A Random Walk Approach to First-Order Stochastic Convex Optimization

Sattar Vakili[*] and Qing Zhao[†]

[*] Prowler.io, Cambridge, UK, *sattar@prowler.io*
[†]School of Electrical & Computer Engineering, Cornell University, Ithaca, NY, *qz16@cornell.edu*

Abstract

Online minimization of an unknown convex function over a convex and compact set is considered under first-order stochastic bandit feedback, which returns a random realization of the gradient of the function at each query point. Without knowing the distribution of the random gradients, a learning algorithm sequentially chooses query points with the objective of minimizing regret defined as the expected cumulative loss of the function values at the query points in excess to the minimum value of the function. An active search strategy based on devising a biased random walk on an infinite-depth tree constructed through successive partitioning of the domain of the function is developed. It is shown that the biased random walk moves toward the optimal point in a geometric rate, leading to an order-optimal regret performance of $\tilde{\mathcal{O}}(\sqrt{T})$. The structural properties of this random-walk based strategy admits detailed finite-time regret analysis. By localizing data processing to small subsets of the input domain based on the tree structure, it enjoys $\mathcal{O}(1)$ computation and memory complexity per query and allows dynamic allocation of limited data storage.

1 Introduction

1.1 Stochastic Convex Optimization

Stochastic convex optimization is concerned with the minimization of a random loss function $f(x)$ over a convex and compact set $\mathcal{X}$. The stochastic model of f is unknown, except the knowledge that $\mathbb{E}[f(x)]$ is convex. At each time t, the decision maker chooses a query point $x_t \in \mathcal{X}$, and either a random sample of the function value $f(x_t)$ or a random sample of the derivative $g(x_t)$ of the function at the query point x_t is revealed. These two feedback models are commonly referred to, respectively, as the zeroth-order and the first-order stochastic optimization.

A traditionally adopted objective of the problem, as in the pioneering work by Robbins and Monro [1] and Kiefer and Wolfowitz [2] in the early 1950s, is to

1

approximate the minimizer $x^* = \arg\min_{x \in \mathcal{X}} \mathbb{E}[f(x)]$ of the loss function. Also known as stochastic approximation, this line of work focuses on the asymptotic convergence of the end point x_T to the optimal point x^* (or $f(x_T)$ to $f(x^*)$) over a growing horizon of length T. We refer to this formulation of the problem as the offline setting, for the reason that the losses $\{f(x_t)\}_{t=1}^{T-1}$ during the query process are inconsequential and the query process is chosen for the sole purpose of outputting a desired end point x_T. Extensive studies exist under this formulation (see an overview article by Lai [3]).

The online counterpart of the problem adopts the measure of regret defined as the expected cumulative loss at the query points in excess to the minimum loss: $R(T) = \sum_{t=1}^{T} \mathbb{E}[f(x_t) - f(x^*)]$. Under this objective, the query process needs to balance the exploration of the input space $\mathcal{X}$ in search for x^* and the associated loss incurred during the search process. The behavior of regret $R(T)$ over a growing horizon length T is a finer measure than the convergence of x_T or $f(x_T)$. Specifically, a policy with a sublinear regret order in T implies that the sequence of samples $\{f(x_t)\}_{t=1}^{\infty}$ converges to the optimum value $f(x^*)$. The converse, however, is not true. In particular, the convergence of x_T to x^* (or $f(x_T)$ to $f(x^*)$) does not imply a sublinear, let alone an optimal, order of the regret.

In this paper, we focus on online first-order stochastic convex optimization, where the observations are random gradients at the query points and the objective is to minimize regret. A prevailing approach to this problem is based on stochastic gradient descent (SGD), where the next query point x_{t+1} is chosen in the opposite direction of the observed gradient (with a properly chosen step size that shrinks in t) while ensuring $x_{t+1} \in \mathcal{X}$ via a projection operation. It is known that this approach offers $\mathcal{O}(\sqrt{T})$ asymptotic regret order, matching the lower bound [1, 4, 5]. The performance of the SGD approaches, however, depends on careful tuning of a large number of parameters (e.g., the sequence of step sizes). Regret analysis for finite T is also lacking.

With access to the gradient, the problem can also be mapped to stochastic root finding [6] with the objective of locating the root of $\mathbb{E}[g(x)]$ based on random samples of the gradient. One solution to the problem under a one-dimensional input space $\mathcal{X}$ is the probabilistic bisection algorithm (PBA). Assuming a prior distribution of the optimal point x^*, PBA updates the belief (i.e., the posterior distribution) of x^* based on each observation and subsequently probes the median point of the belief. It was shown in [7] that the regret order of PBA is upper bounded by $\mathcal{O}(T^{0.5+\epsilon})$ for a small $\epsilon > 0$, and an $\mathcal{O}(\sqrt{T}\log T)$ regret order was conjectured. We point out that PBA requires a *known* stochastic model of the random gradient function $g(x)$ for the belief update using Bayes rule, and the update and sorting of the belief at each query point can be expensive in terms of computation and memory requirement.

1.2 Main Results

We propose a random walk approach to first-order stochastic convex optimization. Referred to as Random-Walk based Gradient Descent (RWGD), the proposed

policy first constructs a binary tree with infinite depth based on successively refined partitioning of the input space $\mathcal{X}$. Specifically, the root of the tree corresponds to $\mathcal{X}$, which, without loss of generality, is assumed to be $[0, 1]$ for the one-dimensional case. The tree grows to infinite depth based on a binary splitting of each node (i.e., the corresponding interval) that forms the two children of the node at the next level. The key idea of RWGD is to devise a biased random walk on this interval tree that initiates at the root node. Each move of the random walk is guided by a local sequential test based on random gradient realizations drawn from the left boundary, the middle point, and the right boundary of the interval corresponding to the current location of the random walk. The goal of the local sequential test is to determine, with a confidence level greater than $1/2$, whether there is a change of sign in the gradient in the left sub-interval or the right sub-interval of the current node. If either one is true (with the chosen confidence level), the walk moves to the corresponding child that sees the sign change. If neither is true, the walk moves back to the parent of the current node. The stopping rule and the output of the local sequential test are based on properly constructed lower and upper confidence bounds of the empirical mean of the observed gradient realizations. A greater than $1/2$ bias of the random walk is sufficient to ensure convergence to the optimal point x^* at a geometric rate, as shown in Sec. 3.

