Journal of Machine Learning Research 23 (2022) 1-47 Submitted 9/19; Revised 12/21; Published 3/22

Stochastic Zeroth-Order Optimization under Nonstationarity and Nonconvexity

Abhishek Roy ABROY@UCDAVIS.EDU
Department of Statistics
University of California
Davis, CA 95616, USA

Krishnakumar Balasubramanian KBALA@UCDAVIS.EDU
Department of Statistics
University of California
Davis, CA 95616, USA

Saeed Ghadimi SGHADIMI@UWATERLOO.CA
Department of Management Sciences
University of Waterloo
Waterloo, ON N2L 3G1, Canada

Prasant Mohapatra PMOHAPATRA@UCDAVIS.EDU
Department of Computer Science
University of California
Davis, CA 95616, USA

Editor: Andreas Krause

Abstract

Stochastic zeroth-order optimization algorithms have been predominantly analyzed under the assumption that the objective function being optimized is time-invariant. Motivated by dynamic matrix sensing and completion problems, and online reinforcement learning problems, in this work, we propose and analyze stochastic zeroth-order optimization algorithms when the objective being optimized changes with time. Considering general nonconvex functions, we propose nonstationary versions of regret measures based on first-order and second-order optimal solutions, and provide the corresponding regret bounds. For the case of first-order optimal solution based regret measures, we provide regret bounds in both the low- and high-dimensional settings. For the case of second-order optimal solution based regret, we propose zeroth-order versions of the stochastic cubic-regularized Newton's method based on estimating the Hessian matrices in the bandit setting via second-order Gaussian Stein's identity. Our nonstationary regret bounds in terms of second-order optimal solutions have interesting consequences for avoiding saddle points in the nonstationary setting.

Keywords: nonstationary and nonconvex optimization, regret measures, stochastic zeroth-order algorithms, online cubic-Newton method

1. Introduction

Consider the canonical optimization problem of minimizing a function $f(x) = \mathbf{E}_\xi[F(x,\xi)]$ using an iterative algorithm. In the stochastic zeroth-order setup, for each iteration t, the optimizer has a guess x_t for the minimum value, based on which we obtain noisy function evaluations of the form $F(x_t, \xi_t)$ at the point x_t. Based on this feedback received, the point x_t is updated so that it is closer to the minimum value of $f(x)$, or the algorithm is terminated if we are already sufficiently close to the minimizer. Such stochastic zeroth-order optimization algorithms have been studied for several decades. We refer the interested reader to Spall (2005); Conn et al. (2009); Brent (2013); Zabinsky (2013); Audet and Hare (2017); Larson et al. (2019) for details regarding more recent progress, and applications to statistical machine learning, simulation-based optimization and operations research. An important aspect of the above stochastic zeroth-order optimization setup is the stationarity aspect – the objective function being optimized stays fixed during the course of the iterative optimization process.

A practical variant of the above setup is that of *nonstationary* stochastic zeroth-order optimization, where we have a sequence of functions $f_t(x) = \mathbf{E}_\xi[F_t(x, \xi)]$ to be optimized with the corresponding minimizers defined as

$$x_t^* := \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ f_t(x) = \mathbf{E}_\xi[F_t(x, \xi)] \right\}. \tag{1}$$

Here, $f_t : \mathbb{R}^d \to \mathbb{R}$ and $\mathcal{X} \subset \mathbb{R}^d$ is convex and compact. In each iteration, the optimizer picks a point x_t and observes (several) noisy function evaluations $F_t(x_t, \xi_t)$ at the picked point, *a posteriori*. The goal of the optimizer is then to select points x_t eventually to minimize the so-called *regret*, which compares the accumulated error over all T rounds, against the error suffered by a certain *oracle optimal rule* that could be computed only knowing all the functions, *a priori*. In the most well-studied setting of this sequential stochastic optimization problem, the functions f_t are assumed to be convex and the *oracle decision rule* compared against, is chosen to be a fixed rule $\bar{x}^* := \operatorname{argmin}_{x \in \mathcal{X}} \sum_{t=1}^T f_t(x)$. In this case, a natural notion of regret is given by $\mathcal{R} = \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(\bar{x}^*)$. It is easy to see that the regret of any non-trivial decision rule should grow sub-linearly in T and several algorithms exists for attaining such regret – we refer the reader to Flaxman et al. (2005); Cesa-Bianchi and Lugosi (2006); Hazan et al. (2007); Agarwal et al. (2010, 2011); Saha and Tewari (2011); Bubeck et al. (2012); Shamir (2013, 2017); Bubeck et al. (2017) for a non-exhaustive overview of such algorithms and their optimality properties under different assumptions on f_t. Another natural way to measure the performance of sequential stochastic optimization algorithms is to compare against the sequence of minimal vectors $\{x_t^*\}_{t=1}^T$ directly. In this case, the *nonstationary regret* is defined as $\mathcal{R} = \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(x_t^*)$ (Bousquet and Warmuth, 2002; Hazan and Seshadhri, 2009; Besbes et al., 2014, 2015; Hall and Willett, 2015; Yang et al., 2016). Indeed, to obtain sub-linear regret in this setting, typically the degree of allowed nonstationarity in terms of either the functions or the minimal vectors needs to be bounded; see, for example Besbes et al. (2015); Yang et al. (2016).

In this paper, we consider stochastic zeroth-order optimization under both *nonstationary* and *nonconvexity*. Several issues arise when considering such problems. The fundamental one is that of defining an appropriate notion of regret under both nonstationarity and non-convexity. Note that when the objective functions are nonconvex, it is computationally hard to obtain a globally optimal value; see for example Murty and Kabadi (1987). Furthermore, even ignoring the computational hardships and allowing for unbounded computational resources, in the stochastic zeroth-order setting, without further assumptions, the number of function queries required to obtain a (approximate) global minimizer scales exponential in the dimensionality (Novak, 2006; Novak and Woźniakowski, 2008). Hence, the notion of function value based regret discussed above for the case of convex functions is not so meaningful from a computational and statistical point of view in this case. As a way forward, it becomes important to define notions of regret that convey computationally and statistically meaningful information by leveraging the structure available in the problem. In this work, we propose local optimality based regret measures – specifically ones that are based on approximate first- or second-order optimal solutions – considering general smooth nonconvex function. Our proposal is motivated by the use of such measures in the stochastic nonconvex optimization literature (Nesterov and Spokoiny, 2017; Ghadimi and Lan, 2013; Balasubramanian and Ghadimi, 2021). In order to obtain meaningful bounds for such local optimality based regret measures, it turns out that, similar to the convex case, controlling the allowed degree of nonstationarity in a delicate manner becomes crucial. It is worth remarking here that Hazan et al. (2017) considered first and second-order optimization under *nonstationarity* and *nonconvexity*, and showed that only trivial regret bounds are possible without controlling the allowed degree of nonstationarity. For the above mentioned notions of *local convergence* based nonstationary regret, in this work, we propose and analyze stochastic zeroth-order algorithms and characterize the precise dependence of the regret bounds of the algorithms on the allowed degree of nonstationarity. To our knowledge, our work provides the first non-trivial regret bounds for stochastic zeroth-order optimization under both *nonstationarity* and *nonconvexity*. We next provide some motivating examples for our proposed regret measures.

Motivating Application I: Dynamic Matrix Completion. Low-rank matrix factorization and completion arise in a variety of signal processing and machine learning applications. In the simplest setting, the problem is to recover an unknown matrix $X \in \mathbb{R}^{n_1 \times n_2}$ which is assumed to be of rank $r \ll \min(n_1, n_2)$, given m observations through a (linear) random operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$. If we let $y \in \mathbb{R}^m$ to be the ouput of this linear operator $\mathcal{A}$, then the matrix X is recovered by minimizing $\|y - \mathcal{A}(X)\|_2^2$ over all rank-r matrices. Due to the nonconvexity of the low-rank constraint, a popular approach is to re-parametrize the problem as $X = UV^\top$ for $U \in \mathbb{R}^{n_1 \times r}$ and $V \in \mathbb{R}^{n_2 \times r}$. Then, several recent works have shown that for the objective $\|y - \mathcal{A}(UV^\top)\|_2^2$, all local-minimzers are approximate global minimizers under certain regularity conditions on the operator $\mathcal{A}$. See Ge et al. (2016); Chi et al. (2019); Zhu et al. (2021); Zhang (2021); Zhang et al. (2021) for details.

A main application of the above problem setup is recommendation systems. Here, n_1 users give ratings to a random subset of n_2 items, which are put together in the form an incomplete matrix. Then using matrix completion techniques, new items are recommended to the users. However, as noted in several works, user's preference change over time (Xu and Davenport, 2016; Gultekin and Paisley, 2018; Xu and Davenport, 2017; Lee et al., 2016; Fattahi et al., 2020). A more practical variant of the above matrix completion problem, which takes explicitly the variations into account, is applicable to recommendation problems arising in practice. In this setup, we are given a sequence of T matrices $X_1, \ldots X_T$ which are observed through time-varying linear random operators $\mathcal{A}_1, \ldots, \mathcal{A}_T$. Denoting the corresponding outputs as $y_1, \ldots, y_T$, the nonconvex and nonstationary version of the matrix completion problem is given an optimization problem where the objective function $f_t = \|y_t - \mathcal{A}_t(U_t V_t^\top)\|_2^2$ changes over time. Indeed, it is natural to assume here that the objective functions f_t do not change too abruptly as user's preferences invariably change smoothly over time. Hence, the problem discussed above fits the setup that we consider in this paper thereby serving as motivating example. The results we provide in Section 3.2, in combination with landscape results discussed above would lead to global sub-linear regret bounds for the time-varying matrix completion problem. A detailed investigation of this is left as future work.

Motivating Application II: Deep Markov Decision Process. Another motivating application for the nonstationary nonconvex setting that we consider is the problem of Markov Decision Process (MDP) that arise in reinforcement learning, a canonical sequential decision making problem (Sutton and Barto, 2018). An MDP M is parametrized by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, c)$. Here, $\mathcal{S} \subset \mathbb{R}^b$ and $\mathcal{A} \subset \mathbb{R}^p$ denote the state and action space respectively, $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$ denotes be the transition probability kernel and $c(s, a) : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ denotes the cost function. The goal of an agent working with the MDP M, at a given time step t, is to choose an action a_t based on data $\{s_i, a_i, c(s_i, a_i)\}_{i=1}^{t-1}$ and s_t. The agent does so by minimizing the cost (given by c) over time. Based on the actions chosen, the process moves to state s_{t+1} with probability $\mathcal{P}(s_{t+1}|a_t, s_t)$. To formulate the problem precisely, we introduce the so-called policy function, $\pi_\theta(a|s) \equiv \pi_\theta(a, s) : \mathcal{A} \times \mathcal{S} \to [0, 1]$, which denotes the probability of taking action a in state s. Here, $\theta \in \mathbb{R}^d$ is a parameter vector of the policy function. Then, the precise formulation of the problem describing the goal of the agent is given by the following offline optimization problem.

$$\theta^* = \min_{\theta \in \Theta} \left\{ J(\theta) = \mathbf{E}_s \left[V_\theta(s) \right] = \mathbf{E}_s \left[\mathbf{E} \left(\sum_{i=1}^{t} c(s_i, a_i) \middle| s_1 = s \right) \right] \right\},$$

where $a_i \sim \pi_\theta(\cdot|s_i)$ and $s_{i+1} \sim \mathcal{P}(\cdot|s_i, a_i)$, for all $1 \le i < t$ and $\mathbf{E}_s$ represents the (fixed) initial distribution of the states. The quantity $V_\theta(s)$ is called as the value function and it is indexed by θ to represent the fact that it depends the policy function π_θ. Policy gradient method (Williams, 1992; Sutton and Barto, 2018) is a popular algorithm for solving the above problem. Recently, it has been realized that parametrizing π_θ by a deep neural network leads to better results empirically; see, for example Haarnoja et al. (2017); Li (2017).

In the online nonstationary version of the MDP problem above, there are two significant changes to the above setup (Neu et al., 2010; Arora et al., 2012; Guan et al., 2014; Dick et al., 2014). First, the cost function c is assumed to change with time and is hence indexed by c_t. Next, the interaction protocol of the agent is changed so that at time t, the agent receives s_t and selects action a_t based on which it receives the cost $c_t(s_t, a_t)$. The probability kernel $\mathcal{P}$ is typically assumed to be known in Online MDP problems (Neu et al., 2010; Dick et al., 2014). The goal in online nonstationary MDP is to come up with a sequence of policies $\pi_{\theta_t^*}$ to minimize an appropriately defined notion of static or dynamic (nonstationary) regret. Clearly this falls under the category of sequential decision making problem described in Equation 1. If the objective function is convex, then existing results on nonstationary online convex optimization could be leveraged to provide regret bounds in this setting. But if the optimization problem involved is nonconvex, there is a lack of a clear notion of regret to work with, to the best of our knowledge. The results we provide in Sections 3.2, in combination with landscape results about neural networks (for example, Kawaguchi et al. (2019)) would lead to global sub-linear regret bounds for nonconvex online MDP problems (which is the case when the policies π_θ are parametrized by deep neural networks). A detailed investigation of this is left as future work.

We end this section with the following remark. For both the motivating examples, it might be possible to obtain refined bounds taking into account and leveraging the structure specific to the respective problem. Our results in this paper are provided for any general functions that satisfy mild smoothness assumptions. The provided results extend similar general results in the stochastic zeroth-order optimization literature to the nonconvex and nonstationary setting.

2. Assumptions and Contributions

We first outline the basic notations that we use in this paper. For a function $f : \mathbb{R}^d \to \mathbb{R}$, we denote the sup-norm as $\|f\|_\infty := \sup_{x\in\mathbb{R}^d}|f(x)|$. Furthermore, we denote the gradient vector and the Hessian matrix at a point $x \in \mathbb{R}^d$ as $\nabla f(x) \in \mathbb{R}^d$ and $\nabla^2 f(x) \in \mathbb{R}^{d\times d}$. For the stochastic function $F(x, \xi)$, we denote its partial gradient and Hessian with respect to x by $\nabla F(x, \xi)$ and $\nabla^2 F(x, \xi)$, respectively. For a vector $a \in \mathbb{R}^d$, we use $\|a\|$, and $\|a\|_*$ to denote a norm and the corresponding dual norm on $\mathbb{R}^d$. For a matrix $A \in \mathbb{R}^{d\times d}$, we use $\|A\|_F$ and $\|A\|_{op}$ to denote the Frobenius norm and operator norm respectively. We also use $\lambda_{\min}(A)$ to denote the minimum eigenvalue of A. In the rest of this section, we first discuss the assumptions we use in this work, after which we introduce the stochastic zeroth-order gradient and Hessian estimators that we use in this work. Next, we introduce the notions of regret that we propose in this work. We conclude the section by highlighting the main contributions that we make in this work regarding the regret bounds.

2.1 Assumptions

We now state the assumption on the stochastic zeroth-order oracle we make in this work.

Assumption 2.1 (Zeroth-order oracle) *For any $x \in \mathbb{R}^d$, the zeroth order oracle outputs an estimator $F(x, \xi)$ of $f(x)$ such that*

$$\mathbf{E}\left[F(x,\xi)\right] = f(x), \quad \mathbf{E}\left[\nabla F(x,\xi)\right] = \nabla f(x), \quad \mathbf{E}\left[\nabla^2 F(x,\xi)\right] = \nabla^2 f(x),$$
$$\mathbf{E}\left[\|\nabla F(x,\xi) - \nabla f(x)\|_*^2\right] \le \sigma^2, \quad \textit{and} \quad \mathbf{E}\left[\|\nabla^2 F(x,\xi) - \nabla^2 f(x)\|_F^4\right] \le \varkappa^4,$$

where all the expectations are w.r.t ξ.

Note that in the deterministic case, we have access to $f(x)$,$\nabla f(x)$, and $\nabla^2 f(x)$ instead of their noisy approximations. Hence, in the deterministic case, $\sigma = 0$, and $\varkappa = 0$. The choice of the (Euclidean) norms will be fixed later in the individual sections. The above set of assumptions are common in the literature of stochastic zeroth-order optimization (Nesterov and Spokoiny, 2017; Ghadimi and Lan, 2013; Duchi et al., 2015; Balasubramanian and Ghadimi, 2021).

We next require the following assumptions, characterizing smoothness properties of the function being optimized, that are standard in the optimization literature Bubeck et al. (2012); Nesterov (2018).

Assumption 2.2 (Lipschitz Function) *The functions F_t are L-Lipschitz, almost surely for any ξ, i.e., $|F_t(x,\xi) - F_t(y,\xi)| \le L\|x-y\|$. We defer the specific choices of the norms to the main results.*

Assumption 2.3 (Lipschitz Gradient) *The functions F_t have Lipschitz continuous gradient, almost surely for any ξ, i.e., $\|\nabla F_t(x,\xi) - \nabla F_t(y,\xi)\| \le L_G\|x-y\|_*$. Note that this also implies $|F_t(y,\xi) - F_t(x,\xi) - \langle \nabla F_t(x,\xi), y-x\rangle| \le \frac{L_G}{2}\|y-x\|^2$. We defer the specific choices of the norms to the main results.*

Note that the above assumption also implies that the function f_t has Lipschitz continuous gradient with the same constant. We now state an analogous assumption for the Hessians which is required for obtaining second-order optimal solution based regret guarantees.

Assumption 2.4 (Lipschitz Hessian) *The functions f_t have Lipschitz continuous Hessian, i.e., $\left\|\nabla^2 f_t(x) - \nabla^2 f_t(y)\right\|_{op} \le L_H\|x-y\|_2$, .*

The above assumptions lead to regret bounds that are polynomially dependent on the dimensions. In order to obtain regret bounds that are only logarithmically dependent on the dimension (thereby facilitating high-dimensional stochastic zeroth-order optimization under nonstationarity), we also make the following *structural sparsity assumption* on the functions.

Assumption 2.5 (Sparse Functions) *The functions $f_t(x)$ are s-sparse. That is, they depend only on s of the d co-ordinates, where $s \ll d$. As a consequence, we have the gradients to be s-sparse as well, i.e., $\|\nabla f_t(x)\|_0 \le s$, where $\|y\|_0$ denotes the number of non-zero coordinates in y.*

Functions that satisfy Assumption 2.5 are common in the fields of constructive approximation (DeVore et al., 2011; Han and Yuan, 2020; Wojtaszczyk, 2011) and non-parametric statistics (Han and Yuan, 2020; Raskutti et al., 2012; Tyagi et al., 2018), as they extend the idea of compressed sensing (Donoho, 2006) to the functional or non-parametric setting. Furthermore, there are also several practical situations in which a function depending only on a few of the coordinates needs to be optimized, for example, hyperparameter tuning in deep learning (Snoek et al., 2012). We also remark that recently Wang et al. (2018) and Balasubramanian and Ghadimi (2021) used similar assumptions in the context of stationary stochastic zeroth-order optimization. Furthermore, sparsity assumptions are also explored in the context of contextual bandits Kim and Paik (2019); Bastani and Bayati (2020); Wang et al. (2020) and reinforcement learning (Hao et al., 2021). Finally, it has been observed in several practical machine learning problems that often times the gradient are approximately sparse (Cai et al., 2020; Elibol et al., 2020). While in this work, we assume exactly sparse functions and hence gradients, it is worth mentioning that the above assumption is extendable to the approximately sparse case in a straightforward manner.

