in terms of the stochastic variance and the adversarial variation of the loss gradients.

## 1 Introduction

Two of the main approaches for solving convex optimization problems under uncertain data are stochastic convex optimization (SCO) [21, 30] and online convex optimization (OCO) [37]. These two models are very different in their assumptions and goals, despite the fact that they share many techniques. In SCO it is assumed that the loss functions follow an independent, identically distributed (i.i.d.) process, and the goal is to minimize the *excess risk*, which is the optimization error under the expected loss. By contrast, in OCO the losses can be chosen adversarially and the goal is to minimize

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

the *cumulative regret*, which is the difference between the cumulative incurred losses over rounds against the best fixed strategy in hindsight. Much less is known about what happens in between, in scenarios that interpolate between the i.i.d. and adversarial settings. This intermediate setting has drawn major attention in the recent years in the expert and bandit setting [14] [2] [36], however, as mentioned in [14], little is known for online convex optimization. Our work studies this in a generalization of the OCO setting, in which nature chooses distributions for the data that may vary arbitrarily over time, and we provide regret bounds in terms of two quantities that measure how adversarial these distributions are. The standard OCO setting corresponds to the case where the distributions are point-masses on adversarial data points.

**Main Contribution.** Our main contribution is a new analysis of *optimistic online algorithms* [25, 26] that takes advantage of smoothness of the expected loss. This analysis allows for a gradual interpolation between worst-case adversarial regret bounds and the best-known expected regret bounds in the stochastic case, and also provides quantifiable improvements for intermediate cases.[1] To capture the full range between i.i.d. and fully adversarial settings, we consider a similar adversarial model as in [27], i.e., nature chooses distributions $\mathcal{D}_t$ in iteration $t$, and the learner suffers loss $f_t(x_t, \xi_t)$ with $\xi_t \sim \mathcal{D}_t$. Importantly, we do not assume that the distributions $\mathcal{D}_t$ are all the same, but they may vary adversarially over time.

To properly quantify the interpolation between the i.i.d. and fully adversarial settings in the regret bound, we introduce two parameters for the loss sequence. Namely the *cumulative variance*, $\overline{\sigma}_T$, which captures the stochastic aspect of the learning task, i.e., the variance of the $\mathcal{D}_t$; and the *cumulative adversarial variation*, $\overline{\Sigma}_T$, which captures the adversarial difficulties of the data, i.e., the difference between $\mathbb{E}_{\xi\sim\mathcal{D}_t}[\nabla f(\,\cdot\,,\xi)]$ and $\mathbb{E}_{\xi\sim\mathcal{D}_{t-1}}[\nabla f(\,\cdot\,,\xi)]$. With these two key quantities, our first main result in Theorem 5 shows that the expected regret $\mathbb{E}[R_T(u)]$, that is the difference between the cumulative losses of the learner and a fixed solution in hindsight, is bounded by

$$\mathbb{E}\left[R_T(u)\right] = O\Big(D(\bar{\sigma}_T + \bar{\Sigma}_T)\sqrt{T} + LD^2\Big)\,, \tag{1}$$

where $L$ is the smoothness constant of the expected functions $F^t = \mathbb{E}_{\xi\sim\mathcal{D}_t}\left[f(\,\cdot\,,\xi)\right]$. If, in addition, the functions $F^t$ are $\mu$-strongly convex, then in Theorem 7 we obtain (below, $\kappa := L/\mu$ is the condition number)

$$\mathbb{E}[R_T(u)] = O\left(\frac{1}{\mu}\left(\sigma_{\max}^2 + \Sigma_{\max}^2\right)\log T + LD^2\kappa\log\kappa\right).$$

Both bounds are tight: we prove matching lower bounds in Theorems 6 and 8. In Section 3.1 we show that our results match the known adversarial regret bounds as well as the best results in the i.i.d. case. For the latter, only the linear case was so far obtained directly via a regret analysis (see Sec. 5.2 in [25], and prior work [12, 4]). Using optimistic mirror descent, they obtained the regret guarantee of $R_T(u) \leqslant O(\sqrt{\sum_{t=1}^T \|\nabla f(x_t,\xi_t) - m_t\|^2})$, where $m_t$ denotes an optimistic guess of the gradient that is chosen before round $t$. In the i.i.d. case with the prediction $m_t = \nabla f(x_{t-1},\xi_{t-1})$, this can be shown to imply that the expected regret is upper bounded by $\mathbb{E}[R_T(u)] = O(\sigma\sqrt{T} + \sqrt{\sum_{t=1}^T \mathbb{E}[\|\nabla F^t(x_t) - \nabla F^t(x_{t-1})\|^2]})$, where $\sigma$ denotes the variance of the stochastic gradients. This simplifies to

$$\mathbb{E}[R_T(u)] = O(\sigma\sqrt{T}) \qquad \text{for the i.i.d. case with linear functions } \nabla F^t,$$

which is a special case of (1), because $\sigma = \bar{\sigma}_T$ and $\bar{\Sigma}_T = 0$ for i.i.d. losses, and $\nabla F^t(x_t) = \nabla F^t(x_{t-1})$ and $L = 0$ for linear functions. It is not immediately obvious how to generalize this result to general convex functions with smoothness $L > 0$, however. In this case, we can guess the appropriate regret bound based on known convergence results for *stochastic accelerated gradient descent* (SAGD) [17, 8]: if we knew in advance that the losses would be i.i.d. and we did not care about computational efficiency, then we could run a new instance of SAGD for each round $t$. Summing the known rate for SAGD over $t$ then gives $\mathbb{E}[R_T(u)] \leqslant O(\sigma\sqrt{T} + L)$ (for more details see the batch-to-online conversion in Appendix D). This raises the question if a similar bound can also be obtained directly via a regret analysis, without assuming i.i.d. observations in advance. This

[1] It is well-known that in the fully adversarial case smoothness does not yield asymptotic improvements on regret [11], whereas for SCO improvements can be obtained only under low-noise [8].

question is then answered by our result (1), which indeed reduces to this rate for general convex smooth i.i.d. functions, matching the aforementioned bound up to constants. We achieve this by using smoothness to bound $\mathbb{E}[\|\nabla F^t(x_t) - \nabla F^t(x_{t-1})\|^2] \leqslant L\,\mathbb{E}[\|x_t - x_{t-1}\|^2]$, which can be canceled by a negative quadratic term that we obtain from an improved analysis of the regret. The use of this negative term in the analysis dates back to [20], who used it to achieve an improved $O(1/T)$ rate on the extra-gradient method.

In addition to unifying the analysis of these two extreme cases and obtaining the best-known results via one algorithm (OFTRL (2) for convex functions and OFTRL on a surrogate loss (5) for strongly convex functions), our results give new insight for intermediate cases. Thus, as a second main contribution we shed light on a setting that is neither fully adversarial nor i.i.d. To illustrate this, we highlight some examples here, which received attention in the recent literature.

**Adversarial corruptions:** Consider i.i.d. functions with adversarial corruptions, as considered in the context of the expert and bandit settings in [14], [2]. If the (cumulative) corruption level is bounded by a constant $C$, in [14] an expected regret bound of $\mathbb{E}\left[R_T(u)\right] = O(R_T^s + \sqrt{CR_T^s})$ was obtained, where $R_T^s$ denotes the regret with respect to the uncorrupted data. In [14], the authors raised the question of whether it is possible to obtain regret bounds with a similar square-root dependence on the corruption level $C$ for online convex optimization. Indeed, for this intermediate model, we derive a regret bound

$$\mathbb{E}\left[R_T(u)\right] \leqslant O(R_T^s + D\sqrt{GC}),$$

for the general convex case from our Theorem 5. We elaborate on this in Section 4.2.

**Random order models** The random order model (ROM) dates back to [18] in combinatorial online learning. It has drawn attention in the online convex optimization community as an elegant relaxation of the adversarial model [31, 6]. Complementary to the results in [31], we show that the dependence on $G$ in the regret bound can be reduced to a dependence of $\sigma$, where $\sigma$ denotes the variance of gradients in the uniform distribution over loss functions $f_1, \dots f_T$. That is,

$$\mathbb{E}\left[R_T(u)\right] \leqslant O\left(D\sigma\sqrt{T\log\left(e\frac{\widetilde{\sigma}}{\sigma}\right)}\right),$$

where $\widetilde{\sigma}$ denotes a slightly weaker notion of variance (see Corollary 10). We derive these results from our main theorem under stronger assumptions than in [31], namely, we require individual convexity of the adversarial choices whereas [31] only assume cumulative convexity. However, we also obtain a better rate with $\sigma$ instead of $G$ as the leading factor, so the results are not directly comparable. We also consider a variant of the random order model, which we call the *multiple pass random order model (multi-pass ROM)*. This is inspired by multiple shuffle SGD and can be considered another intermediate example between adversarial and stochastic data. We elaborate on both examples, i.e., the ROM and multi-pass ROM in Section 4.3.

### 1.1 Related work

As mentioned in the previous section, our work is inspired by results in the *gradual variation* and in the *stochastic approximation* literature. The gradual variation literature dates back to [12], with later extensions by [4] and [25, 26]. In addition to some technical relation to the aforementioned work, there is also a natural relation between our parameters $\bar{\sigma}_T$ and $\bar{\Sigma}_T$ and variational parameters in [33], [12] or [4]. However, as we elaborate in Remark 4, there are some fundamental differences between these variational parameters and $\bar{\sigma}_T, \bar{\Sigma}_T$, which prevent us from directly obtaining a smooth interpolation from these results. It is also interesting to note that there is a relation between $\bar{\Sigma}_T$ and the *path length* parameters considered in *dynamic regret* bounds [35, 34]. However, since their analysis targets a fundamentally different notion of regret, namely dynamic regret, the results are incomparable.

With respect to the results, our findings are fundamentally different from the stochastic approximation literature, since we do not rely on the assumption that the data is following a distribution. However, we were inspired by analysis techniques and the convergence thresholds set by this literature. Our approach of obtaining accelerated rates by negative terms arising from smoothness in a regret bound has previously been used in the context of variational inequalities and saddle-point problems. Using

this idea, [20] obtained improved rates $O(1/T)$ for the extra-gradient method. More recently [17] showed that acceleration in stochastic convex optimization can benefit from negative terms arising in optimistic FTRL via an anytime-online-to-batch conversion [5]. Although an important inspiration for our approach, the techniques of [17] do not directly carry over, because they evaluate gradients at the time-average of the algorithm's iterates, making them much more stable than the last iterate, which comes up when controlling the regret. Algorithms used both in SCO and OCO follow a vast literature on stochastic approximation methods, e.g. [28, 21, 24]. For this work, we are particularly interested in the more recent literature on acceleration in SCO [8, 15, 17]. In this research field efficiency is traditionally measured in terms of excess risk. On the one hand, regret upper bounds can be converted into excess risk bounds, through the so-called *online-to-batch conversions* [3]; on the other hand, excess risk guarantees do not directly lead to regret bounds, and even if they do some key features of the rates might be lost. These latter methods, known as *batch-to-online conversions* are discussed in Appendix D.

**Outline** In Section 2, after setting up notation and basic definitions, we introduce the *stochastically extended adversarial* model, a generalization of the standard adversarial model similar to the model used in *smoothed analysis*. Our main results can be found in Section 3. In Section 4 we illustrate our results by highlighting several special cases, such as the random order model and the adversarially corrupted stochastic model. Finally, in Section 5 we set our findings into a broader context and give perspective for future work.

## 2 Setting

We recall the *online convex optimization (OCO)* problem. Here, we consider a sequence of convex functions $f_1, \dots f_T$ defined over a closed and bounded convex set $\mathcal{X} \subseteq \mathbb{R}^d$, which become available to the learner sequentially. In the standard *adversarial model*, the learner chooses $x_t \in \mathcal{X}$ in round $t$, then function $f_t$ is revealed and the learner suffers *loss* $f_t(x_t)$. The success of the learner is measured against all fixed $u \in \mathcal{X}$. Hence, the goal of the learner is to minimize the *regret*, that is, the difference between their cumulative loss $\sum_{t=1}^T f_t(x_t)$ and that of the best fixed choice in hindsight, namely $\min_{u\in\mathcal{X}} \sum_{t=1}^T f_t(u)$.

Throughout the paper we use the notation $[T] = \{1, \dots, T\}$. We follow the convention that $\delta_c$ denotes a Dirac measure at a point $c$, and $\|\cdot\|$ denotes the Euclidean norm.

### 2.1 Stochastically Extended Adversarial Model

We extend the aforementioned adversarial model by letting nature choose a distribution $\mathcal{D}_t$ from a set of distributions. Then the learner suffers loss $f(x_t, \xi_t)$ where $\xi_t \sim \mathcal{D}_t$. Note that if the set of distributions is sufficiently rich, this model contains the standard adversarial model and the stochastic model as special cases (see Examples 1,2).

We introduce some notation to make this more precise. Let $\mathcal{X} \subset \mathbb{R}^d$ be a closed convex set and $\Xi$ a measurable space. Define $f : \mathcal{X} \times \Xi \to \mathbb{R}$ and assume $f(\cdot, \xi)$ is convex over $\xi \in \Xi$. Suppose $\mathfrak{D}$ is a set of probability distributions over $\Xi$. For any $\mathcal{D} \in \mathfrak{D}$, we denote the gradient mean by $\nabla F^{\mathcal{D}}(x) := \mathbb{E}_{\xi\sim\mathcal{D}}[\nabla f(x, \xi)]$ and the function mean by $F^{\mathcal{D}}(x) := \mathbb{E}_{\xi\sim\mathcal{D}}[f(x, \xi)]$. Furthermore, denote by $\sigma^2_{\mathcal{D}}$ an upper bound on the variance of the gradients

$$\sigma^2_{\mathcal{D}} = \max_{x\in\mathcal{X}} \mathbb{E}_{\xi\sim\mathcal{D}} \left[ \left\| \nabla f(x, \xi) - \nabla F^{\mathcal{D}}(x) \right\|^2 \right].$$

We introduce some shorthand notation when distributions are indexed by rounds. Given $t \in [T]$, we write $F^t$ and $\sigma^2_t$ instead of $F^{\mathcal{D}_t}(x)$ and $\sigma^2_{\mathcal{D}_t}$, respectively. Let us now introduce the *stochastically extended adversary protocol*.