By bounding the sample complexity of the local sequential test and analyzing the trajectory of the biased random walk, we show that RWGD has a regret order of $\mathcal{O}(\sqrt{T} \log T)$, which is optimal up to a logarithmic factor. Furthermore, structural properties of the random walk approach allow a finer tracking of the query points, leading to a finite-time regret analysis, a first such result to our best knowledge.

In addition to its finite-time performance guarantee (in contrast to asymptotic regret analysis), RWGD also enjoys advantages in terms of robustness and computation/memory efficiency over the prevailing approach of SGD. Specifically, with no parameters to tune, RWGD is more robust to model mismatch and offers improved performance as demonstrated in the simulation examples in Sec. 5. By localizing data processing to small subsets of the input domain based on the tree structure, RWGD has $\mathcal{O}(1)$ computation and memory complexity per query and allows dynamic allocation of limited data storage. The projection operation in SGD, however, involves the entire input domain $\mathcal{X}$ at each query and can be computationally expensive.

There may appear to be a connection between RWGD and PBA, since both algorithms involve a certain bisection of the input domain. These two approaches are, however, fundamentally different. First, PBA requires the knowledge on the distribution of the random gradient function, while RWGD operates under unknown models. Second, the belief-based bisection in PBA is on the entire input domain $\mathcal{X}$ at each query and is done dynamically based on each random observation. The interval tree in RWGD based on successively bisection of $\mathcal{X}$ is predetermined, and each move of the random walk leads to a bisection of a *sub-interval* of $\mathcal{X}$ that is shrinking in geometric rate over time with high probability. It is this zooming effect of the biased random walk that leads to a

3

$\mathcal{O}(1)$ computation and memory complexity. For PBA, if $\mathcal{X}$ is discretized to M points for computation and storage, updating and sorting the belief would incur $\mathcal{O}(M \log M)$ computation complexity per sample and linear memory requirement.

1.3 Other Related Work

There are several results on first-order stochastic optimization that adopt stronger assumptions on f than convexity. In particular, assuming strongly convex or exponentially concave f results in a logarithmic regret in time [8].

Under the zeroth-order feedback model where the decision maker has access to the function values, the problem can be viewed as a continuum-armed bandit problem, on which a vast body of results exist. In particular, the work in [9] developed an approach based on the ellipsoid algorithm that achieves an $\mathcal{O}(\sqrt{T}(\log T)^{\frac{3}{2}})$ regret when the objective function f is convex and Lipschitz. The continuum armed bandit under Lipschitz assumption (not necessarily convex) has been studied in [10, 11, 12] where higher orders of regret were shown. The $\mathcal{X}$-armed bandit introduced in [13] considered a Lipschitz function with respect to a dissimilarity function known to the learner. Under the assumption of a finite number of global optima and a particular smoothness property, an $\tilde{\mathcal{O}}(\sqrt{T})$ regret was shown. The proposed policy in [13] uses a tree structure for updating the indexes in a bandit algorithm which is fundamentally different than RWGD and, e.g., does not induce a random walk. This line of work differs from the gradient-based approach considered in this work. Nevertheless, since an $\mathcal{O}(1)$ number of samples from f can be translated to a sample from g under certain regularity assumptions, gradient-based approaches can be extended to cases where samples from f are directly fed into the learning policy.

We mention that the stochastic online learning setting considered here is different, in problem formulation, objective, and techniques, from an adversarial counterpart of the problem where the loss function is adversarially chosen at each time t. On this line of research, see [14, 15] and references therein.

2 Problem Formulation and Preliminaries

Let $f : \mathcal{X} \to \mathbb{R}$ be a differentiable probabilistic loss function and $g = \frac{df}{dx}$. We assume $\mathbb{E}[f] : \mathcal{X} \to \mathbb{R}$ is a convex function. Let $x^* = \operatorname{argmin}_{x \in \mathcal{X}} \mathbb{E}[f(x)]$ denote an optimal point of the objective function $\mathbb{E}[f(x)]$. In this paper, we focus on the one-dimensional case. Without loss of generality let $\mathcal{X} = [0, 1]$. The results can be easily translated to any interval $\mathcal{X} \in \mathbb{R}$.

Without knowing $\mathbb{E}[f(x)]$ or the stochastic models of f or g, an agent sequentially chooses the sampling points $\{x_t\}_{t=1}^{\infty}$. Each sampling time t sees an i.i.d. realization of the loss function $f(x)$. A random cost $f(x_t)$ at the sampling point is incurred, and a random realization of the gradient $g(x_t)$ is observed. The objective is to minimize the cumulative loss. Specifically, the goal is to design a learning policy π that is a mapping from the history of the sampling points and the observations to a new sampling point, $\pi(t) : \mathcal{X}^{t-1} \times \mathbb{R}^{t-1} \to \mathcal{X}$.

The performance of a learning policy π is measured by regret defined as the expected cumulative loss at the chosen sampling points $\{x_{\pi(t)}\}_{t=1}^{\infty}$ in excess to the optimum point x^*,

$$\mathcal{R}_\pi(T) = \mathbb{E}\left[\sum_{t=1}^{T}\left(f(x_{\pi(t)}) - f(x^*)\right)\right], \tag{1}$$

where $\mathbb{E}$ is the expectation operator with respect to the random process $\{f(x_{\pi(t)})\}_{t=1}^{\infty}$ induced by the policy π.

We assume that the distribution of $g(x) - \mathbb{E}[g(x)]$, for all $x \in \mathcal{X}$, is sub-Gaussian with parameter σ^2, i.e., its moment generating function is bounded by that of a Gaussian random variable with variance σ^2:

$$\mathbb{E}\left[\exp(\lambda\left(g(x) - \mathbb{E}[g(x)]\right))\right] \leq \exp(\frac{\lambda^2\sigma^2}{2}).$$

As a result of the Chernoff-Hoedffding bound, we have ([17]), for any $\epsilon > 0$,

$$\Pr[\bar{g}_t(x) - \mathbb{E}[g(x)] > \epsilon] \leq \exp(-\xi t\epsilon^2), \tag{2}$$

where $\bar{g}_t(x)$ is the sample mean obtained from t samples of $g(x)$, and $\xi = \frac{1}{2\sigma^2}$ is a constant depending on the class of distributions.