Next, we define the so-called uncertainty sets corresponding to the functions $\{f_t\}_{t=1}^T$ that capture the degree of nonstationarity allowed, following Besbes et al. (2015).

Definition 1 (Besbes et al. (2015)) *For a given $W_T \geq 0$, the uncertainty set $\mathcal{D}_T$ of functions is defined as*

$$\mathcal{D}_T \stackrel{def}{=} \left\{ \{f_t\}_{t=1}^T : \sum_{t=1}^{T-1} \|f_t - f_{t+1}\|_\infty \leq W_T \right\}. \tag{2}$$

We emphasize here that, though the amount of nonstationarity is bounded by W_T, W_T is allowed to increase with the horizon T. This definition of nonstationarity can accommodate different types of temporal pattern in the data, e.g., variable rates of change, constant changes, periodic changes, and discrete shocks; see Besbes et al. (2015) for details.

2.2 The Zeroth-Order Methodology

We now briefly describe our stochastic zeroth-order gradient and Hessian estimation methodology, both of which are based on Gaussian Stein's identity. Following Spall (1998); Nesterov and Spokoiny (2017); Balasubramanian and Ghadimi (2021); Duchi et al. (2015), we define the Gaussian Stein's identity based gradient estimator of $\nabla f_t(x_t)$ as,

$$G_t^\nu(x_t, u_t, \xi_t) = \frac{F_t(x_t + \nu u_t, \xi_t) - F_t(x_t, \xi_t)}{\nu} u_t, \tag{3}$$

where $u_t \sim N(0, I_d)$. It is well-known (see e.g., Nesterov and Spokoiny (2017)) that $\mathbf{E}\left[G_t^\nu(x_t, u_t, \xi_t)\right] = \nabla f_t^\nu(x)$, where f_t^ν is a Gaussian approximation of f_t defined as

$$f^\nu(x) = \frac{1}{(2\pi)^{d/2}} \int f(x + \nu u)\, e^{-\frac{\|u\|_2^2}{2}}\, du = \mathbf{E}\left[f(x + \nu u)\right]. \tag{4}$$

The results below outline some properties of f^ν and its gradient estimator, and provide some preliminary results on the bias and variance that are used in the rest of the paper.

Lemma 2 (Nesterov and Spokoiny (2017)) *Let f_t^ν and G_t^ν be defined in* (4) *and* (3)*, respectively. If Assumption 2.2 holds with $\|\cdot\| = \|\cdot\|_2$ for $f_t(x)$, for any $x \in \mathbb{R}^d$, we have*

$$|f_t^\nu(x) - f_t(x)| \leq \nu L \sqrt{d}, \quad \text{and} \quad \mathbf{E}\left[\|G_t^\nu(x, u, \xi)\|_2^2\right] \leq L^2 (d+4)^2. \tag{5}$$

Lemma 3 (Nesterov and Spokoiny (2017)) *Let the gradient estimator be defined as* (3) *and let Assumption 2.3 hold with $\|\cdot\| = \|\cdot\|_2$ for $f_t(x)$. Then we have for any $x \in \mathbb{R}^d$,*

$$\|\mathbf{E}\left[G_t^\nu(x, u, \xi)\right] - \nabla f_t(x)\|_2 \leq \frac{\nu}{2} L_G (d+3)^{\frac{3}{2}}, \tag{6}$$

$$\mathbf{E}\left[\|G_t^\nu(x, u, \xi)\|_2^2\right] \leq \frac{\nu^2}{2} L_G^2 (d+6)^3 + 2(d+4)\left(\|\nabla f_t(x)\|_2^2 + \sigma^2\right). \tag{7}$$

The stochastic zeroth-order Hessian estimator is given by

$$H_t^\nu(x_t, u_t, \xi_t) = \frac{F_t(x_t + \nu u_t, \xi_t) + F_t(x_t - \nu u_t, \xi_t) - 2F_t(x_t, \xi_t)}{2\nu^2} \left(u_t u_t^\top - I_d\right), \tag{8}$$

where $I_d \in \mathbb{R}^{d \times d}$ is the identity matrix. The above estimator of the Hessian was proposed recently by Balasubramanian and Ghadimi (2021), and is based on second-order Gaussian

Stein's identity. A theoretical analysis of the bias and variance of the Hessian estimator was also provided in Balasubramanian and Ghadimi (2021) – we defer a discussion of such results to the proofs later, as they are involved.

One-point versus Multi-point feedback: We emphasize that we consider the so-called two- and three-point feedback settings for the zeroth-order gradient and Hessian estimators, respectively in (3) and (8). That is, for a given random vector ξ, we assume that the stochastic function $F(\cdot, \xi)$ could be evaluated at any point $x \in X$, and one or two perturbed points, respectively, for estimating the gradient vector and the Hessian matrix. Multi-point bandit feedback has been explored before extensively in the literature on stochastic zeroth-order optimization and bandit optimization. For example, it is used in online convex/strongly-convex optimization (Agarwal et al., 2010; Liu et al., 2017; Gorbunov et al., 2018; Shames et al., 2019); zeroth-order convex and nonconvex optimization (Conn et al., 2009; Larson et al., 2019; Audet and Hare, 2017; Duchi et al., 2015; Ghadimi and Lan, 2013; Nesterov and Spokoiny, 2017; Balasubramanian and Ghadimi, 2021); online non-stationary convex optimization (Chiang et al., 2013); online linear regression (Yuan et al., 2019); supervised page-rank learning (Bogolubsky et al., 2016); reserve price optimization in the context of auction (Feng et al., 2021); online boosting (Brukhim and Hazan, 2021). This is primarily due to the sub-optimal oracle complexity of one-point feedback based stochastic zeroth-order optimization methods in the stationary setting, either in terms of the approximation accuracy or dimension dependency.

The use of one-point feedback for stochastic zeroth-order optimization could be traced back to Nemirovsky and Yudin (1983). Motivated by this, there have been several works in the machine learning community focusing on obtaining regret bounds for online convex optimization. Specifically, considering the class of convex functions (without any further smoothness assumptions) and adversarial noise (i.e., roughly speaking, with noise vectors not necessarily assumed to be independent and identically distributed (i.i.d.)), Bubeck et al. (2017) proposed a polynomial-time algorithm with a sample complexity of $\mathcal{O}(d^{21}/\epsilon^2)$ which was recently improved to $\mathcal{O}(d^5/\epsilon^2)$ in Lattimore (2020). For Lipschitz smooth convex functions, Belloni et al. (2015) and Gasnikov et al. (2017), in the i.i.d noise case, obtained a sample complexity of $\mathcal{O}(d^{7.5}/\epsilon^2)$ and $\mathcal{O}(d/\epsilon^3)$, respectively. The best known lower bound in this case is known to be $\mathcal{O}(d^2/\epsilon^2)$, which was established by Shamir (2013). Further assuming $(\beta - 1)$ differentiable derivatives, for $\beta > 2$, Bach and Perchet (2016) obtained sample complexities of $\mathcal{O}(d^2/\epsilon^{2\beta/(\beta-1)})$ and $\mathcal{O}(d^2/\epsilon^{(\beta+1)/(\beta-1)})$, respectively for the convex and strongly-convex setting, with i.i.d. noise case. See also Akhavan et al. (2020); Dani et al. (2008); Hu et al. (2016); Saha and Tewari (2011) for other related works with in the one-point feedback setting focusing on online convex optimization. In contrast to the above discussion, with two-point feedback it is possible to obtain much improved oracle complexities (i.e., linear in dimension and optimal in ϵ) for stochastic zeroth-order optimization, as illustrated in Agarwal et al. (2010); Duchi et al. (2015); Ghadimi and Lan (2013); Nesterov and Spokoiny (2017). However, in some practical settings, it might be impossible to work with multi-point feedbacks. Hence, in Section 3.3, we also provide results in the one-point setting.

2.3 Regret Measures

We are now ready to introduce the regret measures that we propose and analyze in this work. Our first proposal is based on the notion of first-order optimal solutions or stationary points, motivated by the use of similar performance measures in stationary nonconvex optimization (Nesterov, 2018).

Definition 4 (Expected Gradient-size Regret) *The expected gradient-size regret of a randomized online algorithm is defined as (with $\|\cdot\|_p$ denoting the L_p norm of a vector)*

$$\mathfrak{R}_G^{(p)}(T) := \sum_{t=1}^{T} \mathbf{E}\left[\|\nabla f_t(x_t)\|_p^2\right], \tag{9}$$

where the expectation is taken with respect to all the randomness in the algorithm.

The above notion of regret is to be considered under the assumption that the functions $\{f_t\}_{t=1}^T$ are general smooth nonconvex functions (that each satisfy Assumptions 2.2 and 2.3), but satisfying the condition in Definition 1. It has been shown in stochastic first-order setting by Hazan et al. (2017) that for general smooth nonconvex functions (that satisfy Assumptions 2.2, and 2.3), under a further boundedness assumption, the order of the above gradient-size based regret is $\Omega(T)$. This motivates us to consider the notion of regret in Definition 4 under a controlled degree of nonstationarity as in Definition 1. It is also worth emphasizing the connection between gradient-size based regret measure in Definition 4 and the path-length of stochastic gradient descent algorithm for offline optimization. Specifically, for offline optimization, when the functions f_t are the same, Oymak and Soltanolkotabi (2018) show that gradient descent follows an almost direct trajectory to the nearest global optima by showing that the path-length is bounded for offline optimization problems. Our results bounds in Theorem 6 on the notion of regret in Definition 4, provides a natural extension of the results of Oymak and Soltanolkotabi (2018) for the online setting, where the functions do change over time. We next consider the following notion of regret, based on second-order optimal solutions.

Definition 5 (Expected Second Order Regret) *The expected second-order regret of a randomized online algorithm is defined as*

$$\mathfrak{R}_{ENC}(T) = \sum_{t=1}^{T} \mathbf{E}\left[r_{NC}(t)\right] = \sum_{t=1}^{T} \mathbf{E}\left[\max\left(\|\nabla f_t(x_t)\|_2, \left(-\frac{2}{L_H}\lambda_{\min}\left(\nabla^2 f_t(x_t)\right)\right)^3\right)\right], \tag{10}$$

where the expectation is taken with respect to all the randomness in the algorithm.

Similar to the case above, this notion of regret is also to be considered under the assumption that the functions $\{f_t\}_{t=1}^T$ are general smooth nonconvex functions (that each satisfy Assumptions 2.2 and 2.4), but satisfying the condition in Definition 1. The scaling $2/L_H$ is used mainly for convenience of theoretical analysis and is also used in stationary nonconvex optimization analysis (Nesterov, 2018). The above regret is motivated by the problem of escaping saddle-points in nonconvex optimization. In other words, considering stationary stochastic nonconvex minimization, while the first-order optimal solutions might include maxima, minima or saddle point, second-order optimal solutions are purely local minima avoiding saddle points. Indeed the second term inside the max function in (10) measure the curvature of the Hessian matrix at the solution thereby characterizing the local minimizers. Such local minimizers turn out to be (near) global minimizers and have favorable statistical properties in a several practical machine learning problems (Haeffele and Vidal, 2015; Kawaguchi, 2016; Jin et al., 2017; Nguyen and Hein, 2017; Roy et al., 2020). The above definition, extends this idea of avoiding saddle points to the nonstationary setting that we consider in this work.

A word is in order regarding our notations of regret above and function-value based regret measures for general nonconvex function from the Multi-Armed Bandits (MAB) literature. As mentioned in the introduction without further assumptions on the function being optimized, finding the global minimizer in the stochastic zeroth-order setting suffers from curse of dimensionality even with unbounded computation (Novak, 2006; Novak and Woźniakowski, 2008). Nevertheless, in the literature on continuum armed bandit problems, function-value based regret bounds are studied in the stationary setting. For example, Bubeck et al. (2011b) shows that when the function f is L-lipschitz continuous where L is unknown, then the minimax optimal order for stationary regret, i.e., regret with respect to the fixed best action, is $\mathcal{O}\left(L^{\frac{d}{d+2}}T^{\frac{d+1}{d+2}}\right)$. Kleinberg et al. (2008) shows that for Lipschitz MAB on metric spaces, the expected stationary regret is upper bounded by $\mathcal{O}\left(T^{\frac{d+1}{d+2}+c}\right)$ where d is the max-min-covering dimension, and $c > 0$ is arbitrarily close to 0. MAB with continuum of arms has also been studied in Kleinberg (2005); Auer et al. (2007); Bubeck et al. (2011a). Tyagi and Gärtner (2013) extends the work of Kleinberg et al. (2008) to the high-deimnsional setting where the reward function is assumed to be ϑ-Hölder continuous and to depend on at most $s < d$ coordinates. They show that the expected stationary regret is $\mathcal{O}\left(T^{\frac{\vartheta+s}{2\vartheta+s}}(\log T)^{\frac{\vartheta}{2\vartheta+s}}C(s,d)\right)$, where $C(s,d)$ depends at most polynomially on s and sub-logarithmically in d. Under same sparsity assumption, in Bayesian optimization literature, Chen et al. (2012) establishes a high-probability bound of $\mathcal{O}\left(\sqrt{T}(\log T)^{\frac{s}{2}+1}\right)$ on the expected stationary regret where the function is assumed to be bounded and sampled from a zero-mean Gaussian Process with squared exponential kernel. When these methods are combined with algorithms designed for nonstationary MAB problems, e.g., Rexp3 (Besbes et al., 2019), Exp3.S (Auer et al., 2002), or other related methods (Besbes et al., 2014; Allesiardo et al., 2017), one can hope to achieve function-value based regret bounds under nonstationarity. Nevertheless, such regret bounds will still suffer from the curse of dimensionality. In this context, our proposed regret measures are based on exploiting the underlying structure of the problem, and are based on meaningful local optimal solutions for which one can obtain improved and practical dependency on the dimensionality of the problem.

2.4 Our Contributions

Our main contribution in this work is on obtaining upper bounds for the above introduced notions of regret. Our regret bounds discussed below have an explicit characterization of

Algorithm (**Reference**)	**Structure**	**Regret bound**	**Regret Notion**
GZGD (Theorem 6)	Asmp. 2.3	$\mathcal{O}\left(\left(d+\sigma^2\right)\ \sqrt{T\,W_T}\right)$	$\mathfrak{R}_G^{(2)}(T)$
	Asmp. 2.2, 2.3	$\mathcal{O}\left(\sqrt{d\ T\ W_T}\ \left(1+\sigma^2\right)\right)$	
GZGD (Theorem 10)	Asmp. 2.3, 2.5	$\mathcal{O}\left(\left((s\log d)^2+\sigma^2\right)\sqrt{T\,W_T}\right)$	$\mathfrak{R}_G^{(1)}(T)$
	Asmp. 2.2, 2.3, 2.5	$\mathcal{O}\left(s\log d\left(1+\sigma^2\right)\sqrt{T\,W_T}\right)$	
Algorithm 1 (Hazan et al. (2017))	Bounded function, Asmp. 2.2, 2.3	$\mathcal{O}(T)$ (Deterministic 1st-Order)	$\mathfrak{R}_G^{(2)}(T)$
OCRN (Theorem 12)	Asmp. 2.3, 2.4	$\mathcal{O}\left(T^{\frac{2}{3}}\ \left(1+W_T^{\frac{1}{3}}\right)+T^{\frac{2}{3}}\ \left(\sigma+\varkappa^3\right)+T^{\frac{5}{9}}\ W_T^{\frac{2}{9}}\ \varkappa^2\right)$	$\mathfrak{R}_{ENC}(T)$
ZCRN (Theorem 15)	Asmp. 2.3, 2.4	$\mathcal{O}\left(T^{\frac{2}{3}}\left(1+W_T^{\frac{1}{3}}\right)\left(1+\sigma+\sigma^{\frac{3}{2}}\right)\right)$	$\mathfrak{R}_{ENC}(T)$
Algorithm 3 (Hazan et al. (2017))	Bounded function, Asmp. 2.2, 2.3, 2.4	$\mathcal{O}(T)$ (Deterministic 2nd-Order)	$\hat{\mathfrak{R}}_{NC}(T)$
One-point GZGD (Theorem 18)	Asmp. 2.3,3.1	$\mathcal{O}\left(\left(1+\sigma^2\right)\ d\,T^{\frac{2}{3}}\ W_T^{\frac{1}{3}}\right)$	$\mathfrak{R}_G^{(2)}(T)$
	Asmp. 2.2, 2.3, 3.1	$\mathcal{O}\left(\left(1+\sigma^2\right)\ d^{\frac{1}{2}}\ T^{\frac{2}{3}}\ W_T^{\frac{1}{3}}\right)$	
One-point ZCRN (Theorem 22)	Asmp. 2.3, 2.4, 3.1, 3.2	$\mathcal{O}\left(T^{\frac{2}{3}}\left(1+W_T^{\frac{1}{3}}\right)\left(1+\sigma^2\left(W_T/T\right)^{\frac{1}{9}}+\sigma\sqrt{\sigma_1}\right)\right)$	$\mathfrak{R}_{ENC}(T)$

Table 1: A list of regret bounds obtained in this work for nonstationary nonconvex optimization. One-point GZGD and One-point ZCRN denote Algorithms GZGD (Algorithm 1) and ZCRN (Algorithm 3) respectively with one-point gradient and hessian estimators instead of two-point gradient and hessian estimators. The results from Hazan et al. (2017) are for the 1st and 2nd order settings, without controlling the degree of nonstationarity. They are provided in the table above just for the sake of comparison. Here $\sigma_1=\max(\sigma^4,\sigma')$ where $\mathbf{E}\left[\|\xi\|_2^4\right]\leq\sigma'$.

the time horizon T, the degree of nonstationarity W_T and the dimensionality of the problem d. The precise rates obtained are summarized in Table 1 for convenience.

- **First-order regret:** Considering nonconvex functions f_t whose degree of nonstationarity is bounded in the sense of Definition 1, in Section 3.1, we first establish sub-linear regret bounds for first-order optimal solution based regret measures proposed in Definition 4. We quantify the dependence of this regret on the dimensionality d (which is polynomial in d). This setting is referred to as the low-dimensional setting. Next, we address the issue of dimensionality in this regret bound. Specifically, we consider the case when the functions f_t depend only on s of the d coordinates (see Assumption 2.5) and provide regret bound that only depends poly-logarithmically on the dimension. We refer to such a scenario as the high-dimensional setting.