**Definition 1** (Stochastically Extended Adversary (SEA))**.** *In each round $t$, the learner chooses $x_t \in \mathcal{X}$, the SEA picks $\mathcal{D}_t \in \mathfrak{D}$. The learner and the SEA then both observe a sample $\xi_t \sim \mathcal{D}_t$, and the learner suffers loss $f(x_t, \xi_t)$.*

We note that the SEA model is closely related to the adversarial model considered in the context of *smoothed analysis* [27, 10, 32]. However, in contrast to this line of work, we do not focus on

distributions with sufficient anti-concentration (c.f., Def. 1.1 in [10]). Indeed, this restriction would exclude, among others, the fully adversarial case as described below. Note also that we assume that the SEA is an adaptive adversary, which has access to the realizations $\xi_t$, so the distribution $\mathcal{D}_{t+1}$ may depend on $\xi_1, \dots, \xi_t$. This assumption is not relevant for the fully adversarial or the i.i.d. settings. In the former, because there is no randomness, and in the latter, because there is no change in distribution. However, it is important for some of the intermediate cases, and in particular for the random order model. The SEA model contains several common settings from the literature as special cases. To illustrate this, we list some examples.

1. **Adversarial Model:** The SEA chooses a Dirac measure $\delta_{c_t} \in \mathfrak{D}$ in each round. Then for any $\xi_t \sim \delta_{c_t}$, the SEA selects $f(\,\cdot\,, \xi_t)$, and the model reduces to an adversary selecting directly the functions $f_t(\,\cdot\,)$.
2. **Stochastic I.I.D. Model:** The SEA chooses a fixed $\mathcal{D} \in \mathfrak{D}$ and selects $\mathcal{D}_t = \mathcal{D}$ at each round $t$.
3. **Adversarially Corrupted I.I.D. Model:** The adversary selects an i.i.d. source $\mathcal{D}$ and perturbs the data with adversarial corruptions. This fits in our framework as follows: given a corruption level $C \geqslant 0$, the SEA chooses distributions $\mathcal{D}_t = \mathcal{D} \otimes \delta_{c_t}$ such that $\sum_{t=1}^{T} \sup_{x \in \mathcal{X}} \| \mathbb{E}_{\xi \sim \mathcal{D}}[\nabla f(x, \xi)] - \mathbb{E}_{\xi' \sim \mathcal{D}_t}[\nabla f(x, \xi')] \| \leqslant C$.
4. **Random Order Model (ROM):** Among a fixed family of losses $\mathcal{F} = (f_i,\, i \in [n])$, the SEA randomly picks functions in $\mathcal{F}$ via sampling without replacement, possibly performing multiple passes over the losses and reshuffling between the passes. Formally, define $\Xi = [n]$, and $f(x, \xi) = f_\xi$ for $\xi \in \Xi$. If $t \in [nk, n(k+1)]$ is in epoch $k \in \mathbb{N}$, then the SEA chooses the distribution $\mathcal{D}_t = \mathrm{Unif}(\Xi \setminus \{\xi_s \,:\, s \in [nk+1, n(t-1)]\})$, which is uniform on the functions that have not yet been sampled in epoch $k$.

To quantify the hardness of the loss sequence, we introduce the *cumulative stochastic variance* and *adversarial variation*; we also define an average of these quantities.

We denote by $\mathbb{E}$ the expectation taken with respect to the joint distribution of $(x_1, \xi_1, \dots, x_T, \xi_T)$. Note that the choice of the adversary $\mathcal{D}_t$ can be random itself, as it depends on the past observations (of both the player's actions and the realizations of the $\xi_t$'s). In this case, $\sigma_t$ is also a random quantity.

**Definition 2** (Cumulative Stochastic Variance and Cumulative Adversarial Variation)**.** *Suppose the SEA chooses distributions $\mathcal{D}_1, \dots, \mathcal{D}_T$. Recall that $\sigma_t^2$ is a shorthand for $\sigma_{\mathcal{D}_t}^2$. The cumulative stochastic variance and the cumulative adversarial variance are defined as*

$$\sigma_{[1:T]}^{(2)} = \mathbb{E}\Bigg[\sum_{t=1}^{T} \sigma_t^2\Bigg] \qquad \textit{and} \qquad \Sigma_{[1:T]}^{(2)} = \mathbb{E}\Bigg[\sum_{t=1}^{T} \sup_{x \in \mathcal{X}} \big\|\nabla F^t(x) - \nabla F^{t-1}(x)\big\|^2\Bigg].$$

*We also let $\bar{\sigma}_T$ and $\bar{\Sigma}_T$ denote the square root of the average stochastic variance or adversarial variation, respectively; that is, $\bar{\sigma}_T^2 = \sigma_{[1:T]}^{(2)}/T$ and $\bar{\Sigma}_T^2 = \Sigma_{[1:T]}^{(2)}/T$.*

Note that in the special case when all $f_t$ are fully adversarial, $\bar{\sigma}_T = 0$. On the contrary, in the stochastic case, i.e., if all for each round $t$, the distribution $\mathcal{D}_t$ is equal to a fixed (but arbitrarily chosen) $\mathcal{D}$, then $\bar{\Sigma}_T = 0$. In this case, $\bar{\sigma}_T$ reduces to the common definition of the gradient variance upper bound in the SCO literature [7, 8]. If however, the SEA chooses one distribution $\mathcal{D}_i$ for the first rounds and then switches to a different distribution $\mathcal{D}_j$, then $\bar{\sigma}_T$ can only be upper bounded by $\max(\sigma_i, \sigma_j)$. This upper bound can be pessimistic, however, for some results it gives a better intuition. For this purpose we also define the *maximal stochastic variance* and *maximal adversarial variation*.

**Definition 3** (Maximal Stochastic Variance and Maximal Adversarial Variation)**.** *Let $\sigma_{\max}^2$ be an upper bound on all variances $\sigma_t^2$. That is,*

$$\sigma_{\max}^2 = \max_{t \in [T]} \mathbb{E}\big[\sigma_t^2\big] \quad \textit{and} \quad \Sigma_{\max}^2 = \max_{t \in [T]} \mathbb{E}\Bigg[\sup_{x \in \mathcal{X}} \big\|\nabla F^t(x) - \nabla F^{t-1}(x)\big\|^2\Bigg].$$

**Remark 4.** *As we mentioned in the introduction, the cumulative stochastic variance and the adversarial variation have some similarities with parameters in gradual variation regret bounds. For linear functions $\langle \mu_t, \cdot \rangle$, the bounds in [12] involve the parameter $\mathrm{Var}_T = \sum_{t=1}^{T} \|\mu_t - \bar{\mu}_T\|^2$ where $\bar{\mu}_T$ is*

*the average of the gradients. For OCO with general convex functions, [4] provide upper bounds on the regret in terms of the $L_p$-deviation $D_p = \sum_{t=1}^{T} \sup_{x\in\mathcal{X}} \|\nabla f_t(x) - \nabla f_{t-1}(x)\|_p^2$. In Lemmas 13 and 14 in Appendix A, we show that in the SEA framework, both of these types of bounds are generally worse than ours, and that the difference can be arbitrarily large.*

*In [26] the regret is bounded in terms of $\sum_{t=1}^{T} \|g_t - M_t\|^2$. As mentioned in the introduction, unless the loss functions are linear or the learner has knowledge of the gradient mean, $\sum_{t=1}^{T} \|g_t - M_t\|^2$ cannot directly be reduced to $\sigma^{(2)}_{[1:T]}$ or $\Sigma^{(2)}_{[1:T]}$.*

### 2.2 Assumptions

In our analysis, we will frequently use several of the following additional assumptions. Some of these were already mentioned in the introduction. We keep them all together here, for the convenience of the reader and for clear reference. For any $\mathcal{D} \in \mathfrak{D}$:

- **(A0)** the adversary has access to independent samples $\xi \sim \mathcal{D}$.
- **(A1)** the function $f(\,\cdot\,,\xi)$ is convex, and gradients are bounded by $G$ a.s. when $\xi \sim \mathcal{D}$.
- **(A2)** the expected function $F^{\mathcal{D}}$ is $L$-smooth, i.e, $\nabla F^{\mathcal{D}}$ is $L$-Lipschitz continuous.
- **(A3)** for any $x \in \mathcal{X}$, the variance $\mathbb{E}_{\xi\sim\mathcal{D}}[\|\nabla f(x,\xi) - \nabla F^{\mathcal{D}}(x)\|^2]$ is finite.
- **(A4)** the expected function $F^{\mathcal{D}}(\,\cdot\,)$ is $\mu$-strongly convex.

We assume that **(A0)** always holds. Assumptions **(A1)** ,**(A2)** and **(A3)** are standard in stochastic optimization, and are similar to common assumptions for online convex optimization. There, it is typically assumed that the adversarial samples $f_t(\,\cdot\,)$ are convex (or even linear) and the gradient norms $\|\nabla f_t(\cdot)\|$ are bounded. Note that we only require gradient Lipschitz continuity and strong convexity to hold for the expected loss.

## 3 Algorithms and Regret Bounds

### 3.1 Convex Smooth Functions

We use *Optimistic Follow-the-Regularised-Leader* (OFTRL) (see, e.g., [16, 25]) to minimize regret. Let $(\eta_t)_{t\in[T]}$ be a non-decreasing and positive sequence of stepsizes, possibly tuned adaptively with the observations. At each step $t$, the learner makes an optimistic prediction $M_t \in \mathbb{R}^d$ and updates its iterates as

$$x_t = \operatorname*{argmin}_{x\in\mathcal{X}} \left\{ \left\langle x, M_t + \sum_{s=1}^{t-1} g_s \right\rangle + \frac{\|x\|^2}{\eta_t} \right\}, \tag{2}$$

where we denoted by $g_t = \nabla f(x_t, \xi_t)$ the observed gradient at time $t$. To state our results, we denote by $\mathbb{E}\,[\,\cdot\,]$ the expectation with respect to the joint distribution of $(x_1, \xi_1, \ldots, x_T, \xi_T)$. Our objective is to bound the average regret:

$$\mathbb{E}\,[R_T(u)] := \mathbb{E}\left[ \sum_{t=1}^{T} \langle g_t, x_t - u\rangle \right].$$

The following theorem, proved in Appendix B.1, is our main result for convex functions.

**Theorem 5.** *Fix a user-specified parameter $\nu > 0$. Under assumptions **(A1)** ,**(A2)** ,**(A3)** , OFTRL, with $M_t = g_{t-1}$ and adaptive step-size $\eta_t = D^2/(\nu + \sum_{s=1}^{t-1} \eta_s \|g_s - M_s\|^2)$, has regret*

$$\mathbb{E}\,[R_T(u)] \leqslant D\big(6\,\bar{\sigma}_T + 3\sqrt{2}\bar{\Sigma}_T\big)\sqrt{T} + \frac{3\sqrt{2}DG}{2} + \nu + \frac{1}{\nu}\left(4D^2G^2 + 9L^2D^4\right). \tag{3}$$

*The algorithm needs to know only $D$. If $G$ and $L$ are also known, one can tune $\nu = LD^2 + D^2G^2$ to get*

$$\mathbb{E}\,[R_T(u)] \leqslant O\big(D\big(\bar{\sigma}_T + \bar{\Sigma}_T\big)\sqrt{T} + DG + LD^2\big).$$

*Moreover, if only convexity of the individual losses holds **(A1)** , then tuning $\nu = 2DG$ ensures the (deterministic) bound $R_T(u) \leqslant 3\sqrt{2}DG\sqrt{T} + 4DG$ .*

Without prior knowledge of the smoothness parameter, the best the player can do is to tune $\nu$ according to a guessed value $L_0$. This affects the constants in the bound by an additive term of order $(L_0 + L^2/L_0)D^2$; it would be interesting to determine if this is an inevitable price to pay for the lack of knowledge of $L$. A similar discussion can be held for $G$. Note that the worst-case regret bound of order $DG\sqrt{T}$ always holds every time OFTRL is used in this article, even without smoothness. To avoid distraction, we will not recall this fact in the applications.

The algorithm and analysis dwell on two ideas: the adaptive tuning of the learning rate à la AdaHedge/AdaFTRL [19, 23] with optimism, together with the fact that we keep a negative Bregman divergence term in the analysis, which is crucial to obtain our bound.

**Lower Bound.** The upper bound in Theorem 5 is tight up to additive constants, as the following result shows.

**Theorem 6.** *For any learning algorithm, and for any pair of positive numbers $(\sigma, \Sigma)$ there exists a function $f : \mathcal{X} \times \Xi \to \mathbb{R}$ and a sequence of distributions satisfying assumptions **(A1)** , **(A2)** ,**(A3)** with $\bar{\sigma}_T \geqslant \sigma$ and $\bar{\Sigma}_T \geqslant \Sigma$ such that*

$$\mathbb{E}\left[R_T(u)\right] \geqslant \Omega\big(D\left(\bar{\sigma}_T + \bar{\Sigma}_T\right)\sqrt{T}\big).$$

The proof, in Appendix B.2 relies on a lower bound from stochastic optimization [1, 21] together with the fact that we can construct a sequence of convex and $L$-smooth loss functions such that $\bar{\Sigma}_T$ is in the order of the gradient norms $G$. Combining these insights with the lower bound $\Omega(DG\sqrt{T})$ [23] gives the desired result.

### 3.2 Strongly Convex and Smooth Functions

Up to this point, we have only considered functions which satisfy the weaker set of assumptions **(A1)** ,**(A2)** ,**(A3)** . In this section, we show what improvements can be achieved if strong convexity also holds, that is, if **(A4)** is satisfied with some known parameter $\mu > 0$. For $g_t = \nabla f(x_t, \xi_t)$, define the surrogate loss function

$$\ell_t(x) = \langle g_t, x - x_t\rangle + \frac{\mu}{2}\left\|x - x_t\right\|^2. \tag{4}$$

We use Optimistic Follow-the-Leader (OFTL) on the surrogate losses. For each step $t$, the learner makes an optimistic prediction of the next gradient $M_t \in \mathbb{R}^d$ and selects

$$x_t = \operatorname*{argmin}_{x\in\mathcal{X}}\left\{\sum_{s=1}^{t-1}\ell_s(x) + \langle M_t, x\rangle\right\}. \tag{5}$$

The next theorem is analogous to Theorem 5 for curved losses, and will be our main tool in establishing results for strongly convex losses; see Appendix B.3 for a proof.