Extensions of both the proposed policy and its regret analysis to more general families of distributions, including heavy-tailed distributions are relatively straightforward, as discussed in subsequent sections.

3 Random-Walk based Gradient Descent on a Tree

The proposed policy is based on an infinite-depth binary tree with nodes representing a subinterval of $\mathcal{X}$ and edges representing the subset relation. The 2^l nodes at depth l ($l = 0, 1, 2, \ldots$) of the tree correspond to the intervals resulting from an equal-length partition of $\mathcal{X}$, with each interval of length $\Delta_l = 2^{-l}$. Each node at depth l has two children corresponding to its equal-length subintervals at depth $l+1$. Let $N_{k,l}$ ($k = 1, \ldots, 2^l$, $l = 0, 1, \ldots$) denote the kth node at depth l. We use the terms node and its corresponding interval interchangeably.

The basic structure of the proposed RWGD is to carry out a biased random walk on the interval tree $\mathcal{T}$. The walk starts at the root of the tree. Each move of the random walk is to one of the three adjacent nodes (i.e., the parent and the two children with the parent of the root defined as itself) of the current location. It is guided by the outputs of a confidence-bound based sequential test carried over the two boundary points and the middle point of the interval currently being visited by the random walk.

We now specify the sequential test carried out on a generic sampling point $x \in [0, 1]$. The goal is to determine, at a confidence level $1 - \alpha$, whether $\mathbb{E}[g(x)]$ is
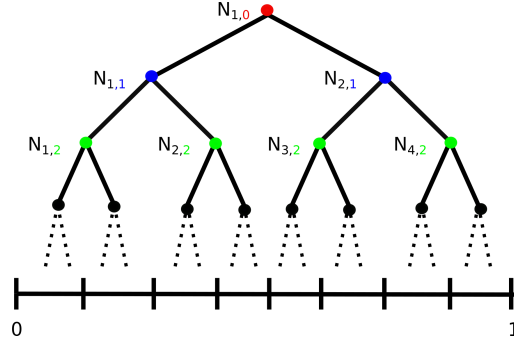
Figure 1: The binary tree $\mathcal{T}$ representing the subintervals of $[0,1]$. At level 0, $N_{1,0}$ corresponds to the interval $[0,1]$; at level 1, $N_{1,1}$ and $N_{2,1}$, respectively, correspond to the intervals $[0,0.5]$ and $[0.5,1]$; at level 2, $N_{1,2}$, $N_{2,2}$, $N_{3,2}$, and $N_{4,2}$, respectively, correspond to the intervals $[0,0.25]$, $[0.25,0.5]$, $[0.5,0.75]$, and $[0.75,1]$; and so on.

negative or positive. If the former is true, the test module outputs -1, indicating the target x^* is more likely to lie on the right of the current sampling point x; if the latter is true, the test module outputs 1, indicating the target x^* is more likely to lie on the left of x. Specifically, the local test module sequentially collects samples from $g(x)$. After collecting each sample, it determines whether to terminate the test and if yes, which value to output. As specified below, the stopping rule and the corresponding output are determined by comparing the upper and lower confidence bounds on the sample mean $\bar{g}_s(x)$ of s random realizations of $g(x)$.

> ▷ If $\bar{g}_s(x) - \sqrt{\frac{\log \frac{s^2}{\alpha}}{\xi s}} > 0$, terminate and output 1.
>
> ▷ If $\bar{g}_s(x) + \sqrt{\frac{\log \frac{s^2}{\alpha}}{\xi s}} < 0$, terminate and output -1.
>
> ▷ Otherwise, continue taking samples of $g(x)$.
>
> where $\bar{g}_s(x)$ denotes the sample mean of $g(x)$ obtained form s observations at point x, $\alpha \in (0,1)$ is a constant, and ξ is the distribution parameter specified in (2).

Figure 2: The sequential test at a sampling point x

By convention, we define the output of the test at the boundary points $x = 0$ to be -1, and at $x = 1$ to be 1, without performing the test.

We now specify the random walk on the tree based on the outputs of the

local tests. The algorithm consists of the following loop. Let $N_{k,l}$ denote the current location of the random walk which is initially set at the root node. The boundary points and the middle point of the interval corresponding to $N_{k,l}$ are probed by the local test module with parameter α, where α can be set to any constant in $(0, 1 - \frac{1}{\sqrt[3]{2}})$ to ensure a greater than $1/2$ bias of the random walk. Based on the output sequence at the left boundary point $((k-1)\Delta_l)$, the middle point $((2k-1)\Delta_{l+1})$, and the right boundary point $(k\Delta_l)$ of the current node, the random walk chooses one of the three neighboring nodes of $N_{k,l}$—its two children and its parent—to move to. The procedure is specified in the pseudo-code given in Algorithm 1.

Algorithm 1 The random walk module of RWGD.

Initialization: initial location of the random walk $N_{k,l} = N_{1,0}$, $\alpha \in (0, 1 - \frac{1}{\sqrt[3]{2}})$.

loop

Test the boundaries and the middle point of the interval corresponding to $N_{k,l}$ by local sequential test.

if the output sequence on the left boundary, middle point and the right boundary, in order, is $\{-1, 1, 1\}$ **then**

move to the left child of $N_{k,l}$: $N_{k,l} \leftarrow N_{2k-1,l+1}$.

else if the output sequence on the left boundary, middle point and the right boundary, in order, is $\{-1, -1, 1\}$ **then**

move to the right child of $N_{k,l}$: $N_{k,l} \leftarrow N_{2k,l+1}$.

else

move to the parent of $N_{k,l}$: $N_{k,l} \leftarrow N_{\lceil \frac{k}{2} \rceil, l-1}$.

end if

end loop

For example, at the root node the output of the test at the left and the right boundary points are -1 and 1, respectively, by convention. We carry out the local test module at the middle point $x = 0.5$. If the output of the test at $x = 0.5$ is 1 (indicating the derivative at $x = 0.5$ is likely to be positive), the random walk moves to the left child corresponding to the interval $[0, 0.5]$. If the output of the test at $x = 0.5$ is -1, the random walk moves to the right child corresponding to the interval $[0.5, 1]$.