- **Second-order regret:** Next, we consider the notion of second-order optimal solution based regret in the sense of Definition 5, when the nonconvex functions f_t are assumed to be nonstationary in the sense of Definition 1. In Section 3.2, we then propose and analyze online and bandit versions of cubic-regularized Newton method and establish sub-linear bounds for the above mentioned regret measures. To the best of our knowledge, we provide the first analysis of cubic-regularized Newton method for stochastic zeroth-order optimization under both *nonconvexity* and *nonstationarity*, and demonstrate sub-linear regret bounds. The proposed stochastic zeroth-order cubic-Newton method is motivated by the recently proposed zeroth-order Hessian estimator with three-point feedback mechanism from Balasubramanian and Ghadimi (2021) and is based on second-order Gaussian Stein's identity.
- **One-point setting:** While the above contributions are in the multi-point feedback setting, in Section 3.3, we also provide the corresponding regret bounds in the one-point setting and highlight the subtle differences that occur in this setting.

3. Main Results on Regret Bounds

In this section, we present out main results on the regret bounds. We first focus on the first-order optimal solution based regret measure as in Definition 4 and provide regret bounds in both the low and high-dimensional setting. We next focus on the second-order optimal solution based regret measure as in Definition 5. For this case, we first provide a regret bounds in the stochastic second order setting assuming access to both noisy gradients and Hessians, after which we provide our regret bound in the stochastic zeroth-order setting. These regret bounds are based on online versions of cubic-Newton method. We next focus on a normalized version of regret to account for the multi-point feedback that we use in our gradient and Hessian estimators, and show that the normalized version of regret is of the same order as the one in Definitions 4 and 5. Finally, we focus on the one-point setting and show how the bounds deteriorate gracefully in this setting.

3.1 Nonstationary First-order Regret Bounds

In order to establish the gradient size or first-order optimal solutions based regret bounds, we consider a zeroth-order version of stochastic gradient descent algorithm adapted to handle nonstationarity. The detailed method is given in Algorithm 1. In each iteration, the gradient is computed based on the zeroth-order gradient estimator defined in Equation 3. We remark Algorithm 1 or its variants is widely used in the literature on stochastic zeroth-order optimization and bandit convex optimization.

Algorithm 1 Gaussian Zeroth-order Gradient Descent (GZGD)

Input: Horizon T, η and ν.

for $t = 1$ to T do

Sample standard Gaussian vector $u_t \sim N(0, I_d)$

Query the function f_t at points x_t and $x_t + \nu u_t$ and receive noisy evaluations $F_t(x_t, \xi_t)$ and $F_t(x_t + \nu u_t, \xi_t)$

Estimate the gradient as

$$G_t^\nu (x_t, u_t, \xi_t) = \frac{F_t(x_t + \nu u_t, \xi_t) - F_t(x_t, \xi_t)}{\nu} u_t.$$

Update

$$x_{t+1} = x_t - \eta G_t(x_t, u_t, \xi_t),$$

end for

3.1.1 Low-dimensional Setting

We now provide the regret bounds achieved by Algorithm 1 in the low dimensional setting in Theorem 6 below.

Theorem 6 *Let $\{x_t\}_{t=1}^T$ be generated by Algorithm 1, and Assumption 2.3 holds with $\|\cdot\| = \|\cdot\|_2$ for any sequence of $\{f_t\}_1^T \in \mathcal{D}_T$.*

(a) Choosing

$$\nu = \frac{1}{\sqrt{TL_G}(d+6)}, \qquad \eta = \frac{\sqrt{W_T}}{4L_G(d+4)\sqrt{T}}, \tag{11}$$

we have

$$\mathfrak{R}_G^{(2)}(T) \leq \mathcal{O}\left(\left(d + \sigma^2\right)\sqrt{TW_T}\right). \tag{12}$$

In the deterministic case, as $\sigma = 0$, choosing $\eta = \frac{1}{4L_G(d+4)}$, we get

$$\mathfrak{R}_G^{(2)}(T) \leq \mathcal{O}(dW_T). \tag{13}$$

(b) Additionally, if Assumption 2.2 holds with $\|\cdot\| = \|\cdot\|_2$, by choosing

$$\nu = \min\left\{\frac{L}{L_G(d+6)}, \frac{1}{(TL_G^3 d^5)^{\frac{1}{4}}}\right\}, \qquad \eta = \frac{\sqrt{W_T}}{L\sqrt{TL_G(d+4)}}, \tag{14}$$

we have

$$\mathfrak{R}_G^{(2)}(T) \leq \mathcal{O}\left(\sqrt{dTW_T}\left(1 + \sigma^2\right)\right). \tag{15}$$

For the deterministic case $\sigma = 0$.

Remark 7 *Theorem 6, shows that as long as $W_T \leq o(T)$ it is possible to achieve a sub-linear regret for $\mathfrak{R}_G^{(2)}(T)$. In other words, one could obtain meaningful regret bounds even in the nonstationary setting, as long as the degree of nonstationarity grows sub-linearly. It is worth recalling that Hazan et al. (2017) showed that in the (deterministic) first-order setting it is impossible to achieve sub-linear rate for $\mathfrak{R}_G^{(2)}(T)$, when the degree of nonstationarity is arbitrary.*

We now present the high-level outline of the proof of Theorem 6 and defer the detailed proof to Appendix A. First we show that at every time point t, if we ignore the bias and the variance of the gradient estimator which was introduced in (3), the term $\eta \|\nabla f_t(x_t)\|_2^2/2$ is upper bounded by $f_t(x_t) - f_t(x_{t+1})$ in expectation. Now we want to form a telescopic sum to bound the sum of the gradient sizes. But due to nonstationarity, at every time step t, an additional term appears of the form $f_t(x_t) - f_{t-1}(x_t)$. Observe that the sum of these additional terms can be bounded by W_T. Hence, in the general case we use Lemma 2 to bound the variance, and Lemma 3 to bound the bias of the gradient estimator. Finally, we pick η, and ν suitably to establish the rates.

3.1.2 High-dimensional Setting

We now bound the gradient size based regret for the high-dimensional case under the assumption that the functions being optimized have s-sparse gradient. We need the following two results similar to Lemma 3 to control the ∞-norm of the bias and the second moment of the gradient estimator.

Lemma 8 *Let Assumption 2.2 be satisfied with $\|\cdot\| = \|\cdot\|_\infty$. Then for any x, and some constant $C > 0$ we have,*

$$|f_t^\nu(x) - f_t(x)| \leq \nu LC\sqrt{2\log d}, \tag{16}$$

$$\mathbf{E}\left[\|G_t^\nu(x,u,\xi)\|_\infty^2\right] \leq 4CL^2(\log d)^2. \tag{17}$$

Proof [Proof of Lemma 8] Using Assumption 2.2 with respect to ∞-norm, we have,

$$|f_t^\nu(x) - f_t(x)| \leq \mathbf{E}\left[|f_t(x+\nu u) - f_t(x)|\right] \leq \nu L \mathbf{E}\left[\|u\|_\infty\right], \tag{18}$$

$$\|G_t^\nu(x,u,\xi)\|_\infty^2 = \left\|\frac{f_t(x+\nu u) - f_t(x)}{\nu} u\right\|_\infty^2 \leq L^2\|u\|_\infty^4, \tag{19}$$

which together with the fact that $\mathbf{E}\left[\|u\|_\infty^k\right] \leq C(2\log d)^{k/2}$ due to Balasubramanian and Ghadimi (2021), imply the result. ■

Lemma 9 (Balasubramanian and Ghadimi (2021)) *Let Assumption 2.3 be satisfied with $\|\cdot\| = \|\cdot\|_\infty$. Then for any x, and some constant $C > 0$ we have,*

$$\|\mathbf{E}\left[G_t^\nu(x,u,\xi)\right] - \nabla f_t(x)\|_\infty \leq C\nu L_G\sqrt{2}\,(\log d)^{\frac{3}{2}}, \tag{20}$$

$$\mathbf{E}\left[\|G_t^\nu(x,u,\xi)\|_\infty^2\right] \leq 4C(\log d)^2\left[\nu^2 L_G^2\log d + 4\|\nabla f_t(x)\|_1^2\right]. \tag{21}$$

Now we present the main result on the gradient size regret bound in the high-dimensional sparse setting.

Theorem 10 *Let Assumption 2.3 be satisfied with* $\|\cdot\| = \|\cdot\|_\infty$ *and Assumption 2.5 hold for any sequence of* $\{f_t\}_1^T \in \mathcal{D}_T$*. Then for Algorithm 1,*

(a) By choosing

$$\nu = \frac{1}{\sqrt{2T}} \min\left\{\sqrt{\frac{1}{CL_G \log d}}, s\sqrt{\frac{C \log d}{L_G}}\right\}, \qquad \eta = \frac{\sqrt{W_T}}{32CL_G s\left(\log d\right)^2 \sqrt{T}}, \tag{22}$$

we have

$$\mathfrak{R}_G^{(1)}(T) \le \mathcal{O}\left(\left(\left(s \log d\right)^2 + \sigma^2\right)\sqrt{TW_T}\right). \tag{23}$$

In the deterministic case, setting $\sigma = 0$*, we get*

$$\mathfrak{R}_G^{(1)}(T) \le \mathcal{O}\left(\left(s \log d\right)^2 \sqrt{TW_T}\right). \tag{24}$$

(b) If, in addition, Assumption 2.2 holds with $\|\cdot\| = \|\cdot\|_\infty$*, by choosing*

$$\nu = \left[\frac{1}{2Ts^2C^3L_GL^2\left(\log d\right)^4}\right]^{\frac{1}{4}}, \qquad \eta = \frac{\sqrt{W_T}}{2\sqrt{TCL_G}L \log d}, \tag{25}$$

we obtain

$$\mathfrak{R}_G^{(1)}(T) \le \mathcal{O}\left(s \log d\left(1 + \sigma^2\right)\sqrt{TW_T}\right). \tag{26}$$

In the deterministic case, setting $\sigma = 0$*, we get*

$$\mathfrak{R}_G^{(1)}(T) \le \mathcal{O}\left(s \log d\sqrt{TW_T}\right). \tag{27}$$

Remark 11 *Compared to the regret bounds obtained in Theorem 6, the ones in Theorem 10 have only a poly-logarithmic dependency on the dimension* d*. The dependency on the sparsity parameter* s*, is quadratic and linear without and with Assumption 2.2 respectively. We believe this dependency could potentially be improved (to linear and square-root respectively), however it seems to be outside the scope of our current proof technique, which is provided Appendix A.*

3.2 Nonstationary Second-Order Regret Bounds

While gradient-size based regret (in Definition 4) controls first-order optimal solutions, it does not allows us to avoid saddle-points that are prevalent in nonconvex optimization problems arising in machine learning and game theory Dauphin et al. (2014); Hazan et al. (2017). Hence, we now consider the notion of second-order optimal solution based regret (Definition 5), and propose online and bandit versions of cubic regularized Newton method to obtain the respective nonstationary regret bounds.

Algorithm 2 Online Cubic-Regularized Newton Algorithm (OCRN)

Input: Horizon T, M, m_t, b_t
for $t = 1$ to T do
Set $\bar{G}_t = \frac{1}{m_t}\sum_{i=1}^{m_t} \nabla F_t\left(x_t, \xi_{i,t}^G\right)$
Set $\bar{H}_t = \frac{1}{b_t}\sum_{i=1}^{b_t} \nabla^2 F_t\left(x_t, \xi_{i,t}^H\right)$
Update

$$x_{t+1} = \operatorname*{argmin}_y \tilde{f}_t\left(x_t, y, \bar{G}_t, \bar{H}_t, M\right), \tag{28}$$

where

$$\tilde{f}_t\left(x_t, y, \bar{G}_t, \bar{H}_t, M\right) = \bar{G}_t^\top\left(y - x_t\right) + \frac{1}{2}\langle \bar{H}_t\left(y - x_t\right), \left(y - x_t\right)\rangle + \frac{M}{6}\left\|y - x_t\right\|_2^3. \tag{29}$$

end for

3.2.1 Online Cubic-regularized Newton Method

The standard cubic-regularized Newton method Nesterov and Polyak (2006) has been recently extended to the stochastic setting in Tripuraneni et al. (2018) and to the zeroth-order setting in Balasubramanian and Ghadimi (2021). In Algorithm 2, we consider it in the online setting. Note that Hazan et al. (2007) used online Newton method previously in the context of online convex optimization to obtain logarithmic regret bounds under certain assumptions and Hazan et al. (2017) used a modified online Newton method in the context of online nonconvex optimization. The following theorem provides a regret bound for $\mathfrak{R}_{ENC}(T)$ using the online cubic-regularized Newton method.

Theorem 12 *Let us choose the parameters for Algorithm 2 as follows:*

$$M = L_H\left(\frac{T}{W_T}\right)^{\frac{2}{9}}, \qquad m_t = \left(\frac{T}{W_T}\right)^{\frac{8}{9}}, \qquad b_t = \left(\frac{T}{W_T}\right)^{\frac{2}{9}}. \tag{30}$$

Moreover, suppose that Assumption 2.3 with $\|\cdot\| = \|\cdot\|_2$, and Assumption 2.4 hold for any sequence of functions $\{f_t\}_1^T \in \mathcal{D}_T$. Then, for all $W_T \leq T$, Algorithm 2 with the choice of $M \geq L_H$ produces updates such that

$$\mathfrak{R}_{ENC}(T) \leq \mathcal{O}\left(T^{\frac{2}{3}}\left(1 + W_T^{\frac{1}{3}}\right) + T^{\frac{2}{3}}\left(\sigma + \varkappa^3\right) + T^{\frac{5}{9}} W_T^{\frac{2}{9}} \varkappa^2\right), \tag{31}$$

where the second-order regret $\mathfrak{R}_{ENC}$ is defined in (10)*. As $\sigma = \varkappa = 0$ in the deterministic case, $m_t = b_t = 1$ is sufficient to get*

$$\mathfrak{R}_{ENC}(T) \leq \mathcal{O}\left(T^{\frac{2}{3}}\left(1 + W_T^{\frac{1}{3}}\right)\right). \tag{32}$$

The proof outline is similar to that of the more general Theorem 15 which we discuss in Section 3.2.2. The detailed proof of Theorem 12 is in Appendix B.

Remark 13 *The total number of function calls* $\sum_{t=1}^{T}(m_t+b_t)$ *over a horizon* T *is upper bounded as* $\mathcal{O}\left(T^{\frac{17}{9}}+T^{\frac{11}{9}}\right)$.

Remark 14 *We now compare our second-order regret bound to that in Hazan et al. (2017), which is given by*

$$\hat{\mathfrak{R}}_{NC}(T) := \sum_{t=1}^{T}\hat{r}_{NC}(t) = \sum_{t=1}^{T}\max\left(\|\nabla f_t(x_t)\|_2^2, -\frac{4L_G}{3L_H^2}\lambda_{min}(\nabla^2 f_t(x_t))^3\right) \leq \mathcal{O}(T). \quad (33)$$

This bound is obtained by assuming each loss function f_t *is bounded instead of assuming their total gradual variation is bounded as we have in Definition 1. Noting that* $r_{NC}(t) \leq \mathcal{O}\left(\sqrt{\hat{r}_{NC}(t)}+\hat{r}_{NC}(t)\right)$, *we can bound our regret by using the second-order method in Hazan et al. (2017) such that*

$$\mathfrak{R}_{ENC}(T) = \mathfrak{R}_{NC}(T) := \sum_{t=1}^{T} r_{NC} \leq \mathcal{O}\left(\sqrt{T\hat{\mathfrak{R}}_{NC}(T)}+\hat{\mathfrak{R}}_{NC}(T)\right) \leq \mathcal{O}(T),$$

where the first equality and inequality follow under the deterministic setting and from Hölder's inequality, respectively. We immediately see that an improved second-order regret bound in achieved in (31), *in comparison to Hazan et al. (2017).*

3.2.2 Zeroth-order Cubic-regularized Newton Method

We now extend the online cubic-regularized Newton method to the zeroth-order setting. In order to do so, we leverage the three-point feedback based Hessian estimation technique, proposed in Balasubramanian and Ghadimi (2021), which is based on Gaussian Stein's identity. The zeroth-order cubic-regularized Newton method is provided in Algorithm 3. The following theorem states the bound for expected second order regret using zeroth-order cubic regularized Newton method.

Theorem 15 *Let us choose the parameters for Algorithm 3 as follows:*

$$M = L_H\left(\frac{T}{W_T}\right)^{\frac{2}{9}}, \qquad \nu_G=\nu_H=\nu=\frac{1}{T^{\frac{4}{9}}d^{\frac{5}{2}}},$$
$$m_t=(d+5)\left(\frac{T}{\max(1,W_T)}\right)^{\frac{8}{9}}, \qquad b_t = 4(1+2\log 2d)(d+16)^4\left(\frac{T}{\max(1,W_T)}\right)^{\frac{4}{9}}. \quad (37)$$

Moreover, let Assumption 2.2—2.3 with $\|\cdot\|=\|\cdot\|_2$, *and Assumption 2.4 hold. Then, for all* $W_T \leq T$, *for any sequence of such functions* $\{f_t\}_1^T \in \mathcal{D}_T$, *Algorithm 3 produces updates for which* $\mathfrak{R}_{ENC}(T)$ *is bounded by,*

$$\mathfrak{R}_{ENC}(T) \leq \mathcal{O}\left(T^{\frac{2}{3}}\left(1+W_T^{\frac{1}{3}}\right)\left(1+\sigma+\sigma^{\frac{3}{2}}\right)\right). \quad (38)$$

In the deterministic case, setting $\sigma=0$, *we obtain*

$$\mathfrak{R}_{ENC}(T) \leq \mathcal{O}\left(T^{\frac{2}{3}}\left(1+W_T^{\frac{1}{3}}\right)\right), \quad (39)$$

Algorithm 3 Zeroth-order Cubic Regularized Newton Algorithm (BCRN)

Input: Horizon T, M,m_t,b_t

for $t = 1$ to T **do**

Generate $u_t^{G(H)} = \left[u_{t,1}^{G(H)}, u_{t,2}^{G(H)}, \cdots, u_{t,m_t(b_t)}^{G(H)}\right]$ where $u_{t,i}^{G(H)} \sim N\left(0, I_d\right)$

Set

$$\bar{G}_t = \frac{1}{m_t}\sum_{i=1}^{m_t} \frac{F_t\left(x_t + \nu u_{t,i}^G, \xi_{t,i}^G\right) - F_t\left(x_t, \xi_{t,i}^G\right)}{\nu_G} u_{t,i}^G \tag{34}$$

Set

$$\bar{H}_t = \frac{1}{b_t}\sum_{i=1}^{b_t} \frac{F_t\left(x_t + \nu u_{t,i}^H, \xi_{t,i}^H\right) + F_t\left(x_t - \nu u_{t,i}^H, \xi_{t,i}^H\right) - 2F_t\left(x_t, \xi_{t,i}^H\right)}{2\nu_H^2} \left(u_{t,i}^H \left(u_{t,i}^H\right)^\top - I_d\right) \tag{35}$$

Update

$$x_{t+1} = \underset{y}{\operatorname{argmin}}\, \tilde{f}_t\left(x_t, y, \bar{G}_t, \bar{H}_t, M\right), \tag{36}$$

where $\tilde{f}_t\left(x_t, y, \bar{G}_t, \bar{H}_t, M\right)$ is defined in Equation 29.

end for

Here we present here the high-level outline of the proof of Theorem 15 while deferring the detailed proof to Appendix A. First we show that at every time point t, if we ignore the terms $\|\nabla_t - \bar{G}_t\|_2$, $\|\nabla_t^2 - \bar{H}_t\|_{op}^2$, and $\|\nabla_t^2 - \bar{H}_t\|_{op}^3$, $r_{NC}(t) = \max\left(\|\nabla_t\|_2, -\frac{8}{L_H^3}\lambda_{t,\min}^3\right)$ is upper bounded by the cube of the ℓ_2 norm of the difference of consecutive iterates $x_{t+1} - x_t$, $\|h_t\|^3$ (Lemma 32). Then we show that $M\|h_t\|^3/36$ is bounded by $f_t(x_t) - f_t(x_{t+1})$ ignoring the terms $\|\nabla_t - \bar{G}_t\|_2^{3/2}$, and $\|\nabla_t^2 - \bar{H}_t\|_{op}^3$. Now we want to form a telescopic sum to bound $R_{ENC}(t) = \sum_{t=1}^T r_{NC}(t)$ in expectation. But since this is a nonstationary environment, similar to the proof of Theorem 6, at every time step t an additional term appears of the form $f_t(x_t) - f_{t-1}(x_t)$. Note that the sum of these additional terms can be bounded by W_T. Now observe that the terms we have been ignoring so far are different moments of gradient and hessian estimation error. We use Lemma 30, and Lemma 31 to bound there moments. Finally, one needs to choose M, ν, m_t, and b_t suitably to establish the rates. The detailed proof of Theorem 15 is in Appendix C.