**Theorem 7.** *Under assumptions **(A1)** –**(A4)** , the expected regret of OFTL with $M_t = \nabla f(x_{t-1}, \xi_{t-1})$ on surrogate loss functions $\ell_t$ defined in* (4) *is bounded as*

$$\begin{aligned}\mathbb{E}\left[R_T(u)\right] &\leqslant \frac{1}{\mu}\sum_{t=1}^{T}\frac{1}{t}\left(8\sigma_{\max}^2 + 4\mathbb{E}\left[\sup_{x\in\mathcal{X}}\left\|\nabla F^t(x) - \nabla F^{t-1}(x)\right\|^2\right]\right) + \frac{4D^2L^2}{\mu}\log\left(1 + \frac{16L}{\mu}\right)\\ &\leqslant \frac{1}{\mu}\left(8\sigma_{\max}^2 + 4\Sigma_{\max}^2\right)\log T + \frac{4D^2L^2}{\mu}\log\left(1 + \frac{16L}{\mu}\right).\end{aligned}$$

Note that OFTL requires no tuning besides the strong convexity parameter used in the surrogate losses. In particular, it is adaptive to the smoothness $L$.

**Lower Bound** The bound in Theorem 7 is tight, as the next result, proved in Appendix B.4 shows.

**Theorem 8.** *For any learning algorithm, and for any pair of positive numbers $(\sigma, \Sigma)$ there exists a function $f : \mathcal{X} \times \Xi \to \mathbb{R}$ and a sequence of distributions satisfying assumptions **(A1)** ,**(A2)** ,**(A3)** and **(A4)** with $\sigma_{\max} \geqslant \sigma$ and $\Sigma_{\max} \geqslant \Sigma$ such that*

$$\mathbb{E}\left[R_T(u)\right] \geqslant \Omega\left(\frac{1}{\mu}\left(\sigma_{\max}^2 + \Sigma_{\max}^2\right)\log T\right).$$

# 4 Implications

We derive consequences of our results from Section 3. Further examples can be found in Appendix E.

## 4.1 Interpolating Known Results: Fully Adversarial and i.i.d. Data

A first implication of our analysis is that we recover both the adversarial and i.i.d. rates, *via a single adaptive algorithm*.

**Convex Case** For adversarial data, $\sigma_t = 0$ for all $t$, and $\Sigma^2_{[1:T]} \leqslant G\sqrt{2T}$. Thus, Theorem 5 guarantees a bound $R_T(u) \leqslant O(DG\sqrt{T})$, which is known to be the optimal rate up to the additive constants, cf. [37] (note that the expectation does not act on the regret in this case). Simultaneously, if the data is i.i.d., then Theorem 5 guarantees that

$$\mathbb{E}\left[R_T(u)\right] \leqslant O\big(D\sigma\sqrt{T} + LD^2 + DG\big). \tag{6}$$

From standard online-to-batch conversion, this implies an excess risk for the related SCO problem of order $O(D\sigma/\sqrt{T} + D(LD+G)/T)$, which matches the well-known result by [7] up to lower order terms. On the other hand, using batch-to-online conversion (see Appendix D) with the best known accelerated convergence result in SCO, gives $O(D\sigma\sqrt{T} + LD^2)$ regret. Therefore, up to a constant, our result coincides with the best-known results from SCO. Note that also generalizes the improvement obtained for linear functions in the i.i.d. setting in [25, Section 6.2].

**Strongly Convex Case** The adaptive interpolation between i.i.d. and adversarial rates also holds in the strongly convex case. For adversarial data, the bound of Theorem 7 is of order $(G^2/\mu)\log T$, which is known to be the optimal worst-case rate, cf. [13]. For i.i.d. data, the dependence on $G^2$ improves to $\sigma^2$, yielding a bound of order $O((\sigma^2/\mu)\log T + LD^2\kappa\log\kappa)$. This improvement is akin to improvements obtained by *accelerated stochastic gradient descent* in the context of stochastic optimization [8, 17]. In fact, applying batch-to-online conversions and summing the optimization rates would yield a regret bound similar to ours; c.f. Appendix D.

## 4.2 Adversarially Corrupted Stochastic Data

We consider a natural generalization to online convex optimization of the corruption model considered in the bandit literature [29, 36], also recently studied in [14] for prediction with expert advice. There, the author obtains a regret bound that is the sum of the i.i.d. rate and of a term of order $\sqrt{C}$ where $C$ is the total amount of perturbation. They then raise the open question of whether similar results could be obtained for general convex losses. We provide a positive answer to this question in this section, with the regret bound in Corollary 9.

In this model, the generating process of the losses is decomposed as a combination of losses coming from i.i.d. data, with a small additive adversarial perturbation. This fits in the framework by setting $\xi_t = (\xi_{\mathrm{iid},t}, c_t) \sim \mathcal{D}_t = \mathcal{D} \otimes \delta_{c_t}$ and

$$f(x, \xi_t) = h(x, \xi_{\mathrm{iid},t}) + c_t(x)$$

where $c_t$ is the adversarial part of the losses selected by the adversary, and $\xi_{\mathrm{iid},t} \sim \mathcal{D}$ is a sequence of identically distributed random variables. Note that, similarly to our inspirations [14, 29], and contrary to other corruption models for prediction with expert advice [2], we measure the regret against the perturbed data. Define $F = \mathbb{E}_{\xi\sim\mathcal{D}}[h(\cdot,\xi)]$, so that $F^t(x) = F(x) + c_t(x)$ . The amount of perturbation is measured by a parameter $C > 0$ bounding

$$\sum_{t=1}^{T} \max_{x\in\mathcal{X}} \|\nabla c_t(x)\| \leqslant C\,,$$

which is a natural measure of perturbation on the feedback used by the player (note that adding a constant to the perturbations does not change the regret). In this case, the adversarial perturbation on the loss does not affect the variance and $\sigma^2_{\mathcal{D}_t} = \sigma^2_{\mathcal{D}}$. The perturbation appears in the loss variation as for any $t \geqslant 2$, for any $x \in \mathcal{X}$,

$$\begin{aligned}\|\nabla F^t(x) - \nabla F^{t-1}(x)\|^2 &\leqslant 2G\|\nabla F^t(x) - \nabla F^{t-1}(x)\| \\ &= 2G\|\nabla c_t(x) - \nabla c_{t-1}(x)\| \leqslant 2G\big(\|\nabla c_t(x)\| + \|\nabla c_{t-1}(x)\|\big)\,.\end{aligned}$$

Upon taking the supremum over $x \in \mathcal{X}$ and summing over $t$, we get (with the convention that $c_0 \equiv 0$),

$$\Sigma^{(2)}_{[1:T]} = \sum_{t=1}^{T} \sup_{x\in\mathcal{X}} \|\nabla c_t(x) - \nabla c_{t-1}(x)\|^2 \leqslant 4GC\,.$$

Hence, Theorem 5 combined with the bounds on $\bar{\sigma}_T$ and $\bar{\Sigma}_T$ yields the following regret guarantee.

**Corollary 9.** *In the adversarially corrupted stochastic model, adaptive OFTRL enjoys the bound*

$$\mathbb{E}\left[R_T(u)\right] = O\big(D\sigma\sqrt{T} + D\sqrt{GC}\,\big).$$

This regret bound is the sum of the i.i.d. rate for the unperturbed source with a term sublinear in the amount of perturbations $C$, achieved without prior knowledge of $C$. This provides an answer to a question by [14]. The construction of the lower bound Theorem 6 also provides a partial lower bound for the corruption model, proving some worst-case optimality of our upper bound, see Appendix B.2.1.

An interesting open question that remains would be to extend these results to strongly convex losses.

**Regret against unperturbed losses** Depending on modeling choices, one might consider the perturbations $c_t$ to be a true part of the data that deviate from an i.i.d. model, or one may see them as adversarial noise which should be discarded. In the latter case, regret against the i.i.d. part of the data is most relevant. As noticed by [2, Observation 4] and [14, Remark 1], regret against the unperturbed losses and against the perturbed losses are closely related. Indeed, Corollary 9 implies the bound

$$\mathbb{E}\left[\sum_{t=1}^{T} F(x_t) - F(u)\right] \leqslant \mathbb{E}[R_T(u)] + 2D\mathbb{E}\left[\sum_{t=1}^{T} \|\nabla c_t(w_t)\|\right] \leqslant \mathcal{O}(D\sigma\sqrt{T} + D\sqrt{GC} + 2DC)\,.$$

### 4.3 Random Order Models

We apply our results from Section 3 to the Random Order model. The online ROM was introduced by [6] as a way of restricting the power of the adversary in OCO. Note that the ROM is a popular model in combinatorial online optimization, e.g., in the context of profit maximization [9]. Our results highlight that the rates in the ROM model, which is not i.i.d., are almost the same as the rates of the i.i.d. model obtained via sampling in the same set of losses with replacement.

**Corollary 10.** *In the single-pass ROM with convex and L-smooth losses $(f_k)_{k\in[T]}$, OFTRL (c.f.* (2)*) enjoys the regret bound*

$$\mathbb{E}[R_T(u)] \leqslant O\left(D\sigma_1\sqrt{\log\left(e\frac{\widetilde{\sigma}_1}{\sigma_1}\right)T} + LD^2 + DG\right),$$

*where*

$$\sigma_1^2 = \max_{x\in\mathcal{X}} \frac{1}{T}\sum_{t=1}^{T}\left\|\nabla f_t(x) - \frac{1}{T}\sum_{s=1}^{T}\nabla f_s(x)\right\|^2 \text{ and } \widetilde{\sigma}_1^2 = \frac{1}{T}\sum_{t=1}^{T}\max_{x\in\mathcal{X}}\left\|\nabla f_t(x) - \frac{1}{T}\sum_{s=1}^{T}\nabla f_s(x)\right\|^2.$$

We remark that $\sigma_1 \leqslant \widetilde{\sigma}_1 \leqslant \min\{4G^2, T\sigma_1\}$. Hence, the logarithm of the ratio $\log(\widetilde{\sigma}_1/\sigma_1)$ is always under control. The proof of Corollary 10 consists in controlling the adversarial variation and the cumulative variance thanks to the following lemma, proved in Appendix C.1.

**Lemma 11.** *In the single-pass ROM, we have $\Sigma^{(2)}_{[1:T]} \leqslant 8G^2$ and $\bar{\sigma}^{(2)}_{[1:T]} \leqslant T\sigma_1^2\log(2e^2\widetilde{\sigma}_1^2/\sigma_1^2)$.*

We would like to emphasize that our results are complementary to those of [6, 31]. The focus of these works is to relax the assumption that individual losses are convex, and to only require convexity of the average loss function, leading to very different technical challenges. Inquiring if our results can also be achieved under the weaker assumptions of [31] would be an interesting direction for future work.

We also consider the multi-pass ROM. Let $P \in \mathbb{N}$ denote the number of passes. From Lemma 11 and Corollary 10 we directly obtain

$$\mathbb{E}\left[R_T(u)\right] \leqslant \mathcal{O}\left(D\sigma_1\sqrt{\log\left(e\frac{\widetilde{\sigma}_1}{\sigma_1}\right)T} + DG\sqrt{P} + LD^2\right).$$

Combining Lemma 11 with Theorem 7 also gives the following corollary for strongly convex functions; see Appendix C.2 for a proof.

**Corollary 12.** *Under the same assumption as in Theorem 7, the expected regret of the ROM is bounded by*

$$\mathbb{E}\left[R_T(u)\right] \leqslant O\left(\frac{\sigma_1^2}{\mu}\log T + \frac{G^2}{\mu} + LD^2\kappa\log\kappa\right).$$

*For multi-pass ROM with $P$ passes, we obtain*

$$\mathbb{E}\left[R_T(u)\right] \leqslant O\left(\frac{\sigma_1^2}{\mu}\log T + \frac{G^2}{\mu} + \frac{G^2\log P}{n\mu} + LD^2\kappa\log\kappa\right).$$

## 5 Conclusion and future work

As we showed, the exploitation of smoothness of the expected loss functions reduces the dependence of the regret bound on the maximal gradient norm to a dependence on the cumulative stochastic variance and the adversarial variation. Furthermore, we took a step towards a deeper theoretical understanding of the practically relevant intermediate scenarios. Our approach also opens several interesting new research directions. For instance, in the ROM, as mentioned in Section 4.3, an interesting question is whether a regret bound with dependence on $\sigma$ instead of $G$ can also be achieved with weaker assumptions as in [31].

Another interesting question is whether it is possible to unify the analyses and algorithms for the strongly convex and convex case. So far our analyses of these cases were intrinsically different and the choice of the algorithm requires the knowledge of strong convexity constant $\mu$. Since this knowledge might not be available, it is of practical interest to design an adaptive method that can automatically get the best rate without manually tuning for $\mu$. Furthermore, it would be interesting to investigate if similar or better improvements of the regret bound can be achieved with different optimistic predictions $M_t$. Our matching lower bounds show that there are instances where $M_t = g_{t-1}$ captures the right adversarial behavior. However, there are cases where other measures of regularity would be more relevant. One such example would be an adversary oscillating between two distributions. Interestingly, using $M_t = g_{t-2}$ would then recover optimal bounds. An interesting future research direction would be to investigate what can be achieved when using linear combinations of past gradients as optimistic predictions.

Social impact: this work is theoretical and does not give rise to any direct societal concerns.