We point out that the extension to all distributions models can be easily handled. The only required change is in the local sequential test module of the proposed policy, while the global random walk module remains the same. The local test module needs to be modified based on corresponding concentration results. Specifically, for light-tailed distributions, Chernoff-Hoeffding bounds similar to the ones for Sub-Gaussian distributions exist. For heavy-tailed distributions, more delicate techniques such as truncated sample mean or median estimates with similar concentration inequalities can be employed. We omit the details, given that such extensions are quite standard (e.g., see [18]).

4   Regret Analysis

In this section, we analyze the performance of RWGD. We show an $\mathcal{O}(\sqrt{T}\log T)$ upper bound on the regret along with a finite-time analysis, as formalized in Theorem 1 below.

Theorem 1. *Let* $\alpha \in (0, 1 - \frac{1}{\sqrt[3]{2}})$ *be the chosen parameter of the RWGD policy. Let* $G = \max_{x \in \mathcal{X}} \mathbb{E}[g(x)]$. *The regret of RWGD satisfies*

$$\mathcal{R}_{RWGD}(T) \leq$$
$$\frac{12}{2(1-\alpha)^3 - 1} \sqrt{\frac{3}{\xi} T \log T \log(8 \sqrt{\frac{1}{\alpha} \frac{T(2(1-\alpha)^3 - 1)^2}{9\xi \log T}})}$$
$$+ G(\log T + 4). \tag{3}$$

Proof of Theorem 1 is based on following two lemmas. Lemma 1 gives upper bounds on the sample complexity and error probabilities of the local sequential test. Lemma 2 establishes the geometric rate of the biased random walk of RWGD to the optimum point x^*.

Lemma 1. *Let* $\{Z_t\}_{t=1}^{\infty}$ *denote the time indexed samples from the random variable* $g(x)$ *for a fixed* $x \in \mathcal{X}$. *Let* μ *denote the expected value of the i.i.d. Sub-Gaussian random process* $\{Z_t\}_{t=1}^{\infty}$ *and* $\overline{Z}_t = \frac{1}{t}\sum_{s=1}^{t} Z_s$ *denote the sample mean obtained from the first t observations. Let* τ *be the stopping time of the local sequential test with parameter* α *applied to* $\{Z_t\}_{t=1}^{\infty}$. *We have,*

$$\mathbb{E}[\tau] \leq \frac{16}{\mu^2 \xi} \log \frac{8\sqrt{\frac{1}{\alpha}}}{\mu^2 \xi} + 2, \tag{4}$$

in the case of $\mu > 0$, $\tag{5}$

$$\mathbb{P}[\overline{Z}_\tau + \sqrt{\frac{\log \frac{\tau^2}{\alpha}}{\xi \tau}} < 0] \leq \alpha,$$

in the case of $\mu < 0$, $\tag{6}$

$$\mathbb{P}[\overline{Z}_\tau - \sqrt{\frac{\log \frac{\tau^2}{\alpha}}{\xi \tau}} > 0] \leq \alpha.$$

Proof: See Appendix A.

The probability that the output of the local test carried out on a point $x \in \mathcal{X}$ is incorrect (i.e. it is 1 when $\mathbb{E}[g(x)] < 0$ and -1 when $\mathbb{E}[g(x)] > 0$) is upper bounded by α. The condition for the random walk to move in the right direction is that the output of all three tests carried out on the boundary points and the middle point of the current interval are correct. Thus, the probability p that the random walk moves in the right direction satisfies $p \geq (1 - \alpha)^3$ which indicates $p > \frac{1}{2}$ by the choice of $\alpha \in (0, 1 - \frac{1}{\sqrt[3]{2}})$. This ensures that the random walk moves toward x^* at a geometric rate with a probabilistic guarantee as specified in Lemma 2.

Let $x_{(m)}$ denote the sampling point at the mth time that the local test is called by RWGD. For example, the first step of the random walk is taken based on the first three tests carried out on points $x_{(1)} = 0, x_{(2)} = 0.5, x_{(3)} = 1$; the second step of the random walk is taken based on the three tests carried out on points $x_{(4)}, x_{(5)}, x_{(6)}$, and so on. Let $\Delta_x = |x - x^*|$ denote the distance between $x \in \mathcal{X}$ and x^*. Lemma 2 establishes an upper bound on $\Delta_{x_{(m)}}$ after n steps are taken by the random walk.

Lemma 2. *For the sampling points at step $n + 1$ of the random walk in RWGD ($m = 3n + 1, 3n + 2, 3n + 3$), with probability at least $1 - \exp(-\frac{n(2p-1)^2}{2})$,*

$$\Delta_{x_{(m)}} \leq 2^{-\frac{n(2p-1)}{2}} \tag{7}$$

Proof: See Appendix B.

Proof of Theorem 1. Let τ_m denote the number of samples taken in the mth time that the local sequential test is carried out; τ_1 samples are taken at point $x_{(1)}$, τ_2 samples are taken at point $x_{(2)}$ and so on. Let t_n denote the time at the end of the nth step of the random walk: $t_n = \sum_{m=1}^{3n} \tau_m$. Notice that both τ_m and t_n are random variables were the randomness comes from the randomness in the samples of g. Define $n_0 = \max\{n \in \mathbb{N} : n \leq \frac{3}{(2p-1)^2} \log t_n, t_n \leq T\}$ and $t_0 = t_{n_0}$. The definition of n_0 and t_0 indicates that at $t > t_0$, the random walk has taken more than $\frac{3}{(2p-1)^2} \log t$ steps. We analyze the regret incurred up to time t_0, and after that, separately.

$$
\begin{aligned}
&\mathcal{R}_{\mathrm{RWGD}}(T) \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \left(f(x_{\pi(t)}) - f(x^*))\right)\right] \\
&= \underbrace{\mathbb{E}\left[\sum_{t=1}^{t_0} \left(f(x_{\pi(t)}) - f(x^*))\right)\right]}_{R_1} \\
&\quad + \underbrace{\mathbb{E}\left[\sum_{t=t_0+1}^{T} \left(f(x_{\pi(t)}) - f(x^*))\right)\right]}_{R_2}.
\end{aligned}
\tag{8}
$$

Next, we establish an upper bound on each term of the regret.