Remark 16 *Although, the bound obtained in Theorem 15 is independent of dimension, we emphasize that we are sampling the function at multiple points during each time step. The total number of function calls is hence,* $\sum_{t=1}^T (m_t + b_t)$ *over a horizon* T *is upper bounded as* $\mathcal{O}\left(d(T/\max(1, W_T))^{\frac{17}{9}} + (\log d)\, d^4 (T/\max(1, W_T))^{\frac{13}{9}}\right)$. *Reducing dimension dependency of this query-complexity is a challenging open-problem.*

Remark 17 *Recall that our results are based on estimating gradients and Hessian matrix based on Gaussian Stein's identities. It is common in the literature to also consider gradient estimators based on random vectors in the unit sphere; see for example Nemirovsky and Yudin (1983); Flaxman et al. (2005). Hence, it is natural to ask if Hessian estimators could be constructed based on random vectors on the unit sphere. Here we provide an approach for estimating Hessian matrix of a deterministic function; we leave the analysis and algorithmic applications of such estimators as future work. Let* $\mathbb{S}^{d-1}$, *and* $\mathbb{B}^d$ *denote the unit d dimensional ball, and the unit d-sphere respectively. We will use* $\mathbb{S}$, *and* $\mathbb{B}$ *instead of* $\mathbb{S}^{d-1}$, *and* $\mathbb{B}^d$ *respectively where the dimension is understood clearly. Let* u_1, *and* u_2 *are chosen randomly on* $\mathbb{S}^{d-1}$ *and* v_1, *and* v_2 *are chosen randomly from* $\mathbb{B}^d$.

$$\begin{aligned}\mathbf{E}\left[f\left(x+\nu u_1+\nu u_2\right)u_1 u_2^\top\right] =& C_1 \iint\limits_{\mathbb{S}} f\left(x+\nu u_1+\nu u_2\right)u_1 u_2^\top \, du_1\, du_2 \\ =& C_2 \int\limits_{\mathbb{S}} \int\limits_{\nu\mathbb{S}} f\left(x+\nu u_2+z_1\right) z_1 \, dz_1 \, u_2^\top \, du_2 \\ =& C_3 \int\limits_{\mathbb{S}} \nabla \int\limits_{\nu\mathbb{B}} f\left(x+\nu u_2+v_1\right)\, dv_1 \, u_2^\top \, du_2.\end{aligned}$$

The last equality follows from Stoke's theorem. Now, let

$$\nabla \int\limits_{\mathbb{B}} f\left(x+\nu u_2+\nu v_1\right)\, dv_1 = \left[g_1\left(x+\nu u_2\right), g_2\left(x+\nu u_2\right), \cdots, g_d\left(x+\nu u_2\right)\right]^\top,$$

and $x = \begin{bmatrix} x_1, & x_2, & \cdots & , x_d \end{bmatrix}^\top$. *Then, using Stoke's theorem again, we have*

$$\begin{aligned}\int\limits_{\mathbb{S}} g_1\left(x+\nu u_2\right) u_2^\top \, du_2 =& C_4 \nabla \int\limits_{\nu\mathbb{B}} g_1\left(x+v_2\right)\, dv_2 \\ =& C_5 \nabla \mathbf{E}_{v_2}\left[g_1\left(x+\nu v_2\right)\right] \\ =& C_6 \nabla \mathbf{E}_{v_2}\left[\frac{\partial}{\partial x_1}\mathbf{E}_{v_1}\left[f\left(x+\nu v_1+\nu v_2\right)\right]\right] \\ =& C_7 \nabla \frac{\partial}{\partial x_1}\mathbf{E}_{v_2}\left[\mathbf{E}_{v_1}\left[f\left(x+\nu v_1+\nu v_2\right)\right]\right].\end{aligned}$$

So we can write,

$$\nabla^2 \mathbf{E}\left[f\left(x+\nu v_1+\nu v_2\right)\right] = \mathbf{E}\left[C_7 f\left(x+\nu u_1+\nu u_2\right) u_1 u_2^\top\right],$$

where C_i *for* $i = 1, 2, \cdots, 7$ *are constants. Hence, we have a bandit Hessian estimator, as this relates the Hessian of the function to point queries of the function.*

3.3 Regret bounds under One-Point Feedback

While estimating the gradient as in (3), we assume that the function can be evaluated at both the points $x_t+\nu u_t$ and x_t with the same realization of the noise ξ_t. But when the noise is additive, i.e., we have $F(x,\xi) = f(x) + \xi$, then the above assumption implies we have a noise-free gradient estimator because then $F(x_t + \nu u, \xi_t) - F(x_t, \xi_t) = f(x_t + \nu u) - f(x_t)$. A similar observation is also true for the Hessian estimator in (8). In this section we consider the case where the noise for each function evaluation is different. Following Bach and Perchet (2016), we refer to this setting as the one-point setting. As we demonstrate next, the variance of the gradient and Hessian estimator in this one-point setting is higher than the previous setting. To counteract this, in this one-point setting we use a mini-batch gradient and hessian estimator:

$$\bar{G}_t = \frac{1}{m_t} \sum_{i=1}^{m_t} \frac{F_t\left(x_t + \nu_G u_{t,i}, \xi_{t,i}\right) - F_t\left(x_t, {\xi'}_{t,i}\right)}{\nu_G} u_{t,i}, \tag{40}$$

$$\bar{H}_t = \frac{1}{b_t} \sum_{i=1}^{b_t} \frac{F_t\left(x_t + \nu_H u_{t,i}, \xi^{''}_{t,i}\right) + F_t\left(x_t - \nu_H u_{t,i}, \xi^{'''}_{t,i}\right) - 2F_t\left(x_t, \xi^{''''}_{t,i}\right)}{2\nu_H^2} \left(u_{t,i}\left(u_{t,i}\right)^\top - I_d\right), \tag{41}$$

where $\xi_{t,i}, {\xi'}_{t,i}, \xi^{''}_{t,i}, \xi^{'''}_{t,i}, \xi^{''''}_{t,i}$ are independent. We also require the following additional assumption to control the variance of the gradient estimator.

Assumption 3.1 (Lipschitz Function) *The functions $F_t(x,\cdot)$ are L'-Lipschitz for any x, i.e., $|F_t(x,\xi) - F_t(x,\xi')| \le L' \|\xi - \xi'\|_2$.*

Using Lemma 4.1 of Roy et al. (2021) and young's inequality, in this setting (7) changes to

$$\mathbf{E}\left[\left\|\bar{G}_t\right\|_2^2\right] \le \frac{\nu^2}{2} L_G^2 (d+6)^3 + 2(d+4)\|\nabla f_t(x)\|_2^2 + \frac{4d\sigma^2 {L'}^2}{m\nu^2}. \tag{42}$$

Correspondingly, we have the following analogue to Theorem 6, the proof of which is in Appendix D.

Theorem 18 *Let $\{x_t\}_1^T$ be generated by Algorithm 1 with one-point gradient estimator, and Assumption 2.3 with $\|\cdot\| = \|\cdot\|_2$, and Assumption 3.1 hold for any sequence of $\{f_t\}_1^T \in \mathcal{D}_T$.*

(a) Choosing

$$\nu = \frac{W_T^{\frac{1}{6}}}{T^{\frac{1}{6}}(d+6)\sqrt{L_G}}, \qquad \eta = \frac{W_T^{\frac{2}{3}}}{4L_G(d+4)T^{\frac{2}{3}}}, \qquad m = d, \tag{43}$$

we have

$$\mathfrak{R}_G^{(2)}(T) \le \mathcal{O}\left(dT^{\frac{2}{3}} W_T^{\frac{1}{3}}\left(1+\sigma^2\right)\right). \tag{44}$$

(b) Additionally, if Assumption 2.2 holdswith $\|\cdot\| = \|\cdot\|_2$, by choosing

$$\nu = \min\left\{\frac{L}{L_G(d+6)}, \frac{W_T^{\frac{1}{6}}}{T^{\frac{1}{6}}(L_G^3 d^5)^{\frac{1}{4}}}\right\}, \qquad \eta = \frac{W_T^{\frac{2}{3}}}{LT^{\frac{2}{3}}\sqrt{L_G(d+4)}}, \qquad m = d^{\frac{5}{2}} \tag{45}$$

we have

$$\mathfrak{R}_G^{(2)}(T) \leq \mathcal{O}\left(\left(1+\sigma^2\right) d^{\frac{1}{2}} T^{\frac{2}{3}} W_T^{\frac{1}{3}}\right). \tag{46}$$

For the deterministic case $\sigma = 0$.

Remark 19 *In the one-point setting we cannot choose the ν parameter to be as small as the two-point setting in Theorem 6 since the variance of the gradient estimator is not longer monotnically decreasing with ν. Also note that in the one-point setting, we need a minibatch gradient estimator with mini-batch size $m = d$ to average out the larger variance of the gradient estimator whereas in the two-point setting, $m = 1$. Not that if one allows m_t to depend on T, and W_T, it is possible to match the regret bounds obtained for the two-point setting in Theorem 6 at the expense of larger oracle complexity.*

Assumption 3.2 (Bounded fourth moment of the noise) *The noise satisfied the following fourth moment condition: $\mathbf{E}\left[\|\xi\|_2^4\right] \leq \sigma'$.*

In the one-point setting, we have the following two lemmas which establish bounds on the moments of gradient and Hessian estimation errors in the one-point setting.

Lemma 20 *Let Assumptions 2.2—2.3, and Assumption 3.1 be true. Then we have*

$$\mathbf{E}\left[\|\bar{G}_t - \nabla_t\|_2^2\right] \leq \frac{4d\sigma^2 L'^2}{m_t \nu^2} + 3\nu^2 L_G^2 (d+3)^3 + \frac{8\left(L^2+\sigma^2\right)(d+5)}{m_t}. \tag{47}$$

The proof of Lemma 20 is in Appendix D.

Lemma 21 *For $b_t \geq 4(1 + 2\log 2d)$, under Assumption 2.3 with $\|\cdot\| = \|\cdot\|_2$, Assumption 2.4, Assumption 3.1, and Assumption 3.2 we have*

$$\mathbf{E}\left[\|\tilde{H}_t - \nabla_t^2\|_{op}^2\right] \leq 6L_H^2 (d+16)^5 \nu^2 + \frac{256(1+2\log 2d)(d+16)^4 L_G^2}{3b_t} + \frac{256 L^2\sigma^2 d(1+2\log 2d)}{b_t \nu^4}. \tag{48a}$$

$$\begin{aligned}\mathbf{E}\left[\|\tilde{H}_t - \nabla_t^2\|_{op}^3\right] \leq & 44L_H^3 (d+16)^{\frac{15}{2}} \nu^3 + \frac{160\sqrt{1+2\log 2d}\,(d+16)^6 L_G^3}{b_t^{\frac{3}{2}}} \\ & + \left(\frac{K_2 L'^6 \sigma_1 \sigma^2 (1+2\log 2d) d^3}{b_t^3 \nu^{12}}\right)^{\frac{1}{2}}.\end{aligned} \tag{48b}$$

where $\sigma_1 = \max(\sigma^4, \sigma')$.

We remark here that Lemma 20, and Lemma 21 are crucial to our proofs of the results in the one-point setting. We present an overview of the proofs of these two Lemma here and defer the detailed proofs to Appendix D. The proof of Lemma 20 follows from the following decomposition

$$\begin{aligned}\mathbf{E}\left[\|\bar{G}_t-\nabla_t\|_2^2\right]\leq\ & 2\mathbf{E}\left[\left\|\frac{1}{m_t}\sum_{i=1}^{m_t}\frac{F_t\left(x_t+\nu_G u_{t,i},\xi'_{t,i}\right)-F_t\left(x_t,{\xi'}_{t,i}\right)}{\nu_G}u_{t,i}-\nabla_t\right\|_2^2\right]\\ &+2\mathbf{E}\left[\frac{1}{m_t^2}\sum_{i=1}^{m_t}\left\|\frac{F_t\left(x_t+\nu u_{t,i},\xi_{t,i}\right)-F_t\left(x_t+\nu u_{t,i},\xi'_{t,i}\right)}{\nu}u_{t,i}\right\|_2^2\right],\end{aligned}$$

where the first term on the RHS is bounded using the two-point setting bounds and the second term can be bounded using Lipschitz continuity of F_t in the variable $\xi_{t,i}$ (Assumption 3.1), and the fact that $u_{t,i}$ is a standard gaussian random vector.

To establish the bounds on the moments of hessian estimation error we first consider the following decomposition:

$$\tilde{H}=\frac{1}{b_t}\sum_{i=1}^{b_t}\frac{F\left(x+\nu u_i,\xi_{i,+}\right)+F\left(x-\nu u_i,\xi_{i,-}\right)-2F\left(x,\xi_{i,0}\right)}{2\nu^2}\left(u_iu_i^\top-I_d\right)=\bar{H}+\bar{\tau},$$

where

$$\tau_i=\frac{F\left(x+\nu u_i,\xi_{i,+}\right)+F\left(x-\nu u_i,\xi_{i,-}\right)-F\left(x+\nu u_i,\xi_{i,0}\right)-F\left(x-\nu u_i,\xi_{i,0}\right)}{2\nu^2}\left(u_iu_i^\top-I_d\right),$$

and $\bar{\tau}=\sum_{i=1}^{b_t}\tau_i/b_t$. Then one can write

$$\|\tilde{H}-\nabla^2\|_{op}^2=\|\bar{H}+\bar{\tau}-\nabla^2\|_{op}^2\leq 2\|\bar{H}-\nabla^2\|_{op}^2+2\|\bar{\tau}\|_{op}^2. \tag{49}$$

Note that τ_i and τ_j are independent for all $i\neq j$. Also,

$$\mathbf{E}\left[\tau_i\right]=\mathbf{E}\left[\mathbf{E}\left[\tau_i|x,u_i\right]\right]=0. \tag{50}$$

Use of Theorem 1 of Tropp (2016) followed by some simple calculation shows,

$$\mathbf{E}\left[\|\bar{\tau}\|_{op}^2\right]\leq\frac{32{L'}^2\sigma^2dC(d)}{b_t\nu^4}, \tag{51}$$

where $C(d)=4(1+2\log 2d)$. To upper bound the third moment we use the following inequality,

$$\mathbf{E}\left[\|\bar{\tau}\|_{op}^3\right]\leq\mathbf{E}\left[\|\bar{\tau}\|_{op}\|\bar{\tau}\|_F^2\right]\leq\left(\mathbf{E}\left[\|\bar{\tau}\|_{op}^2\right]\mathbf{E}\left[\|\bar{\tau}\|_F^4\right]\right)^{\frac{1}{2}}. \tag{52}$$

Now we upper bound $\mathbf{E}\left[\|\bar{\tau}\|_F^4\right]$ to get the final bound on $\mathbf{E}\left[\|\bar{\tau}\|_{op}^3\right]$. Now combining the above results gives the final bounds on $\mathbf{E}\left[\|\tilde{H}_t-\nabla_t^2\|_{op}^2\right]$, and $\mathbf{E}\left[\|\tilde{H}_t-\nabla_t^2\|_{op}^3\right]$.

Lemma 20 and Lemma 21 show that in the one-point setting, the variance of the gradient and hessian estimators does not decrease monotonically with ν unlike the two-point setting. In (47), (48a), and (48b), one needs to select the mini-batch size correctly to counter the effect of ν in the denominator. Since, the dependence of the variance of $\bar{G}_t$ and $\tilde{H}_t$ on ν are different, and the minibatch size required depends on ν, for the one-point setting, we allow for different ν parameters, ν_G and ν_H for $\bar{G}_t$ and $\tilde{H}_t$ respectively. We now present the bound on the second-order non stationary regret in the one-point setting.

Theorem 22 *Let us choose the parameters for Algorithm 3 as follows:*

$$\begin{aligned} M &= L_H \left(\frac{T}{W_T}\right)^{\frac{2}{9}}, \qquad \nu_G = \frac{W_T^{\frac{1}{3}}}{T^{\frac{4}{9}} d^{\frac{3}{2}}}, \qquad \nu_H = \frac{W_T^{\frac{1}{9}}}{T^{\frac{1}{9}} d^{\frac{5}{2}}}, \\ m_t &= \frac{T^{\frac{16}{9}}}{\max(1, W_T)^{\frac{4}{3}}} d^4, \qquad b_t = \left(\frac{T}{\max(1, W_T)}\right)^{\frac{2}{3}} d^{11} \left(1 + 2\log 2d\right). \end{aligned} \tag{53}$$

Moreover, let Assumption 2.3 with $\|\cdot\| = \|\cdot\|_2$, Assumption 2.4, Assumption 3.1, and Assumption 3.2 hold. Then, for all $W_T \le T$, for any sequence of such functions $\{f_t\}_1^T \in \mathcal{D}_T$, Algorithm 3 with one-point gradient and hessian estimators produces updates for which $\mathfrak{R}_{ENC}(T)$ is bounded by,

$$\mathfrak{R}_{ENC}(T) \le \mathcal{O}\left(T^{\frac{2}{3}}\left(1 + W_T^{\frac{1}{3}}\right)\left(1 + \sigma^2\left(\frac{W_T}{T}\right)^{\frac{1}{9}} + \sigma\sqrt{\sigma_1}\right)\right). \tag{54}$$

In the deterministic case, setting $\sigma = 0$, we obtain

$$\mathfrak{R}_{ENC}(T) \le \mathcal{O}\left(T^{\frac{2}{3}}\left(1 + W_T^{\frac{1}{3}}\right)\right). \tag{55}$$

Given the variance bounds in Lemma 21, proof of Theorem 22 is very similar to Theorem 15.