## Acknowledgments and Disclosure of Funding

Sachs, Hadiji, and Van Erven were supported by the Netherlands Organization for Scientific Research (NWO) under grant number VI.Vidi.192.095. Guzmán's research is partially supported by the INRIA Associate Teams project, FONDECYT 1210362 grant and ANID Anillo ACT210005 grant, and National Center for Artificial Intelligence CENIA FB210017, Basal ANID. Part of this work was done while CG was at the University of Twente.

## References

[1] Alekh Agarwal, Peter L. Bartlett, Pradeep Ravikumar, and Martin J. Wainwright. Information-theoretic lower bounds on the oracle complexity of stochastic convex optimization. *IEEE Transactions on Information Theory*, 58(5):3235–3249, 2012.

[2] Idan Amir, Idan Attias, Tomer Koren, Yishay Mansour, and Roi Livni. Prediction with corrupted expert advice. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, pages 14315–14325, 2020.

[3] N. Cesa-Bianchi, A. Conconi, and C. Gentile. On the generalization ability of online learning algorithms. In *Advances in Neural Information Processing Systems 14 (NeurIPS)*, volume 14, 2002.

[4] Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pages 6.1–6.20. PMLR, 25–27 Jun 2012.

[5] Ashok Cutkosky. Anytime online-to-batch, optimism and acceleration. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1446–1454. PMLR, 09–15 Jun 2019.

[6] Dan Garber, Gal Korcia, and Kfir Levy. Online convex optimization in the random order model. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3387–3396. PMLR, 13–18 Jul 2020.

[7] S. Ghadimi and G. Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM J. Optim.*, 23:2341–2368, 2013.

[8] Saeed Ghadimi and Guanghui Lan. Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization i: A generic algorithmic framework. *SIAM Journal on Optimization*, 22(4):1469–1492, 2012.

[9] Anupam Gupta, Ruta Mehta, and Marco S. Molinaro. Maximizing profit with convex costs in the random-order model. In *ICALP*, 2018.

[10] Nika Haghtalab, Tim Roughgarden, and Abhishek Shetty. Smoothed analysis with adaptive adversaries. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 942–953. IEEE, 2022.

[11] Elad Hazan. Introduction to online convex optimization. *Found. Trends Optim.*, 2(3?4):157?325, aug 2016.

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

[13] Elad Hazan and Satyen Kale. Beyond the regret minimization barrier: an optimal algorithm for stochastic strongly-convex optimization. In *Proceedings of the 24th Annual Conference on Learning Theory*, volume 19 of *Proceedings of Machine Learning Research*, pages 421–436. PMLR, 09–11 Jun 2011.

[14] Shinji Ito. On optimal robustness to adversarial corruption in online decision problems. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, volume 34, pages 7409–7420, 2021.

[15] Prateek Jain, Sham M Kakade, Rahul Kidambi, Praneeth Netrapalli, and Aaron Sidford. Accelerating stochastic gradient descent for least squares regression. In *Conference On Learning Theory*, pages 545–604. PMLR, 2018.

[16] Pooria Joulani, András György, and Csaba Szepesvári. A modular analysis of adaptive (non-) convex optimization: Optimism, composite objectives, and variational bounds. In *International Conference on Algorithmic Learning Theory (ALT)*, pages 681–720. PMLR, 2017.

[17] Pooria Joulani, Anant Raj, Andras Gyorgy, and Csaba Szepesvári. A simpler approach to accelerated optimization: iterative averaging meets optimism. In *International Conference on Machine Learning*, pages 4984–4993. PMLR, 2020.

[18] Claire Kenyon. Best-fit bin-packing with random order. In *In 7th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 359–364, 1997.

[19] Brendan McMahan. Follow-the-regularized-leader and mirror descent: Equivalence theorems and l1 regularization. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 525–533. PMLR, 11–13 Apr 2011.

[20] Arkadi Nemirovski. Prox-method with rate of convergence o(1/t) for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. on Optimization*, 15(1):229–251, jan 2005.

[21] A.S. Nemirovsky and D.B. Yudin. Problem complexity and method efficiency in optimization. *SIAM Review*, 27(2):264–265, 1985.

[22] Francesco Orabona. A modern introduction to online learning. arXiv preprint: 1912.13213, 2021.

[23] Francesco Orabona and Dávid Pál. Scale-free online learning. *Theor. Comput. Sci.*, 716:50–69, 2018.

[24] Boris T Polyak and Anatoli B Juditsky. Acceleration of stochastic approximation by averaging. *SIAM journal on control and optimization*, 30(4):838–855, 1992.

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

[26] Alexander Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems 26 (NeurIPS)*, pages 3066–3074, 2013.

[27] Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: stochastic, constrained, and smoothed adversaries. In *Advances in Neural Information Processing Systems 24 (NeurIPS)*, pages 1764–1772, 2011.

[28] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

[29] Yevgeny Seldin and Aleksandrs Slivkins. One practical algorithm for both stochastic and adversarial bandits. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1287–1295. PMLR, 22–24 Jun 2014.

[30] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory, Second Edition*. Society for Industrial and Applied Mathematics, USA, 2014.

[31] Uri Sherman, Tomer Koren, and Yishay Mansour. Optimal rates for random order online optimization. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, volume 34, 2021.

[32] Daniel A Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.

[33] Tianbao Yang, Mehrdad Mahdavi, Rong Jin, and Shenghuo Zhu. Regret bounded by gradual variation for online convex optimization. *Machine Learning*, 95(2):183–223, oct 2013.

[34] Peng Zhao, Yu-Jie Zhang, Lijun Zhang, and Zhi-Hua Zhou. Dynamic regret of convex and smooth functions. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, pages 12510–12520, 2020.

[35] Peng Zhao, Yu-Jie Zhang, Lijun Zhang, and Zhi-Hua Zhou. Adaptivity and non-stationarity: Problem-dependent dynamic regret for online convex optimization. arXiv preprint arXiv:2112.14368, 2021.

[36] Julian Zimmert and Yevgeny Seldin. An optimal algorithm for stochastic and adversarial bandits. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 467–475. PMLR, 16–18 Apr 2019.

[37] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, ICML'03, pages 928–935. AAAI Press, 2003.

## Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] (see Conclusion and Future Work, Section 5)
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A] Justification: This is theoretical work.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

## A Proofs of Section 2

Note that $\mathrm{Var}_T = \sum_{t=1}^T \|\nabla f_t(x) - \mu_T^*\|^2$ can be understood as an empirical approximation of $\sigma_{[1:T]}^{(2)}$. The following lemma shows the relation of $\mathrm{Var}_T$ to parameters $\sigma_{[1:T]}^{(2)}$ and $\Sigma_{[1:T]}^{(2)}$.

**Lemma 13.** *Define* $\mathrm{Var}_T = \sum_{t=1}^T \|\nabla f_t(x) - \mu_T^*\|^2$ *with* $\mu_T^* = \frac{1}{T}\sum_{t=1}^T \nabla f_t(x)$. *In expectation with respect to distributions* $\mathcal{D}_1, \dots \mathcal{D}_T$,

$$\mathbb{E}\left[\mathrm{Var}_T\right] \geqslant \frac{1}{5}\left(\sigma_{[1:T]}^{(2)} + \Sigma_{[1:T]}^{(2)}\right).$$

*Furthermore, there exists distributions such that* $\mathbb{E}\left[\mathrm{Var}_T\right]$ *is arbitrarily larger than* $\sigma_{[1:T]}^{(2)} + \Sigma_{[1:T]}^{(2)}$.

*Proof.* Since the distribution mean minimizes the least squares error, $\mathbb{E}\left[\mathrm{Var}_T\right] \geqslant \sigma_{[1:T]}^{(2)}$ always holds. Using the same argument, we have

$$\Sigma_{[1:T]}^{(2)} = \sum_{t=1}^T \left\|\nabla F^t - \nabla F^{t-1}\right\|^2 \leqslant 4\sum_{t=1}^T \left\|\nabla f_t - \nabla F^t\right\|^2 \leqslant 4\sum_{t=1}^T \|\nabla f_t - \mu_T^*\|^2 = 4\,\mathrm{Var}_T\,.$$

Consider $\mathcal{X} = [-1, 1]$ and $f(x, \xi) = \xi x$. Now suppose the SEA chooses truncated normal distribution with mean $-2$ for the first $T/2$ rounds, then truncated normal distribution with mean 2 for the remaining rounds. Assume in both cases the variance is $\sigma^2$ and truncation is in the range $[-G, G]$ for $G > 0$. Hence, for sufficiently large $T$, $\mu_T = 0$ and $\mathrm{Var}_T = \sum_{t=1}^T \|f_t\|^2 \propto TG^2$. However, $\sigma_{[1:T]}^{(2)}$ is equal to $T\sigma^2$ which can be considerably smaller than $TG^2$. The price for the distribution switch is captured with a small constant overhead by $\Sigma_{[1:T]}^{(2)} \leqslant 2G^2$. Thus, $\sigma_{[1:T]}^{(2)} + \Sigma_{[1:T]}^{(2)} \propto T\sigma^2 + 2G^2$. □

**Lemma 14.** *Define* $D_p = \sum_{t=1}^T \sup_{x\in\mathcal{X}} \|\nabla f_t(x) - \nabla f_{t-1}(x)\|_p^2$ . *In the SEA framework,*

$$\mathbb{E}\left[D_2\right] \geqslant \frac{1}{2}\left(\sigma_{[1:T]}^{(2)} + \Sigma_{[1:T]}^{(2)}\right).$$

*Furthermore, there exist instances such that* $\mathbb{E}\left[D_2\right] \gg \sigma_{[1:T]}^{(2)} + \Sigma_{[1:T]}^{(2)}$.

*Proof.* We shall in fact prove that $\mathbb{E}[D_2] \geqslant \max(\sigma_{[1:T]}^{(2)}, \Sigma_{[1:T]}^{(2)})$, which directly implies the first part of the statement. Fix a time $t \geqslant 2$ and let $\mathcal{G}_{t-1}$ denote the $\sigma$-algebra generated by $(x_1, \xi_1, \dots, \xi_{t-1}, x_t)$. Then for any $x \in \mathcal{X}$, the variable $F^t(x) = \mathbb{E}[f_t(x)|\mathcal{G}_{t-1}]$ is $\mathcal{G}_{t-1}$-measurable, and we have

$$\begin{aligned}\mathbb{E}\left[\sup_{x\in\mathcal{X}} \|\nabla f_t(x) - \nabla f_{t-1}(x)\|^2 \,\middle|\, \mathcal{G}_{t-1}\right] &\geqslant \sup_{x\in\mathcal{X}} \mathbb{E}\Big[\|\nabla f_t(x) - \nabla f_{t-1}(x)\|^2 \mid \mathcal{G}_{t-1}\Big] \\ &\geqslant \sup_{x\in\mathcal{X}} \mathbb{E}\Big[\|\nabla f_t(x) - \nabla F^t(x)\|^2 \mid \mathcal{G}_{t-1}\Big] = \sigma_t^2\end{aligned}$$

since $f_{t-1} = f(\,\cdot\,, \xi_{t-1})$ is $\mathcal{G}_{t-1}$-measurable, and $\nabla F^t(x) = \mathbb{E}\big[\nabla f^t(x) \,|\, \mathcal{G}_{t-1}\big]$. Therefore, by conditioning on $\mathcal{G}_{t-1}$ at time step $t$, and applying the tower rule, we obtain

$$\mathbb{E}\left[\sum_{t=1}^T \sup_{x\in\mathcal{X}} \|\nabla f_t(x) - \nabla f_{t-1}(x)\|^2\right] \geqslant \mathbb{E}\left[\sum_{t=1}^T \sigma_t^2\right] = \sigma_{[1:T]}^{(2)}\,.$$

The lower bound by $\Sigma_{[1:T]}^{(2)}$ holds by a direct application of Jensen's inequality, and by swapping suprema and expectations.

For the second part of the lemma, consider the $d$-dimensional euclidean ball $\mathcal{X} = B_d(1) \subset \mathbb{R}^d$, and $\Xi = [d]$. Define

$$f(x, i) = x_i^2/2\,.$$

Consider a fully stochastic (i.i.d.) SEA picking $\xi_t \in [d]$ uniformly at random at every time step. Then $\Sigma^{(2)}_{[1:T]} = 0$. We shall now see that $\sigma^{(2)}_{[1:T]} \leqslant T/d$. Indeed, for any $x \in \mathcal{X}$ and $t \in [T]$, then $F^t$ does not depend on $t$ and its value is

$$F(x) = \mathbb{E}_{I\sim\mathcal{D}}[f(x,I)] = \frac{1}{2d}\sum_{i=1}^{d} x_i^2\,,$$

which is a convex and smooth function. We can upper bound the variance, as for any $x \in \mathcal{X}$,

$$\mathbb{E}_{I\sim\mathcal{D}}\big[\|\nabla f(x,I) - \nabla F(x)\|^2\big] \leqslant \mathbb{E}_{I\sim\mathcal{D}}\big[\|\nabla f(x,I)\|^2\big] = \mathbb{E}_{I\sim\mathcal{D}}\big[\|x_I e_I\|^2\big] = \frac{1}{d}\sum_{i=1}^{d} x_i^2 \leqslant \frac{1}{d}\,.$$

Therefore, after the taking the supremum over $x \in \mathcal{X}$, we see that $\sigma_t^2 \leqslant 1/d$. On the other hand, for any $I, J \in [d]$, we have

$$\|\nabla f(x,I) - \nabla f(x,J)\|^2 = \|x_I e_I - x_J e_J\|^2 = (x_I^2 + x_J^2)\mathbf{1}\{I \neq J\}\,.$$

The maximum in the ball of this difference is reached at $x = \sqrt{2}/2(e_I + e_J)$ and

$$\max_{x\in\mathcal{X}} \|\nabla f(x,I) - \nabla f(x,J)\|^2 = \mathbf{1}\{I \neq J\}\,.$$

Therefore, if $I$ and $J$ are independent and uniformly distributed over $[d]$, then

$$\mathbb{E}_{(I,J)\sim\mathcal{D}\otimes\mathcal{D}}\Big[\max_{x\in\mathcal{X}} \|\nabla f(x,I) - \nabla f(x,J)\|^2\Big] = \mathbb{P}_{(I,J)\sim\mathcal{D}\otimes\mathcal{D}}[I \neq J] = \Big(1 - \frac{1}{d}\Big)^2 \geqslant 1/4\,.$$

Summarizing the above inequalities, we have built an example in which

$$\mathbb{E}[D_2] \geqslant \frac{T}{4} \gg \frac{T}{d} \geqslant \sigma^{(2)}_{[1:T]} + \Sigma^{(2)}_{[1:T]}\,.$$

In particular, the expectation of the variation $D_2$ can be arbitrarily larger than the cumulative variance, and our bounds are then tighter than those obtained via a direct application of known results. □

# B Proofs of Section 3

## B.1 Proof of Theorem 5

To prove Theorem 5, we need the following well-known result from the literature.