Upper bound on the first term R_1. From Lemma 1, we have that τ_m satisfies

$$\mathbb{E}[\tau_m] \leq \frac{16}{\mathbb{E}[g(x_{(m)})]^2 \xi} \log \frac{8\sqrt{\frac{1}{\alpha}}}{\xi \mathbb{E}[g(x_{(m)})]^2} + 2. \tag{9}$$

Based on this upper bound on $\mathbb{E}[\tau_m]$ and by convexity of f we have

$$R_1 \le \sum_{m=1}^{3n_0} \left(\frac{16|\mathbb{E}[g(x_{(m)})]||\Delta_{x_{(m)}}}{\mathbb{E}[g(x_{(m)})]^2\xi} \log \frac{8\sqrt{\frac{1}{\alpha}}}{\xi\mathbb{E}[g(x_{(m)})]^2} \right.$$
$$\left. + 2|\mathbb{E}[g(x_{(m)})]||\Delta_{x_{(m)}} \right). \tag{10}$$

Noticing the constraint that $\sum_{m=1}^{3n_0}\mathbb{E}[\tau_m] = \mathbb{E}[t_0] \le T$ and using $\Delta_x \le 1$, the following constrained optimization problems gives us an upper bound on R_1.

$$maximize_{\{\mathbb{E}[g(x_{(m)})]\}_{m=1}^{3n_0}} \sum_{m=1}^{3n_0} \left(\right.$$
$$\left. \frac{16|\mathbb{E}[g(x_{(m)})]|}{\mathbb{E}[g(x_{(m)})]^2\xi} \log \frac{8\sqrt{\frac{1}{\alpha}}}{\xi\mathbb{E}[g(x_{(m)})]^2} + 2|\mathbb{E}[g(x_{(m)})]| \right),$$
$$subject\ to: \sum_{m=1}^{3n_0}\mathbb{E}[\tau_m] \le T. \tag{11}$$

We can prove that the values of $\mathbb{E}[g(x_{(m)})] = \sqrt{\frac{48n_0\log(8\sqrt{\frac{1}{\alpha}}\frac{T}{3n_0\xi})}{\xi T}}$ (for all m) gives an upper bound on the above optimization problem which by $n_0 \le \frac{3}{(2p-1)^2}\log t_0 \le \frac{3}{(2p-1)^2}\log T$ results in

$$R_1 \le \frac{12}{2p-1}\sqrt{\frac{3}{\xi}T\log T \log(8\sqrt{\frac{1}{\alpha}}\frac{T(2p-1)^2}{9\xi\log T})}. \tag{12}$$

Upper bound on the second term R_2. At time $t : t_0 < t \le T$, by definition of t_0 we know that the random walk has taken more than $\frac{3}{(2p-1)^2}\log t$ steps. From Lemma 2, we have with probability at least $1 - \exp(-\frac{\frac{3}{(2p-1)^2}\log(t)(2p-1)^2}{2}) = 1 - \left(\frac{1}{t}\right)^{\frac{3}{2}}$,

$$\Delta_{x_{\pi(t)}} \le 2^{-\frac{\frac{3}{(2p-1)^2}\log(t)(2p-1)}{2}}$$
$$= \left(\frac{1}{t}\right)^{\frac{3\log 2}{2(2p-1)}}$$
$$\le \frac{1}{t}, \tag{13}$$

where the last inequality is obtained by $p < 1$ and $\frac{3\log 2}{2} < 1$.

The second term R_2 in the regret is upper bounded as follows.

$$
\begin{aligned}
R_2 &= \mathbb{E}\left[\sum_{t=t_0+1}^{T}\left(f(x_{\pi(t)}-f(x^*))\right)\right] \\
&= \mathbb{E}\left[\sum_{t=t_0+1}^{T}\left(f(x_{\pi(t)}-f(x^*))\right)\mathbb{I}[\Delta_{x_{\pi(t)}}\leq\frac{1}{t}]\right] \\
&\quad + \mathbb{E}\left[\sum_{t=t_0+1}^{T}\left(f(x_{\pi(t)}-f(x^*))\right)\mathbb{I}[\Delta_{x_{\pi(t)}}>\frac{1}{t}]\right] \\
&\leq \sum_{t=1}^{T}G\left(\frac{1}{t}+\left(\frac{1}{t}\right)^{\frac{3}{2}}\right) \\
&\leq G(\log T + 4).
\end{aligned}
\tag{14}
$$

In the above inequalities $\mathbb{I}$ is the indicator function. We used the convexity of $\mathbb{E}[f]$ and $\Delta_x \leq 1$ to arrive at (14).

From (12) and (14), we have

$$
\mathcal{R}_{\mathrm{RWGD}}(T) \leq
$$

$$
\frac{12}{2p-1}\sqrt{\frac{3}{\xi}T\log T\log(8\sqrt{\frac{1}{\alpha}\frac{T(2p-1)^2}{9\xi\log T}})} + G(\log T + 4),
$$

which concludes the proof. $\qquad\square$

Remark 1. *Let $\bar{x}_{\pi(t)} = \frac{1}{t}\sum_{s=1}^{t}x_{\pi(s)}$. By Jensen's inequality*

$$
\mathbb{E}[f(\bar{x}_{\pi(t)}) - f(x^*)] \leq \frac{1}{t}\sum_{s=1}^{t}\mathbb{E}[f(x_{\pi(s)}) - f(x^*)].
\tag{15}
$$

Theorem 1 thus shows that $f(\bar{x}_{\pi(T)})$ converges to $f(x^*)$, at an $\tilde{\mathcal{O}}(\frac{1}{\sqrt{T}})$ rate.