Remark 23 *As pointed out earlier, in the one-point setting ν cannot be made as small as the two-point setting because the variance of the gradient estimator does not monotonically decrease with ν in the on-point setting unlike the two-point setting. Indeed in Theorem 15 for the two-point setting we choose $\nu_G = \nu_H = \nu = 1/\left(T^{4/9} d^{5/2}\right)$ whereas in the one-point setting we choose $\nu_G = W_T^{1/3}/\left(T^{4/9} d^{3/2}\right), \nu_H = W_T^{1/9}/\left(T^{1/9} d^{5/2}\right)$. Moreover, to control the larger variance the oracle-complexity also increases to*

$$\sum_{t=1}^{T}(m_t + b_t) = \mathcal{O}\left(d^4 (T^{25/9}/\max(1, W_T)^{7/3}) + (\log d)\, d^{11} (T/\max(1, W_T))^{5/3}\right)$$

whereas in the two-point setting it was $\mathcal{O}\left(d(T/\max(1, W_T))^{17/9} + (\log d)\, d^4 (T/\max(1, W_T))^{13/9}\right)$.

3.4 A Note on Multi-point Feedbacks and Our Regret Measures

Note that our Algorithms 1 and 3 use multiple function evaluations in each iteration. Indeed Algorithm 1 uses m noisy evaluations ($m = 2$ for Theorem 6, and Theorem 10) and Algorithm 3.2.2 uses $2m_t + 3b_t$ noisy evaluations in each iteration. A potential source of contention regarding the regret measure that we proposed in Definitions 4 and 5 could be that they are not taking into account that our the algorithms are using multi-point feedbacks. However, as we show below, a normalized version of the regret measure that takes into account explictly the effect of multi-point feedback is of the same order as our original regret measure in Definitions 4 and 5. Specifically, our normalized regret measure, following Agarwal et al. (2010) are as follows:

$$\bar{\mathfrak{R}}_G^{(2)}(T) = \sum_{t=1}^{T} \frac{1}{2m_t} \sum_{j=1}^{m_t} \left(\mathbf{E}\left[\|\nabla f_t(x_t + \nu u_{t,j})\|_2^2 \right] + \mathbf{E}\left[\|\nabla f_t(x_t)\|_2^2 \right] \right).$$

Recall that for the zeroth-order cubic-regularized Newton method we call the zeroth-order oracle $2m_t$ times to evaluate $\bar{G}_t$, and $3b_t$ times to evaluate $\bar{H}_t$. Then, expected second-order regret can also be defined as

$$\begin{aligned}\bar{\mathfrak{R}}_{ENC}(T) = &\sum_{t=1}^{T} \frac{1}{2m_t + 3b_t} \left((m_t + b_t)\mathbf{E}\left[\max\left(\|\nabla f_t(x_t)\|_2, \left(-\frac{2}{L_H} \lambda_{\min}\left(\nabla^2 f_t(x_t)\right)\right)^3 \right) \right] \right. \\ &\left. + \sum_{j=1}^{m_t + 2b_t} \mathbf{E}\left[\max\left(\|\nabla f_t(x_t + \nu u_{t,j})\|_2, \left(-\frac{2}{L_H} \lambda_{\min}\left(\nabla^2 f_t(x_t + \nu u_{t,j})\right)\right)^3 \right) \right] \right).\end{aligned}$$

Proposition 24 *Consider the setup, and choice of ν of Theorem 6, and Theorem 18. In both cases we have,*

$$\bar{\mathfrak{R}}_G^{(2)}(T) = \mathcal{O}\left(\mathfrak{R}_G^{(2)}(T) \right).$$

Under the setup, and choice of ν of Theorem 10, we have,

$$\bar{\mathfrak{R}}_G^{(1)}(T) = \mathcal{O}\left(\mathfrak{R}_G^{(1)}(T) \right).$$

Under the setup, and choice of ν_G, and ν_H of Theorem 15, and Theorem 22, we have,

$$\bar{\mathfrak{R}}_{ENC}(T) = \mathcal{O}\left(\mathfrak{R}_{ENC}(T) \right).$$

The proof of Proposition 24 is in Appendix E. From Proposition 24 one can see that when the first-order and second-order regrets are adapted to the multi-point feedbacks, the same regret bounds holds.

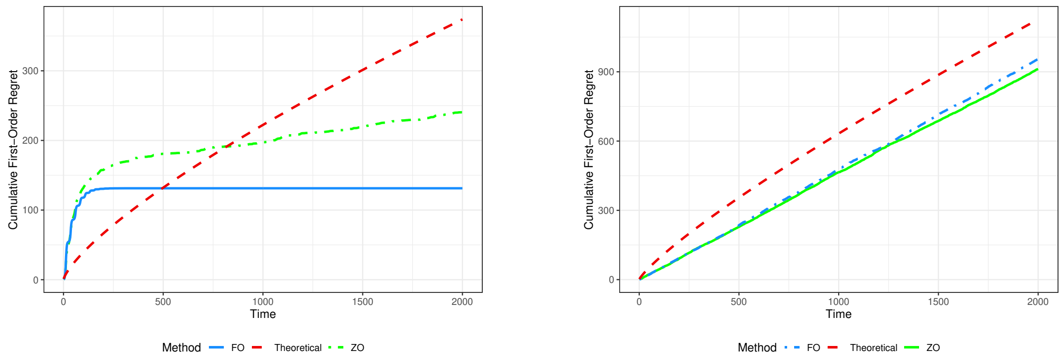
4. Experiments

In this section we illustrate our results on nonstationary nonconvex optimization through simulation, and compare the performance of zeroth-order methods with their higher-order counterparts. For $i = 1, 2, 3, \cdots$, consider the functions

$$f_t(x) = \begin{cases} \frac{S_I W_T}{2T} \sin(\omega x) & 2(i-1)S_I + 1 \le t \le (2i-1)S_I \\ -\frac{S_I W_T}{2T} \sin(\omega x) & (2i-1)S_I + 1 \le t \le 2iS_I \end{cases}$$

These functions belong to the uncertainty set $\mathcal{D}_T$ since,

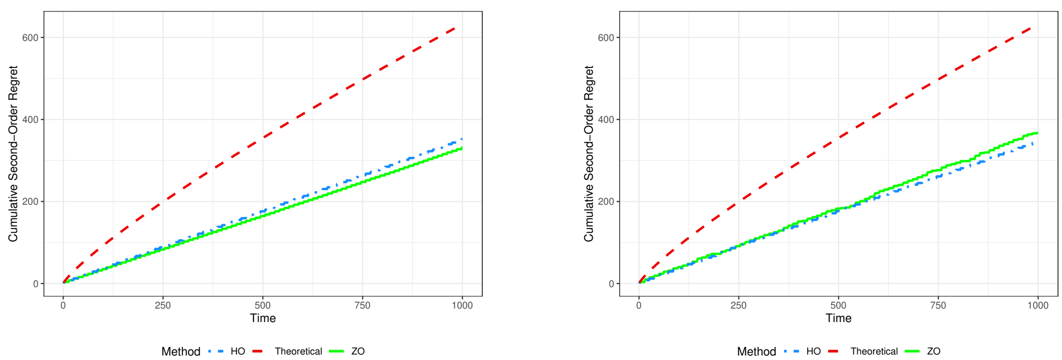
$$\sum_{t=1}^{T-1} \|f_t - f_{t+1}\|_2 = \frac{2S_I W_T}{2T} \times \frac{T}{S_I} = W_T.$$



(a) Two-point Setting

(b) One-point Setting

Figure 1: Cumulative first-order regret $\mathfrak{R}_G^{(2)}(T)$ using Algorithm 1. Here, $S_I = 12$, $\omega = 22$, $x_0 = 0.078$, and $\sigma = 0.5$. Here FO and ZO stand for First-Order and Zeroth-Order respectively.



(a) Two-point Setting

(b) One-point Setting

Figure 2: Cumulative second-order regret $\mathfrak{R}_{ENC}(T)$ using Algorithm 2 and Algorithm 3. Here, $S_I = 12$, $\omega = 22$, $x_0 = 0.078$, and $\sigma = 0.5$. Here HO and ZO stand for Higher-Order and Zeroth-Order respectively.

For the two-point setting, we assume that the multiplicative noise ξ in the function evaluation is sampled from a uniform distribution $U[1-\sigma, 1+\sigma]$. It is easy to see that $F_t(x, \xi) = \xi f_t(x)$ satisfies Assumptions 2.1–2.4 $\forall t = 1, 2, \cdots, T$. For the one-point setting, we sample the additive noise ξ from $\mathcal{N}(0, \sigma^2)$. Then Assumption 3.1, in addition to Assumptions 2.1–2.4, holds true for $F_t(x, \xi) = f_t(x) + \xi$. For this experiment we set $S_I = 12,\ \omega = 22,\ x_0 = 0.078,\ \sigma = 0.5$. Since in this paper we are interested in the expected regret, Figure 1 and Figure 2 show curves averaged over 50 simulations. In Figure 1 we show the evolution of $\mathfrak{R}_G^{(2)}(T)$ over a horizon of $T = 2000$ with $W_T = \sqrt{T}$ using Algorithm 1. Figure 1a and 1b present the results for the two-point and one-point settings respectively. All the algorithm parameters are chosen as described in Theorem 6 and Theorem 18 in the corresponding cases. In both cases the performance of Algorithm 1 is comparable to the first-order variant, i.e., when the stochastic gradient is available. Theoretical rates obtained in Theorem 6 and Theorem 18 are also shown (red, dashed line) for comparison purpose. In Figure 2 we show $\mathfrak{R}_{ENC}(T)$ over a horizon of $T = 1000$ with $W_T = \sqrt{T}$ using Algorithm 2 and 3. Figure 2a and 2b presents the result for the two-point and one-point setting respectively. For the two-point setting, all the algorithm parameters are chosen as described in Theorem 12 and Theorem 15 in the higher-order(HO) and zeroth-order(ZO) cases respectively. For the one-point setting the algorithm parameters are chosen as stated in Theorem 22. One can see that the performances of Algorithm 3 is comparable to Algorithm 2 where the higher-order information is available. Theoretical rate for $\mathfrak{R}_{ENC}(T)$ obtained in Theorem 12, Theorem 15 and Theorem 22 is also shown (red, dashed line) for comparison purpose.

5. Discussion

In this paper, we provide regret bounds for nonstationary nonconvex optimization problems in the stochastic zeroth-order setting. We make the following specific contributions: (i) provide low and high-dimensional regret bounds in terms of gradient-size for general nonconvex function with bounded nonstationarity, (ii) provide, to the best of our knowledge, the first analysis of cubic regularized Newton method for bounding second-order stationary solution based nonstationary regret in the zeroth-order and higher-order settings, and (iii) explore the relationship between the multi-point and one-point feedbacks for the above regret bounds.

There are several avenues for future work: (i) obtaining lower bounds for the regrets considered is challenging, (ii) defining other notions of uncertainty set that capture more general nonstationary environment is also interesting, (iii) obtaining parameter-free algorithms, similar to the convex setting (see for example, Jadbabaie et al. (2015); Luo and Schapire (2015); Cheung et al. (2018); Auer et al. (2019)) is intriguing and (iv) establishing connections between online nonparametric regression and nonstationary regret bounds (see, for example, Baby and Wang (2019)) is interesting.

Appendix A. Proofs of Section 3.1

Proof [Proof of Theorem 6] First note that under Assumption 2.3 with $\|\cdot\| = \|\cdot\|_2$ we have

$$
\begin{aligned}
f_t(x_{t+1}) \leq & f_t(x_t) + \nabla f_t(x_t)^\top (x_{t+1} - x_t) + \frac{L_G}{2}\|x_{t+1} - x_t\|_2^2 \\
= & f_t(x_t) - \eta \nabla f_t(x_t)^\top G_t^\nu(x_t, u_t, \xi_t) + \frac{\eta^2 L_G}{2}\|G_t^\nu(x_t, u_t, \xi_t)\|_2^2 \\
= & f_t(x_t) - \eta\|\nabla f_t(x_t)\|_2^2 + \eta \nabla f_t(x_t)^\top (\nabla f_t(x_t) - G_t^\nu(x_t, u_t, \xi_t)) + \frac{\eta^2 L_G}{2}\|G_t^\nu(x_t, u_t, \xi_t)\|_2^2 .
\end{aligned}
$$

Define $\mathcal{F}_t$ to be the σ-algebra generated by the randomness in the iterates till time-step t. Taking conditional expectation on both sides of the above equation, we obtain

$$
\begin{aligned}
\mathbf{E}[f_t(x_{t+1}) | \mathcal{F}_t] \leq & f_t(x_t) - \eta\|\nabla f_t(x_t)\|_2^2 + \eta\|\nabla f_t(x_t)\|_2 \|\nabla f_t(x_t) - \mathbf{E}[G_t^\nu(x_t, u_t, \xi_t) | \mathcal{F}_t]\|_2 \\
& + \frac{\eta^2 L_G}{2}\mathbf{E}\left[\|G_t^\nu(x_t, u_t, \xi_t)\|_2^2 | \mathcal{F}_t\right].
\end{aligned}
$$

Now, by invoking Young's inequality we have

$$\mathbf{E}\left[f_t\left(x_{t+1}\right)|\mathcal{F}_t\right] \leq f_t\left(x_t\right) - \eta\left\|\nabla f_t\left(x_t\right)\right\|_2^2 + \frac{\eta}{2}\left\|\nabla f_t\left(x_t\right)\right\|_2^2 + \frac{\eta}{2}\left\|\nabla f_t\left(x_t\right) - \mathbf{E}\left[G_t^\nu\left(x_t, u_t, \xi_t\right)|\mathcal{F}_t\right]\right\|_2^2$$
$$+ \frac{\eta^2 L_G}{2}\mathbf{E}\left[\left\|G_t^\nu\left(x_t, u_t, \xi_t\right)\right\|_2^2|\mathcal{F}_t\right]. \tag{56}$$

Note that the third, and the fourth term of (56) are the bias and the second moment of the gradient estimator. Re-arranging the terms and noting Lemma 3, we obtain

$$\frac{\eta}{2}\left\|\nabla f_t\left(x_t\right)\right\|_2^2 \leq f_t\left(x_t\right) - \mathbf{E}\left[f_t\left(x_{t+1}\right)|\mathcal{F}_t\right] + \frac{\eta}{8}\nu^2 L_G^2\left(d+3\right)^3$$
$$+ \frac{\eta^2 L_G}{2}\left(\frac{\nu^2}{2}L_G^2\left(d+6\right)^3 + 2\left(d+4\right)\left(\left\|\nabla f_t\left(x_t\right)\right\|_2^2 + \sigma^2\right)\right).$$

Summing from $t=1$ to T, and using Definition 1 we get

$$\sum_{t=1}^{T}\mathbf{E}\left[\left\|\nabla f_t\left(x_t\right)\right\|_2^2\right] \leq \frac{2}{\eta}\left(f_1\left(x_1\right) - \mathbf{E}\left[f_T\left(x_{T+1}\right)\right] + W_T\right) + \frac{T}{4}\nu^2 L_G^2\left(d+3\right)^3$$
$$+ \eta T\frac{\nu^2}{2}L_G^3\left(d+6\right)^3 + 2\eta L_G\left(d+4\right)\sum_{t=1}^{T}\mathbf{E}\left[\left\|\nabla f_t\left(x_t\right)\right\|_2^2 + \sigma^2\right]. \tag{57}$$

Now we split the proof in two parts corresponding to the parts in Theorem 6.

(a) From (57) by rearranging terms we obtain

$$\sum_{t=1}^{T}\left(1 - 2\eta L_G\left(d+4\right)\right)\mathbf{E}\left[\left\|\nabla f_t\left(x_t\right)\right\|_2^2\right] \leq \frac{2}{\eta}\left(f_1\left(x_1\right) - \mathbf{E}\left[f_T\left(x_{T+1}\right)\right] + W_T\right) + \frac{T}{4}\nu^2 L_G^2\left(d+3\right)^3$$
$$+ \eta T\frac{\nu^2}{2}L_G^3\left(d+6\right)^3 + 2\eta T L_G\left(d+4\right)\sigma^2.$$

Now by choosing ν and η according to (11), we get (12).

(b) It is possible to improve the dependence of the regret bound on the problem dimension assuming that the loss functions are Lipschitz continuous. In this case, we have $\left\|\nabla f_t\left(x_t\right)\right\| \leq L$ which together with (57), imply that

$$\sum_{t=1}^{T}\mathbf{E}\left[\left\|\nabla f_t\left(x_t\right)\right\|_2^2\right] \leq \frac{2}{\eta}\left(f_1\left(x_1\right) - \mathbf{E}\left[f_T\left(x_{T+1}\right)\right] + W_T\right) + \frac{T}{4}\nu^2 L_G^2\left(d+3\right)^3$$
$$+ \eta T L_G\left(\frac{\nu^2}{2}L_G^2\left(d+6\right)^3 + 2\left(d+4\right)\left(L^2 + \sigma^2\right)\right). \tag{58}$$

Now by choosing ν and η according to (14), we obtain (15).

■

Proof [Proof of Theorem 10] Under Assumption 2.3 w.r.t l_∞-norm and similar to (56), we get

$$\mathbf{E}\left[f_t\left(x_{t+1}\right)|\mathcal{F}_t\right] \leq f_t\left(x_t\right) - \eta\left\|\nabla f_t\left(x_t\right)\right\|_2^2 + \frac{\eta}{2s}\left\|\nabla f_t\left(x_t\right)\right\|_1^2 + \frac{\eta s}{2}\left\|\nabla f_t\left(x_t\right) - \mathbf{E}\left[G_t^\nu\left(x_t, u_t, \xi_t\right)|\mathcal{F}_t\right]\right\|_\infty^2$$
$$+ \eta^2\frac{L_G}{2}\mathbf{E}\left[\left\|G_t^\nu\left(x_t, u_t, \xi_t\right)\right\|_\infty^2 |\mathcal{F}_t\right].$$

Noting Lemma 9, the fact that $\|\nabla f_t(x_t)\|_1 \leq \sqrt{s}\|\nabla f_t(x_t)\|_2$ under Assumption 2.5 and after re-arranging the terms, we obtain

$$\frac{\eta}{2s}\left[1 - 16C\eta L_G s(\log d)^2\right]\left\|\nabla f_t\left(x_t\right)\right\|_1^2 \leq f_t\left(x_t\right) - \mathbf{E}\left[f_t\left(x_{t+1}\right)|\mathcal{F}_t\right]$$
$$+ C\eta L_G\left(\log d\right)^2\left[s\nu^2 C L_G \log d + 2\eta\left(\nu^2 L_G^2\log d + 4\sigma^2\right)\right].$$

Summing up both sides of the above inequality, noting (22) and Definition 1 we get (23). Noting Lemma 8 under Assumption 2.2, part (b) follows similarly. ■

Appendix B. Proof of Section 3.2.1

In order to prove Theorem 12, we require the following result from Nesterov and Polyak (2006).