**Lemma 15.** *Suppose $f_t(\,\cdot\,)$ are convex for all $t \in [T]$ and $\psi_t(\,\cdot\,) = \frac{2}{\eta_t}\,\|\cdot\|^2$. Further, let $g_t \in \partial f_t(x_t)$ and assume $\eta_t \geqslant \eta_{t+1}$. Then the regret for OFTRL is bounded by*

$$R_T \leqslant \frac{D^2}{\eta_T} + \sum_{t=1}^{T}\Big(\langle g_t - M_t,\, x_t - x_{t+1}\rangle - \frac{1}{\eta_t}\|x_{t+1} - x_t\|^2\Big). \tag{7}$$

*Proof.* Denote $F_t(x) := \psi_t(x) + \sum_{s=1}^{t-1} f_s(x)$. Note that $F_t$ is $\frac{2}{\eta_t}$-strongly convex. Thus, Thm 7.29 in [22] gives

$$\begin{aligned} R_T(u) \leqslant\; & \psi_{T+1}(u) - \psi_1(x_1) \\ & + \sum_{t=1}^{T}\Big(\langle g_t - M_t, x_t - x_{t+1}\rangle - \frac{1}{\eta_t}\,\|x_t - x_{t+1}\|^2 + \psi_t(x_{t+1}) - \psi_{t+1}(x_{t+1})\Big). \end{aligned}$$

Since $\psi_t(x) - \psi_{t+1}(x) \leqslant 0$ and $\psi_{T+1}(u) \leqslant \frac{D^2}{\eta_T}$, this gives

$$\leqslant \frac{D^2}{\eta_T} + \sum_{t=1}^{T}\Big(\langle g_t - M_t,\, x_t - x_{t+1}\rangle - \frac{1}{\eta_t}\|x_{t+1} - x_t\|^2\Big)\,. \qquad \square$$

**Theorem 5.** *Fix a user-specified parameter $\nu > 0$. Under assumptions **(A1)** ,**(A2)** ,**(A3)** , OFTRL, with $M_t = g_{t-1}$ and adaptive step-size $\eta_t = D^2/(\nu + \sum_{s=1}^{t-1} \eta_s \|g_s - M_s\|^2)$, has regret*

$$\mathbb{E}\left[R_T(u)\right] \leqslant D\big(6\,\bar{\sigma}_T + 3\sqrt{2}\bar{\Sigma}_T\big)\sqrt{T} + \frac{3\sqrt{2}DG}{2} + \nu + \frac{1}{\nu}\left(4D^2G^2 + 9L^2D^4\right). \tag{3}$$

*The algorithm needs to know only $D$. If $G$ and $L$ are also known, one can tune $\nu = LD^2 + D^2G^2$ to get*

$$\mathbb{E}\left[R_T(u)\right] \leqslant O\big(D\big(\bar{\sigma}_T + \bar{\Sigma}_T\big)\sqrt{T} + DG + LD^2\big).$$

*Moreover, if only convexity of the individual losses holds **(A1)** , then tuning $\nu = 2DG$ ensures the (deterministic) bound $R_T(u) \leqslant 3\sqrt{2}DG\sqrt{T} + 4DG$ .*

*Proof of Theorem 5.* Write $g_t = \nabla f(x_t, \xi_t)$ and denote by $\mathbb{E}$ the expectation with respect to all the randomness. Using the Optimistic FTRL bound from Lemma 15,

$$\begin{aligned}\sum_{t=1}^{T}\langle g_t, x_t - u\rangle &\leqslant \frac{D^2}{\eta_T} + \sum_{t=1}^{T}\Big(\langle g_t - M_t,\, x_t - x_{t+1}\rangle - \frac{1}{2\eta_t}\|x_t - x_{t+1}\|^2\Big)\\ &\leqslant \frac{D^2}{\eta_T} + \sum_{t=1}^{T}\frac{\eta_t}{2}\|g_t - M_t\|^2 - \sum_{t=1}^{T}\frac{1}{2\eta_t}\|x_t - x_{t+1}\|^2,\end{aligned}$$

where the last inequality is obtained by separating the negative norm term in two parts, and keeping half of it in the regret bound. Let us plug in the value $\eta_t$,

$$\eta_t = D^2\left(\nu + \sum_{s=1}^{t-1}\eta_s\|g_s - M_s\|^2\right)^{-1},$$

and use the fact that $\eta_t \leqslant D^2/C$ to further upper bound the deterministic regret by

$$\nu + \frac{3}{2}\sum_{t=1}^{T}\eta_t\|g_t - M_t\|^2 - \frac{\nu}{2D^2}\sum_{t=1}^{T}\|x_t - x_{t+1}\|^2. \tag{8}$$

To bound the second term above, we first compute

$$\begin{aligned}\left(\sum_{t=1}^{T}\eta_t\|g_t - M_t\|^2\right)^2 &= \sum_{t=1}^{T}\sum_{s=1}^{T}\eta_s\|g_s - M_s\|^2\eta_t\|g_t - M_t\|^2\\ &= 2\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\eta_s\|g_s - M_s\|^2\right)\eta_t\|g_t - M_t\|^2 + \sum_{t=1}^{T}\eta_t^2\|g_t - M_t\|^4.\end{aligned}$$

Since $\eta_t \leqslant D^2/(\sum_{s=1}^{t-1}\eta_s\|g_s - M_s\|^2)$

$$\leqslant 2D^2\sum_{t=1}^{T}\|g_t - M_t\|^2 + \left(\frac{\sum_{t=1}^{T}\eta_t^2\|g_t - M_t\|^4}{\sum_{t=1}^{T}\eta_t\|g_t - M_t\|^2}\right)\sum_{t=1}^{T}\eta_t\|g_t - M_t\|^2.$$

Now we use the fact $X^2 \leqslant 2A + BX$ implies $X \leqslant \sqrt{2A} + B$ for $A, B > 0$. Hence

$$\sum_{t=1}^{T}\eta_t\|g_t - M_t\|^2 \leqslant D\sqrt{2\sum_{t=1}^{T}\|g_t - M_t\|^2 + \frac{\sum_{t=1}^{T}\eta_t^2\|g_t - M_t\|^4}{\sum_{t=1}^{T}\eta_t\|g_t - M_t\|^2}}.$$

Next we use $\eta_t \leqslant D^2/\nu$ and $\|g_t - M_t\|^2 \leqslant 4G^2$ to bound the last term. The sum satisfies $\sum_{t=1}^{T}\eta_t^2\|g_t - M_t\|^4 \leqslant (4G^2D^2/\nu)\sum_{t=1}^{T}\eta_t\|g_t - M_t\|^2$ and we can further bound the term above

$$\leqslant D\sqrt{2\sum_{t=1}^{T}\|g_t - M_t\|^2 + \frac{4D^2G^2}{\nu}}.$$

All in all, we have

$$\sum_{t=1}^{T}\langle g_t, x_t - u\rangle \leqslant \frac{3\sqrt{2}D}{2}\sqrt{\sum_{t=1}^{T}\|g_t - M_t\|^2 + \nu + \frac{4D^2G^2}{\nu}} - \frac{\nu}{2D^2}\sum_{t=1}^{T}\|x_{t+1} - x_t\|^2 . \tag{9}$$

Note that we have not used any assumption on the expected $F^t$'s, and in particular not the smoothness. Therefore, even if the expected losses are not smooth, our analysis already entails that if $\|M_t\| \leqslant G$

$$\sum_{t=1}^{T}\langle g_t, x_t - u\rangle \leqslant \frac{3\sqrt{2}DG}{2}\sqrt{4T} + \nu + \frac{4D^2G^2}{\nu} , \tag{10}$$

proving the final claim of the statement.

Let us now proceed with the proof of the finer results. We use the value of $M_t$, together with the fact that (by convexity of $a \mapsto \|a\|^2$), for any $t \geqslant 2$,

$$\begin{aligned}
\|g_t - g_{t-1}\|^2 &\leqslant 4\left\|g_t - \nabla F^t(x_t)\right\|^2 + 4\left\|\nabla F^t(x_t) - \nabla F^t(x_{t-1})\right\|^2 \\
&\quad + 4\left\|\nabla F^t(x_{t-1}) - \nabla F^{t-1}(x_{t-1})\right\|^2 + 4\left\|\nabla F^{t-1}(x_{t-1}) - g_{t-1}\right\|^2 \\
&\leqslant 4\left\|g_t - \nabla F^t(x_t)\right\|^2 + 4L^2\left\|x_t - x_{t-1}\right\|^2 \\
&\quad + 4\left\|\nabla F^t(x_{t-1}) - \nabla F^{t-1}(x_{t-1})\right\|^2 + 4\left\|\nabla F^{t-1}(x_{t-1}) - g_{t-1}\right\|^2 .
\end{aligned}$$

Therefore, using the inequality $\sqrt{a+b} \leqslant \sqrt{a} + \sqrt{b}$, as well as $\|g_1\| \leqslant G$ and reorganizing the terms,

$$\begin{aligned}
\sqrt{\sum_{t=1}^{T}\|g_t - g_{t-1}\|^2} &\leqslant G + \sqrt{8\sum_{t=2}^{T}\|g_t - \nabla F^t(x_t)\|^2} \\
&\quad + 2L\sqrt{\sum_{t=2}^{T}\|x_t - x_{t-1}\|^2} + 2\sqrt{\sum_{t=2}^{T}\|\nabla F^t(x_{t-1}) - \nabla F^{t-1}(x_{t-1})\|^2} .
\end{aligned}$$

The sum of the variations of $x_t$'s can be canceled thanks to the negative term in (9), as

$$\frac{3\sqrt{2}D}{2}2L\sqrt{\sum_{t=1}^{T-1}\|x_{t+1} - x_t\|^2} - \frac{\nu}{2D^2}\sum_{t=1}^{T}\|x_{t+1} - x_t\|^2 \leqslant \sup_{X \geqslant 0}\left\{3\sqrt{2}LDX - \frac{\nu}{2D^2}X^2\right\} = \frac{9L^2D^4}{\nu} .$$

After replacing these bounds in (9), we have obtained the regret bound

$$\begin{aligned}
\sum_{t=1}^{T}\langle g_t, x_t - u\rangle &\leqslant 6D\sqrt{\sum_{t=1}^{T}\|g_t - \nabla F^t(x_t)\|^2} + 3\sqrt{2}D\sqrt{\sum_{t=1}^{T-1}\|\nabla F^{t+1}(x_t) - \nabla F^t(x_t)\|^2} \\
&\quad + \frac{3\sqrt{2}DG}{2} + \nu + \frac{4D^2G^2}{\nu} + \frac{9L^2D^4}{\nu} .
\end{aligned} \tag{11}$$

We will then take expectations in the inequality above. To bound the right-hand side, let us denote by $\mathcal{G}_t = \sigma(x_1, \xi_1, \ldots, x_{t-1}, \xi_{t-1}, x_t)$, then $\xi_t$ is distributed according to $\mathcal{D}_t$ given $\mathcal{G}_t$, and since $x_t$ is $\mathcal{G}_t$-measurable, therefore

$$\begin{aligned}
\mathbb{E}\big[\|\nabla f(x_t, \xi_t) - \nabla F^t(x_t)\|^2 \mid \mathcal{G}_t\big] &\leqslant \mathbb{E}_{\xi\sim\mathcal{D}_t}\big[\|\nabla f(x_t, \xi) - \nabla F^t(x_t)\|^2\big] \\
&\leqslant \sup_{x\in\mathcal{X}}\mathbb{E}_{\xi\sim\mathcal{D}_t}\big[\|\nabla f(x, \xi) - \nabla F^t(x)\|^2\big] = \sigma_t^2 .
\end{aligned}$$

Therefore, by the tower rule,

$$\mathbb{E}\big[\|\nabla f(x_t, \xi_t) - \nabla F^t(x_t)\|^2\big] = \mathbb{E}\Big[\mathbb{E}\big[\|\nabla f(x_t, \xi_t) - \nabla F^t(x_t)\|^2 \mid \mathcal{G}_t\big]\Big] \leqslant \mathbb{E}\big[\sigma_t^2\big] . \tag{12}$$

The final result follows from taking expectations in (11), applying Jensen's inequality, incorporating (12) and using the definitions of $\bar{\sigma}_T$ and $\bar{\Sigma}_T$. □

### B.2 Proof of Theorem 6

**Theorem 6.** *For any learning algorithm, and for any pair of positive numbers $(\sigma, \Sigma)$ there exists a function $f : \mathcal{X} \times \Xi \to \mathbb{R}$ and a sequence of distributions satisfying assumptions **(A1)** , **(A2)** ,**(A3)** with $\bar{\sigma}_T \geqslant \sigma$ and $\bar{\Sigma}_T \geqslant \Sigma$ such that*

$$\mathbb{E}\left[R_T(u)\right] \geqslant \Omega\big(D\left(\bar{\sigma}_T + \bar{\Sigma}_T\right)\sqrt{T}\big).$$

*Proof of Theorem 6.* Suppose we are given two parameters $\hat{\sigma}_T$ and $\hat{\Sigma}_T$, we show that there exists a sequence of distributions $\mathcal{D}_1, \dots \mathcal{D}_T$ such that the expected regret is at least $\Omega(D(\hat{\sigma}_T + \hat{\Sigma}_T)\sqrt{T})$. Let $1 \leqslant a < b$ be constants such that $a \geqslant \frac{1}{2}b$. Since for any closed convex set, there exist an affine transformation which maps it to the interval $[a, b]$, we assume without loss of generality that $\mathcal{X} = [a, b]$.