5 Simulation

In this section, we compare the performance of RWGD and the standard SGD in simulations. The expected value of the functions is chosen as $\mathbb{E}[f(x)] = a|x-x^*|^b$. The gradient g is thus in the form of $g(x) = ab\mathrm{SIGN}(x-x^*)|x-x^*|^{b-1} + \mathcal{N}_x(0,1)$, where $\mathcal{N}_x(0,1)$ is a random variable with a normal distribution with mean 0 and variance 1 for each $x \in \mathcal{X}$. We vary the signal to noise ratio in $g(x)$ using parameter a (this is equivalent to changing the magnitude of noise).
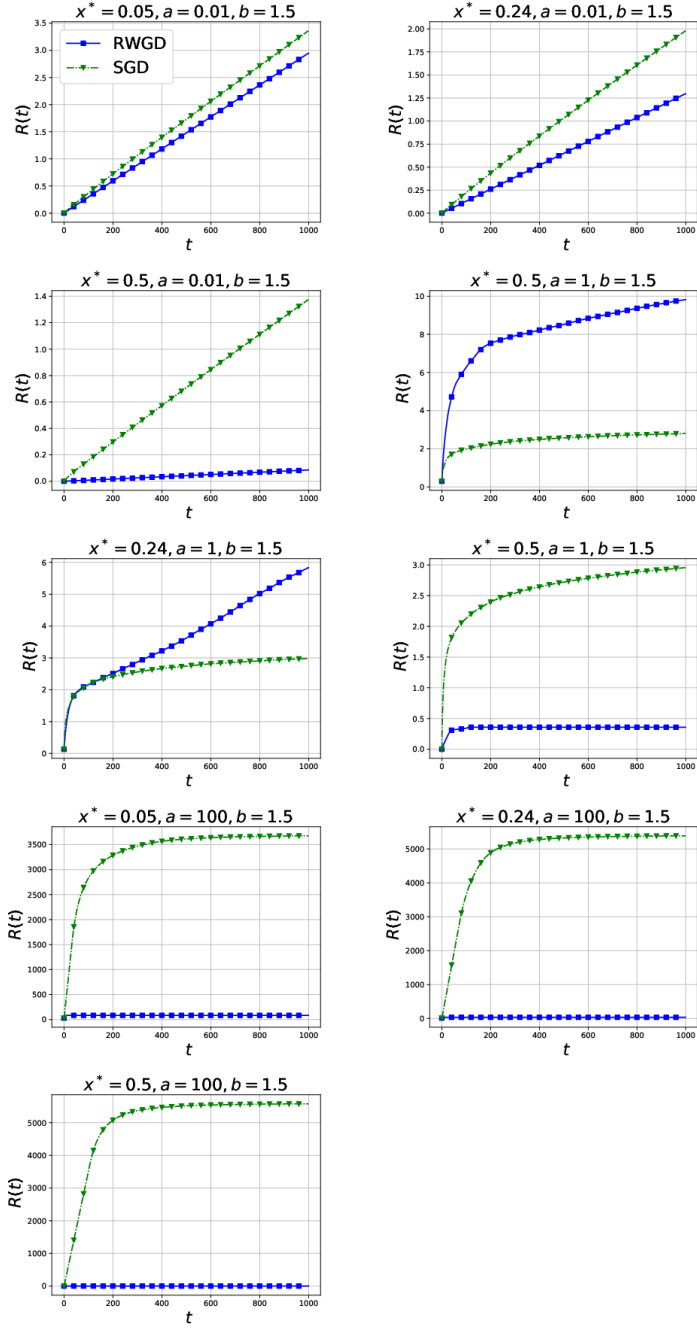
11

Figure 3: Comparison of the performance of RWGD and SGD ($b = 1.5$).

In SGD, a sequence of sampling points $\{y_t\}_{t=1}^{\infty}$ is generated according to

the following rule: $y_t = y_{t-1} - \theta_t g(y_{t-1})^1$, where y_0 is randomly chosen, and $\{\theta_t\}_{t=1}^{\infty}$ is a sequence of chosen step sizes. The convergence of $\{y_t\}_{t=1}^{\infty}$ is strongly dependent on $\{\theta_t\}_{t=1}^{\infty}$. In particular, in addition to the the the assumptions on the distribution of $g(x)$, for convergence of $\{y_t\}_{t=1}^{\infty}$, it is required that $\sum_{t=1}^{\infty} \theta_t \to \infty$ and $\sum_{t=1}^{\infty} \theta_t < C$ for some constant $C \in \mathbb{R}$ [6]. In our simulations, we use $\theta_t = \frac{1}{t}$ for SGD, which satisfies the requirement and we set $y_0 = 0.5$ (SGD does not show considerable sensitivity to the initial point in these examples). The parameter α in RWGD is set to 0.1.

---

[1]If the value of y_t in SGD is not in $\mathcal{X}$, it is projected to $\hat{y}_t = \operatorname{argmin}_{y \in \mathcal{X}} \{|y - y_t|\}$.
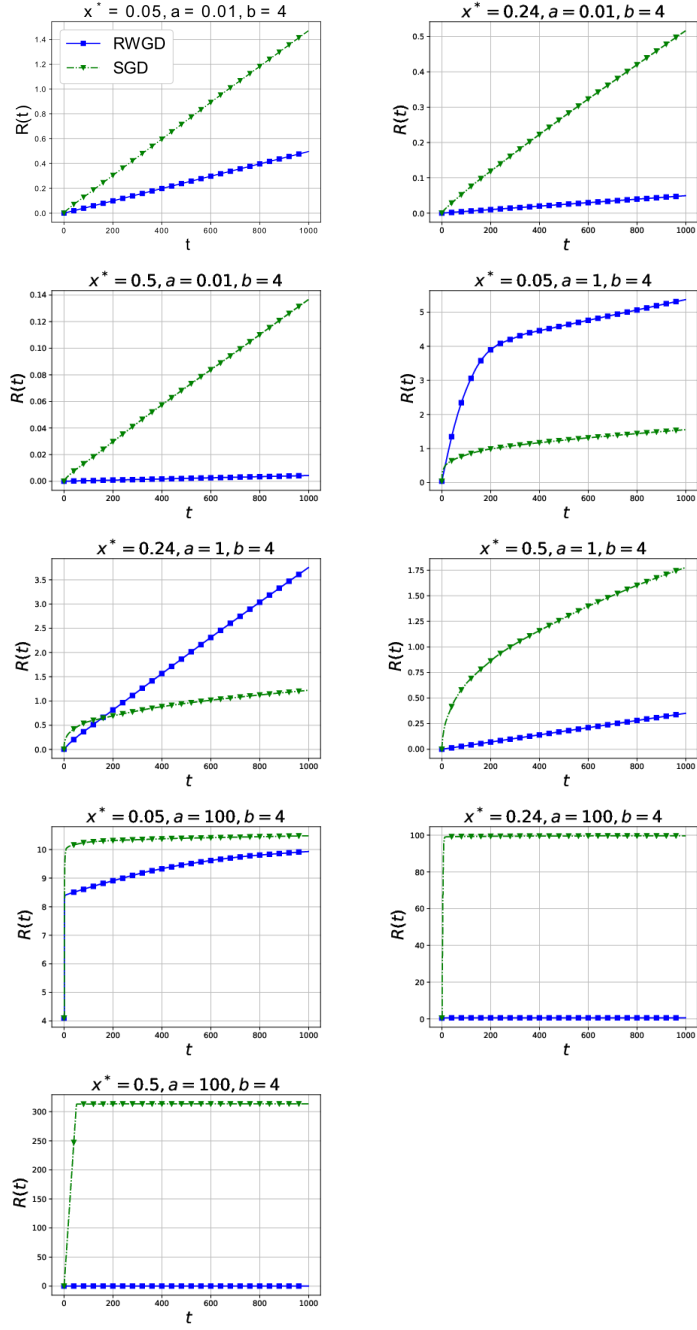
Figure 4: Comparison of the performance of RWGD and SGD ($b = 4$).

