Proceedings of Machine Learning Research vol 75:1–34, 2018 31st Annual Conference on Learning Theory

Online Variance Reduction for Stochastic Optimization

Zalán Borsos ZALAN.BORSOS@INF.ETHZ.CH
Andreas Krause KRAUSEA@INF.ETHZ.CH
Kfir Y. Levy YEHUDA.LEVY@INF.ETHZ.CH
Department of Computer Science
ETH Zurich

Editors: Sébastien Bubeck, Vianney Perchet and Philippe Rigollet

Abstract

Modern stochastic optimization methods often rely on uniform sampling which is agnostic to the underlying characteristics of the data. This might degrade the convergence by yielding estimates that suffer from a high variance. A possible remedy is to employ non-uniform *importance sampling* techniques, which take the structure of the dataset into account. In this work, we investigate a recently proposed setting which poses variance reduction as an online optimization problem with bandit feedback. We devise a novel and efficient algorithm for this setting that finds a sequence of importance sampling distributions competitive with the best fixed distribution in hindsight, the first result of this kind. While we present our method for sampling data points, it naturally extends to selecting coordinates or even blocks of thereof. Empirical validations underline the benefits of our method in several settings.

Keywords: importance sampling, variance reduction, bandit feedback, empirical risk minimization

1. Introduction

Empirical risk minimization (ERM) is among the most important paradigms in machine learning, and is often the strategy of choice due to its generality and statistical efficiency. In ERM, we draw a set of samples $\mathcal{D} = \{x_1, \ldots, x_n\} \subset \mathcal{X}$ from the underlying data distribution, and we aim to find a solution $w \in \mathcal{W}$ that minimizes the empirical risk,

$$\min_{w \in \mathcal{W}} L(w) := \frac{1}{n} \sum_{i=1}^{n} \ell(x_i, w), \tag{1}$$

where $\ell : \mathcal{X} \times \mathcal{W} \to \mathbb{R}$ is a given loss function, and $\mathcal{W} \subseteq \mathbb{R}^d$ is usually a compact domain.

In this work we are interested in sequential procedures for minimizing the ERM objective, and relate to such methods as *ERM solvers*. More concretely, we focus on the regime where the number of samples n is very large, and it is therefore desirable to employ ERM solvers that only require few passes over the dataset. There exists a rich arsenal of such efficient solvers which have been investigated throughout the years, with the canonical example from this category being Stochastic Gradient Descent (SGD).

Typically, such methods require an unbiased estimate of the loss function at each round, which is usually generated by sampling a few points uniformly at random from the dataset. However, by employing uniform sampling, these methods are insensitive to the intrinsic structure of the data. In case of SGD, for example, some data points might produce large gradients, but they are nevertheless assigned the same probability of being sampled as any other point. This ignorance often results in high-variance estimates, which is likely to degrade the performance.

The above issue can be mended by employing non-uniform importance sampling. And indeed, we have recently witnessed several techniques to do so: Zhao and Zhang (2015) and similarly Needell et al. (2014), suggest using prior knowledge on the gradients of each data point in order to devise predefined importance sampling distributions. Stich et al. (2017) devise adaptive sampling techniques guided by a robust optimization approach. These are only a few examples of a larger body of work (Bouchard et al., 2015; Alain et al., 2015; Csiba and Richtárik, 2016).

Interestingly, the recent works of Namkoong et al. (2017) and Salehi et al. (2017) formulate the task of devising importance sampling distributions as an online learning problem with bandit feedback. In this context, they think of the algorithm, which adaptively chooses the distribution, as a player that competes against the ERM solver. The goal of the player is to minimize the cumulative variance of the resulting (gradient) estimates. Curiously, both methods rely on some form of the "linearization trick"[1] to resort to the analysis of the EXP3 (Auer et al., 2002).

On the other hand, the theoretical guarantees of the above methods are somewhat limited. Strictly speaking, none of them provides regret guarantees with respect to the best fixed distribution in hindsight: Namkoong et al. (2017) only compete with the best distribution among a *subset* of the simplex (around the uniform distribution). Conversely, Salehi et al. (2017) compete against a solution which might perform worse than the best in hindsight up to a multiplicative factor of 3.

In this work, we adopt the above mentioned online learning formulation, and design novel importance sampling techniques. Our adaptive sampling procedure is simple and efficient, and in contrast to previous work, we are able to provide regret guarantees with respect to the best fixed point among the simplex. As our contribution, we

- motivate theoretically why regret minimization is meaningful in this setting,
- propose a novel bandit algorithm for variance reduction ensuring regret of $\tilde{\mathcal{O}}(n^{1/3}T^{2/3})$,
- empirically validate our method, and provide an efficient implementation[2].

On the technical side, we do not rely on a "linearization trick" but rather directly employ a scheme based on the classical Follow-the-Regularized-Leader approach. Our analysis entails several technical challenges, most notably handling unbounded cost functions while only receiving partial (bandit) feedback. Our design and analysis draws inspiration from the seminal works of Auer et al. (2002) and Abernethy et al. (2008). Although we present our method for choosing *data points*, it naturally applies to choosing *coordinates in coordinate descent* or even *blocks* of thereof (Allen-Zhu et al., 2016; Perekrestenko et al., 2017; Nesterov, 2012; Necoara et al., 2011). More broadly, the proposed algorithm can be incorporated in *any sequential algorithm* that relies on an unbiased estimation of the loss. A prominent application of our method is variance reduction for SGD, which can be achieved by considering gradient norms as losses, i.e., replacing $\ell(w, x_i) \leftrightarrow \|\nabla \ell(w, x_i)\|$. With this modification, our method is minimizing the cumulative variance of the gradients throughout the optimization process.

The paper is organized as follows. In Section 2, we formalize the online learning setup of variance reduction, and motivate why regret is a suitable performance measure. As the first step of

1. By "linearization trick" we mean that these methods update according to a first order approximation of the costs rather than the costs themselves.
2. The source code is available at https://github.com/zalanborsos/online-variance-reduction

our analysis, we investigate the full information setting in Section 3, which serves as a mean for studying the bandit setting in Section 4. Finally, we validate our method empirically, and provide the detailed discussion of the results in Appendix F.

2. Motivation and Problem Definition

Typical sequential solvers for ERM usually require a fresh unbiased estimate $\tilde{L}_t$ of the loss L_t at each round, which is obtained by repeatedly sampling from the dataset. The template of Figure 1 captures a rich family of such solvers such as SGD, SAGA (Defazio et al., 2014), SVRG (Johnson and Zhang, 2013), and online k-Means (Bottou and Bengio, 1995).

Sequential Optimization Procedure for ERM

Input: Dataset $\mathcal{D} = \{x_1, \ldots, x_n\}$
Initialize: $w_1 \in \mathcal{W}$
for $t = 1, \ldots, T$ **do**
  Draw samples from $\mathcal{D}$ using $p_t \in \Delta$ to generate $\tilde{L}_t(\cdot)$, an unbiased estimate for $L(\cdot)$.
  Update solution: $w_{t+1} \leftarrow \mathcal{A}(w_t, \tilde{L}_t(\cdot))$.
end for

Figure 1: Template of a sequential procedure for minimizing the ERM objective. At each round, we devise a fresh unbiased estimate $\tilde{L}_t(\cdot)$ of the empirical loss, then we update the solution based on the previous solution w_t and $\tilde{L}_t(\cdot)$.

A natural way to devise the unbiased estimates $\tilde{L}_t$ is to sample $i_t \in \{1, \ldots, n\}$ uniformly at random and return $\tilde{L}_t(w) = \ell(x_{i_t}, w)$. Indeed, uniform sampling is the common practice when applying SGD, SAGA, SVRG and online k-Means. Nevertheless, *any distribution* p in the probability simplex Δ induces an unbiased estimate. Concretely, sampling an index $i \sim p$ induces the estimate

$$\tilde{L}(w) := \frac{1}{n \cdot p(i)} \cdot \ell(x_i, w) \tag{2}$$

and it is immediate to show that $\mathbb{E}_{x_i \sim p}[\tilde{L}(w)] = L(w)$. This work is concerned with efficient ways of choosing a "good" sequence of sampling distributions $\{p_1(\cdot), \ldots, p_T(\cdot)\}$.

It is well known that the performance of typical solvers (e.g. SGD/SAGA/SVRG) improves as the variance of the estimates $\tilde{L}_t(w_t)$ is becoming smaller. Thus, a natural criterion for measuring the performance of a sampling distribution p is the variance of the induced estimate

$$\mathrm{Var}_p(\tilde{L}(w)) = \frac{1}{n^2} \sum_{i=1}^{n} \frac{\ell^2(x_i, w)}{p(i)} - L^2(w).$$

Denoting $\ell_t(i) := \ell(x_i, w_t)$ and noting that the second term is independent of p, we may now cast the task of sequentially choosing the sampling distributions as the online optimization problem shown in Figure 2. In the above protocol, we treat the sequential solver as an adversary that chooses a sequence of loss vectors $\{\ell_t\}_{t\in[T]} \subset \mathbb{R}^n$, where $t \in [T]$ denotes $t \in \{1, \ldots, T\}$. Each loss

Online Variance Reduction Protocol

Input: Dataset $\mathcal{D} = \{x_1, \dots, x_n\}$
for $t = 1, \dots, T$ **do**
Player chooses $p_t \in \Delta$.
Adversary chooses $\ell_t \in \mathbb{R}^n$, which induces a cost function $f_t(p) := \sum_{i=1}^n \frac{\ell_t^2(i)}{p(i)}$.
Player draws a sample $I_t \sim p_t$.
Player incurs a cost $\frac{1}{n^2} f_t(p_t)$, and receives $\ell_t(I_t)$ as (bandit) feedback.
end for

Figure 2: Online variance reduction protocol with bandit feedback

vector is a function of w_t, the solution chosen by the solver in the corresponding round (note that we abstract out this dependence of ℓ_t in w_t). The cost[3] $\frac{1}{n^2} f_t(p_t)$ that the player incurs at round t is the second moment of the loss estimate, which is induced by the distribution chosen by the player at round t.

Next, we define the regret, which is our performance measure for the player,

$$\text{Regret}_T = \frac{1}{n^2} \left(\sum_{t=1}^T f_t(p_t) - \min_{p \in \Delta} \sum_{t=1}^T f_t(p) \right).$$

Our goal is to devise a no-regret algorithm such that $\lim_{T \to \infty} \text{Regret}_T / T = 0$, which in turn guarantees that we recover asymptotically the best fixed sampling distribution. In the bandit feedback setting, the player aims to minimize its expected regret $\mathbb{E}[\text{Regret}_T]$, where the expectation is taken with respect to the randomized choices of the player and the adversary. Note that we allow the choices of the adversary to depend on the past choices of the player.

There are few noteworthy comments regarding the above setup. First, it is immediate to verify that the cost functions $f_1, \dots, f_T$ are convex in Δ, therefore this is an online convex optimization problem. Secondly, the cost functions are unbounded in Δ, which poses a challenge in ensuring no-regret. Finally, notice that the player receives a *bandit feedback*, i.e., he is allowed to inspect the losses only at the coordinate I_t chosen at time t. To the best of our knowledge, this is the first natural setting where, as we will show, it is possible to provide no regret guarantees despite bandit feedback and unbounded costs.

Throughout this work, we assume that the losses are bounded, $l_t^2(i) \leq L$ for all $i \in [n]$ and $t \in [T]$. Note that our analysis may be extended to the case where the bounds are instance-dependent, i.e., $l_t^2(i) \leq L_i$ for all $i \in [n]$ and $t \in [T]$. In practice, it can be beneficial to take into account the different L_i's, as we demonstrate in our experiments.

2.1. Is Regret a Meaningful Performance Measure?

Let us focus on the family of ERM solvers depicted in Figure 1. As discussed above, devising loss estimates such that $\tilde{L}_t(w_t)$ has low variance is beneficial for such solvers — in case of SGD, this is due to strong connection between the cumulative variance of gradients and the quality of

3. We use the term "cost function" to refer to f in order to distinguish it from the loss ℓ.

optimization that we discuss in more detail in Appendix A. Translating this observation into the online variance reduction setting suggests a natural performance measure: rather than competing with the best *fixed* distribution in hindsight, we would like to compete against the *sequence* of best distributions per-round $\{p_t^* \leftarrow \arg\min \sum_{i=1}^n \ell_t^2(i)/p(i)\}_{t\in[T]}$. This optimal sequence ensures *zero* variance in every round, and is therefore the ideal baseline to compete against. This also raises the question whether regret guarantees, which compare against the best *fixed* distribution in hindsight, are at all meaningful in this context. Note that regret minimization is meaningful in stochastic optimization, when we assume that the losses are generated i.i.d. from some fixed distribution (Cesa-Bianchi et al., 2004). Yet, this certainly does not apply in our case since losses are non-stationary and non-oblivious.