Suppose $f : \mathcal{X} \times \Xi \to \mathbb{R}$ and let $z^{(\sigma)}, z^{(\Sigma)} \in \mathbb{R}$. Assume the gradients have the form

$$\nabla f(x, \xi) = z^{(\sigma)} \qquad \text{or} \qquad \nabla f(x, \xi) = z^{(\Sigma)}.$$

Assume SEA chooses each case with probability $1/2$. The idea is to construct two sequences $\{z_t^{(\sigma)}\}_{t\in[T]}$ and $\{z_t^{(\Sigma)}\}_{t\in[T]}$ such that these sequences have at least $\Omega(D\hat{\sigma}_T\sqrt{T})$ and $\Omega(D\hat{\Sigma}_T\sqrt{T})$ expected regret, respectively. Therefore, let $x_t$ denote the learners choice in round $t$ and define linearised regret with respect to $\{z_t^{(\sigma)}\}_{t\in[T]}$ and $\{z_t^{(\Sigma)}\}_{t\in[T]}$.

$$R_T^{\sigma} = \min_{u\in\mathcal{X}} \sum_{t\in[T]} \left\langle z_t^{(\sigma)}, x_t - u \right\rangle \quad \text{and} \quad R_T^{\Sigma} = \min_{u\in\mathcal{X}} \sum_{t\in[T]} \left\langle z_t^{(\Sigma)}, x_t - u \right\rangle.$$

**Case $R_T^{\Sigma}$:** Let $G = \hat{\Sigma}_T$. Define $z : \mathcal{X} \to \mathbb{R}$, $z(x) = \frac{1}{4b}Gx^2$. Then $z$ is $G$-Lipschitz, smooth and $z'(x) \in [\frac{1}{2}G, G]$ for any $x \in \mathcal{X}$. Let $\{\varepsilon_t\}_{t\in[T]}$ be an i.i.d. sequence of Rademacher random variables, that is, $\mathbb{P}[\varepsilon_t = 1] = \mathbb{P}[\varepsilon_t = -1] = 1/2$. The sequence $\{z_t^{(\Sigma)}\}_{t\in[T]}$ is defined as

$$z_t^{(\Sigma)} = \begin{cases} 0 & \text{if } t \text{ even} \\ \varepsilon_t\, z'(x_t) & \text{if } t \text{ odd.} \end{cases}$$

Using that $\mathbb{E}[\varepsilon] = 0$ together with the definition of $z_t^{(\Sigma)}$ gives

$$\begin{aligned}
\mathbb{E}_{\varepsilon\sim\text{Rad}}\left[R_T^{\Sigma}\right] &= \mathbb{E}_{\varepsilon\sim\text{Rad}}\left[\min_{u\in\mathcal{X}} \sum_{t=1}^{T} \left\langle z_t^{(\Sigma)}, x_t - u \right\rangle\right] \\
&= \mathbb{E}_{\varepsilon\sim\text{Rad}}\left[\max_{u\in\mathcal{X}} \sum_{\substack{t=1 \\ t \text{ odd}}}^{T} \left\langle \varepsilon_t\, z'(x_t), u \right\rangle\right] \\
&\geqslant \frac{G}{4}\, \mathbb{E}_{\varepsilon\sim\text{Rad}}\left[\max_{u\in\mathcal{X}} \sum_{t=1}^{\frac{T}{2}} \varepsilon_t u\right].
\end{aligned}$$

Now use that for a linear function $l(x)$, $\max_{x\in[a,b]} l(x) = \max_{x\in\{a,b\}} l(x) = \frac{l(a+b)}{2} + \frac{|l(a-b)|}{2}$.

$$\begin{aligned}
&= \frac{G}{4}\, \mathbb{E}_{\varepsilon\sim\text{Rad}}\left[\max_{u\in\{a,b\}} \sum_{t=1}^{\frac{T}{2}} \varepsilon_t u\right] \\
&= \frac{G}{8}\, \mathbb{E}_{\varepsilon\sim\text{Rad}}\left[\sum_{t=1}^{T/2} \varepsilon_t (a+b)\right] + \frac{G}{8}\, \mathbb{E}_{\varepsilon\sim\text{Rad}}\left[\left|\sum_{t=1}^{T/2} G\varepsilon_t (a-b)\right|\right] \\
&= \frac{G}{16}\, \mathbb{E}_{\varepsilon\sim\text{Rad}}\left[\left|\sum_{t=1}^{T} \varepsilon_t (a-b)\right|\right].
\end{aligned}$$

Where we have used $\mathbb{E}[\varepsilon] = 0$ again. Now we use that by definition $D = \sup_{x,y\in\mathcal{X}} \|x - y\|$.

$$= \frac{GD}{16}\, \mathbb{E}_{\varepsilon\sim\text{Rad}} \left[ \left| \sum_{t=1}^{T} \varepsilon_t \right| \right] \geqslant \frac{1}{32} D\sqrt{G^2 T}.$$

In the last step, we have used the Khintchine inequality. Now note that $G^2 \geqslant \frac{1}{2}\sup_{x\in\mathcal{X}} \|z'(x)\|^2 = \frac{1}{2}\sup_{x\in\mathcal{X}} \|z'(x) - 0\|^2$. Due to the definition of the sequence $\{z_t^{(\Sigma)}\}_{t\in[T]}$, if $\|\nabla f(x,\xi^t)\| \neq 0$, then $\left\|\nabla f(x,\xi^{t-1})\right\| = 0$. Thus, $\frac{1}{2}\sup_{x\in\mathcal{X}} \|z'(x) - 0\|^2 = \frac{1}{2}\sup_{x\in\mathcal{X}} \left\|\nabla f(x,\xi^t) - \nabla f(x,\xi^{t-1})\right\|^2$ for any $t \in [T]$.

Thus, $\sqrt{G^2 T} = \sqrt{T/(2T)\sum_{t=1}^{T} \sup_{x\in\mathcal{X}} \|\nabla f(x,\xi^t) - \nabla f(x,\xi^{t-1})\|^2} = \bar{\Sigma}_T\sqrt{T/2}$. Setting the value $G = \hat{\Sigma}_T$ completes this part of the proof.

**Case $R_T^\sigma$:** We will show this part by contradiction. Suppose that $\mathcal{D}$ is a distribution such that the variance of the gradients $\sigma$ is equal to $\hat{\sigma}_T$. Suppose the SEA picks this distribution every round and assume for contradiction that $\mathbb{E}[R_T^\sigma] \leqslant o(D\sigma\sqrt{T})$. Using online-to-batch conversion gives a convergence bound of order $o(D\sigma/\sqrt{T})$ which contradicts well-known lower bounds from stochastic optimization (c.f., [1, 21] Section 5). □

### B.2.1 Lower Bound for Adversarially Corrupted Stochastic Data

A careful look at the proof of Theorem 6 shows that the same construction applies to the corrupted data model, providing some worst-case optimality of our bounds.

**Proposition 16.** *For any algorithm, for any $C > 0$, and $\sigma > 0$ there exists an instance of the Adversarially Corrupted model with variance $\sigma$, corruption level $C$, and gradients of norm $G = C/T$ such that*

$$\mathbb{E}[R_T] \geqslant \Omega(D\sigma\sqrt{T} + D\sqrt{GC})\,.$$

*Proof.* Case I. If $\sqrt{GC} \geqslant \sigma\sqrt{T}$. The first construction from the proof of Theorem 6 guarantees the existence of a sequence of linear losses with gradient $G$ against which the regret is $\Omega(DG\sqrt{T}) = \Omega(D\sqrt{GC})$. This sequence can be seen as adversarial perturbations on identically stochastic data. Case II. If $\sigma\sqrt{T} \geqslant \sqrt{GC}$. Then the stochastic lower bound in the second part of the proof of Theorem 6 provides a lower bound on the expected regret of $\Omega(D\sigma\sqrt{T})$. □

### B.3 Proof of Theorem 7

We first need a well-known result for OFTL for strongly convex loss functions.

**Lemma 17.** *Suppose $f_t(\cdot)$ are $\mu$-strongly convex for all $t \in [T]$ and $\psi_t(\cdot) = 0$. Further, let $m_t : \mathcal{X} \to \mathbb{R}$ denote*

*the optimistic prediction, and $g_t \in \partial f_t(x_t)$, $M_t \in \partial m_t(x_t)$. Then the regret for OFTRL is bounded by*

$$R_T(u) \leqslant \sum_{t=1}^{T} \Big( \langle g_t - M_t, x_t - x_{t+1}\rangle - \frac{t\mu}{2}\|x_t - x_{t+1}\|^2 \Big).$$

This is a well-known result and can be found in the literature, e.g., [22]. We include a short proof for completeness.

*Proof.* Let $\bar{F}_t(x) = \sum_{s=1}^{t-1} f_t(x)$ and $G_t \in \partial \bar{F}_{t+1}(x_t)$. Note that $\bar{F}_t$ is $[(t-1)\mu]$-strongly convex. From standard analysis (see, e.g., [22] Lem. 7.1) we obtain

$$\begin{aligned}
\sum_{t=1}^{T}[f_t(x_t) - f_t(u)] &= \underbrace{\bar{F}_{T+1}(x_{T+1}) - \bar{F}_{T+1}(u)}_{\leqslant 0} + \sum_{t=1}^{T}[\underbrace{\bar{F}_t(x_t) + f_t(x_t)}_{=\bar{F}_{t+1}(x_t)} - \bar{F}_{t+1}(x_{t+1})]. \\
&\leqslant \sum_{t=1}^{T}[\bar{F}_{t+1}(x_t) - \bar{F}_{t+1}(x_{t+1})] \\
&\leqslant \sum_{t=1}^{T}\Big( \langle G_t, x_t - x_{t+1}\rangle - \frac{t\mu}{2}\|x_t - x_{t+1}\|\Big).
\end{aligned}$$

Due to convexity, $G_t \in \partial \bar{F}_{t+1}(x_t) = \partial \bar{F}_t(x_t) \cap \partial f_t(x_t)$ and due to update operation $0 \in \partial \bar{F}_t(x_t) \cap \partial m_t(x_t)$. Thus, there exist $g_t \in \partial f_t(x_t)$ and $M_t \in \partial m_t(x_t)$ such that $G_t = g_t - M_t$, which completes the proof. □

**Theorem 7.** *Under assumptions **(A1)** –**(A4)** , the expected regret of OFTL with $M_t = \nabla f(x_{t-1}, \xi_{t-1})$ on surrogate loss functions $\ell_t$ defined in* (4) *is bounded as*

$$\begin{aligned}
\mathbb{E}\left[R_T(u)\right] &\leqslant \frac{1}{\mu}\sum_{t=1}^{T}\frac{1}{t}\left(8\sigma_{\max}^2 + 4\mathbb{E}\left[\sup_{x\in\mathcal{X}}\left\|\nabla F^t(x) - \nabla F^{t-1}(x)\right\|^2\right]\right) + \frac{4D^2L^2}{\mu}\log\left(1 + \frac{16L}{\mu}\right) \\
&\leqslant \frac{1}{\mu}\left(8\sigma_{\max}^2 + 4\Sigma_{\max}^2\right)\log T + \frac{4D^2L^2}{\mu}\log\left(1 + \frac{16L}{\mu}\right).
\end{aligned}$$

*Proof of Theorem 7.* Thanks to the strong convexity assumption **(A4)** ,

$$F^t(x_t) - F^t(x) \leqslant \langle x_t - x, \nabla F^t(x_t)\rangle - \frac{\mu}{2}\|x - x_t\|^2.$$

Taking expectation and using the definition of $\ell_t$ gives

$$\begin{aligned}
\mathbb{E}\left[F^t(x_t) - F^t(x)\right] &\leqslant \mathbb{E}\left[\left\langle x_t - x, \nabla F^t(x_t)\right\rangle - \frac{\mu}{2}\|x - x_t\|^2\right] \\
&= \mathbb{E}\left[\langle x_t - x, \nabla f(x_t, \xi_t)\rangle - \frac{\mu}{2}\|x - x_t\|^2\right] \\
&= \mathbb{E}\left[\langle x_t - x, g_t\rangle - \frac{\mu}{2}\|x - x_t\|^2\right] \\
&= \mathbb{E}\left[\ell_t(x_t) - \ell_t(x)\right].
\end{aligned}$$

Now each function $\ell_t$ is $\mu$-strongly convex, and $\nabla \ell_t(x_t) = g_t$. Thus we can apply Lemma 17

$$\begin{aligned}
\sum_{t=1}^{T}\ell_t(x_t) - \ell_t(x) &\leqslant \sum_{t=1}^{T}\Big( \langle g_t - M_t, x_t - x_{t+1}\rangle - \frac{\mu t}{2}\|x_t - x_{t+1}\|^2\Big). \\
&\leqslant \sum_{t=1}^{T}\left(\frac{1}{\mu t}\|g_t - M_t\|^2 + \left(\frac{\mu t}{4} - \frac{\mu t}{2}\right)\|x_t - x_{t+1}\|^2\right).
\end{aligned}$$

where we used the inequality $\langle a, b\rangle \leqslant \frac{1}{2c}\|a\|^2 + \frac{c}{2}\|b\|^2$. Once again, keeping the negative norm term is crucial. Indeed, using the convexity of $x \mapsto \|x\|^2$ and the smoothness assumption on $\nabla F^t$, we get that for $t \geqslant 2$

$$\begin{aligned}
\|g_t - g_{t-1}\|^2 &\leqslant 4\left\|g_t - \nabla F^t(x_t)\right\|^2 + 4\left\|\nabla F^t(x_t) - \nabla F^t(x_{t-1})\right\|^2 \\
&\quad + 4\left\|\nabla F^t(x_{t-1}) - \nabla F^{t-1}(x_{t-1})\right\|^2 + 4\left\|\nabla F^{t-1}(x_{t-1}) - g_{t-1}\right\|^2 \\
&\leqslant 4\left\|g_t - \nabla F^t(x_t)\right\|^2 + 4L^2\|x_t - x_{t-1}\|^2 \\
&\quad + 4\left\|\nabla F^t(x_{t-1}) - \nabla F^{t-1}(x_{t-1})\right\|^2 + 4\left\|\nabla F^{t-1}(x_{t-1}) - g_{t-1}\right\|^2.
\end{aligned}$$