As shown in Figures 3 and 4, RWGD outperforms SGD in most cases. In

contrast to SGD, whose performance strongly depends on the sequence of the step sizes, RWGD always performs well.

6 Conclusion

We introduced a novel policy for the stochastic convex online learning problem based on constructing a search tree and inducing an efficient random walk on the tree. We established a finite-time regret analysis of the proposed policy which ensures an $\tilde{\mathcal{O}}(\sqrt{T})$ regret at any finite time T. The low computational complexity and memory requirement of the proposed policy make it suitable for a variety of applications. A potential future direction is the applicability of the proposed policy to non-convex online learning problems.

References

[1] H. Robbins and S. Monro "A stochastic approximation method" *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951.

[2] J. Kiefer, J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function," *Ann. Math. Statist.*, vol. 23, no. 3, pp. 462-466, 1952.

[3] T. Lai, "Stochastic Approximation," *Annuals of Statistics,* no 31, pp. 391–406, 2003.

[4] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," *in ICML '03: Proceedings of the 20th International Conference on Machine Learning*, pp. 928–936, 2003

[5] E. Cope, "Regret and Convergence Bounds for a Class of Continuum-Armed Bandit Problems," *IEEE TRANSACTIONS ON AUTOMATIC CONTROL,* vol. 54, no. 6, 2009.

[6] R. Pasupathy, S. Kim, "The stochastic root-finding problem: overview, solutions, and open questions," *ACM Transactions on Modeling and Computational Simulations*, vol. 21, no. 3, pp. 19, 2011.

[7] P. I. Frazier, S. G. Henderson, R. Waeber, "Probabilistic bisection converges almost as quickly as stochastic approximation", available at *arXiv:1612.03964v1 [math.PR]*, 2016.

[8] E. Hazan, A. Agrawal, S. Kale, "Logarithmic regret algorithms for online convex optimization," *Machine Learning*, vol. 69, no. 2-3, pp. 169-192, 2007.

[9] A. Agrawal, D.P. Foster, D. Hsu, S.M. Kakade, A. Rakhlin, "Stochastic convex optimization with bandit feedback," *Advances in Neural Information Processing Systems, 24*, 2011.

[10] R. Agrawal, "The continuum-armed bandit problem," *SIAM journal on control and optimization*, no. 33, pp. 1926, 1995.

[11] R. Kleinberg, "Nearly tight bounds for the continuum-armed bandit problem," *Advances in Neural Information Processing Systems, 18*, 2005.

[12] R. Kleinberg, A. Slivkins, and E. Upfal, "Multi-armed bandits in metric spaces," *In Proceedings of the 40th annual ACM symposium on Theory of computing,* pp. 681–690. ACM, 2008.

[13] S. Bubeck, R. Munos, G. Stolz, C. Szepesvari, "$\mathcal{X}$-armed Bandits," *Journal of Machine Learning Research 12.* pp. 1655-1695. 2011.

[14] E. Hazan, "Introduction to Online Convex Optimization," *Foundations and Trends in Optimization*, vol. 2, no. 3-4, pp 157-325, 2016.

[15] S. Shalev-Shwartz, "Online Learning and Online Convex Optimization," *Foundations and Trends in Optimization*, vol. 4, no. 2, pp 107-194, 2012.

[16] S. Vakili, Q. Zhao, C. Liu, C.N. Chuah, "Hierarchical Heavy Hitter Detection under Unknown Models," *International Conference on Acoustics, Speech, and Signal Processing*, 2018.

[17] R. G. Antonioni, Y. Kozachenko, A. Volodin, "Convergence of series of dependent ϕ-subgaussian random variables," *Journal of Mathematical Analysis and Applications*, vol. 338, no. 2, pp. 1188-1203, 2008.

[18] S. Bubeck, N. Cesa-Bianchi, G. Lugosi, "Bandits with Heavy Tail," *IEEE Transaction on Information Theory*, vol. 59, pp. 7711-7717, 2013.

Appendix A

Proof of Lemma 1. The proof of Lemma 1 is based on concentration inequalities for Sub-Gaussian distributions. We prove inequalities (4) and (5). The same results under the other case, $\mu < 0$, can be proven similarly. Let $s_0 = \lceil \frac{16}{\mu^2 \xi} \log \frac{8\sqrt{\frac{1}{\alpha}}}{\mu^2 \xi} \rceil$. We have,

$$
\begin{aligned}
\sqrt{\frac{\log \frac{s_0^2}{\alpha}}{\xi s_0}} \;&=\; \sqrt{\frac{\log \frac{\lceil \frac{16}{\mu^2 \xi} \log \frac{8\sqrt{\frac{1}{\alpha}}}{\mu^2 \xi} \rceil^2}{\alpha}}{\xi \lceil \frac{16}{\mu^2 \xi} \log \frac{8\sqrt{\frac{1}{\alpha}}}{\mu^2 \xi} \rceil}} \\[2ex]
&\leq\; \sqrt{\frac{\log \frac{(\frac{16}{\mu^2 \xi} \log \frac{8\sqrt{\frac{1}{\alpha}}}{\mu^2 \xi})^2}{\alpha}}{\xi (\frac{16}{\mu^2 \xi} \log \frac{8\sqrt{\frac{1}{\alpha}}}{\mu^2 \xi})}} \qquad (16) \\[2ex]
&=\; \sqrt{\frac{2\log \frac{8\sqrt{\frac{1}{\alpha}}}{\mu^2 \xi} + 2\log\log(\frac{8\sqrt{\frac{1}{\alpha}}}{\mu^2 \xi})^2}{\frac{16}{\mu^2} \log \frac{8\sqrt{\frac{1}{\alpha}}}{\mu^2 \xi}}} \\[2ex]
&=\; \sqrt{\frac{\log(\frac{8\sqrt{\frac{1}{\alpha}}}{\mu^2 \xi})^2 + \log(\frac{8\sqrt{\frac{1}{\alpha}}}{\mu^2 \xi})^2}{\frac{8}{\mu^2} \log(\frac{8\sqrt{\frac{1}{\alpha}}}{\mu^2 \xi})^2}} \qquad (17) \\[2ex]
&=\; \sqrt{\frac{\mu^2}{4}} \\[1ex]
&=\; \frac{\mu}{2}. \qquad (18)
\end{aligned}
$$