Unfortunately, ensuring guarantees compared to the sequence of best distributions per-round seems generally hard. However, as we show next, regret is still a meaningful measure for sequential ERM solvers. Concretely, recall that our ultimate goal is to minimize the ERM objective. Thus, we are only interested in ERM solvers that actually converge to a (hopefully good) solution for the ERM problem. More formally, let us define $\ell_*(i)$ as follows,

$$\ell_*(i) := \lim_{t\to\infty} \ell_t(i),$$

where we recall that $\ell_t(i) := \ell(x_i, w_t)$, and assume the above limit to exist for every $i \in [n]$. We will also denote $L_* := \frac{1}{n}\sum_{i=1}^n \ell_*(i)$. Moreover, let us assume that the asymptotic solution is better on average than any of the sequential solutions in the following sense,

$$\frac{1}{T}\sum_{t=1}^T L(w_t) \ge L_* \ , \quad \forall T \ge 1$$

where $L(w_t) := \frac{1}{n}\sum_{i=1}^n \ell(x_i, w_t)$. This assumption naturally holds when the ERM solver converges to the optimal solution for the problem, which applies for SGD in the convex case. The next lemma shows that under these mild assumptions, competing against the best *fixed* distribution in hindsight is not far from competing against the ideal baseline.

Lemma 1 *Consider the online variance reduction setting, and for any* $i \in [n]$ *denote* $V_T(i) = \sum_{t=1}^T (\ell_t(i) - \ell_*(i))^2$. *Assuming that the losses,* $l_t(i)$*, are non-negative for all* $i \in [n]$, $t \in [T]$*, the following holds for any* $T \ge 1$,

$$\frac{1}{n^2}\min_{p\in\Delta}\sum_{t=1}^T f_t(p) \le \frac{1}{n^2}\sum_{t=1}^T \min_{p\in\Delta} f_t(p) \ + \ 2\sqrt{T}L_* \cdot \frac{1}{n}\sum_{i=1}^n \sqrt{V_T(i)} + \left(\frac{1}{n}\sum_{i=1}^n \sqrt{V_T(i)}\right)^2 .$$

Thus, the above lemma connects the convergence rate of the ERM solver to the benefit that we get by regret minimization. It shows that the benefit is larger if the ERM solver converges faster. As an example, let us assume that $|\ell_t(i) - \ell_*(i)| \le \mathcal{O}(1/\sqrt{t})$, which loosely speaking holds for SGD. This assumption implies $V_T(i) \le \mathcal{O}(\log(T))$, hence by Lemma 1 the regret guarantees translate into guarantees with respect to the ideal baseline, with an additional cost of $\tilde{\mathcal{O}}(\sqrt{T})$.

3. Full Information Setting

In this section, we analyze variance reduction with full-information feedback. We henceforth consider the same setting as in Fig. 2, with the difference that in each round the player receives as a feedback the loss vector at all points $(l_t(1), l_t(2), \dots, l_t(n))$ instead of only $l_t(I_t)$. We introduce a new algorithm based on the FTRL approach, and establish an $\mathcal{O}(\sqrt{T})$ regret bound for our method in Theorem 3. While this setup in itself has little practical relevance, it later serves as a mean for investigating the bandit setting.

Follow-the-Regularized-Leader (FTRL) is a powerful approach to online learning problems. According to FTRL, in each round, one selects a point that minimizes the cost functions over past rounds plus a regularization term, i.e., $p_t \leftarrow \arg\min_{p\in\Delta} \sum_{\tau=1}^{t-1} f_\tau(p) + \mathcal{R}(p)$. The regularizer $\mathcal{R}$ usually assures that the choices do not change abruptly over the rounds. We choose $\mathcal{R}(p) = \gamma \sum_{i=1}^n \frac{1}{p(i)}$ which allows to write FTRL as

$$p_t \leftarrow \arg\min_{p\in\Delta} \sum_{\tau=1}^{t-1} f_\tau(p) + \gamma \sum_{i=1}^{n} \frac{1}{p(i)} . \tag{3}$$

The regularizer $\mathcal{R}(p) = \gamma \sum_{i=1}^n 1/p(i)$ is a natural candidate in our setting, since it has the same structural form as the cost functions. It also prevents FTRL from assigning vanishing probability mass to any component, thus ensuring that the incurred costs never explode. Moreover, $\mathcal{R}$ assures a closed form solution to the FTRL as the following lemma shows.

Lemma 2 *Denote $l_{1:t}^2(i) := \sum_{\tau=1}^t \ell_\tau^2(i)$. The solution to Eq. (3) is given by $p_t(i) \propto \sqrt{\ell_{1:t-1}^2(i) + \gamma}$.*

Proof sketch Recalling $f_t(p) = \sum_{i=1}^n \frac{\ell_t^2(i)}{p(i)}$, allows to write the FTRL objective as follows,

$$\sum_{\tau=1}^{t-1} f_\tau(p) + \gamma \sum_{i=1}^{n} 1/p(i) = \sum_{i=1}^{n} (\ell_{1:t-1}^2(i) + \gamma)/p(i) .$$

It is immediate to validate that the offered solution satisfies the first order optimality conditions in Δ. Global optimality follows since the FTRL objective is convex in the simplex. ■

We are interested in the regret incurred by our method. The following theorem shows that, despite the non-standard form of the cost functions, we can obtain $\mathcal{O}(\sqrt{T})$ regret.

Theorem 3 *Setting $\gamma = L$, the regret of the FTRL scheme proposed in Equation (3) is:*

$$\textit{Regret}_T \le \frac{27\sqrt{L}}{n} \left(\sum_{i=1}^{n} \sqrt{\ell_{1:T}^2(i)} \right) + 44L.$$

Furthermore, since $\ell_t^2(i) \le L$ we have $\textit{Regret}_T \le 27L\sqrt{T} + 44L$.

Before presenting the proof, we briefly describe it. Trying to apply the classical FTRL regret bounds, we encounter a difficulty, namely that the regularizer in Equation (3) can be unbounded. To overcome this issue, we first consider competing with the optimal distribution on a restricted simplex where $\mathcal{R}(\cdot)$ is bounded. Then we investigate the cost of considering the restricted simplex instead of the full simplex.

Along the lines described above, consider the simplex Δ and the restricted simplex $\Delta' = \{p \in \Delta |\ p(i) \geq p_{\min}, \forall i \in [n]\}$ where $p_{\min} \leq 1/n$ is to be defined later. We can now decompose the regret as follows,

$$n^2 \cdot \text{Regret}_T = \underbrace{\sum_{t=1}^T f_t(p_t) - \min_{p\in\Delta'} \sum_{t=1}^T f_t(p)}_{\text{(A)}} + \underbrace{\min_{p\in\Delta'} \sum_{t=1}^T f_t(p) - \min_{p\in\Delta} \sum_{t=1}^T f_t(p)}_{\text{(B)}}. \tag{4}$$

We continue by separately bounding the above terms. To bound (A), we will use standard tools which relate the regret to the stability of the FTRL decision sequence (FTL-BTL lemma). Term (B) is bounded by a direct calculation of the minimal values in Δ and Δ'. The following lemma bounds term (A).

Lemma 4 *Setting $\gamma = L$, we have:*

$$\sum_{t=1}^T f_t(p_t) - \min_{p\in\Delta'} \sum_{t=1}^T f_t(p) \leq 22n\sqrt{L} \cdot \left(\sum_{i=1}^n \sqrt{\ell_{1:T}^2(i)} \right) + 22n^2 L + \frac{nL}{p_{min}}.$$

Proof sketch of Lemma 4 The regret of FTRL may be related to the stability of the online decision sequence as shown in the following lemma due to Kalai and Vempala (2005) (proof can also be found in Hazan (2011) or in Shalev-Shwartz et al. (2012)):

Lemma 5 *Let $\mathcal{K}$ be a convex set and $\mathcal{R} : \mathcal{K} \mapsto \mathbb{R}$ be a regularizer. Given a sequence of cost functions $\{f_t\}_{t\in[T]}$ defined over $\mathcal{K}$, then setting $p_t = \arg\min_{p\in\Delta} \sum_{\tau=1}^{t-1} f_\tau(p) + \mathcal{R}(p)$ ensures,*

$$\sum_{t=1}^T f_t(p_t) - \sum_{t=1}^T f_t(p) \leq \sum_{t=1}^T (f_t(p_t) - f_t(p_{t+1})) + (\mathcal{R}(p) - \mathcal{R}(p_1)), \quad \forall p \in \mathcal{K}$$

Notice that $\mathcal{R}(p) = L\sum_{i=1}^n 1/p(i)$ is non-negative and bounded by $nL/p_{\min}$ over Δ'. Thus, applying the above lemma implies that $\forall\ p \in \Delta'$,

$$\sum_{t=1}^T f_t(p_t) - \sum_{t=1}^T f_t(p) \leq \sum_{t=1}^T (f_t(p_t) - f_t(p_{t+1})) + \frac{nL}{p_{\min}} \leq \sum_{t=1}^T \sum_{i=1}^n \ell_t^2(i) \left(\frac{1}{p_t(i)} - \frac{1}{p_{t+1}(i)} \right) + \frac{nL}{p_{\min}}.$$

Using the closed form solution for the p_t's (see Lemma. 2) enables us to upper bound the last term as follows,

$$\sum_{t=1}^T \sum_{i=1}^n \ell_t^2(i) \left(\frac{1}{p_t(i)} - \frac{1}{p_{t+1}(i)} \right) \leq 22n\sqrt{L} \sum_{i=1}^n \sqrt{\ell_{1:T}^2(i) + L}\,. \tag{5}$$

Combining the above with $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ completes the proof. ■

The next lemma bounds term (B).

Lemma 6

$$\min_{p\in\Delta'} \sum_{t=1}^T f_t(p) - \min_{p\in\Delta} \sum_{t=1}^T f_t(p) \leq 6n \cdot p_{min} \cdot \left(\sum_{i=1}^n \sqrt{\ell_{1:T}^2(i)} \right)^2$$

Proof sketch of Lemma 6 Using first order optimality conditions we are able show that the minimal value of the $\sum_{t=1}^{T} f_t(p)$ over Δ is exactly $\left(\sum_{i=1}^{n} \sqrt{\ell_{1:t}^2(i)}\right)^2$. Similar analysis allows to extract a closed form solution to the best in hindsight over Δ'. This in turn enables to upper bound the minimal value over Δ' by $\left(\sum_{i=1}^{n} \sqrt{\ell_{1:t}^2(i)}\right)^2 / (1 - n \cdot p_{\min})^2$. Combining these bounds together with $p_{\min} \leq 1/2n$ we are able to prove the lemma. ■

Proof of Theorem 3 Combining Lemma 4 and 6, we have after dividing by n^2,

$$\text{Regret}_T \leq \frac{22\sqrt{L}}{n} \cdot \left(\sum_{i=1}^{n} \sqrt{\ell_{1:T}^2(i)}\right) + 22L + \frac{L}{n \cdot p_{\min}} + \frac{6 \cdot p_{\min}}{n} \cdot \left(\sum_{i=1}^{n} \sqrt{\ell_{1:T}^2(i)}\right)^2$$

Since the choice of $p_{\min}$ is arbitrary and is relevant only for the theoretical analysis, we can set it to $p_{\min} = \min\left\{1/(2n),\, \sqrt{L}/\left(\sqrt{6}\sum_{i=1}^{n} \sqrt{\ell_{1:T}^2(i)}\right)\right\}$ that yields the final result. ■

4. The Bandit Setting

In this section, we investigate the bandit setting (see Figure 2) which is of great practical appeal as we described in Section 2. Our method for the bandit setting is depicted in Algorithm 3, and it ensures a bound of $\tilde{O}(n^{1/3}T^{2/3})$ on the expected regret (see Theorem 8). Importantly, this bound holds even for non-oblivious adversaries. The design and analysis of our method builds on some of the ideas that appeared in the seminal work of Auer et al. (2002).

Algorithm 3 is using the bandit feedback in order to design an unbiased estimate of the true loss $(\ell_t(1), \ldots, \ell_t(n))$ in each round. These estimates are then used instead of the true losses by the full information FTRL algorithm that was analyzed in the previous section. We do not directly play according to the FTRL predictions but rather mix them with a uniform distribution. Mixing is necessary in order to ensure that the loss estimates are bounded, which is a crucial condition used in the analysis. Next we elaborate on our method and its analysis.

Algorithm 3 samples an arm $I_t \sim \tilde{p}_t$ at every round, and receives a bandit feedback $\ell_t(I_t)$. This may be used in order to construct an estimate of the true (squared) loss as follows,

$$\tilde{\ell}_t^2(i) := \frac{\ell_t^2(i)}{\tilde{p}_t(i)} \cdot \mathbb{1}_{I_t = i} \ ,$$

and it is immediate to validate that the above is unbiased in the following sense,

$$\mathbb{E}[\tilde{\ell}_t^2(i) | \tilde{p}_t, \ell_t] = \ell_t^2(i), \quad \forall i \in [n].$$

Analogously to the previous section it is natural to define modified cost functions as

$$\tilde{f}_t(p) = \sum_{i=1}^{n} \tilde{\ell}_t^2(i)/p(i) \ .$$

Clearly, $\tilde{f}_t$ is an unbiased estimate of the true cost, $\mathbb{E}[\tilde{f}_t(p) | \tilde{p}_t, \ell_t] = f_t(p)$. From now on we omit the conditioning on $\tilde{p}_t, \ell_t$ for notational brevity. Having devised an unbiased estimate, we could return

Variance Reducer Bandit (VRB)

Input: θ, L, n
Initialize $w(i) = 0$ for all $i \in [n]$.
for $t = 1, \dots, T$ **do**
  $p_t(i) \propto \sqrt{w(i) + L \cdot n/\theta}$
  $\tilde{p}_t(i) = (1-\theta) \cdot p_t(i) + \theta/n$, for all $i \in [n]$
  Draw $I_t \sim \tilde{p}_t$ and play I_t.
  Receive feedback $l_t(I_t)$, and update $w(I_t) \leftarrow w(I_t) + l_t^2(I_t)/\tilde{p}_t(I_t)$.
end for

Figure 3: Variance Reducer Bandit

to the full information analysis of FTRL with the modified losses. However, this poses a difficulty, since the modified losses can possibly be unbounded. We remedy this by mixing the FTRL output, p_t, with a uniform distribution. Mixing encourages exploration, and in turn gives a handle on the possibly unbounded modified losses. Let $\theta \in [0,1]$, and define,

$$\tilde{p}_t(i) = (1-\theta) \cdot p_t(i) + \theta/n.$$

Since $\tilde{p}_t(i) \geq \theta/n$, we have $\tilde{\ell}_t^2(i) \leq nL/\theta$. We summarize the resulting method[4] in Figure 3.