So that, upper bounding the first term $\ell_1(x_1) - \ell_1(u) \leqslant GD$ we get

$$\sum_{t=1}^{T} \ell_t(x_t) - \ell_t(u)$$
$$\leqslant \sum_{t=2}^{T} \frac{1}{\mu t}\Big(4\|g_t - \nabla F^t(x_t)\|^2 + 4\|g_{t-1} - \nabla F^{t-1}(x_{t-1})\|^2 + 4\|\nabla F^t(x_{t-1}) - \nabla F^{t-1}(x_t)\|^2\Big)$$
$$+ \sum_{t=1}^{T}\Big(\frac{4L^2}{\mu(t+1)} - \frac{\mu t}{4}\Big)\|x_t - x_{t+1}\|^2 + GD\,.$$

The indices can be simplified by noting that,

$$\sum_{t=2}^{T} \frac{4}{\mu t}\|g_{t-1} - \nabla F^{t-1}(x_{t-1})\|^2 \leqslant \sum_{t=2}^{T} \frac{4}{\mu(t-1)}\|g_{t-1} - \nabla F^{t-1}(x_{t-1})\|^2 \leqslant \sum_{t=1}^{T} \frac{4}{\mu t}\|g_t - \nabla F^t(x_t)\|^2\,.$$

To recover

$$\sum_{t=1}^{T} \ell_t(x_t) - \ell_t(u) \leqslant \sum_{t=1}^{T} \frac{8}{\mu t}\|g_t - \nabla F^t(x_t)\|^2 + \sum_{t=2}^{T} \frac{4}{\mu t}\|\nabla F^t(x_{t-1}) - \nabla F^{t-1}(x_t)\|^2$$
$$+ \sum_{t=1}^{T}\Big(\frac{4L^2}{\mu t} - \frac{\mu t}{4}\Big)\|x_t - x_{t+1}\|^2 + GD\,. \quad (13)$$

Define the condition number $\kappa = L/\mu$. Then, for $t \geqslant 16\kappa$, we have $\frac{4L^2}{\mu t} - \frac{\mu t}{4} \leqslant 0$. Therefore the second term can be bounded independently of $T$

$$\sum_{t=1}^{\lceil 16\kappa \rceil}\Big(\frac{4L^2}{\mu t} - \frac{\mu t}{4}\Big)D^2 \leqslant \frac{4L^2D^2}{\mu}\sum_{t=1}^{\lceil 16\kappa \rceil}\frac{1}{t} \leqslant \frac{4L^2D^2}{\mu}\log(1+16\kappa)\,.$$

Combining all bounds, and incorporating the definition of $\sigma_{\max}$ and $\Sigma_{\max}$,

$$\mathbb{E}\left[R_T(u)\right] \leqslant \frac{1}{\mu}\left(8\sigma_{\max}^2 + 4\Sigma_{\max}^2\right)\log T + 4D^2L\kappa\log(1+16\kappa) + GD\,. \qquad \square$$

### B.4 Proof of Theorem 8

**Theorem 8.** *For any learning algorithm, and for any pair of positive numbers $(\sigma, \Sigma)$ there exists a function $f : \mathcal{X} \times \Xi \to \mathbb{R}$ and a sequence of distributions satisfying assumptions **(A1)** ,**(A2)** ,**(A3)** and **(A4)** with $\sigma_{\max} \geqslant \sigma$ and $\Sigma_{\max} \geqslant \Sigma$ such that*

$$\mathbb{E}\left[R_T(u)\right] \geqslant \Omega\left(\frac{1}{\mu}\left(\sigma_{\max}^2 + \Sigma_{\max}^2\right)\log T\right).$$

*Proof of Theorem 8.* Let $\hat{\sigma}_{\max}, \hat{\Sigma}_{\max}$ be given parameters and set $G = \max(\hat{\sigma}_{\max}, \hat{\Sigma}_{\max}/2)$. We want to show that there exists sequence of distributions $\mathcal{D}_1, \dots \mathcal{D}_T$ such that

1. $\sigma_{\max} = \hat{\sigma}_{\max}$ and $\Sigma_{\max} = \hat{\Sigma}_{\max}$.
2. $\mathbb{E}\left[R_T(u)\right] \geqslant c\frac{1}{\mu}(\Sigma_{\max}^2 + \sigma_{\max}^2)\log T$ for some constant $c > 0$.
3. $F_1, \dots, F_T$ are $\mu$-strongly convex.

Consider the iterations up to $T-3$. From Corollary 20 in [13] we obtain an $\Omega(\frac{1}{\mu}G^2\log(T-3))$ lower bound on the expected regret. Thus, there exist a realization $\xi_1, \dots, \xi_{T-3}$ and corresponding $\mu$-strongly convex functions $f(\,\cdot\,, \xi_1), \dots, f(\,\cdot\,, \xi_{T-3})$ such that with respect to this realization, $R_T(u) \geqslant \Omega(\frac{1}{\mu}G^2\log(T-3))$. We now let $\delta_1, \dots \delta_{T-3}$ be the Dirac measure corresponding to this realization. Then, $\sigma_{\max} = 0$ and $\max_{t\in[T-3]}\sup_{x\in\mathcal{X}}\|\nabla F^t(x) - \nabla F^{t-1}(x)\| \leqslant 2G$. But we do not necessarily have that $\sigma_{\max} = \hat{\sigma}_{\max}$ and $\max_{t\in[T-3]}\sup_{x\in\mathcal{X}}\|\nabla F^t(x) - \nabla F^{t-1}(x)\| = \hat{\Sigma}_{\max}$. To guarantee this we want to choose $\mathcal{D}_{T-2}, \mathcal{D}_{T-1}, \mathcal{D}_T$, such that

1. $\sup_{x\in\mathcal{X}} \|\nabla F^{T-2}(x) - \nabla F^{T-1}(x)\| = \hat{\Sigma}_{\max}$ and $\|\nabla F^{T-i}(x)\| \leqslant G$ for $i = 1, 2$.
2. $\sigma_T = \hat{\sigma}_{\max}$ and $\|\nabla F^T(x)\| \leqslant G$.

To satisfy Condition 1, let $\mathcal{D}_{T-2}, \mathcal{D}_{T-1}$ be Dirac measures, such that $\nabla F^{T-2}(x) = -\nabla F^{T-1}(x)$ and $\|\nabla F^{T-2}(x)\| = \frac{1}{2}\hat{\Sigma}_{\max}$. Then, by definition of $G$, we know that $\|\nabla F^{T-2}(x)\| \leqslant G$ and $\sup_{x\in\mathcal{X}} \|\nabla F^{T-2}(x) - \nabla F^{T-1}(x)\| = \hat{\Sigma}_{\max}$. Condition 2 can be satisfied by setting $\mathcal{D}_T$ to be any distribution with sufficient variance. This gives

$$\mathbb{E}\left[R_T(u)\right] \geqslant c\frac{1}{\mu}\left(\Sigma_{\max}^2 + \sigma_{\max}^2\right)\log(T-3) - \mathbb{E}\left[\sum_{t=T-2}^{T} f(x_t,\xi_t) - f(u,\xi_t)\right]. \tag{14}$$

Now it remains to show that the last term is negligible. Indeed, from the upper bound, we know

$$\mathbb{E}\left[\sum_{t=T-2}^{T}\left[f(x_t,\xi_t) - f(u,\xi_t)\right]\right] \leqslant \frac{3}{(T-2)\mu}\left(\Sigma_{\max}^2 + \sigma_{\max}^2\right).$$

Hence, for any $T \geqslant 10$, we get $\frac{3}{(T-2)\mu}\left(\Sigma_{\max}^2 + \sigma_{\max}^2\right) \leqslant \frac{1}{2\mu}\left(\Sigma_{\max}^2 + \sigma_{\max}^2\right)\log(T)$ which together with (14) completes the proof. □

# C Missing Proofs of Section 4

We first show the following general property of the variance for the ROM. This proposition will useful for showing the claims of this section.

**Proposition 18.** *For any $t \in [T]$, the variance of the ROM with respect to $\mathcal{D}_t$ satisfies*

$$\mathbb{E}_{\xi\sim\mathcal{D}_t}\left[\|\nabla f(x,\xi) - \nabla F^t(x)\|^2\right] \leqslant \frac{T}{T-t+1}\sigma_1^2,$$

*for any $x \in \mathcal{X}$.*

*Proof.* For any $x \in \mathcal{X}$, since $\nabla F^t(x) = \mathbb{E}_{\xi\sim\mathcal{D}_t}[\nabla f(x,\xi)]$, we have

$$\mathbb{E}_{\xi\sim\mathcal{D}_t}\left[\|\nabla f(x,\xi) - \nabla F^t(x)\|^2\right] \leqslant \mathbb{E}_{\xi\sim\mathcal{D}_t}\left[\|\nabla f(x,\xi) - \nabla F^1(x)\|^2\right].$$

Now, let $\mathcal{T}_t \subseteq [T]$ denote a subset of indices of gradients which remain to be selected in round $t$, and let $k_t \in \mathcal{T}_{t-1} \setminus \mathcal{T}_t$ be the index selected at round $t$.

For any $x \in \mathcal{X}$

$$\begin{aligned}\mathbb{E}_{\xi\sim\mathcal{D}_t}\left[\|\nabla f(x,\xi) - \nabla F^1(x)\|^2\right] &= \frac{1}{T-t+1}\sum_{\xi\in\mathcal{T}_t}\|\nabla f(x,\xi) - \nabla F^1(x)\|^2 \\ &\leqslant \frac{1}{T-t+1}\sum_{\xi\in[n]}\|\nabla f(x,\xi) - \nabla F^1(x)\|^2 \leqslant \frac{T}{T-t+1}\sigma_1^2, \end{aligned} \tag{15}$$

which is the claimed result. □

## C.1 Proof of Lemma 11

Note that in any case, $\tilde{\sigma}_1 \leqslant T\sigma_1$, and therefore $\log(\tilde{\sigma}_1/\sigma_1) \leqslant \log(T)$. Thus Lemma 11 directly yields $\sigma_{[1:T]}^{(2)} \leqslant \sigma_1 T\log(T)$. This means in particular that the rate of OFTRL in the ROM is never more than a factor $\sqrt{\log T}$ worse than the i.i.d. sampling with replacement rate of $\sigma_1\sqrt{T}$; the next bound can often be much tighter.

**Lemma 11.** *In the single-pass ROM, we have $\Sigma_{[1:T]}^{(2)} \leqslant 8G^2$ and $\bar{\sigma}_{[1:T]}^{(2)} \leqslant T\sigma_1^2\log(2e^2\tilde{\sigma}_1^2/\sigma_1^2)$.*

*Proof of Lemma 11.* Let us begin with the adversarial variation. We will show that deterministically (that is, for any order in which the losses are selected), for any $x \in \mathcal{X}$,

$$\left\|\nabla \mathrm{F}^t(x) - \nabla \mathrm{F}^{t-1}(x)\right\|^2 \leqslant \frac{4G^2}{(T-t+2)^2}\,.$$

With the same notation as in Proposition 18, recall that we denote by $\mathcal{T}_t \subseteq [T]$ the support of $\mathcal{D}_t$ and $k_t = \mathcal{T}_{t-1} \setminus \mathcal{T}_t$. We have $|\,\mathcal{T}_t\,| = T - t + 1$ and for any $x \in \mathcal{X}$,

$$\begin{aligned}
\left\|\nabla \mathrm{F}^t(x) - \nabla \mathrm{F}^{t-1}(x)\right\|^2 &= \left\|\frac{1}{T-t+1}\sum_{s\in\mathcal{T}_t}\nabla f_s(x) - \frac{1}{T-t+2}\sum_{s\in\mathcal{T}_{t-1}}\nabla f_s(x)\right\|^2 \\
&= \left\|\frac{1}{(T-t+1)(T-t+2)}\sum_{s\in\mathcal{T}_t}\nabla f_s(x) - \frac{1}{T-t+2}\nabla f_{k_t}(x)\right\|^2 \\
&\leqslant \frac{2}{(T-t+2)^2}\left\|\frac{1}{(T-t+1)}\sum_{s\in\mathcal{T}_t}\nabla f_s(x)\right\|^2 + \frac{2}{(T-t+2)^2}\left\|\nabla f_{k_t}(x)\right\|^2\,.
\end{aligned}$$

Thus, after maximising over $x \in \mathcal{X}$, and taking expectations (note that the inequality holds almost surely) and summing over rounds $t \in [T]$,

$$\Sigma^{(2)}_{[1:T]} = \mathbb{E}\left[\sum_{t=1}^{T}\sup_{x\in\mathcal{X}}\left\|\nabla \mathrm{F}_t(x) - \nabla \mathrm{F}_{t-1}(x)\right\|^2\right] \leqslant \sum_{t=1}^{T}\frac{4G^2}{(T-t+2)^2} \leqslant 8G^2.$$

**Variance.** From Proposition 18, we know that

$$\sigma_t^2 \leqslant \frac{T}{T-t+1}\sigma_1^2$$

Moreover, one can see that

$$\begin{aligned}
\mathbb{E}[\sigma_t^2] &\leqslant \mathbb{E}\Big[\max_{x\in\mathcal{X}}\mathbb{E}_{\xi\sim\mathcal{D}_t}\left[\|\nabla f(x,\xi) - \nabla F^1(x)\|^2\right]\Big] \\
&\leqslant \mathbb{E}\Big[\mathbb{E}_{\xi\sim\mathcal{D}_t}\Big[\max_{x\in\mathcal{X}}\|\nabla f(x,\xi) - \nabla F^1(x)\|^2\Big]\Big] = \tilde{\sigma}_1^2\,. \qquad (16)
\end{aligned}$$

Let us introduce a threshold time step $\tau \in [T]$, of which we will set the value later. We upper bound $\mathbb{E}[\sigma_t^2]$ by (15) for the rounds before $\tau$ and by (16) for the other rounds:

$$\mathbb{E}\Big[\sum_{t=1}^{T}\sigma_t^2\Big] \leqslant \mathbb{E}\Big[\sum_{t=1}^{\tau}\sigma_t^2\Big] + \mathbb{E}\Big[\sum_{t=\tau+1}^{T}\sigma_t^2\Big] \leqslant \sum_{t=1}^{\tau}\frac{T}{T-t+1}\sigma_1^2 + (T-\tau)\tilde{\sigma}_1^2$$

Now using standard bounds on the harmonic series,

$$\sum_{t=1}^{\tau}\frac{1}{T-t+1} = \sum_{n=T-\tau+1}^{T}\frac{1}{n} \leqslant 1 + \log\frac{T}{T-\tau+1}\,.$$

Therefore for any $\tau \in [T-1]$, we get

$$\mathbb{E}\Big[\sum_{t=1}^{T}\sigma_t^2\Big] \leqslant T\sigma_1^2\Big(1 + \log\frac{T}{T-\tau+1}\Big) + (T-\tau)\tilde{\sigma}_1^2\,. \qquad (17)$$

We now conclude by setting the appropriate value for $\tau$. If $T\sigma_1^2/\tilde{\sigma}_1^2 \leqslant 2$, then $\log T \leqslant \log(2\tilde{\sigma}_1^2/\sigma_1^2)$, and taking $\tau = T$ gives a bound of $T\sigma_1^2(1+\log T) \leqslant T\sigma_1^2(1+\log(2\tilde{\sigma}_1^2/\sigma_1^2))$, which is (better than) the claimed result.