Lemma 25 (Nesterov and Polyak (2006)) *Let $\{x_t\}$ be generated by Algorithm 2 with $M \geq L_H$. Then, we have*

$$\bar{G}_t + \bar{H}_t h_t + \frac{M}{2}\left\|h_t\right\|_2 h_t = 0. \tag{59a}$$
$$\bar{H}_t + \frac{M}{2}\left\|h_t\right\|_2 I_d \succcurlyeq 0. \tag{59b}$$
$$\bar{G}_t^\top h_t \leq 0. \tag{59c}$$

The following two lemma bounds the variance of the gradient and the Hessian estimators.

Lemma 26 *Under Assumption 2.2 with $\|\cdot\| = \|\cdot\|_2$, and Assumption 2.3 with $\|\cdot\| = \|\cdot\|_2$ we have*

$$\mathbf{E}\left[\|\bar{G}_t - \nabla_t\|_2^2\right] \leq \frac{\sigma^2}{m_t}. \tag{60}$$

Lemma 27 *Under Assumption 2.3 with $\|\cdot\| = \|\cdot\|_2$, and Assumption 2.4 we have*

$$\mathbf{E}\left[\|\bar{H}_t - \nabla_t^2\|_{op}^2\right] \leq \frac{\varkappa^2}{b_t}. \tag{61a}$$
$$\mathbf{E}\left[\|\bar{H}_t - \nabla_t^2\|_{op}^3\right] \leq \frac{2\varkappa^3}{b_t^{\frac{3}{2}}}. \tag{61b}$$

Lemma 26 and 27 are essentially simplified versions of Lemma 2.1, and Lemma 4.4 from Balasubramanian and Ghadimi (2021), and hence their proofs are omitted here.

In the rest of the proof we use ∇_t, ∇_t^2, h_t, and $\lambda_{t,\min}$ to denote $\nabla f_t\left(x_t\right)$, $\nabla^2 f_t\left(x_t\right)$, $(x_{t+1}-x_t)$, and the minimum eigenvalue of $\nabla^2 f_t(x_t)$ respectively. First, in Lemma 28 we show that $M\|h_t\|^3/36$ is bounded by $f_t(x_t)-f_t(x_{t+1})$ ignoring the terms $\|\nabla_t-\bar{G}_t\|_2^{3/2}$, and $\|\nabla_t^2-\bar{H}_t\|_{op}^3$. Then, in (66) we show that at every time point t, if we ignore the terms $\|\nabla_t-\bar{G}_t\|_2$, $\|\nabla_t^2-\bar{H}_t\|_{op}^2$, and $\|\nabla_t^2-\bar{H}_t\|_{op}^3$, $r_{NC}(t)=\max\left(\|\nabla_t\|_2, -\frac{8}{L_H^3}\lambda_{t,\min}^3\right)$ is upper bounded by a cubic polynomial of the ℓ_2 norm of the difference of consecutive iterates $h_t=x_{t+1}-x_t$. Combining these two results we want to form a telescopic sum to bound $R_{ENC}(t)=\sum_{t=1}^T r_{NC}(t)$ in expectation. But since this is a nonstationary environment, similar to the proof of Theorem 6, at every time step t an additional term appears of the form $f_t(x_t)-f_{t-1}(x_t)$ (see (69), (70)). Note that the sum of these additional terms can be bounded by W_T. Now observe that the terms we have been ignoring so far are different moments of gradient and hessian estimation error. We use Lemma 26, and Lemma 27 to bound there moments. Finally, one needs to choose M, ν, m_t, and b_t as in (30) to establish the rates.

Lemma 28 *Under Assumption 2.3 with $\|\cdot\|=\|\cdot\|_2$, and Assumption 2.4, for $M\geq L_H$, the points generated by Algorithm 3 satisfy the following*

$$\frac{M}{36}\left\|h_t\right\|_2^3\leq f_t\left(x_t\right)-f_t\left(x_{t+1}\right)+\frac{4}{\sqrt{3M}}\|\nabla_t-\bar{G}_t\|_2^{\frac{3}{2}}+\frac{24}{M^2}\|\nabla_t^2-\bar{H}_t\|_{op}^3. \tag{62}$$

Proof [Proof of Lemma 28] If $M\geq L_H$, using Assumption 2.4 we obtain

$$\begin{aligned}
f_t\left(x_{t+1}\right)\leq & f_t\left(x_t\right)+\nabla_t^\top h_t+\frac{1}{2}\langle\nabla_t^2 h_t,h_t\rangle+\frac{M}{6}\left\|h_t\right\|_2^3\\
\leq & f_t\left(x_t\right)+\bar{G}_t^\top h_t+\frac{1}{2}\langle\bar{H}_t h_t,h_t\rangle+\|\nabla_t-\bar{G}_t\|_2\|h_t\|_2+\frac{1}{2}\|\nabla_t^2-\bar{H}_t\|_{op}\|h_t\|_2^2+\frac{M}{6}\left\|h_t\right\|_2^3.
\end{aligned}$$

Now by using (59a) we hence obtain

$$f_t\left(x_{t+1}\right)\leq f_t\left(x_t\right)-\frac{1}{2}\langle\bar{H}_t h_t,h_t\rangle+\|\nabla_t-\bar{G}_t\|_2\|h_t\|_2+\frac{1}{2}\|\nabla_t^2-\bar{H}_t\|_{op}\|h_t\|_2^2-\frac{M}{3}\left\|h_t\right\|_2^3. \tag{63}$$

Combining (59a), and (59c) we get

$$-\frac{1}{2}\langle\bar{H}_t h_t,h_t\rangle-\frac{M}{3}\left\|h_t\right\|_2^3\leq-\frac{M}{12}\left\|h_t\right\|_2^3,$$

which combined with (63) gives

$$f_t\left(x_{t+1}\right)\leq f_t\left(x_t\right)+\|\nabla_t-\bar{G}_t\|_2\|h_t\|_2+\frac{1}{2}\|\nabla_t^2-\bar{H}_t\|_{op}\|h_t\|_2^2-\frac{M}{12}\left\|h_t\right\|_2^3.$$

Rearranging terms we then obtain

$$\frac{M}{12}\left\|h_t\right\|_2^3\leq f_t\left(x_t\right)-f_t\left(x_{t+1}\right)+\|\nabla_t-\bar{G}_t\|_2\|h_t\|_2+\frac{1}{2}\|\nabla_t^2-\bar{H}_t\|_{op}\|h_t\|_2^2.$$

Now by using Young's inequality, we have

$$\begin{aligned}
&\frac{M}{12}\|h_t\|_2^3 \le f_t(x_t) - f_t(x_{t+1}) + \frac{4}{\sqrt{3M}}\|\nabla_t - \bar{G}_t\|_2^{\frac{3}{2}} + \frac{24}{M^2}\|\nabla_t^2 - \bar{H}_t\|_{op}^3 + \frac{M}{18}\|h_t\|_2^3 \\
\implies &\frac{M}{36}\|h_t\|_2^3 \le f_t(x_t) - f_t(x_{t+1}) + \frac{4}{\sqrt{3M}}\|\nabla_t - \bar{G}_t\|_2^{\frac{3}{2}} + \frac{24}{M^2}\|\nabla_t^2 - \bar{H}_t\|_{op}^3,
\end{aligned}$$

which completes the proof. ■

Proof [Proof of Theorem 12] First note that using Assumption 2.3 with $\|\cdot\| = \|\cdot\|_2$, we have

$$\|\nabla f_t(x_t)\|_2 - \|\nabla f_t(x_{t+1})\|_2 \le \|\nabla f_t(x_{t+1}) - \nabla f_t(x_t)\|_2 \le L_G\|h_t\|_2. \tag{64}$$

Using, Assumption 2.4, (59a), (64), and Young's inequality, we also have

$$\begin{aligned}
&\|\nabla f_t(x_{t+1}) - \nabla_t - \nabla_t^2 h_t\|_2 \le \frac{L_H}{2}\|h_t\|_2^2, \\
&\|\nabla f_t(x_{t+1})\|_2 \le \|\nabla_t - \bar{G}_t\|_2 + \|\nabla_t^2 - \bar{H}_t\|_{op}\|h_t\|_2 + \frac{L_H + M}{2}\|h_t\|_2^2, \\
&\|\nabla_t\|_2 \le L_G\|h_t\|_2 + \|\nabla_t - \bar{G}_t\|_2 + \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}}{2(L_H + M)} + (L_H + M)\|h_t\|_2^2.
\end{aligned}$$

Now, from (59b), we obtain

$$\begin{aligned}
&-2\lambda_{t,\min} \le M\|h_t\|_2 + 2\|\nabla_t^2 - \bar{H}_t\|_{op} \\
\implies &\left(-\frac{2}{L_H}\lambda_{t,\min}\right)^3 \le \frac{32}{L_H^3}\|\bar{H}_t - \nabla_t^2\|_{op}^3 + \frac{4M^3}{L_H^3}\|h_t\|_2^3.
\end{aligned} \tag{65}$$

Now by combining (64), and (65), we obtain

$$\begin{aligned}
r_{NC}(t) = \max\left(\|\nabla_t\|, -\frac{8}{L_H^3}\lambda_{t,\min}^3\right) \le & L_G\|h_t\|_2 + \frac{1}{2}(L_H + M)\|h_t\|_2^2 + \frac{4M^3}{L_H^3}\|h_t\|_2^3 \\
&+ \frac{32}{L_H^3}\|\bar{H}_t - \nabla_t^2\|_{op}^3 + \|\nabla_t - \bar{G}_t\|_2 + \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2(L_H + M)}.
\end{aligned} \tag{66}$$

Now, let us consider two cases $\|h_t\|_2 \le T^{-\frac{1}{3}}W_T^{\frac{1}{3}}$, and $\|h_t\|_2 > T^{-\frac{1}{3}}W_T^{\frac{1}{3}}$.

1. When $\|h_t\|_2 \le T^{-\frac{1}{3}}W_T^{\frac{1}{3}}$, we have

$$\begin{aligned}
r_{NC}(t) \le & L_G T^{-\frac{1}{3}}W_T^{\frac{1}{3}} + \frac{1}{2}(L_H + M)T^{-\frac{2}{3}}W_T^{\frac{2}{3}} + \frac{4M^3}{L_H^3}T^{-1}W_T \\
&+ \|\nabla_t - \bar{G}_t\|_2 + \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2(L_H + M)} + \frac{32}{L_H^3}\|\bar{H}_t - \nabla_t^2\|_{op}^3.
\end{aligned} \tag{67}$$

2. When $\|h_t\|_2 > T^{-\frac{1}{3}} W_T^{\frac{1}{3}}$, by using (63) we get,

$$
\begin{aligned}
r_{NC}(t) \leq & \left(L_G T^{\frac{2}{3}} W_T^{-\frac{2}{3}} + \frac{1}{2}(L_H + M) T^{\frac{1}{3}} W_T^{-\frac{1}{3}} + \frac{4M^3}{L_H^3} \right) \|h_t\|_2^3 + \|\nabla_t - \bar{G}_t\|_2 \\
& + \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2(L_H + M)} + \frac{32}{L_H^3} \|\bar{H}_t - \nabla_t^2\|_{op}^3 \\
\leq & \left(36\frac{L_G}{M} + 18\left(\frac{L_H}{M} + 1\right) T^{-\frac{1}{3}} W_T^{-y} + \frac{144M^2}{L_H^3} T^{-\frac{2}{3}} W_T^{-2y} \right) T^{\frac{2}{3}} W_T^{2y} \left(f_t(x_t)\right. \\
& \left. - f_t(x_{t+1}) + \frac{4}{\sqrt{3M}} \|\nabla_t - \bar{G}_t\|_2^{\frac{3}{2}} + \frac{24}{M^2} \|\nabla_t^2 - \bar{H}_t\|_{op}^3 \right) + \|\nabla_t - \bar{G}_t\|_2 \\
& + \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2(L_H + M)} + \frac{32}{L_H^3} \|\bar{H}_t - \nabla_t^2\|_{op}^3.
\end{aligned}
\tag{68}
$$

Combining (67), and (68), we then have

$$
\begin{aligned}
r_{NC}(t) \leq & \left(L_G T^{-\frac{1}{3}} W_T^{\frac{1}{3}} + \frac{1}{2}(L_H + M) T^{-\frac{2}{3}} W_T^{\frac{2}{3}} + \frac{4M^3}{L_H^3} T^{-1} W_T \right) \\
& + \frac{32}{L_H^3} \|\bar{H}_t - \nabla_t^2\|_{op}^3 + \|\nabla_t - \bar{G}_t\|_2 + \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2(L_H + M)} \\
& + \left(36\frac{L_G}{M} + 18\left(\frac{L_H}{M} + 1\right) T^{-\frac{1}{3}} W_T^{\frac{1}{3}} + \frac{144M^2}{L_H^3} T^{-\frac{2}{3}} W_T^{\frac{2}{3}} \right) \\
& T^{\frac{2}{3}} W_T^{-\frac{2}{3}} \left(f_t(x_t) - f_t(x_{t+1}) + \frac{4}{\sqrt{3M}} \|\nabla_t - \bar{G}_t\|_2^{\frac{3}{2}} + \frac{24}{M^2} \|\nabla_t^2 - \bar{H}_t\|_{op}^3 \right).
\end{aligned}
\tag{69}
$$

Summing both sides from $t = 1$, to T, taking expectation on both sides and using Definition 5 we get

$$
\begin{aligned}
& \mathfrak{R}_{ENC}(T) = \sum_{t=1}^{T} \mathbf{E}\left[r_{NC}(t)\right] \\
\leq & \left(L_G T^{\frac{2}{3}} W_T^{\frac{1}{3}} + \frac{1}{2}(L_H + M) T^{\frac{1}{3}} W_T^{\frac{2}{3}} + \frac{4M^3}{L_H^3} W_T \right) + \sum_{t=1}^{T} \left(\frac{32}{L_H^3} \mathbf{E}\left[\|\bar{H}_t - \nabla_t^2\|_{op}^3\right] \right. \\
& \left. + \mathbf{E}\left[\|\nabla_t - \bar{G}_t\|_2\right] + \frac{\mathbf{E}\left[\|\nabla_t^2 - \bar{H}_t\|_{op}^2\right]}{2(L_H + M)} \right) \\
& + \left(36\frac{L_G}{M} + 18\left(\frac{L_H}{M} + 1\right) T^{-\frac{1}{3}} W_T^{\frac{1}{3}} + \frac{144M^2}{L_H^3} T^{-\frac{2}{3}} W_T^{\frac{2}{3}} \right) T^{\frac{2}{3}} W_T^{-\frac{2}{3}} \left(f_1(x_1) - f_T(x_{T+1}) + W_T \right. \\
& \left. + \frac{4}{\sqrt{3M}} \sum_{t=1}^{T} \mathbf{E}\left[\|\nabla_t - \bar{G}_t\|_2^{\frac{3}{2}}\right] + \frac{24}{M^2} \sum_{t=1}^{T} \mathbf{E}\left[\|\nabla_t^2 - \bar{H}_t\|_{op}^3\right] \right).
\end{aligned}
\tag{70}
$$

Now choosing M, m_t, and b_t as in (30) and Lemma 26 and Lemma 27 we get,

$$
\mathfrak{R}_{ENC}(T) \leq \mathcal{O}\left(T^{\frac{2}{3}} \left(1 + W_T^{\frac{1}{3}}\right) + T^{\frac{2}{3}} \left(\sigma + \varkappa^3\right) + T^{\frac{5}{9}} W_T^{\frac{2}{9}} \varkappa^2 \right),
\tag{71}
$$

which completes the proof. ■

Appendix C. Proofs of Section 3.2.2

Before we prove Theorem 15, we state some preliminary results that are required for the proof.

Lemma 29 *Let* $x_{t+1} = \text{argmin}_y \tilde{f}_t\left(x_t, y, \bar{G}_t, \bar{H}_t, h_t, M\right)$ *and* $M \geq L_H$. *Then, we have*

$$\bar{G}_t + \bar{H}_t h_t + \frac{M}{2}\left\|h_t\right\|_2 h_t = 0. \tag{72a}$$

$$\bar{H}_t + \frac{M}{2}\left\|h_t\right\|_2 I_d \succcurlyeq 0. \tag{72b}$$

$$\bar{G}_t^\top h_t \leq 0. \tag{72c}$$

Lemma 29 is essentially the same as Lemma 25 but we restate it here to emphasize that it holds for bandit cubic-regularized Newton method as well. The following two lemma bounds the variance of the gradient and the Hessian estimators.

Lemma 30 (Balasubramanian and Ghadimi (2021)) *Under Assumption 2.2 with* $\|\cdot\| = \|\cdot\|_2$, *and Assumption 2.3 with* $\|\cdot\| = \|\cdot\|_2$ *we have*

$$\mathbf{E}\left[\|\bar{G}_t - \nabla_t\|_2^2\right] \leq \frac{3\nu^2}{2}L_G^2\left(d+3\right)^3 + \frac{4\left(L^2+\sigma^2\right)\left(d+5\right)}{m_t}. \tag{73}$$

Lemma 31 (Balasubramanian and Ghadimi (2021)) *For* $b_t \geq 4\left(1 + 2\log 2d\right)$, *under Assumption 2.3 with* $\|\cdot\| = \|\cdot\|_2$, *and Assumption 2.4 we have*

$$\mathbf{E}\left[\|\bar{H}_t - \nabla_t^2\|_{op}^2\right] \leq 3L_H^2\left(d+16\right)^5\nu^2 + \frac{128\left(1+2\log 2d\right)\left(d+16\right)^4 L_G^2}{3b_t}. \tag{74a}$$

$$\mathbf{E}\left[\|\bar{H}_t - \nabla_t^2\|_{op}^3\right] \leq 21L_H^3\left(d+16\right)^{\frac{15}{2}}\nu^3 + \frac{160\sqrt{1+2\log 2d}\left(d+16\right)^6 L_G^3}{b_t^{\frac{3}{2}}}. \tag{74b}$$

Before proceeding with the proof, we highlight the main steps. First we show that at every time point t, if we ignore the terms $\|\nabla_t - \bar{G}_t\|_2$, $\|\nabla_t^2 - \bar{H}_t\|_{op}^2$, and $\|\nabla_t^2 - \bar{H}_t\|_{op}^3$, $r_{NC}(t) = \max\left(\|\nabla_t\|_2, -\frac{8}{L_H^3}\lambda_{t,\min}^3\right)$ is upper bounded by the cube of the ℓ_2 norm of the difference of consecutive iterates $x_{t+1}-x_t$, $\|h_t\|^3$ (Lemma 32). Then we show that $M\|h_t\|^3/36$ is bounded by $f_t(x_t) - f_t(x_{t+1})$ ignoring the terms $\|\nabla_t - \bar{G}_t\|_2^{3/2}$, and $\|\nabla_t^2 - \bar{H}_t\|_{op}^3$. Now we want to form a telescopic sum to bound $R_{ENC}(t) = \sum_{t=1}^T r_{NC}(t)$ in expectation. But since this is a nonstationary environment, similar to the proof of Theorem 6, at every time step t an additional term appears of the form $f_t(x_t) - f_{t-1}(x_t)$. Note that the sum of these additional terms can be bounded by W_T. Now observe that the terms we have been ignoring so far are different moments of gradient and hessian estimation error. We use Lemma 30, and Lemma 31 to bound there moments. Finally, one needs to choose M, ν, m_t, and b_t suitably to establish the rates. The detailed proof of Theorem 15 is in Appendix C.