We start with analyzing the *pseudo-regret* of our algorithm, where we compare the cost incurred by the algorithm to the cost incurred by the optimal distribution *in expectation*. The pseudo-regret is defined below,

$$\frac{1}{n^2} \min_{p \in \Delta} \mathbb{E}\left[\sum_{t=1}^{T} f_t(\tilde{p}_t) - \sum_{t=1}^{T} f_t(p)\right], \tag{6}$$

where the expectation is taken with respect to both the player's choices and the loss realizations. The pseudo-regret is only a lower bound for the *expected regret*, with an equality when the adversary is oblivious, i.e., does not take the past choices of the player into account.

Theorem 7 *Let $\theta = (n/T)^{1/3}$. Assuming $T \geq n$, the algorithm in Figure 3 ensures the following bound:*

$$\frac{1}{n^2} \min_{p \in \Delta} \mathbb{E}\left[\sum_{t=1}^{T} f_t(\tilde{p}_t) - \sum_{t=1}^{T} f_t(p)\right] \leq 74 L n^{\frac{1}{3}} T^{\frac{2}{3}}.$$

Proof sketch of Theorem 7 Using the unbiasedness of the modified costs we have

$$\min_{p \in \Delta} \mathbb{E}\left[\sum_{t=1}^{T} f_t(\tilde{p}_t) - \sum_{t=1}^{T} f_t(p)\right] = \min_{p \in \Delta} \mathbb{E}\left[\sum_{t=1}^{T} \tilde{f}_t(\tilde{p}_t) - \sum_{t=1}^{T} \tilde{f}_t(p)\right].$$

4. The sampling and update in the presented form have a complexity of $\mathcal{O}(n)$. There is a standard way to improve this issue based on segment trees that gives $\mathcal{O}(\log n)$ for sampling and update. A detailed description of this idea can be found in section A.4. of Salehi et al. (2017)

We can decompose $\frac{1}{n^2}\min_{p\in\Delta}\mathbb{E}\left[\sum_{t=1}^T \tilde{f}_t(\tilde{p}_t) - \sum_{t=1}^T \tilde{f}_t(p)\right]$ into the following terms:

$$\underbrace{\frac{1}{n^2}\mathbb{E}\left[\sum_{t=1}^T \tilde{f}_t(\tilde{p}_t) - \sum_{t=1}^T \tilde{f}_t(p_t)\right]}_{\text{(A)}} + \underbrace{\frac{1}{n^2}\min_{p\in\Delta}\mathbb{E}\left[\sum_{t=1}^T \tilde{f}_t(p_t) - \sum_{t=1}^T \tilde{f}_t(p)\right]}_{\text{(B)}}$$

where (A) is the cost we incur by mixing, and (B) is upper bounded by the regret of playing FTRL with the modified losses. Now we inspect each term separately.

An upper bound of θLT on (A) results from the simple observation that $1/\tilde{p}_t(i) - 1/p_t(i) \le n\theta$. For bounding (B), notice that p_t is performing FTRL over the modified cost sequence. Combining this together the bound $\tilde{\ell}_t^2(i) \le nL/\theta$ allows us to apply Theorem 3 and get,

$$\frac{1}{n^2}\left(\sum_{t=1}^T \tilde{f}_t(p_t) - \min_{p\in\Delta}\sum_{t=1}^T \tilde{f}_t(p)\right) \le 27\sqrt{\frac{L}{n\theta}}\left(\sum_{i=1}^n \sqrt{\tilde{\ell}_{1:T}^2(i)}\right) + \frac{44nL}{\theta}\,. \tag{7}$$

Due to Jensen's inequality we have $\mathbb{E}\left[\sum_{i=1}^n \sqrt{\tilde{\ell}_{1:T}^2(i)}\right] \le \sum_{i=1}^n \sqrt{\mathbb{E}\left[\tilde{\ell}_{1:T}^2(i)\right]} = \sum_{i=1}^n \sqrt{\ell_{1:T}^2(i)}$. Finally, we get an upper bound on the pseudo-regret which we can optimize in terms of θ:

$$\frac{1}{n^2}\min_{p\in\Delta}\mathbb{E}\left[\sum_{t=1}^T f_t(\tilde{p}_t) - \sum_{t=1}^T f_t(p)\right] \le \theta LT + 27\sqrt{\frac{L}{n\theta}}\left(\sum_{i=1}^n \sqrt{\ell_{1:T}^2(i)}\right) + \frac{44nL}{\theta}.$$

Using the bound $\sum_{i=1}^n \sqrt{\ell_{1:T}^2(i)} \le n\sqrt{LT}$ and the assumption $T \ge n$, we set $\theta = (n/T)^{1/3}$ to get the result. Note that θ is dependent on knowing T in advance. If we do not assume that this is possible, we can use the "doubling trick" starting from $T = n$, and incur an additional constant multiplier in the regret. ■

Ultimately, we are interested in the *expected regret*, where the adversary is allowed to make decisions by taking into account the player's past choices, i.e., to be *non-oblivious*. We present the main result of this paper, which establishes a $\tilde{\mathcal{O}}(n^{1/3}T^{2/3})$ regret bound, where the $\tilde{\mathcal{O}}$ notation hides the logarithmic factors.

Theorem 8 *Assuming $T \ge n$, the following holds for the expected regret:*

$$\frac{1}{n^2}\mathbb{E}\left[\sum_{t=1}^T f_t(\tilde{p}_t) - \min_{p\in\Delta}\sum_{t=1}^T f_t(p)\right] \le \tilde{\mathcal{O}}\left(Ln^{\frac{1}{3}}T^{\frac{2}{3}}\right).$$

Proof sketch of Theorem 8 Using the unbiasedness of the modified costs allows to decompose the regret as follows,

$$\begin{aligned} n^2\mathbb{E}\left[\mathrm{Regret}_T\right] &= \mathbb{E}\left[\sum_{t=1}^T \tilde{f}_t(\tilde{p}_t) - \min_{p\in\Delta}\sum_{t=1}^T \tilde{f}_t(p)\right] + \mathbb{E}\left[\min_{p\in\Delta}\sum_{t=1}^T \tilde{f}_t(p) - \min_{p\in\Delta}\sum_{t=1}^T f_t(p)\right] \\ &\le n^2\mathcal{O}(Ln^{1/3}T^{2/3}) + \mathbb{E}\left[\underbrace{\left(\sum_{i=1}^n \sqrt{\tilde{\ell}_{1:T}^2(i)}\right)^2 - \left(\sum_{i=1}^n \sqrt{\ell_{1:T}^2(i)}\right)^2}_{\text{(A)}}\right], \end{aligned} \tag{8}$$

where the last line uses Equation (7) together with Jensen's inequality (similarly to the proof of Theorem 7). We have also used the closed form solution for the minimal values of $\sum_t f_t(p)$ and $\sum_t \tilde{f}_t(p)$ over the simplex.

Our approach to bounding the remaining term is to establish high probability bound for (A). In order to do so we shall bound the following differences $\tilde{\ell}^2_{1:T}(i) - \ell^2_{1:T}(i)$. This can be done by applying the appropriate concentration results described below.

Bounding $\tilde{\ell}^2_{1:T}(i) - \ell^2_{1:T}(i)$. Fix $i \in [n]$ and define $Z_{t,i} := \tilde{\ell}^2_t(i) - \ell^2_t(i)$. Recalling that $\mathbb{E}[\tilde{\ell}^2_t(i)|\tilde{p}_t, \ell_t] = \ell^2_t(i)$, we have that $\{Z_{t,i}\}_{t\in[T]}$ is a martingale difference sequence with respect to the filtration $\{\mathcal{F}_t\}_{t\in[T]}$ associated with the history of the strategy. This allows us to apply a version of Freedman's inequality (Freedman, 1975), which bounds the sum of differences with respect to their cumulative conditional variance. Loosely speaking, Freedman's inequality implies that w.p. $\geq 1-\delta$,

$$\tilde{\ell}^2_{1:T}(i) - \ell^2_{1:T}(i) \leq \tilde{\mathcal{O}}\left(\sqrt{\sum_{t=1}^T \text{Var}(Z_{t,i}|\mathcal{F}_{t-1})}\right).$$

Importantly, the sum of conditional variances can be related to the regret. Indeed let p^* be the best distribution in hindsight, i.e., $p^* = \arg\min \sum_{t=1}^T f_t(p)$, and define

$$n^2\text{Regret}_T(i) = \sum_{t=1}^T \frac{\ell^2_t(i)}{\tilde{p}_t(i)} - \sum_{t=1}^T \frac{\ell^2_t(i)}{p^*(i)}$$

Then the following can be shown,

$$\sum_{t=1}^T \text{Var}(Z_{t,i}|\mathcal{F}_{t-1}) = \tilde{\mathcal{O}}\left(n^2 L \cdot \text{Regret}_T(i) + \frac{\ell^2_{1:T}(i)}{p^*(i)}\right).$$

To simplify the proof sketch, ignore the second term. Plugging this back into Freedman's inequality we get,

$$\tilde{\ell}^2_{1:T}(i) - \ell^2_{1:T}(i) \leq \tilde{\mathcal{O}}\left(\sqrt{n^2 L \cdot \text{Regret}_T(i)}\right). \tag{9}$$

Final bound. Combining the above with the definition of (A) one can to show that w.p. $\geq 1-\delta$,

$$(\text{A}) \leq \tilde{\mathcal{O}}\left(n\sqrt{LT}\sum_{i=1}^n \left(n^2 L \cdot \text{Regret}_T(i)\right)^{\frac{1}{4}}\right).$$

Since (A) is bounded by $\text{poly}(n,T)$, we can take a small enough $\delta = 1/\text{poly}(n,T)$ such that,

$$\begin{aligned}
\mathbb{E}\left[(\text{A})\right] &\leq \tilde{\mathcal{O}}\left(n^{3/2}L^{3/4}T^{1/2} \cdot \mathbb{E}\left[\sum_{i=1}^n \left(\text{Regret}_T(i)\right)^{1/4}\right]\right) \\
&\leq \tilde{\mathcal{O}}\left(n^{3/2}L^{3/4}T^{1/2} \cdot \sum_{i=1}^n \left(\mathbb{E}\left[\text{Regret}_T(i)\right]\right)^{1/4}\right) \\
&\leq \tilde{\mathcal{O}}\left(n^{9/4}L^{3/4}T^{1/2} \cdot \left(\mathbb{E}\left[\text{Regret}_T\right]\right)^{1/4}\right)
\end{aligned}$$

where the second line uses Jensen's inequality with respect to the concave function $h(u) = u^{1/4}$, and the last line uses $\sum_{i=1}^n \text{Regret}_T(i) = \text{Regret}_T$ together with $\sum_{i=1}^n x_i^{1/4} \le n^{3/4} \left(\sum_{i=1}^n x_i\right)^{1/4}$, which is also a consequence of Jensen's inequality since $\frac{1}{n}\sum_{i=1}^n x_i^{1/4} \le \left(\frac{1}{n}\sum_{i=1}^n\right)^{1/4}$. Plugging the above bound back into Eq. (8) we are able to establish the proof. The full proof is deferred to Appendix E. Note that in the full proof we do not explicitly relate the conditional variances to the regret, but this is rather more implicit in the analysis. ■

Experiments. We also validate our method empirically on the tasks of classifying images with logistic regression and mini-batch k-Means. A detailed description of the experiments can be found in Appendix F. In both cases we observe that our method (VRB) produces significant gains compared to uniform sampling and compares favorably to other variance reduction methods of similar nature (Salehi et al., 2017; Namkoong et al., 2017).

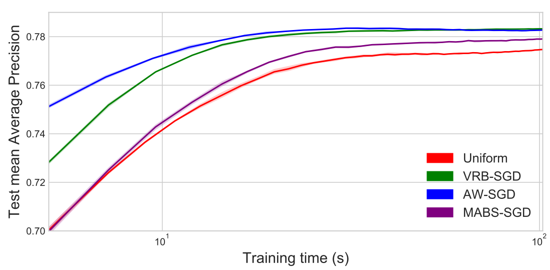

Figure 4: Mean Average Precision scores achieved on the test part of VOC 2007.

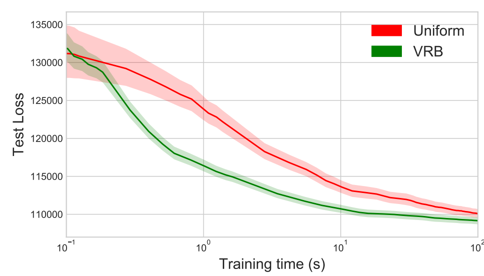

Figure 5: k-Means test loss evolution on the CSN dataset.