Inequality (16) holds because $\frac{\log \sqrt{\frac{x^2}{\alpha}}}{x}$ is a decreasing function in x for $x \geq e\sqrt{\alpha}$ and (17) holds because for all $x > 0$, $\log\log x < \frac{1}{2} \log x$. We thus have

$$
\begin{aligned}
\mathbb{P}&\left[\overline{Z}_\tau + \sqrt{\frac{\log \frac{\tau^2}{\alpha}}{\xi \tau}} < 0 \right] \\[1ex]
&\leq\; \sum_{s=2}^{\infty} \mathbb{P}\left[\overline{Z}_s + \sqrt{\frac{\log \frac{s^2}{\alpha}}{\xi s}} < 0, \forall s' \leq s : \overline{Z}_{s'} - \sqrt{\frac{\log \frac{s'^2}{\alpha}}{\xi s'}} < 0 \right] \\[1ex]
&\leq\; \sum_{s=2}^{s_0-1} \mathbb{P}\left[\overline{Z}_s + \sqrt{\frac{\log \frac{s^2}{\alpha}}{\xi s}} < 0 \right] + \mathbb{P}\left[\overline{Z}_{s_0} - \sqrt{\frac{\log \frac{s_0^2}{\alpha}}{\xi s_0}} < 0 \right] \\[1ex]
&\leq\; \sum_{s=2}^{s_0-1} \exp(-\log \frac{s^2}{\alpha}) + \exp(-\log \frac{s_0^2}{\alpha}) \qquad (19) \\[1ex]
&\leq\; \sum_{s=2}^{s_0} \frac{\alpha}{s^2} \\[1ex]
&\leq\; \sum_{s=2}^{\infty} \frac{\alpha}{s^2} \\[1ex]
&\leq\; \alpha,
\end{aligned}
$$

17

where inequity (19) is based on the Chernoff-Hoeffding bound and (18). For $n \geq s_0$,

$$
\begin{aligned}
\mathbb{P}[\tau > n] & \\
\leq \quad & \Pr\left[\forall s \leq n : \overline{Z}_s + \sqrt{\frac{\log \frac{s^2}{\alpha}}{\xi s}} > 0, \text{ and } \overline{Z}_s - \sqrt{\frac{\log \frac{s^2}{\alpha}}{\xi s}} < 0\right] \\
\leq \quad & \Pr\left[\forall s \leq n : \overline{Z}_s - \sqrt{\frac{\log \frac{s^2}{\alpha}}{\xi s}} < 0\right] \\
\leq \quad & \mathbb{P}\left[\overline{Z}_n - \mu < -\frac{\mu}{2}\right] \quad (20) \\
\leq \quad & \exp(-\xi n \mu^2 / 4). \quad (21)
\end{aligned}
$$

The inequality (20) holds based on (18), for $n \geq s_0$, and the inequality (21) holds based on the Chernof-Hoeffding bound. We can write $\mathbb{E}[\tau]$ in terms of the sum of the tail probabilities $\mathbb{P}[\tau \geq n]$ as

$$
\begin{aligned}
\mathbb{E}[\tau] \quad = \quad & \sum_{n=0}^{\infty} \mathbb{P}[\tau \geq n] \\
= \quad & s_0 + \sum_{n=s_0+1}^{\infty} \mathbb{P}[\tau \geq n] \\
\leq \quad & s_0 + \sum_{n=s_0+1}^{\infty} \exp(-\xi n \mu^2 / 4) \\
\leq \quad & s_0 + \frac{4}{\xi \mu^2} \exp(-\xi s_0 \mu^2 / 4) \\
\leq \quad & s_0 + 1,
\end{aligned}
$$

which completes the proof by substituting s_0.

$\square$

Appendix B

Proof of Lemma 2. We define the value W_n of each step n of the random walk as $W_n = 1$ if the random walk moves in the right direction, i.e., the random walk moves to the child who contains x^* or moves to the parent if neither of the children contain x^*; and $W_n = -1$ otherwise. We also use $V_n = \sum_{m=1}^{n} W_m$ to denote the cumulative value of the steps.

The condition for $W_n = 1$ is that the result of all three local sequential tests at step n are true. Thus, as a result of Lemma 1, we have $W_n = 1$ with probability $p \geq (1 - \alpha)^3 > \frac{1}{2}$ which indicates $\mathbb{E}[W_n] \geq 2p - 1 > 0$. The positive expected value of each step in the random walk indicates the random walk is

more likely to move closer to x^* rather than away from it. In particular,

$$\Pr[V_n \leq \frac{n(2p-1)}{2}]$$

$$= \Pr[V_n - n(2p-1) \leq -\frac{n(2p-1)}{2}]$$

$$\leq \Pr[V_n - \mathbb{E}[V_n] \leq -\frac{n(2p-1)}{2}]$$

$$\leq \Pr[\frac{V_n}{n} - \frac{\mathbb{E}[V_n]}{n} \leq -\frac{(2p-1)}{2}]$$

$$\leq \exp(-\frac{n(2p-1)^2}{2}).$$

The last inequality is based on Hoeffding inequality on independent Bernoulli random variables W_n. Each time the random walk moves one step in the right direction the value of $\Delta_{x_{(m)}}$ in the local sequential tests is divided by half. For example when $V_n = 0$ we have, trivially, $\Delta_{x_{(m)}} \leq 1$. If the random walk moves to the child who contains x^* then we have $\Delta_{x_{(m)}} \leq 0.5$, and so on. Thus, for $V_n > \frac{n(2p-1)}{2}$, we have, for $m = 3n+1, 3n+2, 3n+3$,

$$\Delta_{(m)} \leq 2^{-V_n}$$

$$\leq 2^{-\frac{n(2p-1)}{2}},$$

which completes the proof.

$\square$