Lemma 32 *Under Assumption 2.2 with $\|\cdot\| = \|\cdot\|_2$, Assumption 2.3 with $\|\cdot\| = \|\cdot\|_2$, and Assumption 2.4, the points generated by Algorithm 3 satisfy the following*

$$r_{NC}(t) = \max\left(\|\nabla_t\|_2, -\frac{8}{L_H^3}\lambda_{t,\min}^3\right) \leq L_G\|h_t\|_2 + \frac{1}{2}\left(L_H + M\right)\|h_t\|_2^2 + \frac{4M^3}{L_H^3}\|h_t\|_2^3 + \frac{32}{L_H^3}\|\bar{H}_t - \nabla_t^2\|_{op}^3 + \|\nabla_t - \bar{G}_t\|_2 + \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2\left(L_H + M\right)}. \tag{75}$$

Proof Under Assumption 2.4, using (72a) and Young's inequality, we have

$$\begin{aligned}
&\|\nabla f_t(x_{t+1}) - \nabla_t - \nabla_t^2 h_t\|_2 \leq \frac{L_H}{2}\|h_t\|_2^2 \\
\implies &\|\nabla f_t(x_{t+1})\|_2 \leq \|\nabla_t - \bar{G}_t\|_2 + \|\nabla_t^2 - \bar{H}_t\|_{op}(x_{t+1} - x_t) + \frac{L_H + M}{2}\|h_t\|_2^2 \\
&\leq \|\nabla_t - \bar{G}_t\|_2 + \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2\left(L_H + M\right)} + \left(L_H + M\right)\|h_t\|_2^2.
\end{aligned}$$

Under Assumption 2.3, we get

$$\|\nabla_t\|_2 \leq L_G\|h_t\|_2 + \|\nabla_t - \bar{G}_t\|_2 + \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2\left(L_H + M\right)} + \left(L_H + M\right)\|h_t\|_2^2. \tag{76}$$

From (72b) we get,

$$\begin{aligned}
&-2\lambda_{t,\min} \leq M\|h_t\|_2 + 2\|\nabla_t^2 - \bar{H}_t\|_{op} \\
\implies &\left(-\frac{2}{L_H}\lambda_{t,\min}\right)^3 \leq \frac{32}{L_H^3}\|\bar{H}_t - \nabla_t^2\|_{op}^3 + \frac{4M^3}{L_H^3}\|h_t\|_2^3.
\end{aligned} \tag{77}$$

Combining (76), and (77), and choosing $M = L_H$ we get

$$r_{NC}(t) = \max\left(\|\nabla_t\|_2, -\frac{8}{L_H^3}\lambda_{t,\min}^3\right) \leq L_G\|h_t\|_2 + \frac{1}{2}\left(L_H + M\right)\|h_t\|_2^2 + \frac{4M^3}{L_H^3}\|h_t\|_2^3 + \frac{32}{L_H^3}\|\bar{H}_t - \nabla_t^2\|_{op}^3 + \|\nabla_t - \bar{G}_t\|_2 + \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2\left(L_H + M\right)},$$

which completes the proof. ■

Lemma 33 *Under Assumption 2.3, and Assumption 2.4, for $M \geq L_H$, the points generated by Algorithm 3 satisfy the following*

$$\frac{M}{36}\|h_t\|_2^3 \leq f_t(x_t) - f_t(x_{t+1}) + \frac{4}{\sqrt{3M}}\|\nabla_t - \bar{G}_t\|_2^{\frac{3}{2}} + \frac{24}{M^2}\|\nabla_t^2 - \bar{H}_t\|_{op}^3. \tag{78}$$

Proof If $M \geq L_H$, using Assumption 2.4

$$f_t(x_{t+1}) \leq f_t(x_t) + \nabla_t^\top h_t + \frac{1}{2}\langle\nabla_t^2 h_t, h_t\rangle + \frac{M}{6}\|h_t\|_2^3$$

$$\leq f_t(x_t) + \bar{G}_t^\top h_t + \frac{1}{2}\langle \bar{H}_t h_t, h_t\rangle + \|\nabla_t - \bar{G}_t\|_2 \|h_t\|_2 + \frac{1}{2}\|\nabla_t^2 - \bar{H}_t\|_{op}\|h_t\|_2^2 + \frac{M}{6}\|h_t\|_2^3.$$

Using (72a) we get

$$f_t(x_{t+1}) \leq f_t(x_t) - \frac{1}{2}\langle \bar{H}_t h_t, h_t\rangle + \|\nabla_t - \bar{G}_t\|_2 \|h_t\|_2 + \frac{1}{2}\|\nabla_t^2 - \bar{H}_t\|_{op}\|h_t\|_2^2 - \frac{M}{3}\|h_t\|_2^3. \tag{79}$$

Combining (72a), and (72c) we get

$$-\frac{1}{2}\langle \bar{H}_t h_t, h_t\rangle - \frac{M}{3}\|h_t\|_2^3 \leq -\frac{M}{12}\|h_t\|_2^3,$$

which combined with (63) gives

$$f_t(x_{t+1}) \leq f_t(x_t) + \|\nabla_t - \bar{G}_t\|_2 \|h_t\|_2 + \frac{1}{2}\|\nabla_t^2 - \bar{H}_t\|_{op}\|h_t\|_2^2 - \frac{M}{12}\|h_t\|_2^3.$$

Rearranging terms we get

$$\frac{M}{12}\|h_t\|_2^3 \leq f_t(x_t) - f_t(x_{t+1}) + \|\nabla_t - \bar{G}_t\|_2 \|h_t\|_2 + \frac{1}{2}\|\nabla_t^2 - \bar{H}_t\|_{op}\|h_t\|_2^2.$$

Using Young's inequality

$$\begin{aligned}
&\frac{M}{12}\|h_t\|_2^3 \leq f_t(x_t) - f_t(x_{t+1}) + \frac{4}{\sqrt{3M}}\|\nabla_t - \bar{G}_t\|_2^{\frac{3}{2}} + \frac{24}{M^2}\|\nabla_t^2 - \bar{H}_t\|_{op}^3 + \frac{M}{18}\|h_t\|_2^3 \\
\implies &\frac{M}{36}\|h_t\|_2^3 \leq f_t(x_t) - f_t(x_{t+1}) + \frac{4}{\sqrt{3M}}\|\nabla_t - \bar{G}_t\|_2^{\frac{3}{2}} + \frac{24}{M^2}\|\nabla_t^2 - \bar{H}_t\|_{op}^3,
\end{aligned}$$

which completes the proof. ■

Proof [Proof of Theorem 15] Let us consider two cases $\|h_t\|_2 \leq (W_T/T)^{\frac{1}{3}}$, and $\|h_t\|_2 > (W_T/T)^{\frac{1}{3}}$.

1. When $\|h_t\|_2 \leq T^{-\frac{1}{3}} W_T^{\frac{1}{3}}$, we have

$$\begin{aligned}
r_{NC}(t) \leq &L_G T^{-\frac{1}{3}} W_T^{\frac{1}{3}} + \frac{1}{2}(L_H + M) T^{-\frac{2}{3}} W_T^{\frac{2}{3}} + \frac{4M^3}{L_H^3} T^{-1} W_T \\
&+ \|\nabla_t - \bar{G}_t\|_2 + \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2(L_H + M)} + \frac{32}{L_H^3}\|\bar{H}_t - \nabla_t^2\|_{op}^3.
\end{aligned} \tag{80}$$

2. When $\|h_t\|_2 > T^{-\frac{1}{3}} W_T^{\frac{1}{3}}$, using (79) we obtain,

$$r_{NC}(t) \leq \left(L_G T^{\frac{2}{3}} W_T^{-\frac{2}{3}} + \frac{1}{2}(L_H + M) T^{\frac{1}{3}} W_T^{-\frac{1}{3}} + \frac{4M^3}{L_H^3}\right)\|h_t\|_2^3 + \|\nabla_t - \bar{G}_t\|_2$$

$$
\begin{aligned}
&+ \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2\left(L_H + M\right)} + \frac{32}{L_H^3}\|\bar{H}_t - \nabla_t^2\|_{op}^3 \\
&\leq \left(36\frac{L_G}{M} + 18\left(\frac{L_H}{M} + 1\right)T^{-\frac{1}{3}}W_T^{-y} + \frac{144M^2}{L_H^3}T^{-\frac{2}{3}}W_T^{-2y}\right)T^{\frac{2}{3}}W_T^{2y}\left(f_t\left(x_t\right)\right. \\
&\left. - f_t\left(x_{t+1}\right) + \frac{4}{\sqrt{3M}}\|\nabla_t - \bar{G}_t\|_2^{\frac{3}{2}} + \frac{24}{M^2}\|\nabla_t^2 - \bar{H}_t\|_{op}^3\right) + \|\nabla_t - \bar{G}_t\|_2 \\
&+ \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2\left(L_H + M\right)} + \frac{32}{L_H^3}\|\bar{H}_t - \nabla_t^2\|_{op}^3. \qquad (81)
\end{aligned}
$$

Now by combining (80), and (81), we have

$$
\begin{aligned}
r_{NC}\left(t\right) \leq &\left(L_G T^{-\frac{1}{3}}W_T^{\frac{1}{3}} + \frac{1}{2}\left(L_H + M\right)T^{-\frac{2}{3}}W_T^{\frac{2}{3}} + \frac{4M^3}{L_H^3}T^{-1}W_T\right) \\
&+ \frac{32}{L_H^3}\|\bar{H}_t - \nabla_t^2\|_{op}^3 + \|\nabla_t - \bar{G}_t\|_2 + \frac{\|\nabla_t^2 - \bar{H}_t\|_{op}^2}{2\left(L_H + M\right)} \\
&+ \left(36\frac{L_G}{M} + 18\left(\frac{L_H}{M} + 1\right)T^{-\frac{1}{3}}W_T^{\frac{1}{3}} + \frac{144M^2}{L_H^3}T^{-\frac{2}{3}}W_T^{\frac{2}{3}}\right) \\
&T^{\frac{2}{3}}W_T^{-\frac{2}{3}}\left(f_t\left(x_t\right) - f_t\left(x_{t+1}\right) + \frac{4}{\sqrt{3M}}\|\nabla_t - \bar{G}_t\|_2^{\frac{3}{2}} + \frac{24}{M^2}\|\nabla_t^2 - \bar{H}_t\|_{op}^3\right). \qquad (82)
\end{aligned}
$$

Summing both sides from $t = 1$, to T, taking expectation on both sides and using Definition 5 we get

$$
\begin{aligned}
&\mathfrak{R}_{ENC}\left(T\right) = \sum_{t=1}^{T}\mathbf{E}\left[r_{NC}\left(t\right)\right] \\
&\leq \left(L_G T^{\frac{2}{3}}W_T^{\frac{1}{3}} + \frac{1}{2}\left(L_H + M\right)T^{\frac{1}{3}}W_T^{\frac{2}{3}} + \frac{4M^3}{L_H^3}W_T\right) + \sum_{t=1}^{T}\left(\frac{32}{L_H^3}\mathbf{E}\left[\|\bar{H}_t - \nabla_t^2\|_{op}^3\right]\right. \\
&\left. + \mathbf{E}\left[\|\nabla_t - \bar{G}_t\|_2\right] + \frac{\mathbf{E}\left[\|\nabla_t^2 - \bar{H}_t\|_{op}^2\right]}{2\left(L_H + M\right)}\right) \\
&+ \left(36\frac{L_G}{M} + 18\left(\frac{L_H}{M} + 1\right)T^{-\frac{1}{3}}W_T^{\frac{1}{3}} + \frac{144M^2}{L_H^3}T^{-\frac{2}{3}}W_T^{\frac{2}{3}}\right)T^{\frac{2}{3}}W_T^{-\frac{2}{3}}\left(f_1\left(x_1\right) - f_T\left(x_{T+1}\right) + W_T\right. \\
&\left. + \frac{4}{\sqrt{3M}}\sum_{t=1}^{T}\mathbf{E}\left[\|\nabla_t - \bar{G}_t\|_2^{\frac{3}{2}}\right] + \frac{24}{M^2}\sum_{t=1}^{T}\mathbf{E}\left[\|\nabla_t^2 - \bar{H}_t\|_{op}^3\right]\right). \qquad (83)
\end{aligned}
$$

Now choosing ν, M, m_t, and b_t as in (37) and Lemma 30 and Lemma 31 we get

$$
\mathfrak{R}_{ENC}\left(T\right) \leq \mathcal{O}\left(T^{\frac{2}{3}}\left(1 + W_T^{\frac{1}{3}}\right)\left(1 + \sigma + \sigma^{\frac{3}{2}}\right)\right), \qquad (84)
$$

which completes the proof. ■

Appendix D. Proofs of Section 3.3

Proof [Proof of Theorem 18] In this setting (57) changes to the following

$$\sum_{t=1}^{T} \mathbf{E}\left[\|\nabla f_t(x_t)\|_2^2\right] \leq \frac{2}{\eta}\left(f_1(x_1) - \mathbf{E}\left[f_T(x_{T+1})\right] + W_T\right) + \frac{T}{4}\nu^2 L_G^2 (d+3)^3 + \eta T \frac{\nu^2}{2} L_G^3 (d+6)^3 + 2\eta L_G (d+4) \sum_{t=1}^{T} \mathbf{E}\left[\|\nabla f_t(x_t)\|_2^2 + \frac{\sigma^2}{m\nu^2}\right]. \tag{85}$$

The rest of the proof is similar to that of Theorem 6 and is hence omitted. ■

Proof [Proof of Lemma 20] First note that we have

$$\begin{aligned}
&\mathbf{E}\left[\|\bar{G}_t - \nabla_t\|_2^2\right] \\
=&\mathbf{E}\left[\left\|\frac{1}{m_t}\sum_{i=1}^{m_t}\frac{F_t(x_t+\nu u_{t,i},\xi_{t,i}) - F_t\left(x_t,\xi'_{t,i}\right)}{\nu}u_{t,i} - \nabla_t\right\|_2^2\right] \\
\leq&2\mathbf{E}\left[\left\|\frac{1}{m_t}\sum_{i=1}^{m_t}\frac{F_t(x_t+\nu u_{t,i},\xi_{t,i}) - F_t\left(x_t+\nu u_{t,i},\xi'_{t,i}\right)}{\nu}u_{t,i}\right\|_2^2\right] \\
&+2\mathbf{E}\left[\left\|\frac{1}{m_t}\sum_{i=1}^{m_t}\frac{F_t\left(x_t+\nu u_{t,i},\xi'_{t,i}\right) - F_t\left(x_t,\xi'_{t,i}\right)}{\nu}u_{t,i} - \nabla_t\right\|_2^2\right] \\
=&2\mathbf{E}\left[\frac{1}{m_t^2}\sum_{i=1}^{m_t}\left\|\frac{F_t(x_t+\nu u_{t,i},\xi_{t,i}) - F_t\left(x_t+\nu u_{t,i},\xi'_{t,i}\right)}{\nu}u_{t,i}\right\|_2^2\right] + 2\mathbf{E}\left[\left\|\bar{G}_t - \nabla_t\right\|_2^2\right] \\
=&\frac{4d\sigma^2 L'^2}{m_t\nu^2} + 3\nu^2 L_G^2 (d+3)^3 + \frac{8\left(L^2+\sigma^2\right)(d+5)}{m_t}.
\end{aligned}$$

The last inequality follows from (73), thereby completing the proof. ■

Proof [Proof of Lemma 21] First note that we have

$$\tilde{H} = \frac{1}{b_t}\sum_{i=1}^{b_t}\frac{F(x+\nu u_i,\xi_{i,+}) + F(x-\nu u_i,\xi_{i,-}) - 2F(x,\xi_{i,0})}{2\nu^2}\left(u_i u_i^\top - I_d\right) \tag{86}$$

$$= \frac{1}{b_t}\sum_{i=1}^{b_t}\frac{F(x+\nu u_i,\xi_{i,0}) + F(x-\nu u_i,\xi_{i,0}) - 2F(x,\xi_{i,0})}{2\nu^2}\left(u_i u_i^\top - I_d\right) + \frac{1}{b_t}\sum_{i=1}^{b_t}\tau_i \tag{87}$$

$$= \bar{H} + \bar{\tau}, \tag{88}$$

where

$$\tau_i = \frac{F\left(x+\nu u_i,\xi_{i,+}\right)+F\left(x-\nu u_i,\xi_{i,-}\right)-F\left(x+\nu u_i,\xi_{i,0}\right)-F\left(x-\nu u_i,\xi_{i,0}\right)}{2\nu^2}\left(u_i u_i^\top - I_d\right),$$

and $\bar{\tau} = \sum_{i=1}^{b_t} \tau_i / b_t$. Then one can write

$$\|\tilde{H} - \nabla^2\|_{op}^2 = \|\bar{H} + \bar{\tau} - \nabla^2\|_{op}^2 \leq 2\|\bar{H} - \nabla^2\|_{op}^2 + 2\|\bar{\tau}\|_{op}^2. \tag{89}$$

Note that τ_i and τ_j are independent for all $i \neq j$. Also, we have

$$\mathbf{E}\left[\tau_i\right] = \mathbf{E}\left[\mathbf{E}\left[\tau_i | x, u_i\right]\right] = 0. \tag{90}$$

Hene, we obtain

$$\begin{aligned}
\mathbf{E}\left[\|\bar{\tau}\|_{op}^2\right] \leq& \frac{4C(d)}{b_t^2}\sum_{i=1}^{b_t}\mathbf{E}\left[\|\tau_i\|_{op}^2\right] &(91)\\
\leq& \frac{2L'^2 C(d)}{b_t^2}\sum_{i=1}^{b_t}\mathbf{E}\left[\frac{|\xi_{i,+}-\xi_{i,0}|^2 + |\xi_{i,-}-\xi_{i,0}|^2}{\nu^4}\|u_i u_i^\top - I_d\|_{op}^2\right] &(92)\\
\leq& \frac{8L'^2\sigma^2 C(d)}{b_t^2 \nu^4}\sum_{i=1}^{b_t}\mathbf{E}\left[\|u_i u_i^\top - I_d\|_{op}^2\right] &(93)\\
\leq& \frac{32 L'^2 \sigma^2 d C(d)}{b_t \nu^4}, &(94)
\end{aligned}$$

where $C(d) = 4(1 + 2\log 2d)$. Furthermore, we also have

$$\mathbf{E}\left[\|\bar{\tau}\|_{op}^3\right] \leq \mathbf{E}\left[\|\bar{\tau}\|_{op}\|\bar{\tau}\|_F^2\right] \leq \left(\mathbf{E}\left[\|\bar{\tau}\|_{op}^2\right]\mathbf{E}\left[\|\bar{\tau}\|_F^4\right]\right)^{\frac{1}{2}}. \tag{95}$$