5. Conclusion and Future Work

We presented a novel importance sampling technique for variance reduction in an online learning formulation. First, we motivated why regret is a sensible measure of performance in this setting. Despite the bandit feedback and the unbounded costs, we provided an expected regret guarantee of $\tilde{\mathcal{O}}(n^{1/3}T^{2/3})$, where our reference is the best fixed sampling distribution in hindsight. We confirmed the theoretical findings with empirical validation.

Among the many possible future directions stands the question of the tightness of the expected regret bound of the algorithm. Another naturally arising idea is the theoretical analysis of the method when employed in conjunction with advanced stochastic solvers such as SVRG and SAGA.

Acknowledgments

The authors would like to thank Hasheminezhad Seyedrouzbeh for useful discussions during the course of this work. This research was supported by SNSF grant 407540_167212 through the NRP 75 Big Data program. K.Y.L. is supported by the ETH Zürich Postdoctoral Fellowship and Marie Curie Actions for People COFUND program.

References

Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *COLT*, pages 263–274, 2008.

Guillaume Alain, Alex Lamb, Chinnadhurai Sankar, Aaron Courville, and Yoshua Bengio. Variance reduction in sgd by distributed importance sampling. *arXiv preprint arXiv:1511.06481*, 2015.

Zeyuan Allen-Zhu, Zheng Qu, Peter Richtárik, and Yang Yuan. Even faster accelerated coordinate descent using non-uniform sampling. In *International Conference on Machine Learning*, pages 1110–1119, 2016.

David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

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

Leon Bottou and Yoshua Bengio. Convergence properties of the k-means algorithms. In *Advances in neural information processing systems*, pages 585–592, 1995.

Guillaume Bouchard, Théo Trouillon, Julien Perez, and Adrien Gaidon. Online learning to sample. *arXiv preprint arXiv:1506.09016*, 2015.

Nicolo Cesa-Bianchi, Alex Conconi, and Claudio Gentile. On the generalization ability of on-line learning algorithms. *IEEE Transactions on Information Theory*, 50(9):2050–2057, 2004.

Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

Dominik Csiba and Peter Richtárik. Importance sampling for minibatches. *arXiv preprint arXiv:1602.02283*, 2016.

Aaron Defazio, Francis Bach, and Simon Lacoste-Julien. Saga: A fast incremental gradient method with support for non-strongly convex composite objectives. In *Advances in Neural Information Processing Systems*, pages 1646–1654, 2014.

John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159, 2011.

Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2): 303–338, 2010.

Matthew Faulkner, Michael Olson, Rishi Chandy, Jonathan Krause, K Mani Chandy, and Andreas Krause. The next big one: Detecting earthquakes and other rare events from community-based sensors. In *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on*, pages 13–24. IEEE, 2011.

David A Freedman. On tail probabilities for martingales. *the Annals of Probability*, pages 100–118, 1975.

Elad Hazan. A survey: The convex optimization approach to regret minimization. *Optimization for machine learning*, pages 287–302, 2011.

Rie Johnson and Tong Zhang. Accelerating stochastic gradient descent using predictive variance reduction. In *Advances in neural information processing systems*, pages 315–323, 2013.

Sham M Kakade and Ambuj Tewari. On the generalization ability of online strongly convex programming algorithms. In *Advances in Neural Information Processing Systems*, pages 801–808, 2009.

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

KDD Cup 2004. KDD Cup 2004. Protein Homology Dataset. http://osmot.cs.cornell.edu/kddcup/, 2004. Accessed: 10.11.2016.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

H Brendan McMahan and Matthew Streeter. Adaptive bound optimization for online convex optimization. *COLT 2010*, page 244, 2010.

Hongseok Namkoong, Aman Sinha, Steve Yadlowsky, and John C. Duchi. Adaptive sampling probabilities for non-smooth optimization. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2574–2583, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.

I Necoara, Y Nesterov, and F Glineur. A random coordinate descent method on large optimization problems with linear constraints. 2011.

Deanna Needell, Rachel Ward, and Nati Srebro. Stochastic gradient descent, weighted sampling, and the randomized kaczmarz algorithm. In *Advances in Neural Information Processing Systems*, pages 1017–1025, 2014.

Yu Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization*, 22(2):341–362, 2012.

Dmytro Perekrestenko, Volkan Cevher, and Martin Jaggi. Faster coordinate descent via adaptive importance sampling. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54. PMLR, 2017.

F. Salehi, L. E. Celis, and P. Thiran. Stochastic Optimization with Bandit Sampling. *ArXiv e-prints*, August 2017.

Farnood Salehi, Patrick Thiran, and L Elisa Celis. Stochastic dual coordinate descent with bandit sampling. *arXiv preprint arXiv:1712.03010*, 2017.

David Sculley. Web-scale k-means clustering. In *Proceedings of the 19th international conference on World wide web*, pages 1177–1178. ACM, 2010.

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

Abhinav Shrivastava, Abhinav Gupta, and Ross Girshick. Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 761–769, 2016.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *ICLR*, 2015.

Sebastian U Stich, Anant Raj, and Martin Jaggi. Safe adaptive importance sampling. In *Advances in Neural Information Processing Systems 30*, pages 4384–4394. Curran Associates, Inc., 2017.

Peilin Zhao and Tong Zhang. Stochastic optimization with importance sampling for regularized loss minimization. In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pages 1–9, 2015.

Appendix A. Cumulative Variance of the Gradients and Quality of Optimization

The relationship between cumulative second moment of the gradients and quality of optimization has been demonstrated in several works. Since the difference between the second moment and the variance is independent of the sampling distribution p_t, the guarantees of our method also translate to guarantees with respect to the cumulative second moments of the gradient estimates. Here we provide two concrete references.

For the following, assume that we would like to minimize a convex objective,

$$\min_{w \in \mathcal{W}} F(w) := \mathbb{E}_{z \sim \mathcal{D}}[f(w; z)]$$

and we assume that we are able to draw i.i.d. samples from the unknown distribution $\mathcal{D}$. Thus, given a point $w \in \mathcal{W}$ we are able to design an unbiased estimate for $\nabla F(w)$ by sampling $z \sim \mathcal{D}$ and taking $g := \nabla f(w; z)$ (clearly, $\mathbb{E}[g|w] = \nabla F(w)$). Now assume a gradient-based update rule, i.e.,

$$w_{t+1} = \Pi_{\mathcal{W}}(w_t - \eta_t g_t), \quad \text{where} \quad \mathbb{E}[g_t|w_t] = \nabla F(w_t) \tag{10}$$

and $\Pi_{\mathcal{W}}(u) := \arg\min_{w \in \mathcal{W}} \|u - w\|$. Next we show that for two very popular gradient based-methods — AdaGrad and SGD for strongly-convex functions, the performance is directly related to the cumulative second moment of the gradient estimates, $\sum_{t=1}^T \mathbb{E}\|g_t\|^2$. The latter is exactly the objective of our online variance reduction method.

The AdaGrad algorithm employs the same rule as in Eq. (10) using $\eta_t = D/\sqrt{2\sum_{\tau=1}^t \|g_t\|^2}$. The next theorem substantiates its guarantees.

Theorem 9 (Duchi et al. (2011)) *Assume that the diameter of $\mathcal{W}$ is bounded by D. Then:*

$$\mathbb{E}\left[F\left(\frac{1}{T}\sum_{t=1}^T w_t\right)\right] - \min_{w \in \mathcal{W}} F(w) \leq \frac{2D}{T}\sqrt{\sum_{t=1}^T \mathbb{E}\|g_t\|^2}$$

The SGD algorithm for μ-strongly-convex objectives employs the same rule as in Eq. (10) using $\eta_t = \frac{2}{\mu t}$. The next theorem substantiates its guarantees.

Theorem 10 (Salehi et al. (2017)) *Assume that F is μ-strongly convex, then:*

$$\mathbb{E}\left[F\left(\frac{2}{T(T+1)}\sum_{t=1}^T t \cdot w_t\right)\right] - \min_{w \in \mathcal{W}} F(w) \leq \frac{2}{\mu T(T+1)}\sum_{t=1}^T \mathbb{E}\|g_t\|^2$$

Appendix B. Proof of Lemma 1

Proof Denote $\ell_{1:t}^2(i) = \sum_{\tau=1}^{t} \ell_\tau^2(i)$. Next, we bound the cumulative loss per point $i \in [n]$,

$$
\begin{aligned}
\ell_{1:T}^2(i) = \sum_{t=1}^{T} \ell_t^2(i) &= \sum_{t=1}^{T} \left(\ell_*(i) + (\ell_t(i) - \ell_*(i))\right)^2 \\
&\leq T \cdot \ell_*^2(i) + 2\ell_*(i) \sum_{t=1}^{T} |\ell_t(i) - \ell_*(i)| + \sum_{t=1}^{T} (\ell_t(i) - \ell_*(i))^2 \\
&\leq T \cdot \ell_*^2(i) + 2\ell_*(i)\sqrt{T \cdot V_T(i)} + V_T(i) \\
&= T\left(\ell_*(i) + \sqrt{\frac{V_T(i)}{T}}\right)^2 \qquad (11)
\end{aligned}
$$

where the second line uses $\ell_*(i) \geq 0$ and the third line uses the definition of $V_T(i)$ together with the inequality $\|u\|_1 \leq \sqrt{T}\|u\|_2, \ \forall u \in \mathbb{R}^T$.

We require the following lemma:

Lemma 11 *Let $a_1, \ldots, a_n \geq 0$. Then the following holds,*

$$
\min_{p \in \Delta} \sum_{i=1}^{n} \frac{a_i}{p(i)} = \left(\sum_{i=1}^{n} \sqrt{a_i}\right)^2 .
$$

The proof of the lemma is analogous to the proof of Lemma 2, which is given in the next section. Notice that according to this lemma and using the non-negativity of losses we have,

$$
\frac{1}{n^2} \min_{p \in \Delta} \sum_{i=1}^{n} \frac{\ell_t^2(i)}{p(i)} = \left(\frac{1}{n}\sum_{i=1}^{n} \ell_t(i)\right)^2 := L^2(w_t) . \qquad (12)
$$

We are now ready to bound the value of best fixed point in hindsight,

$$
\begin{aligned}
\min_p \frac{1}{n^2} \sum_{t=1}^{T} \sum_{i=1}^{n} \frac{\ell_t^2(i)}{p(i)} &= \min_p \frac{1}{n^2} \sum_{i=1}^{n} \frac{\ell_{1:T}^2(i)}{p(i)} \\
&= \frac{1}{n^2} \left(\sum_{i=1}^{n} \sqrt{\ell_{1:t}^2(i)}\right)^2 \\
&= T\left(\frac{1}{n}\sum_{i=1}^{n} \ell_*(i) + \frac{1}{n}\sum_{i=1}^{n} \sqrt{\frac{V_T(i)}{T}}\right)^2 \\
&= T \cdot L_*^2 + 2\sqrt{T} L_* \cdot \frac{1}{n}\sum_{i=1}^{n} \sqrt{V_T(i)} + \left(\frac{1}{n}\sum_{i=1}^{n} \sqrt{V_T(i)}\right)^2 ,
\end{aligned}
$$

where in the second line we use Lemma 11, and the third line uses Eq. (11).

We are now left to prove that $T \cdot L_*^2 \le \sum_{t=1}^T \frac{1}{n^2} \min_{p\in\Delta} \sum_{i=1}^n \ell_t^2(i)/p(i)$. Indeed,

$$\begin{aligned} L_*^2 &\le \left(\frac{1}{T}\sum_{t=1}^T L(w_t)\right)^2 \\ &\le \frac{1}{T}\sum_{t=1}^T L^2(w_t) \\ &= \frac{1}{T}\sum_{t=1}^T \frac{1}{n^2}\min_{p\in\Delta}\sum_{i=1}^n \ell_t^2(i)/p(i) \;. \end{aligned}$$

where the first line uses the asuumption about the average optimality of L_*, the second line uses Jensen's inequality, and the last line uses Eq. (12). This concludes the proof. ■

Appendix C. Proofs for the Full Information Setting

C.1. Proof of Lemma 2

Proof We formulate the Lagrangian of the optimization problem in Equation (3):

$$\begin{aligned} \underset{p}{\text{minimize}} \quad & \sum_{i=1}^n \frac{\ell_{1:t-1}^2(i)}{p(i)} + \gamma \sum_{i=1}^n \frac{1}{p(i)} \\ \text{subject to} \quad & \sum_{i=1}^n p(i) = 1 \\ & p(i) \ge 0,\ i = 1, \ldots, n \end{aligned}$$

and get:

$$\mathcal{L}(p, \lambda) = \sum_{i=1}^n \frac{\ell_{1:t-1}^2(i)}{p(i)} + \gamma \sum_{i=1}^n \frac{1}{p(i)} + \alpha \cdot \left(\sum_{i=1}^n p(i) - 1\right) - \sum_{i=1}^n \beta_i \cdot p(i)$$

From setting $\frac{\partial \mathcal{L}(p,\lambda)}{\partial p(i)} = 0$ we have:

$$p(i) = \frac{\sqrt{\ell_{1:t-1}^2(i) + \gamma}}{\sqrt{\alpha - \beta_i}} \tag{13}$$