Otherwise, we take $\tau = T - \lfloor T\sigma_1^2/\tilde{\sigma}_1^2\rfloor$, then

$$(T-\tau)\tilde{\sigma}_1^2 = \lfloor T\sigma_1^2/\tilde{\sigma}_1^2\rfloor\tilde{\sigma}_1^2 \leqslant T\sigma_1^2\,,$$

and the argument of the logarithm can be bounded as

$$\frac{T}{T-\tau+1} \leqslant \frac{T}{\lfloor T\sigma_1^2/\tilde{\sigma}_1^2 \rfloor} \leqslant \frac{1}{\sigma_1^2/\tilde{\sigma}_1^2 - 1/T} \leqslant \frac{2\tilde{\sigma}_1^2}{\sigma_1^2}\,.$$

where we used the fact that $T\sigma_1^2/\widetilde{\sigma}_1^2 > 2$. This yields the final bound

$$\mathbb{E}\left[\sum_{t=1}^{T}\sigma_t^2\right] \leqslant T\sigma_1^2\left(1+\log\frac{2\tilde{\sigma}_1^2}{\sigma_1^2}\right) + T\sigma_1^2 \leqslant T\sigma_1^2\log\left(\frac{2e^2\tilde{\sigma}_1^2}{\sigma_1^2}\right). \qquad \square$$

### C.2 Proof of Corollary 12

**Corollary 12.** *Under the same assumption as in Theorem 7, the expected regret of the ROM is bounded by*

$$\mathbb{E}\left[R_T(u)\right] \leqslant O\left(\frac{\sigma_1^2}{\mu}\log T + \frac{G^2}{\mu} + LD^2\kappa\log\kappa\right).$$

*For multi-pass ROM with $P$ passes, we obtain*

$$\mathbb{E}\left[R_T(u)\right] \leqslant O\left(\frac{\sigma_1^2}{\mu}\log T + \frac{G^2}{\mu} + \frac{G^2\log P}{n\mu} + LD^2\kappa\log\kappa\right).$$

*Proof of Corollary 12.* **Single-pass ROM:** From Theorem 7 we obtain (c.f. (13))

$$\begin{aligned}\mathbb{E}\left[R_T(u)\right] \leqslant\ & \mathbb{E}\left[\sum_{t=1}^{T}\frac{8}{\mu t}\|g_t - \nabla F^t(x_t)\|^2 + \sum_{t=2}^{T}\frac{4}{\mu t}\|\nabla F^t(x_{t-1}) - \nabla F^{t-1}(x_t)\|^2\right] \\ & + GD + \frac{4L^2D^2}{\mu}\log(1+16\kappa)\,.\end{aligned}$$

By Lemma 11, we have

$$\mathbb{E}\left[\sum_{t=2}^{T}\frac{4}{\mu t}\|\nabla F^t(x_{t-1}) - \nabla F^{t-1}(x_t)\|^2\right] \leqslant 8G^2.$$

Furthermore, recall that by Proposition 18 $\mathbb{E}[\sigma_t^2] \leqslant T/(T-t+1)\sigma_1^2$.

$$\mathbb{E}\left[\sum_{t=1}^{T}\frac{8}{\mu t}\|g_t - \nabla F^t(x_t)\|^2\right] \leqslant \frac{8}{\mu}\sum_{t=1}^{T}\frac{T}{t(T-t+1)}\sigma_1^2 \leqslant \frac{8\sigma_1^2}{\mu}(2+2\log(T)).$$

Indeed, using a standard bound on the harmonic series,

$$\sum_{t=1}^{T}\frac{T}{t(T-t+1)} = \sum_{t=1}^{T}\frac{T-t+1+t-1}{t(T-t+1)} \leqslant \sum_{t=1}^{T}\frac{1}{t} + \frac{1}{T-t+1} \leqslant 2+2\log T\,.$$

Combining these bounds gives the first part of the corollary.

**Multi-pass ROM:** The critical term to upper bound, is the differences of the means whenever a pass ends and a new pass starts. Thus, for $P \in \mathbb{N}$ passes, we need to control $\sup_{x\in\mathcal{X}}\left\|\nabla F^t(x) - \nabla F^{t-1}(x)\right\|^2$ for $t = jn+1$, with $j \in [P]$.

Inside the $i$-th pass, for $k \in [n]$ we bound the $k$-th variation by

$$\sup_{x\in\mathcal{X}}\left\|\nabla F^k(x) - \nabla F^{k-1}(x)\right\|^2 \leqslant \frac{4G^2}{(n-k+2)^2}\,,$$

and we bound it by $G^2$ between the passes, so that

$$
\begin{aligned}
&\frac{4}{\mu}\sum_{t=1}^{T}\frac{1}{t}\sup_{x\in\mathcal{X}}\left\|\nabla F^t(x)-\nabla F^{t-1}(x)\right\|^2\\
&\leqslant \frac{4}{\mu}\sum_{i=1}^{P}\sum_{k=1}^{n}\frac{1}{(i-1)n+k}\sup_{x\in\mathcal{X}}\left\|\nabla F^k(x)-\nabla F^{k-1}(x)\right\|^2+\frac{4}{\mu}\sum_{i=1}^{P}\frac{1}{in}\sup_{x\in\mathcal{X}}\left\|\nabla F^1(x)-\nabla F^n(x)\right\|^2\\
&\leqslant \frac{4}{\mu}\sum_{i=1}^{P}\sum_{k=1}^{n}\frac{1}{(i-1)n+k}\frac{2G^2}{(n-k+2)^2}+\frac{4}{\mu}\sum_{i=1}^{P}\frac{1}{in}\sup_{x\in\mathcal{X}}\left\|\nabla F^1(x)-\nabla F^n(x)\right\|^2\\
&\leqslant \frac{16G^2}{\mu}\left(1+2\frac{\log P}{n}\right). \qquad \square
\end{aligned}
$$

# D Batch-to-online Conversion

Consider the stochastic optimization problem $\min_{x\in\mathcal{X}}\ \mathbb{E}_{\xi\sim\mathcal{D}}\left[f(x,\xi)\right]$ and let $x^*$ denote a minimiser for this problem. Further, let $\mathcal{A}$ be any first order stochastic optimization method with convergence guarantee $\mathbb{E}_{\xi\sim\mathcal{D}}\left[f(x_t,\xi)-f(x^*,\xi)\right]\leqslant c(t)$. As input $\mathcal{A}$ takes an initial iterate $x_1$ and a sequence of i.i.d. samples $\{f(\,\cdot\,,\xi_s)\}_{s\in[t]}$. We let $\mathcal{A}(x_1,\{f(\,\cdot\,,\xi_s)\}_{s\in[t]})$ denote the output $x_{t+1}$ of the stochastic optimization algorithm with respect to the given input. Now consider an OCO with $f(\,\cdot\,,\xi_1),\dots f(\,\cdot\,,\xi_T)$ and $\xi_1,\dots\xi_T$ are sampled i.i.d. from a distribution.

**Input:** Stochastic first order method $\mathcal{A}$
**for** $t=1,2,\dots T$ **do**
  play $x_t$ and suffer loss $f(x_t,\xi_t)$;
  restart $\mathcal{A}$ and set $x_{t+1}=\mathcal{A}(x_1,\{f(\,\cdot\,,\xi_s)\}_{s\in[t]})$
**end**

**Algorithm 1:** Batch-to-online

This batch-to-online conversion trivially achieves $\sum_{t=1}^{T}c(t)$ expected regret. However, with this conversion, some aspects of the stochastic convergence bound are lost. Consider for instance a convergence rate $c(t)=O(LD^2/t+D\sigma/\sqrt{t})$, from the the first-order stochastic approximation method in [7] and the accelerated version $c(t)=O(LD^2/t^2+D\sigma/\sqrt{t})$ [8, 17]. In both cases, the functions are assumed to satisfy **(A1)** -**(A3)** . Batch-to-online conversion yields

$$\mathbb{E}\left[R_T(u)\right]\leqslant O(LD^2\log T+D\sigma\sqrt{T}) \qquad \text{and} \qquad \mathbb{E}\left[R_T(u)\right]\leqslant O(LD^2+D\sigma\sqrt{T}).$$

The benefits of acceleration can be seen in the lower-order terms. Now using standard online-to-batch [3] conversion gives the convergence bounds

$$\mathbb{E}\left[f(x_T,\xi)-f(x^*,\xi)\right]\leqslant O\left(\frac{D\sigma}{\sqrt{T}}+LD^2\frac{\log T}{T}\right) \text{ and } \mathbb{E}\left[f(x_T,\xi)-f(x^*,\xi)\right]\leqslant O\left(\frac{D\sigma}{\sqrt{T}}+\frac{LD^2}{T}\right),$$

In the case of accelerated stochastic approximation, the benefits of acceleration are inevitably lost through batch-to-online and online-to-batch conversion.

# E Additional Examples for Intermediate Cases

We provide regret bounds for intermediate cases not discussed in the main body of the paper, namely the cases when the adversary selects slowly shifting distributions and when the adversary switches rarely between distributions.

**Distribution shift:** In this example, the SEA picks $\mathcal{D}_t$ and $\mathcal{D}_{t-1}$, such that $\nabla F^t(x)$ is close to the mean of the previous distribution gradient $\nabla F^{t-1}(x)$. We shall consider two kinds of distribution shifts. Firstly, when the means are close on average, that is, when $(1/T)\sum_{t=1}^{T}\sup_{x\in\mathcal{X}}\|\nabla F^t(x)-\nabla F^{t-1}(x)\|^2\leqslant\varepsilon$, secondly, when this holds for each iteration

$t$, i.e., $\sup_{x\in\mathcal{X}} \|\nabla F^t(x) - \nabla F^{t-1}(x)\|^2 \leqslant \varepsilon$. We refer to the former as the *average distribution shift* case, and to the latter as the *bounded distribution shift* case.

For strongly convex functions, Theorem 7 directly yields the regret bound

$$\mathbb{E}\left[R_T(u)\right] \leqslant O\left(\frac{1}{\mu}(\sigma_{\max}^2 + \varepsilon)\log T + D^2 L\kappa \log \kappa\right).$$

For the considerably weaker assumption of an average distribution shift, we have

$$\sum_{t=1}^{T} \frac{1}{t\mu} \sup_{x\in\mathcal{X}} \left\|\nabla F^t(x) - \nabla F^{t-1}(x)\right\|^2 \leqslant \Sigma_{[1:T]}^{(2)} \sqrt{\sum_{t=1}^{T} \frac{1}{t^2\mu^2}} \leqslant \frac{4}{\mu} T\varepsilon.$$

To obtain the first inequality, we have used the Cauchy-Schwarz inequality together with the fact that $\sqrt{a+b} \leqslant \sqrt{a} + \sqrt{b}$, and the second inequality follows directly from the definition of the averaged distribution shift. Now suppose $\varepsilon \leqslant 1/T$, then we obtain the following regret bound in case of average distribution shift.

$$\mathbb{E}\left[R_T(u)\right] \leqslant O\left(\frac{\sigma_{\max}^2}{\mu}\log T + \frac{1}{\mu} + D^2 L\kappa \log \kappa\right).$$

Since $\Sigma_{[1:T]}^{(2)} \leqslant T\varepsilon$ for the average distribution shift, for convex and smooth functions, Theorem 5 entails that

$$\mathbb{E}\left[R_T(u)\right] \leqslant O\left(D(\sigma_{\max} + \sqrt{\varepsilon})\sqrt{T} + DG + LD^2\right).$$

**Distribution switch:** SEA switches $c$ times between distributions $\mathcal{D}_1, \ldots, \mathcal{D}_c \in \mathfrak{D}$. These switches can happen at any round and the learner does not know when a switch occurs. In this case, we can upper bound $\Sigma_{[1:T]}^{(2)} \leqslant \Sigma_{\max}^2 c$. Thus, for strongly convex functions Theorem 7 directly yields

$$\mathbb{E}\left[R_T(u)\right] \leqslant O\left(\frac{1}{\mu}\left(\sigma_{\max}^2 \log T + \Sigma_{\max}^2 \log c\right) + D^2 L\kappa \log \kappa\right).$$

And for convex smooth functions Theorem 5 gives

$$\mathbb{E}\left[R_T(u)\right] \leqslant O\left(D\sigma_{\max}\sqrt{T} + D\Sigma_{\max}\sqrt{c} + DG + LD^2\right).$$