Note that setting $p(i) = 0$ implies an objective value of infinity due to the regularizer. Thus, at the optimum $p(i) > 0,\ \forall i \in [n]$; which in turn implies that $\beta_i = 0,\ \forall i \in [n]$ (due to complementary slackness). Combining this with $\sum_{i=1}^n p(i) = 1$, we get $\sqrt{\alpha} = \sum_{i=1}^n \sqrt{\ell_{1:t-1}^2(i) + \gamma}$ which gives:

$$p(i) = \frac{\sqrt{\ell_{1:t-1}^2(i) + \gamma}}{\sum_{j=1}^n \sqrt{\ell_{1:t-1}^2(j) + \gamma}} \tag{14}$$

Since the minimization problem is convex for $p \in \Delta$, we obtained a global minimum. ■

C.2. Proof of Lemma 4

Proof The regret of FTRL may be related to the stability of the online decision sequence as shown in the following lemma due to Kalai and Vempala (2005) (proof can be found in Hazan (2011) or in Shalev-Shwartz et al. (2012)):

Lemma 12 *Let $\mathcal{K}$ be a convex set and $\mathcal{R} : \mathcal{K} \mapsto \mathbb{R}$ be a regularizer. Given a sequence of cost functions $\{f_t\}_{t\in[T]}$ defined over $\mathcal{K}$, then setting $p_t = \arg\min_{p\in\Delta} \sum_{\tau=1}^{t-1} f_\tau(p) + \mathcal{R}(p)$ ensures*

$$\sum_{t=1}^T f_t(p_t) - \sum_{t=1}^T f_t(p) \le \sum_{t=1}^T (f_t(p_t) - f_t(p_{t+1})) + (\mathcal{R}(p) - \mathcal{R}(p_1)), \quad \forall p \in \mathcal{K}.$$

Notice that our regularizer $\mathcal{R}(p) = L\sum_{i=1}^n 1/p(i)$ is non-negative and bounded by $nL/p_{\min}$ over Δ'. Thus, applying the above lemma to the FTRL rule of Eq. (3) implies that $\forall p \in \Delta'$,

$$\sum_{t=1}^T f_t(p_t) - \sum_{t=1}^T f_t(p) \le \sum_{t=1}^T (f_t(p_t) - f_t(p_{t+1})) + \frac{nL}{p_{\min}} . \tag{15}$$

We are left to bound the remaining term. Let us first recall the closed from solution for the p_t's as stated in Lemma 2,

$$p_t(i) = \frac{\sqrt{\ell_{1:t-1}^2(i) + L}}{c_t} ,$$

where $c_t = \sum_{i=1}^n \sqrt{\ell_{1:t-1}^2(i) + L}$ is the normalization factor. Noticing that $\{c_t\}_{t\in[T]}$ is a non-decreasing sequence we, are now ready to bound the remaining term,

$$\begin{aligned}
\sum_{t=1}^T (f_t(p_t) - f_t(p_{t+1})) &= \sum_{t=1}^T \sum_{i=1}^n \ell_t^2(i) \cdot \left(\frac{c_t}{\sqrt{\ell_{1:t-1}^2(i) + L}} - \frac{c_{t+1}}{\sqrt{\ell_{1:t}^2(i) + L}} \right) \\
&\le \sum_{t=1}^T \sum_{i=1}^n \ell_t^2(i) \cdot \left(\frac{c_t}{\sqrt{\ell_{1:t-1}^2(i) + L}} - \frac{c_t}{\sqrt{\ell_{1:t}^2(i) + L}} \right) \\
&= \sum_{t=1}^T \sum_{i=1}^n \frac{\ell_t^2(i) \cdot c_t}{\sqrt{\ell_{1:t}^2(i) + L}} \cdot \left(\sqrt{1 + \frac{\ell_t^2(i)}{\ell_{1:t-1}^2(i) + L}} - 1 \right) \\
&\le \frac{c_T}{2} \sum_{t=1}^T \sum_{i=1}^n \frac{\ell_t^4(i)}{\sqrt{\ell_{1:t}^2(i) + L} \cdot (\ell_{1:t-1}^2(i) + L)}
\end{aligned}$$

where in the first inequality we used the fact that $c_t \le c_{t+1}$ and in the last inequality we relied on the fact that $\sqrt{1+x} \le 1 + \frac{x}{2}$ for all $x \ge 0$. Furthermore, we observe that $\sqrt{\ell_{1:t}^2(i) + L} \ge \sqrt{\ell_{1:t}^2(i)}$ and $\ell_{1:t-1}^2(i) + L \ge \ell_{1:t}^2(i)$ in order to get:

$$\sum_{t=1}^T (f_t(p_t) - f_t(p_{t+1})) \le \frac{c_T}{2} \cdot \sum_{t=1}^T \sum_{i=1}^n \frac{\ell_t^4(i)}{\left(\ell_{1:t}^2(i)\right)^{3/2}} = \sqrt{L} \cdot \frac{c_T}{2} \cdot \sum_{i=1}^n \sum_{t=1}^T \frac{\frac{\ell_t^4(i)}{L^2}}{\left(\frac{\ell_{1:t}^2(i)}{L}\right)^{3/2}}$$

For a fixed index i, denote $a_t := \ell_t(i)/\sqrt{L}$ and note that $a_t \in [0,1], \ \forall t \in [T]$. The innermost sum can be therefore written as $\sum_{t=1}^T \frac{a_t^4}{(a_{1:t}^2)^{3/2}}$, which is upper bounded by 44 as stated in lemma below.

Lemma 13 *For any sequence of numbers $a_1, \dots, a_T \in [0,1]$ the following holds:*

$$\sum_{t=1}^{T} \frac{a_t^4}{(a_{1:t}^2)^{3/2}} \leq 44 \, .$$

The proof of the lemma is provided in section C.3. As a consequence,

$$\begin{aligned} \sum_{t=1}^{T} (f_t(p_t) - f_t(p_{t+1})) &\leq \sqrt{L} \cdot \frac{c_T}{2} \cdot \sum_{i=1}^{n} \sum_{t=1}^{T} \frac{\frac{\ell_t^4(i)}{L^2}}{\left(\frac{\ell_{1:t}^2(i)}{L}\right)^{3/2}} \\ &\leq 22n\sqrt{L} \cdot \sum_{i=1}^{n} \sqrt{\ell_{1:T-1}^2(i) + L} \, , \end{aligned} \tag{16}$$

where we have used the expression for c_T.

We get our final result once we plug Equation (16) into Equation (15) and observe that $\sqrt{\ell_{1:T-1}^2(i) + L} \leq \sqrt{\ell_{1:T}^2(i)} + \sqrt{L}$. ■

C.3. Proof of Lemma 13

Proof Without loss of generality assume that $a_1 > 0$ (otherwise we can always start the analysis from the first t such that $a_t > 0$). Let us define the following index sets,

$$\begin{aligned} P_k &= \{t \in [T] : 4^{k-1} a_1^2 < a_{1:t}^2 \leq 4^k a_1^2\}, & \forall k \in \{1, 2, \dots \lceil \log_2(1/a_1) \rceil\} \\ Q_k &= \{t \in [T] : k < a_{1:t}^2 \leq k+1\}, & \forall k \in \{1, 2, \dots\} \end{aligned}$$

The definitions of P_k implies,

$$\sum_{t \in P_k} a_t^4 \leq \left(\sum_{t \in P_k} a_t^2 \right)^2 \leq 4^{2k} a_1^4 \tag{17}$$

The definition of Q_k implies,

$$\sum_{t \in Q_k} a_t^4 \leq \left(\sum_{t \in Q_k} a_t^2 \right)^2 \leq 2^2 = 4 \tag{18}$$

where the second inequality uses $\sum_{t \in Q_k} a_t^2 \leq 2$ which follows from the fact that if a set Q_k is non-empty then so is Q_{k-1} (since $a_t \in [0,1]$), and thus,

$$\begin{aligned} \sum_{t \in Q_k} a_t^2 &= \sum_{t=1}^{T_k} a_t^2 - \sum_{t=1}^{T_{k-1}} a_t^2 \\ &\leq (k+1) - (k-1) \\ &= 2 \, . \end{aligned}$$

where we have defined $T_k := \max\{t \in [T] : t \in Q_k\}$.

Using the definitions of P_k and Q_k together with Equations (17), (18), we get,

$$
\begin{aligned}
\sum_{t=1}^{T} \frac{a_t^4}{(a_{1:t}^2)^{3/2}} &\le a_1 + \sum_{k=1}^{\lceil \log_2(1/a_1) \rceil} \sum_{t \in P_k} \frac{a_t^4}{(a_{1:t}^2)^{3/2}} + \sum_{k=1}^{\infty} \sum_{t \in Q_k} \frac{a_t^4}{(a_{1:t}^2)^{3/2}} \\
&\le a_1 + \sum_{k=1}^{\lceil \log_2(1/a_1) \rceil} \frac{\sum_{t \in P_k} a_t^4}{4^{3(k-1)/2} a_1^3} + \sum_{k=1}^{\infty} \frac{\sum_{t \in Q_k} a_t^4}{k^{3/2}} \\
&\le a_1 + \sum_{k=1}^{\lceil \log_2(1/a_1) \rceil} \frac{4^{2k} a_1^4}{4^{3(k-1)/2} a_1^3} + \sum_{k=1}^{\infty} \frac{4}{k^{3/2}} \\
&\le a_1 \cdot \left(1 + \sum_{k=1}^{\lceil \log_2(1/a_1) \rceil} \frac{4^{2k}}{4^{3(k-1)/2}} \right) + \sum_{k=1}^{\infty} \frac{4}{k^{3/2}} \\
&\le a_1 \cdot \sum_{k=0}^{\lceil \log_2(1/a_1) \rceil} 2^{k+3} + 4 \cdot \sum_{k=1}^{\infty} \frac{1}{k^{3/2}} \\
&\le 16 a_1 \cdot 2^{\lceil \log_2(1/a_1) \rceil} + 4 + 4 \cdot \sum_{k=2}^{\infty} \frac{1}{k^{3/2}} \\
&\le 36 + 4 \cdot \sum_{k=2}^{\infty} \frac{1}{k^{3/2}} \\
&\le 36 + 4 \cdot \int_{x=1}^{\infty} \frac{1}{x^{3/2}} dx \\
&\le 44
\end{aligned}
$$

which concludes the proof. ■

C.4. Proof of Lemma 6

Proof We first look at the loss of the best distribution in hindsight:

$$
\begin{aligned}
\underset{p}{\text{minimize}} \quad & \sum_{t=1}^{T} \sum_{i=1}^{n} \frac{\ell_t^2(i)}{p(i)} \\
\text{subject to} \quad & \sum_{i=1}^{n} p(i) = 1 \\
& p(i) \ge 0, \; i = 1, \ldots, n.
\end{aligned}
$$

Analogous reasoning to the proof of Lemma 2 we get $p(i) \propto \sqrt{\ell_{1:T}^2(i)}$ and as a consequence, the loss of the best distribution in hindsight over the unrestricted simplex is:

$$
\min_{p \in \Delta} \sum_{t=1}^{T} \sum_{i=1}^{n} \frac{\ell_t^2(i)}{p(i)} = \left(\sum_{i=1}^{n} \sqrt{\ell_{1:T}^2(i)} \right)^2 \tag{19}
$$

The next step is to solve the optimization problem over the restricted simplex Δ':

$$\begin{aligned} \underset{p}{\text{minimize}} \quad & \sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\ell_t^2(i)}{p(i)} \\ \text{subject to} \quad & \sum_{i=1}^{n} p(i) = 1 \\ & p(i) \geq p_{\text{min}},\ i = 1,\dots,n. \end{aligned}$$

We start our proof similarly to the proof of Proposition 5 of Namkoong et al. (2017). First, we formulate the Lagrangian:

$$\mathcal{L}(p,\lambda,\theta) = \sum_{i=1}^{n}\frac{\ell_{1:T}^2(i)}{p(i)} + \alpha\cdot\left(\sum_{i=1}^{n}p(i)-1\right) - \sum_{i=1}^{n}\beta_i\cdot(p(i)-p_{\text{min}}) \tag{20}$$

Setting $\frac{\partial\mathcal{L}}{\partial p(i)} = 0$ and using complementary slackness we get:

$$p(i) = \frac{\sqrt{\ell_{1:T}^2(i)}}{\sqrt{\alpha-\beta_i}} = \begin{cases} \frac{\sqrt{\ell_{1:T}^2(i)}}{\sqrt{\alpha}} & \text{if } \sqrt{\ell_{1:T}^2(i)} > \sqrt{\alpha}\cdot p_{\text{min}} \\ p_{\text{min}} & \text{else} \end{cases} \tag{21}$$

Next we determine the value of α. Denoting $I = \{i \mid \sqrt{\ell_{1:T}^2(i)} > \sqrt{\alpha}\cdot p_{\text{min}}\}$, and using $\sum_{i=1}^{n} p(i) = 1$ implies,

$$\sum_{i=1}^{n}p(i) = \sum_{i\in I}p(i) + \sum_{i\in I^C}p(i) = \frac{1}{\sqrt{\alpha}}\sum_{i\in I}\sqrt{\ell_{1:T}^2(i)} + (n-|I|)\cdot p_{\text{min}} = 1$$

From this we get,

$$\sqrt{\alpha} = \frac{\sum_{i\in I}\sqrt{\ell_{1:T}^2(i)}}{1-(n-|I|)\cdot p_{\text{min}}}\ . \tag{22}$$

Now we can plug this into the original problem to get the optimal value:

$$
\begin{aligned}
\sum_{i=1}^{n} \frac{\ell_{1:T}^2(i)}{p(i)} &= \sum_{i\in I} \frac{\ell_{1:T}^2(i)}{p(i)} + \sum_{i\in I^C} \frac{\ell_{1:T}^2(i)}{p(i)} \\
&= \sqrt{\alpha}\cdot\left(\sum_{i\in I}\sqrt{\ell_{1:T}^2(i)}\right) + \frac{1}{p_{\text{min}}}\sum_{i\in I^C}\ell_{1:T}^2(i) && \triangleright \text{ Eq. 21, def. of } p(i) \\
&= \alpha\cdot(1-(n-|I|)\cdot p_{\text{min}}) + \frac{1}{p_{\text{min}}}\sum_{i\in I^C}\ell_{1:T}^2(i) && \triangleright \text{ Eq. 22, replacing } \sum_{i\in I}\sqrt{\ell_{1:T}^2(i)} \\
&\le \alpha\cdot(1-(n-|I|)\cdot p_{\text{min}}) + \alpha\cdot p_{\text{min}}\cdot(n-|I|) && \triangleright \text{ Eq. 21, } \ell_{1:T}^2(i)\le \alpha p_{\text{min}}^2, \forall i\in I^C \\
&= \alpha \\
&= \frac{\left(\sum_{i\in I}\sqrt{\ell_{1:T}^2(i)}\right)^2}{(1-(n-|I|)p_{\text{min}})^2} && \triangleright \text{ Eq. 22} \\
&\le \frac{\left(\sum_{i\in I}\sqrt{\ell_{1:T}^2(i)}\right)^2}{(1-n\cdot p_{\text{min}})^2} \\
&\le \frac{\left(\sum_{i=1}^{n}\sqrt{\ell_{1:T}^2(i)}\right)^2}{(1-n\cdot p_{\text{min}})^2}
\end{aligned}
$$

Combining this result with Equation (19) we obtain,

$$
\min_{p\in\Delta'}\sum_{i=1}^{n}\frac{\ell_{1:T}^2(i)}{p(i)} - \min_{p\in\Delta}\sum_{i=1}^{n}\frac{\ell_{1:T}^2(i)}{p(i)} \le \left(\frac{1}{(1-n\cdot p_{\text{min}})^2}-1\right)\cdot\left(\sum_{i=1}^{n}\sqrt{\ell_{1:T}^2(i)}\right)^2
$$

Using the fact that $\frac{1}{(1-x)^2}-1\le 6x$ for $x\in[0,1/2]$, with which we are assuming that $p_{\text{min}}\le 1/(2n)$, we finally get the claim of the lemma. Note that in the sections following this lemma, all choices of p_{min} respect $p_{\text{min}}\le 1/(2n)$. ■

Appendix D. Proofs for the Pseudo-Regret

D.1. Proofs of Theorem 7

Proof What remains from the proof sketch is to bound the term (A), which we do here. Due to the mixing we always have $\tilde{p}_t(i)\ge\theta/n$ for all $t\in[T]$, $i\in[n]$. Moreover $p_t(i)\ge 1/n$ implies $\tilde{p}_t(i)\ge 1/n$. Next we upper bound $1/\tilde{p}_t(i)-1/p_t(i)$. If $p_t(i)\le 1/n$, then the difference is negative, otherwise,

$$
\frac{1}{\tilde{p}_t(i)}-\frac{1}{p_t(i)} = \theta\cdot\frac{p_t(i)-\frac{1}{n}}{\tilde{p}_t(i)p_t(i)} < \theta\cdot\frac{p_t(i)}{\tilde{p}_t(i)p_t(i)} = \frac{\theta}{\tilde{p}_t(i)} \le n\theta.
$$

As an immediate consequence we obtain a bound on (A),

$$
\begin{aligned}
n^2 \cdot (\mathrm{A}) :&= \mathbb{E}\left[\sum_{t=1}^{T} \tilde{f}_t(\tilde{p}_t) - \sum_{t=1}^{T} \tilde{f}_t(p_t)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} \tilde{\ell}_t^2(i)\left(\frac{1}{\tilde{p}_t(i)} - \frac{1}{p_t(i)}\right)\right] \\
&\leq n\theta \cdot \mathbb{E}\left[\sum_{i=1}^{n} \tilde{\ell}_{1:T}^2(i)\right] \\
&\leq n^2 \theta L T \,,
\end{aligned}
$$

where we used $\ell_t^2(i) \leq L$. The rest of the proof is completed in the proof sketch. ■

Appendix E. Proofs for the Expected Regret

Throughout the proofs we assume $n \leq T$.

E.1. Proof of Theorem 8

Proof Using the unbiasedness of the modified costs allows to decompose the regret as follows,

$$
\begin{aligned}
n^2 \mathbb{E}\left[\mathrm{Regret}_T\right] &= \mathbb{E}\left[\sum_{t=1}^{T} f_t(\tilde{p}_t) - \min_{p\in\Delta}\sum_{t=1}^{T} f_t(p)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \tilde{f}_t(\tilde{p}_t) - \min_{p\in\Delta}\sum_{t=1}^{T} \tilde{f}_t(p)\right] + \mathbb{E}\left[\min_{p\in\Delta}\sum_{t=1}^{T} \tilde{f}_t(p) - \min_{p\in\Delta}\sum_{t=1}^{T} f_t(p)\right] \\
&\leq n^2 \mathcal{O}(L n^{1/3} T^{2/3}) + \mathbb{E}\left[\underbrace{\left(\sum_{i=1}^{n}\sqrt{\tilde{\ell}_{1:T}^2(i)}\right)^2 - \left(\sum_{i=1}^{n}\sqrt{\ell_{1:T}^2(i)}\right)^2}_{(\mathrm{A})}\right],
\end{aligned}
\tag{23}
$$

where the last line uses Equation (7) together with Jensen's inequality (similarly to the proof of Theorem 7). We have also used the closed form solution for the minimal values of the cumulative true/modified costs, i.e,

$$
\min_{p\in\Delta}\sum_{t=1}^{T} f_t(p) = \left(\sum_{i=1}^{n}\sqrt{\ell_{1:T}^2(i)}\right)^2 \text{ and } \min_{p\in\Delta}\sum_{t=1}^{T} \tilde{f}_t(p) = \left(\sum_{i=1}^{n}\sqrt{\tilde{\ell}_{1:T}^2(i)}\right)^2,
$$

the above is established in the proof of Lemma 6.

Thus, in order to establish the theorem, we bound the expectation of (A). The high level idea of the proof is to show that for any small enough $\delta \in [0, 1]$ then w.p. $\geq 1 - \delta$ the term (A) is bounded by $n^2 \mathcal{O}(n^{1/3} T^{2/3} \log(nT/\delta))$. Then, by showing that (A) is bounded *almost surely*, we are able to choose a small enough δ such that $\mathbb{E}\left[(A)\right] = n^2\tilde{\mathcal{O}}(Ln^{1/3}T^{2/3})$. Let us first establish a trivial bound on (A),

$$
\begin{aligned}
(\mathrm{A}) &\leq \left(\sum_{i=1}^{n}\sqrt{\tilde{\ell}^2_{1:T}(i)}\right)^2 \\
&\leq \left(\sum_{i=1}^{n}\sqrt{\frac{\ell^2_{1:T}(i)}{\theta/n}}\right)^2 \\
&= Ln^{8/3}T^{4/3}\,,
\end{aligned}
$$

where we used $\ell^2_{1:T}(i) \leq LT$, and $\theta = (n/T)^{1/3}$. Thus, choosing $1/\delta \geq Ln^{8/3}T^{4/3}$ ensures that $\delta \cdot (\mathrm{A}) \leq 1$ with probability 1. It now remains to establish a high probability bound for (A). To do so, we shall bound the differences $\tilde{\ell}^2_{1:t}(i) - \ell^2_{1:t}(i)$ using a version of Freedman's concentration inequality (Freedman, 1975). Later, this will enable us to bound (A). Next we proceed according to these two steps.

Step 1: bounding $\tilde{\ell}^2_{1:t}(i) - \ell^2_{1:t}(i)$.

Fix $i \in [n]$ and define the following sequence $\{Z_{t,i} := \tilde{\ell}^2_t(i) - \ell^2_t(i)\}_{t\in[T]}$. Recalling that $\mathbb{E}[\tilde{\ell}^2_t(i)|\tilde{p}_t, \ell_t] = \ell^2_t(i)$, we have that $\{Z_{t,i}\}_{t\in[T]}$ is a martingale difference sequence with respect to the filtration $\{\mathcal{F}_t\}_{t\in[T]}$ associated with the history of the strategy. Also notice that due to the mixing $|Z_{t,i}| \leq 2|\tilde{\ell}^2_t(i)| \leq 2nL/\theta$. We may bound the conditional variance of the $Z_{t,i}$ as follows,

$$
\begin{aligned}
\mathrm{Var}(Z_{t,i}|\mathcal{F}_{t-1}) &= \mathbb{E}\left[\left(\frac{\ell^2_t(i)}{\tilde{p}_t(i)}\mathbb{1}_{I_t=i} - \ell^2_t(i)\right)^2 |\mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\frac{\ell^4_t(i)}{\tilde{p}^2_t(i)}\mathbb{1}_{I_t=i} - 2\frac{\ell^4_t(i)}{\tilde{p}_t(i)}\mathbb{1}_{I_t=i} + \ell^4_t(i)|\mathcal{F}_{t-1}\right] \\
&= \frac{\ell^4_t(i)}{\tilde{p}_t(i)} - \ell^4_t(i) \\
&\leq L\frac{\ell^2_t(i)}{\tilde{p}_t(i)}
\end{aligned} \tag{24}
$$

The above characterization of the sequence $\{Z_{t,i}\}_{t\in[T]}$ allows us to apply Freedman's concentration inequality that we state below,

Lemma 14 (Freedman's Inequality (Freedman, 1975; Kakade and Tewari, 2009)) *Suppose $\{Z_t\}_{t\in[T]}$ is a martingale difference sequence with respect to a filtration $\{\mathcal{F}_t\}_{t\in[T]}$, such that $|Z_t| \leq b$. Define, $\mathrm{Var}_t Z_t = \mathrm{Var}\left(Z_t|\mathcal{F}_{t-1}\right)$ and let $\sigma = \sqrt{\sum_{t=1}^T \mathrm{Var}_t Z_t}$ be the sum of conditional variances of Z_t's. Then for any $\delta \leq 1/e$ and $T \geq 3$ we have,*

$$
P\left(\sum_{t=1}^{T} Z_t \geq \max\left\{2\sigma, 3b\sqrt{\log(1/\delta)}\right\}\sqrt{\log(1/\delta)}\right) \leq 4\delta\log(T)
$$

Since $Z_{1,i}, \ldots, Z_{T,i}$ is a martingale difference sequence with $|Z_{t,i}| \leq 2nL/\theta$, we can applying the two-sided extension of Lemma 14 to this sequence. Combined with union bound over all

$i \in [n], t \in [T]$ we have that $\forall i \in [n], t \in [T]$, then w.p.$\geq 1 - 8nT\delta \log(T)$,

$$\begin{aligned}|\tilde{\ell}^2_{1:t}(i) - \ell^2_{1:t}(i)| &= \left|\sum_{\tau=1}^{t} Z_{\tau,i}\right| \\ &\leq \max\left\{2\sqrt{\sum_{\tau=1}^{t} \mathrm{Var}(Z_{\tau,i}|\mathcal{F}_{\tau-1})}, \frac{6nL}{\theta}\sqrt{\log(1/\delta)}\right\}\sqrt{\log(1/\delta)} \\ &\leq \max\left\{2\sigma_i, \frac{6nL}{\theta}\sqrt{\log(1/\delta)}\right\}\sqrt{\log(1/\delta)}. \end{aligned} \tag{25}$$

where we have defined $\sigma_i := \sqrt{\sum_{t=1}^{T} \mathrm{Var}(Z_{t,i}|\mathcal{F}_{t-1})}$. Notice that the last line above uses the fact that $\forall t \in [T] : \sum_{\tau=1}^{t} \mathrm{Var}(Z_{\tau,i}|\mathcal{F}_{\tau-1}) \leq \sigma_i^2$, which holds since the conditional variance is non-negative.

A few remarks are in place before we go on with the proof:

1. Define $\mathcal{B}$ to be the event that the bound stated in Equation (25) holds. Note that $P(\mathcal{B}) \geq 1 - 8nT\delta\log(T)$. From this point on, all of the statements in the proof are conditioned on the event $\mathcal{B}$.
2. For ease of notation we shall ignore the $\log(1/\delta)$ terms appearing in Equation (25). Note that these only affect the final guarantees by a factor of $O(\log(nT))$ for the choice of $\delta = 1/\mathrm{poly}(n, T)$.
3. We denote $M_i := \max\{2\sigma_i, 6nL/\theta\}$. Ignoring $\log(1/\delta)$ factors, Equation (25) can be now restated as follows, $\forall i \in [n], t \in [T]$ w.p.$\geq 1 - 8nT\delta\log(T)$,

$$|\tilde{\ell}^2_{1:t}(i) - \ell^2_{1:t}(i)| \leq M_i \tag{26}$$

We are now ready to go on with the proof. Notice that combining Equations (26) and (24) provides us with a bound on $|\tilde{\ell}^2_{1:t}(i) - \ell^2_{1:t}(i)|$ which depends on the $\tilde{p}_t$'s. The next lemma provides us with a cleaner bound which gets rid of this dependence. The proof of is provided in Section E.2.

Lemma 15 *Conditioning on the event $\mathcal{B}$, the following bound holds,*

$$M_i \leq 10n^{\frac{2}{3}}LT^{\frac{1}{3}} + 4\sqrt{L}\left(\ell^2_{1:T}(i)\right)^{\frac{1}{4}}\left(\sum_{i=1}^{n}\sqrt{\ell^2_{1:T}(i)}\right)^{\frac{1}{2}}, \tag{27}$$

and also

$$M_i \leq 14n^{\frac{1}{2}}LT^{\frac{1}{2}}. \tag{28}$$

Step 2: bounding (A)**.** First, we formulate a helper lemma, with its proof provided in Section E.3.

Lemma 16 *Let $x, a > 0$ then*

$$\sqrt{x+a} - \sqrt{x} \leq \min\{\sqrt{a}, a/\sqrt{x}\}.$$

Equation (26) enables us to bound (A) as follows,

$$
\begin{aligned}
(\mathrm{A}) &= \left(\sum_{i=1}^{n} \sqrt{\tilde{\ell}^2_{1:T}(i)}\right)^2 - \left(\sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i)}\right)^2 \\
&= \left(\sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i) + (\tilde{\ell}^2_{1:T}(i) - \ell^2_{1:T}(i))}\right)^2 - \left(\sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i)}\right)^2 \\
&\leq \left(\sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i) + M_i}\right)^2 - \left(\sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i)}\right)^2 \\
&= \left(\sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i) + M_i} + \sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i)}\right) \cdot \left(\sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i) + M_i} - \sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i)}\right) \\
&\leq 2\sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i) + M_i} \cdot \left(\sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i) + M_i} - \sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i)}\right) \\
&\leq 2\sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i)} \cdot \sum_{i=1}^{n} \min\left\{\sqrt{M_i}, \frac{M_i}{\sqrt{\ell^2_{1:T}(i)}}\right\} + 2\sum_{i=1}^{n} \sqrt{M_i} \cdot \sum_{i=1}^{n} \min\left\{\sqrt{M_i}, \frac{M_i}{\sqrt{\ell^2_{1:T}(i)}}\right\} \\
&\leq \underbrace{2\sum_{i=1}^{n} \sqrt{\ell^2_{1:T}(i)} \cdot \sum_{i=1}^{n} \min\left\{\sqrt{M_i}, \frac{M_i}{\sqrt{\ell^2_{1:T}(i)}}\right\}}_{(*)} + \underbrace{2\left(\sum_{i=1}^{n} \sqrt{M_i}\right)^2}_{(**)}, \qquad (29)
\end{aligned}
$$

where the second-to-last line uses $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, together with Lemma 16.

Let us start with bounding $(**)$,

$$
\begin{aligned}
\left(\sum_{i=1}^{n} \sqrt{M_i}\right)^2 &\leq n^2 \max_i M_i \leq 14Ln^{2+\frac{1}{2}}T^{\frac{1}{2}} \\
&\leq 14Ln^{2+\frac{1}{3}}T^{\frac{2}{3}} \qquad (30)
\end{aligned}
$$

where we have used the second part of Lemma 15; the second line uses $T \geq n$ leading to $(nT)^{1/2} \leq n^{1/3}T^{2/3}$.

The last step of the proof is to bound $(*)$. From Lemma 15, we have the immediate corollary that,

$$
M_i \leq \max\left\{16n^{\frac{2}{3}}LT^{\frac{1}{3}},\, 16\sqrt{L}\left(\ell^2_{1:T}(i)\right)^{\frac{1}{4}}\left(\sum_{i=1}^{n}\sqrt{\ell^2_{1:T}(i)}\right)^{\frac{1}{2}}\right\}. \qquad (31)
$$

Denote $i_* = \arg\max_{i\in[n]} \min\left\{\sqrt{M_i}, \frac{M_i}{\sqrt{\ell^2_{1:T}(i)}}\right\}$. We divide the remainder of the proof into two cases depending on the argument returned by $\max$ of Eq. (31) for the index i_*. If the $\max$ returns the first argument for i_*, i.e. $M_{i_*} = 16n^{\frac{2}{3}}LT^{\frac{1}{3}}$, then

$$\begin{aligned}
(*) &= \sum_{i=1}^{n}\sqrt{\ell^2_{1:T}(i)}\cdot\sum_{i=1}^{n}\min\left\{\sqrt{M_i}, \frac{M_i}{\sqrt{\ell^2_{1:T}(i)}}\right\} \\
&\leq \sum_{i=1}^{n}\sqrt{\ell^2_{1:T}(i)}\cdot n\min\left\{\sqrt{M_{i_*}}, \frac{M_{i_*}}{\sqrt{\ell^2_{1:T}(i_*)}}\right\} \\
&= n\sum_{i=1}^{n}\sqrt{\ell^2_{1:T}(i)}\cdot\sqrt{M_{i_*}} \\
&\leq n^2\sqrt{LT}\cdot\sqrt{16n^{2/3}T^{1/3}} \\
&\leq 4n^{2+\frac{1}{3}}LT^{\frac{2}{3}}. \qquad (32)
\end{aligned}$$

In the other case, we have $M_{i_*} = 16\sqrt{L}\left(\ell^2_{1:T}(i_*)\right)^{\frac{1}{4}}\left(\sum_{i=1}^n\sqrt{\ell^2_{1:T}(i_*)}\right)^{\frac{1}{2}}$. We will need the following lemma, with its proof given Section E.4:

Lemma 17 *Fix $w > 0$; Let $x \in [0, w]$ and $a, b : \mathbb{R}^+ \to \mathbb{R}^+$ functions of x, then*

$$\max_{x\in[0,w]}\min\left\{a(x)\cdot x^{1/8}, b(x)\cdot x^{-1/4}\right\} \leq \max_{x\in[0,w]} a(x)^{2/3}b(x)^{1/3}.$$

Now we can upper bound $(*)$,

$$\begin{aligned}
(*) &= \sum_{i=1}^{n}\sqrt{\ell^2_{1:T}(i)}\cdot\sum_{i=1}^{n}\min\left\{\sqrt{M_i}, \frac{M_i}{\sqrt{\ell^2_{1:T}(i)}}\right\} \\
&\leq n\sum_{i=1}^{n}\sqrt{\ell^2_{1:T}(i)}\cdot\min\left\{\sqrt{M_{i^*}}, \frac{M_{i_*}}{\sqrt{\ell^2_{1:T}(i_*)}}\right\} \\
&\leq n^2L^{\frac{1}{2}}T^{\frac{1}{2}}\cdot\min\left\{\sqrt{M_{i_*}}, \frac{M_{i_*}}{\sqrt{\ell^2_{1:T}(i_*)}}\right\} \\
&= n^2L^{\frac{1}{2}}T^{\frac{1}{2}}\cdot\min\left\{4L^{\frac{1}{4}}\left(\ell^2_{1:T}(i_*)\right)^{\frac{1}{8}}\left(\sum_{i=1}^{n}\sqrt{\ell^2_{1:T}(i_*)}\right)^{\frac{1}{4}}, \frac{16\sqrt{L}\left(\sum_{i=1}^n\sqrt{\ell^2_{1:T}(i_*)}\right)^{\frac{1}{2}}}{\left(\ell^2_{1:T}(i_*)\right)^{\frac{1}{4}}}\right\} \\
&\overset{(\diamond)}{\leq} n^2L^{\frac{1}{2}}T^{\frac{1}{2}}\cdot 16L^{\frac{1}{2}}n^{\frac{1}{3}}T^{\frac{1}{6}} \\
&\leq 16n^{2+\frac{1}{3}}LT^{\frac{2}{3}}, \qquad (33)
\end{aligned}$$

where for $(\diamond)$ we used Lemma 17 with $x = \ell^2_{1:T}(i_*)$, $a(x) = 4L^{\frac{1}{4}}\left(\sum_{i=1}^n\sqrt{\ell^2_{1:T}(i_*)}\right)^{\frac{1}{4}}$, $b(x) = 16\sqrt{L}\left(\sum_{i=1}^n\sqrt{\ell^2_{1:T}(i_*)}\right)^{\frac{1}{2}}$, and therefore $\max_x a(x)^{2/3}b(x)^{1/3} \leq 16L^{1/2}n^{1/3}T^{1/6}$. Combining Equation (29) together with Equations (30),(32), and (33), we may establish the final bound for (A), conditioned on the event $\mathcal{B}$:

$$(A) \leq 64n^{2+\frac{1}{3}}LT^{\frac{2}{3}} \tag{34}$$

Concluding: Combining Equation (23) with (34) and taking an sufficiently small $\delta = 1/\mathrm{poly}(\mathrm{n},\mathrm{T})$ we have proven the theorem. ■

E.2. Proof of Lemma 15

Proof Recalling that $M_i = \max\{2\sigma_i, 6nL/\theta\}$, it is natural to divide the proof into two cases depending on the value of M_i. Since $\theta = (n/T)^{1/3}$, it is immediate to show that the lemma holds for the case where $2\sigma_i \leq 6nL/\theta$, since in this case $M_i = 6nL/\theta = 6Ln^{2/3}T^{1/3}$. The rest of the proof regards the other case where $2\sigma_i > 6nL/\theta$, and therefore $M_i = 2\sigma_i$.

Step (1): Decomposing M_i^2.

$$\begin{aligned}
\frac{1}{4L}M_i^2 &= \frac{1}{L}\sigma_i^2 \\
&\leq \sum_{t=1}^{T} \frac{\ell_t^2(i)}{\tilde{p}_t(i)} \\
&= \sum_{t:\ell_{1:t}^2(i)\leq 2M_i} \frac{\ell_t^2(i)}{\tilde{p}_t(i)} + \sum_{t:\ell_{1:t}^2(i)\geq 2M_i} \ell_t^2(i)\left(\frac{1}{\tilde{p}_t(i)} - \frac{1}{p_t(i)}\right) + \sum_{t:\ell_{1:t}^2(i)\geq 2M_i} \frac{\ell_t^2(i)}{p_t(i)} \\
&\leq \frac{n}{\theta}\cdot \sum_{t:\ell_{1:t}^2(i)\leq 2M_i} \ell_t^2(i) + n\theta\ell_{1:T}^2(i) + \sum_{t:\ell_{1:t}^2(i)\geq 2M_i} \frac{\ell_t^2(i)}{p_t(i)} \\
&\leq \frac{2nM_i}{\theta} + n\theta LT + \underbrace{\sum_{t:\ell_{1:t}^2(i)\geq 2M_i} \frac{\ell_t^2(i)}{p_t(i)}}_{(\star)}, \qquad (35)
\end{aligned}$$

where in the second line we use the definition of σ_i together with the bound of Eq. (24), implying $\sigma_i^2 \leq L\sum_{t=1}^{T}\ell_t^2(i)/\tilde{p}_t(i)$; in the fourth line we use $\tilde{p}_t(i) \geq \frac{\theta}{n}$ (due to mixing), and we also use $\frac{1}{\tilde{p}_t(i)} - \frac{1}{p_t(i)} \leq n\theta$ (see proof of Theorem 7). The last line uses $\ell_{1:T}^2(i) \leq LT$. Next we bound the last term, $(\star)$.

Step (2): Bounding $(\star)$. We shall first bound $1/p_t(i)$ and later use this in order to bound $(\star)$. Notice that the following hold $\forall t \in [T]$ such that $\ell_{1:t}^2(i) \geq 2M_i$,

$$\tilde{\ell}_{1:t}^2(i) \geq \ell_{1:t}^2(i) - M_i \geq \frac{1}{2}\ell_{1:t}^2(i) \tag{36}$$

$$\tilde{\ell}_{1:t}^2(i) \leq \ell_{1:t}^2(i) + M_i \leq \frac{3}{2}\ell_{1:t}^2(i), \qquad \forall i \in [n] \tag{37}$$

where we have used $|\tilde{\ell}_{1:t}^2(i) - \ell_{1:t}^2(i)| \leq M_i$ (see Eq (26)), which follows since we condition on the event $\mathcal{B}$. Combining the above with the definition of p_t (see Lemma 2) and denoting $L' := Ln/\theta$, yields,

$$\begin{aligned}
\frac{1}{p_t(i)} &= \frac{\sum_{i=1}^n \sqrt{\tilde{\ell}^2_{1:t-1}(i) + L'}}{\sqrt{\tilde{\ell}^2_{1:t-1}(i) + L'}} \\
&\leq \frac{\sum_{i=1}^n \sqrt{\tilde{\ell}^2_{1:t}(i) + L'}}{\sqrt{\tilde{\ell}^2_{1:t}(i)}} \\
&\leq \sqrt{2}\frac{\sum_{i=1}^n \sqrt{\frac{3}{2}\ell^2_{1:t}(i) + L'}}{\sqrt{\ell^2_{1:t}(i)}} \\
&\leq 2\frac{\sum_{i=1}^n \sqrt{\ell^2_{1:T}(i)}}{\sqrt{\ell^2_{1:t}(i)}}
\end{aligned}$$

where in the second line we use $\tilde{\ell}^2_t(i) \leq L'$, in the third we employ Equations (36), (37); and the fourth follows by noticing $L' = Ln/\theta \leq M_i \leq \ell^2_{1:t}(i)/2$, and also $\ell^2_{1:t}(i) \leq \ell^2_{1:T}(i),\ \forall t \in [T]$.

Using the above inequality we may now bound **(⋆)**,

$$\begin{aligned}
(\star) &= \sum_{t:\ell^2_{1:t}(i) \geq 2M_i} \frac{\ell^2_t(i)}{p_t(i)} \\
&\leq 2\left(\sum_{i=1}^n \sqrt{\ell^2_{1:T}(i)}\right) \sum_{t:\ell^2_{1:t}(i) \geq 2M_i} \frac{\ell^2_t(i)}{\sqrt{\ell^2_{1:t}(i)}} \\
&\leq 2\left(\sum_{i=1}^n \sqrt{\ell^2_{1:T}(i)}\right) \sum_{t=1}^T \frac{\ell^2_t(i)}{\sqrt{\ell^2_{1:t}(i)}} \\
&\leq 4\left(\sum_{i=1}^n \sqrt{\ell^2_{1:T}(i)}\right) \sqrt{\ell^2_{1:T}(i)}
\end{aligned} \tag{38}$$

where the last inequality uses the following lemma from (McMahan and Streeter, 2010):

Lemma 18 *(McMahan and Streeter, 2010) For any non-negative numbers $a_1, \ldots, a_T$ the following holds:*

$$\sum_{t=1}^T \frac{a_t}{\sqrt{\sum_{\tau=1}^t a_\tau}} \leq 2\sqrt{\sum_{t=1}^T a_t}$$

Step (3): Final bound. Plugging the bound of Equation (38) back into Equation (35) implies,

$$\frac{1}{4L}M_i^2 \leq \frac{2nM_i}{\theta} + n\theta LT + 4\sqrt{\ell^2_{1:T}(i)}\left(\sum_{i=1}^n \sqrt{\ell^2_{1:T}(i)}\right)$$

Denote $a = 1/(4L)$, $b = 2n/\theta$, $c_1 = n\theta LT$, $c_2 = 4\sqrt{\ell^2_{1:T}(i)}\left(\sum_{i=1}^n \sqrt{\ell^2_{1:T}(i)}\right)$. Then, the above inequality can be reformulated as:

$$aM_i^2 - bM_i - c_1 - c_2 \leq 0.$$

Due to the quadratic formula, the largest M_i that satisfies the inequality above is $M_i = \left(b + \sqrt{b^2 + 4a(c_1 + c_2)}\right)/(2a)$. We can get an upper bound on M_i by using $\sqrt{b^2 + 4a(c_1 + c_2)} \leq b + 2\sqrt{ac_1} + 2\sqrt{ac_2}$ to finally get that

$$M_i \leq \frac{8nL}{\theta} + 2n^{\frac{1}{2}} L\theta^{\frac{1}{2}} T^{\frac{1}{2}} + 4\sqrt{L}\left(\ell^2_{1:T}(i)\right)^{\frac{1}{4}} \left(\sum_{i=1}^n \sqrt{\ell^2_{1:T}(i)}\right)^{\frac{1}{2}}.$$

Using $\theta = (n/T)^{1/3}$ we have proven the first claim of the lemma. For the second claim, we use the upper bound $\ell^2_{1:T}(i) \leq LT$ and note that $n^{\frac{2}{3}} T^{\frac{1}{3}} \leq n^{\frac{1}{2}} T^{\frac{1}{2}}$ (since $T \geq n$).

■

E.3. Proof of Lemma 16

Proof

$$(\sqrt{x + a} - \sqrt{x})^2 = 2x + a - 2\sqrt{x^2 + xa} \leq a$$

which proves that $\sqrt{x + a} - \sqrt{x} \leq \sqrt{a}$. On the other hand, we have that $\sqrt{x + a} - \sqrt{x} \leq a/\sqrt{x}$, which can be easily seen by rearranging it as $\sqrt{x + a} \leq a/\sqrt{x} + \sqrt{x}$ and taking the square of both side. Combining these two facts we get the results. ■

E.4. Proof of Lemma 17

Proof Define $F(x) := \min\left\{a(x) \cdot x^{1/8}, b(x) \cdot x^{-1/4}\right\}$. Note that in order to establish the lemma it is sufficient to show that the following holds for any $x \geq 0$,

$$F(x) \leq a(x)^{2/3} b(x)^{1/3}.$$

To do so, fix $x \geq 0$ and divide into two cases.

Case 1: If $a(x) \cdot x^{1/8} \leq b(x) \cdot x^{-1/4}$ then $x \leq (b(x)/a(x))^{8/3}$ implying that $F(x) = a(x) \cdot x^{1/8} \leq a(x)^{2/3} b(x)^{1/3}$.

Case 2: If $a(x) \cdot x^{1/8} \geq b(x) \cdot x^{-1/4}$ then $x \geq (b(x)/a(x))^{8/3}$ implying that $F(x) = b(x) \cdot x^{-1/4} \leq a(x)^{2/3} b(x)^{1/3}$. ■

Appendix F. Experiments

F.1. Image Classification

Training a binary classifier with imbalanced data is a challenging task in machine learning. Practices for dealing with imbalance include optimizing class weight hyperparameters, hard negative mining (Shrivastava et al., 2016) and synthetic minority oversampling (Chawla et al., 2002). Without accounting for imbalance, the minority samples are often misclassified in early stages of the iterative training procedures, resulting in high loss and high gradient norms associated with these points. Importance sampling schemes for reducing the variance of the gradient norms will sample these instances more often at the early phases, offering a way of tackling imbalance.

For verifying this intuition, we perform the image classification experiment of Bouchard et al. (2015). We train one-vs-all logistic regression Pascal VOC 2007 dataset Everingham et al. (2010) with image features extracted from the last layer of the VGG16 (Simonyan and Zisserman, 2015) pretrained on Imagenet. We measure the average precision by reporting its mean over the 20 classes of the test data. The optimization is performed with AdaGrad (Duchi et al., 2011), where learning rate is initialized to 0.1. The losses received by the bandit methods are the norms of the logistic loss gradient. We compare our method, Variance Reducer Bandit (VRB), to:

- uniform sampling for SGD,
- Adaptive Weighted SGD (AW) (Bouchard et al., 2015) — variance reduction by sampling from a chosen distribution whose parameters are optimized alternatingly with the model parameters,
- MABS (Salehi et al., 2017) — bandit algorithm for variance reduction that relies on EXP3 through employing modifies losses.

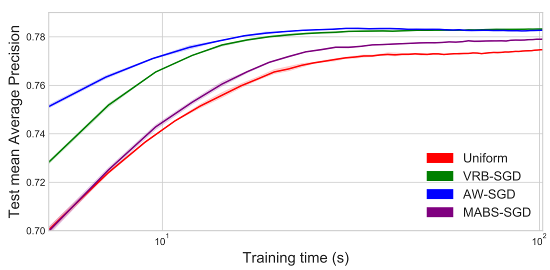

Figure 6: Mean Average Precisions on the test part of VOC 2007.

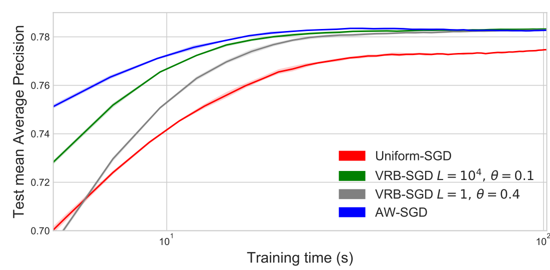

Figure 7: The effect of different hyperparameters on VRB.

The hyperparameters of the methods are chosen based on cross-validation on the validation portion of the dataset. The results can be seen in Figure 6, where the shaded areas represent confidence 95% intervals over 10 runs. The best performing method is AW, but its disadvantage compared to the bandit algorithms is that it requires choosing a family of sampling distributions, which usually incorporates prior knowledge, and calculating the derivative of the log-density. VRB and AW both outperform uniform subsampling with respect to the training time. VRB performs similarly to AW at convergence, and speeds up training 10 times compared to uniform sampling, by attaining a certain score level 10 times faster. We have also experimented with the variance reduction method of Namkoong et al. (2017), but it did not outperform uniform sampling significantly. Since cross-validation is costly, in Figure 7 we show the effect of the hyperparameters of our method. More specifically, we compare the performance of VRB with misspecified regularizer $L = 1$ to the best $L = 10^8$ chosen by cross-validation, and we compensate by using higher mixing coefficient $\theta = 0.4$. The fact that only the early-stage performance is affected is a sign of method's robustness against regularizer misspecification.

We also measure the regret incurred both by the full information and VRB samplers, and show the results in Figure 8. For a fair comparison, we choose an *oblivious* adversary that generates the loss sequences by performing the same optimization process as described above on a subset of 1000 data points from VOC 2007, with uniform sampling. For VRB, we report the average regret over 10 runs.

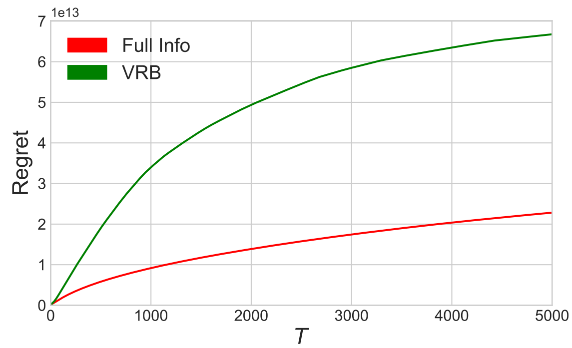

Figure 8: Regret incurred by the full info and VRB.

F.2. k-Means

In this experiment, we show that in some applications it is beneficial to work with per-sample upper bound estimates L_i instead of a single global bound. As an illustrative example, we choose mini-batch k-Means clustering (Sculley, 2010). This is a slight deviation from the presented theory, since we sample multiple points for the batch and update the sampler only once, upon observing the loss for the batch.

In the case of k-Means, the parameters consist of the coordinates of the k centers $Q = \{q_1, q_2, \ldots, q_k\}$. As the cost function for a point $x_i \in \{x_1, x_2, \ldots, x_n\}$ is the squared Euclidean distance to the closest center, the loss received by VRB is the norm of the gradient $\min_{q \in Q} 2 \cdot ||x_i - q||_2$. This lends itself to a natural estimation of L_i: choose a point u randomly from the dataset and define $L_i = 4 \cdot ||x_i - u||_2^2$. For this experiment, we set $\theta = 0.5$.

We solve mini-batch k-Means for $k = 100$ and batch size $b = 100$ with uniform sampling and VRB. The initial centers are chosen with k-Means++ (Arthur and Vassilvitskii, 2007) from a random subsample of 1000 points from the training data and they are shared between the methods. We generate 10 different sets of initial centers and run both algorithms 10 times on each set of centers, with different random seeds for the samplers. We train the algorithm on 80% of the data, and measure the cost of the 20% test portion for the following datasets:

- `CSN` (Faulkner et al., 2011) — cellphone accelerometer with 80,000 observations and 17 features,

- KDD (KDD Cup 2004) — data set used for Protein Homology Prediction KDD competition containing 145,751 observations with 74 features,
- MNIST (LeCun et al., 1998) — 70,000 low resolution images of handwritten characters transformed using PCA with whitening and retaining 10 dimensions.

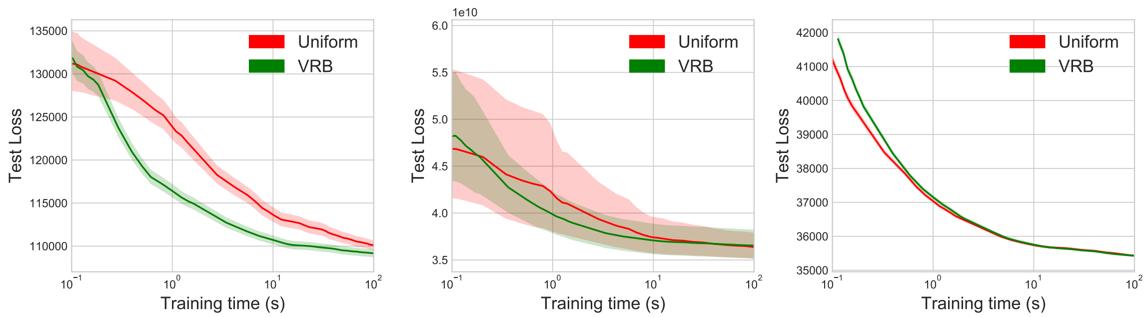

Figure 9: The evolution of the loss of k-Means on the test set. The shaded areas represent 95% confidence intervals over 100 runs.

The evolution of the cost function on the test set with respect to the elapsed training time is shown in Figure 9. The chosen datasets illustrate three observed behaviors of our algorithm. In the case of CSN, our method significantly outperforms uniform subsampling. In the case of KDD, the advantage of our method can be seen in the reduced variance of the cost over multiple runs, whereas on MNIST we observe no advantage. This behavior is highly dependent on intrinsic dataset characteristics: for MNIST, we note that the entropy of the best-in-hindsight sampling distribution is close the entropy of the uniform distribution. We have also compared VRB with the bandit algorithms mentioned in the previous section. Since mini-batch k-Means converges in 1-2 epochs, these methods with uniform initialization do not outperform uniform subsampling significantly. Thus, for this setting, careful initialization is necessary, which is naturally supported by our method.