Now we upper bound $\mathbf{E}\left[\|\bar{\tau}\|_F^4\right]$ as follows. First, we write $\tau_i = a_i(u_i u_i^\top - I_d)$ where

$$a_i = \frac{F\left(x+\nu u_i,\xi_{i,+}\right)+F\left(x-\nu u_i,\xi_{i,-}\right)-F\left(x+\nu u_i,\xi_{i,0}\right)-F\left(x-\nu u_i,\xi_{i,0}\right)}{2\nu^2}.$$

Observe that, a_i are independent for $i = 1, 2, \cdots, b$, $\mathbf{E}\left[a_i\right] = 0$, $\mathbf{E}\left[a_i | u_i\right] = 0$, and

$$\mathbf{E}\left[a_i^2\right] \leq \frac{2L'^2\sigma^2}{\nu^4} \qquad \mathbf{E}\left[a_i^4\right] \leq \frac{16 L'^4 \sigma'}{\nu^8}.$$

Hence, we have

$$\begin{aligned}
&\mathbf{E}\left[\|\sum_{i=1}^{b_t}\tau_i\|_F^4\right]\\
=&\mathbf{E}\left[\|\sum_{i=1}^{b_t} a_i(u_i u_i^\top - I_d)\|_F^4\right]
\end{aligned}$$

$$
\begin{aligned}
=&\mathbf{E}\left[\operatorname{tr}\left(\sum_{i=1}^{b_t}\sum_{j=1}^{b_t} a_i a_j (u_i u_i^\top - I_d)(u_j u_j^\top - I_d)\right)^2\right]\\
=&\mathbf{E}\left[\left(\operatorname{tr}\left(\sum_{i=1}^{b_t} a_i^2 (u_i u_i^\top - I_d)^2\right) + \operatorname{tr}\left(\sum_{i=1}^{b_t}\sum_{j\neq i} a_i a_j (u_i u_i^\top - I_d)(u_j u_j^\top - I_d)\right)\right)^2\right]\\
=&\mathbf{E}\left[\left(\sum_{i=1}^{b_t} a_i^2((\|u_i\|_2^2 - 2)\|u_i\|_2^2 + d) + \sum_{i=1}^{b_t}\sum_{j\neq i} a_i a_j ((u_i^\top u_j)^2 - \|u_i\|_2^2 - \|u_j\|_2^2 + d)\right)^2\right]\\
\leq&\mathbf{E}\left[\underbrace{2b_t \sum_{i=1}^{b_t} a_i^4((\|u_i\|_2^2 - 2)\|u_i\|_2^2 + d)^2}_{I_1} + \underbrace{2\left(\sum_{i=1}^{b_t}\sum_{j\neq i} a_i a_j ((u_i^\top u_j)^2 - \|u_i\|_2^2 - \|u_j\|_2^2 + d)\right)^2}_{I_2}\right].
\end{aligned}
\tag{96}
$$

Now note that $\mathbf{E}[I_1]$ is of the order $b_t\sigma' d^2 {L'}^4/\nu^8$. Furthermore, we have

$$
\begin{aligned}
&\mathbf{E}[I_2]\\
=&2\mathbf{E}\left[\left(\sum_{i=1}^{b_t}\sum_{j\neq i} a_i a_j ((u_i^\top u_j)^2 - \|u_i\|_2^2 - \|u_j\|_2^2 + d)\right)^2\right]\\
=&2\mathbf{E}\left[\underbrace{\sum_{i=1}^{b_t}\sum_{j\neq i} a_i^2 a_j^2 ((u_i^\top u_j)^2 - \|u_i\|_2^2 - \|u_j\|_2^2 + d)^2}_{I_3}\right.\\
&\left.+\underbrace{\sum_{i=1}^{b_t}\sum_{j\neq i}\sum_{m=1}^{b_t}\sum_{\substack{m\neq n\\(m,n)\neq(i,j)}} a_i a_j a_m a_n ((u_i^\top u_j)^2 - \|u_i\|_2^2 - \|u_j\|_2^2 + d)((u_m^\top u_n)^2 - \|u_m\|_2^2 - \|u_n\|_2^2 + d)}_{I_4}\right].
\end{aligned}
\tag{97}
$$

Note that $\mathbf{E}[I_3]$ is of the order ${L'}^4\sigma^4 {b_t}^2 d^2/\nu^8$. Now, we turn to I_4. Note that in each term of I_4, note that there is at least one index n for which a_n is independent from the other terms in the product. Conditioning on the other terms, and taking expectation, one can see that all the terms in $\mathbf{E}[I_4]$ is 0. Combining (96), and (97), we have

$$
\mathbf{E}\left[\|\sum_{i=1}^{b_t}\tau_i\|_F^4\right] = \frac{K_1 {b_t}^2 d^2 {L'}^4 \sigma_1}{\nu^8}. \tag{98}
$$

where $\sigma_1 = \max(\sigma^4, \sigma')$ and K_1 is a constant. From (94), (95), and (98), we have,

$$\mathbf{E}\left[\|\bar{\tau}\|_{op}^3\right] \le \left(\frac{K_2 {L'}^6 \sigma_1 \sigma^2 (1 + 2\log 2d) d^3}{{b_t}^3 \nu^{12}}\right)^{\frac{1}{2}}, \tag{99}$$

where K_2 is a constant, thereby completing the proof. ■

Appendix E. Proof of Section 3.4

Proof [Proof of Proposition 24] First we show that $\bar{\mathfrak{R}}_G^{(2)}(T) = \mathcal{O}\left(\mathfrak{R}_G^{(2)}(T)\right)$. To do so, note that we have

$$\mathbf{E}\left[\|\nabla f_t(x_t + \nu u_{t,j})\|_2^2\right] \le \mathbf{E}\left[2\|\nabla f_t(x_t)\|_2^2 + 2L_G^2\nu^2\|u_{t,j}\|_2^2\right] \le 2\mathbf{E}\left[\|\nabla f_t(x_t)\|_2^2\right] + 2dL_G^2\nu^2.$$

Then, for the choices of ν of Theorem 10, and Theorem 18, we get

$$\bar{\mathfrak{R}}_G^{(2)}(T) = \mathcal{O}\left(\mathfrak{R}_G^{(2)}(T)\right).$$

In case of Theorem 10, we get,

$$\mathbf{E}\left[\|\nabla f_t(x_t + \nu u_{t,j})\|_1^2\right] \le \mathbf{E}\left[2\|\nabla f_t(x_t)\|_1^2 + 2L_G^2\nu^2\|u_{t,j}\|_\infty^2\right] \le 2\mathbf{E}\left[\|\nabla f_t(x_t)\|_1^2\right] + 2L_G^2\nu^2\log d.$$

Here too, one can see that choosing ν as in Theorem 10 gives,

$$\bar{\mathfrak{R}}_G^{(1)}(T) = \mathcal{O}\left(\mathfrak{R}_G^{(1)}(T)\right).$$

Now we show that $\bar{\mathfrak{R}}_{ENC}(T) = \mathcal{O}\left(\mathfrak{R}_{ENC}(T)\right)$. To do so, note that we have

$$\begin{aligned}
&\mathbf{E}\left[\max\left(\|\nabla f_t(x_t + \nu u_{t,j})\|_2, \left(-\frac{2}{L_H}\lambda_{\min}\left(\nabla^2 f_t(x_t + \nu u_{t,j})\right)\right)^3\right)\right] \\
\le& \mathbf{E}\Big[\max\Big(\|\nabla f_t(x_t)\|_2 + L_G\nu\|u_{t,j}\|_2, \\
&\left(|-\frac{2}{L_H}\lambda_{\min}\left(\nabla^2 f_t(x_t + \nu u_{t,j}) - \nabla^2 f_t(x_t)\right)| - \frac{2}{L_H}\lambda_{\min}\left(\nabla^2 f_t(x_t)\right)\right)^3\Big)\Big] \\
\le& \mathbf{E}\left[\max\left(\|\nabla f_t(x_t)\|_2 + L_G\nu\|u_{t,j}\|_2, 32\nu^3\|u_{t,j}\|_2^3 - \left(\frac{32}{L_H^3}\lambda_{\min}\left(\nabla^2 f_t(x_t)\right)\right)^3\right)\right] \\
\le& \mathbf{E}\left[\max\left(\|\nabla f_t(x_t)\|_2 + L_G\nu\|u_{t,j}\|_2 + 32\nu^3\|u_{t,j}\|_2^3, L_G\nu\|u_{t,j}\|_2 + 32\nu^3\|u_{t,j}\|_2^3 - \left(\frac{32}{L_H^3}\lambda_{\min}\left(\nabla^2 f_t(x_t)\right)\right)^3\right)\right] \\
\le& \mathbf{E}\left[\max\left(\|\nabla f_t(x_t)\|_2, -\left(\frac{32}{L_H^3}\lambda_{\min}\left(\nabla^2 f_t(x_t)\right)\right)^3\right) + L_G\nu\|u_{t,j}\|_2 + 32\nu^3\|u_{t,j}\|_2^3\right] \\
\le& \mathbf{E}\left[\max\left(\|\nabla f_t(x_t)\|_2, -\left(\frac{32}{L_H^3}\lambda_{\min}\left(\nabla^2 f_t(x_t)\right)\right)^3\right)\right] + L_G\nu\sqrt{d} + 32\nu^3(d+3)^{\frac{3}{2}}.
\end{aligned}$$

The first and second inequality follows from Lipschitz continuity of gradient and hessian (Assumption 2.3–2.4), and Hölder's inequality. Then,

$$\begin{aligned}\bar{\mathfrak{R}}_{ENC}\left(T\right) \leq &\sum_{t=1}^{T}\mathbf{E}\left[\max\left(\|\nabla f_t(x_t)\|_2, -\left(\frac{32}{L_H^3}\lambda_{\min}\left(\nabla^2 f_t\left(x_t\right)\right)\right)^3\right)\right] + \nu L_G T\sqrt{d} + 32\nu^3 T(d+3)^{\frac{3}{2}} \\ \leq &\mathfrak{R}_{ENC}\left(T\right) + \nu L_G T\sqrt{d} + 32\nu^3 T(d+3)^{\frac{3}{2}}.\end{aligned}$$

For Theorem 15, and Theorem 22, by choosing ν_G, and ν_H as in (37), and (53) respectively, we get in both cases,

$$\bar{\mathfrak{R}}_{ENC}\left(T\right) = \mathcal{O}\left(\mathfrak{R}_{ENC}\left(T\right)\right).$$

■

Acknowledgments

We thank the Action Editor Andreas Krause and the two anonymous reviewers for their insightful comments and pointers to the relevant literature that helped to greatly improve the paper. We also acknowledge support for this project from the National Science Foundation via NSF TRIPODS grant CCF-1934568.

References

Alekh Agarwal, Ofer Dekel, and Lin Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Conference on Learning Theory (COLT)*, pages 28–40, 2010.

Alekh Agarwal, Dean P Foster, Daniel J Hsu, Sham M Kakade, and Alexander Rakhlin. Stochastic convex optimization with bandit feedback. In *Advances in Neural Information Processing Systems*, pages 1035–1043, 2011.

Arya Akhavan, Massimiliano Pontil, and Alexandre Tsybakov. Exploiting higher order smoothness in derivative-free optimization and continuous bandits. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Robin Allesiardo, Raphaël Féraud, and Odalric-Ambrym Maillard. The non-stationary stochastic multi-armed bandit problem. *International Journal of Data Science and Analytics*, 3(4):267–283, 2017.

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

Charles Audet and Warren Hare. *Derivative-free and blackbox optimization.* Springer, 2017.

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.

Peter Auer, Ronald Ortner, and Csaba Szepesvári. Improved rates for the stochastic continuum-armed bandit problem. In *International Conference on Computational Learning Theory*, pages 454–468. Springer, 2007.

Peter Auer, Yifang Chen, Pratik Gajane, Chung-Wei Lee, Haipeng Luo, Ronald Ortner, and Chen-Yu Wei. Achieving optimal dynamic regret for non-stationary bandits without prior information. In *Conference on Learning Theory*, pages 159–163, 2019.

Dheeraj Baby and Yu-Xiang Wang. Online forecasting of total-variation-bounded sequences. *Advances in Neural Information Processing Systems*, 32:11071–11081, 2019.

Francis Bach and Vianney Perchet. Highly-smooth zero-th order online optimization. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 257–283. PMLR, 2016.

Krishnakumar Balasubramanian and Saeed Ghadimi. Zeroth-order Nonconvex Stochastic Optimization: Handling Constraints, High-Dimensionality and Saddle-Points. *Foundations of Computational Mathematics*, 2021.

Hamsa Bastani and Mohsen Bayati. Online decision making with high-dimensional covariates. *Operations Research*, 68(1):276–294, 2020.

Alexandre Belloni, Tengyuan Liang, Hariharan Narayanan, and Alexander Rakhlin. Escaping the local minima via simulated annealing: Optimization of approximately convex functions. In Peter Grunwald, Elad Hazan, and Satyen Kale, editors, *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 240–265. PMLR, 2015.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Stochastic multi-armed-bandit problem with non-stationary rewards. In *Advances in neural information processing systems*, pages 199–207, 2014.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations research*, 63(5):1227–1244, 2015.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Optimal exploration–exploitation in a multi-armed bandit problem with non-stationary rewards. *Stochastic Systems*, 9(4):319–337, 2019.

Lev Bogolubsky, Pavel Dvurechenskii, Alexander Gasnikov, Gleb Gusev, Yurii Nesterov, Andrei M Raigorodskii, Aleksey Tikhonov, and Maksim Zhukovskii. Learning supervised pagerank with gradient-based and gradient-free optimization methods. In *Advances in neural information processing systems*, pages 4914–4922, 2016.

Olivier Bousquet and Manfred K Warmuth. Tracking a small set of experts by mixing past posteriors. *Journal of Machine Learning Research*, 3(Nov):363–396, 2002.

Richard P Brent. *Algorithms for minimization without derivatives.* Courier Corporation, 2013.

Nataly Brukhim and Elad Hazan. Online boosting with bandit feedback. In *Algorithmic Learning Theory*, pages 397–420. PMLR, 2021.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412(19):1832–1852, 2011a.

Sébastien Bubeck, Gilles Stoltz, and Jia Yuan Yu. Lipschitz bandits without the lipschitz constant. In *International Conference on Algorithmic Learning Theory*, pages 144–158. Springer, 2011b.

Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1): 1–122, 2012.

Sébastien Bubeck, Yin Tat Lee, and Ronen Eldan. Kernel-based methods for bandit convex optimization. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 72–85. ACM, 2017.

HanQin Cai, Daniel Mckenzie, Wotao Yin, and Zhenliang Zhang. Zeroth-order regularized optimization (ZORO): Approximately sparse gradients and adaptive sampling, 2020.

Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

B Chen, RM Castro, and A Krause. Joint optimization and variable selection of high-dimensional gaussian processes. In *International Conference on Machine Learning (ICML)*, pages 1423–1430, 2012.

Wang Chi Cheung, David Simchi-Levi, and Ruihao Zhu. Learning to optimize under non-stationarity. *arXiv preprint arXiv:1810.03024*, 2018.

Yuejie Chi, Yue M Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing*, 67(20):5239–5269, 2019.

Chao-Kai Chiang, Chia-Jung Lee, and Chi-Jen Lu. Beating bandits in gradually evolving worlds. In *Conference on Learning Theory (COLT)*, pages 210–227, 2013.

Andrew R Conn, Katya Scheinberg, and Luis N Vicente. *Introduction to derivative-free optimization*. SIAM, 2009.

Varsha Dani, Sham M Kakade, and Thomas P Hayes. The price of bandit information for online optimization. In *Advances in Neural Information Processing Systems*, pages 345–352, 2008.

Yann N Dauphin, Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, Surya Ganguli, and Yoshua Bengio. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In *Advances in neural information processing systems*, pages 2933–2941, 2014.

Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. Approximation of functions of few variables in high dimensions. *Constructive Approximation*, 33(1):125–143, 2011.

Travis Dick, Andras Gyorgy, and Csaba Szepesvari. Online learning in markov decision processes with changing cost sequences. In *International Conference on Machine Learning (ICML)*, pages 512–520, 2014.

David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4): 1289–1306, 2006.

John C Duchi, Michael I Jordan, Martin J Wainwright, and Andre Wibisono. Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, 61(5):2788–2806, 2015.

Melih Elibol, Lihua Lei, and Michael I Jordan. Variance reduction with sparse gradients. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, pages 1058–1070, 2020.

Salar Fattahi, Cedric Josz, Reza Mohammadi, Javad Lavaei, and Somayeh Sojoudi. Absence of spurious local trajectories in time-varying optimization: A control-theoretic perspective. In *2020 IEEE Conference on Control Technology and Applications (CCTA)*, pages 140–147. IEEE, 2020.

Zhe Feng, Sebastien Lahaie, Jon Schneider, and Jinchao Ye. Reserve price optimization for first price auctions in display advertising. In *International Conference on Machine Learning (ICML)*, pages 3230–3239. PMLR, 2021.

Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 385–394. Society for Industrial and Applied Mathematics, 2005.

Alexander V Gasnikov, Ekaterina A Krymova, Anastasia A Lagunovskaya, Ilnura N Usmanova, and Fedor A Fedorenko. Stochastic online optimization. single-point and multi-point non-linear multi-armed bandits. convex and strongly-convex case. *Automation and remote control*, 78(2):224–234, 2017.

Rong Ge, Jason D Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 2981–2989, 2016.

Saeed Ghadimi and Guanghui Lan. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

Eduard Gorbunov, Pavel Dvurechensky, and Alexander Gasnikov. An accelerated method for derivative-free smooth stochastic convex optimization, 2018.

Peng Guan, Maxim Raginsky, and Rebecca M Willett. Online markov decision processes with kullback–leibler control cost. *IEEE Transactions on Automatic Control*, 59(6):1423–1438, 2014.

San Gultekin and John Paisley. Online forecasting matrix factorization. *IEEE Transactions on Signal Processing*, 67(5):1223–1236, 2018.

Tuomas Haarnoja, Haoran Tang, Pieter Abbeel, and Sergey Levine. Reinforcement learning with deep energy-based policies. In *International Conference on Machine Learning (ICML)*, pages 1352–1361, 2017.

Benjamin D Haeffele and René Vidal. Global optimality in tensor factorization, deep learning, and beyond. *arXiv preprint arXiv:1506.07540*, 2015.

Eric C Hall and Rebecca M Willett. Online convex optimization in dynamic environments. *IEEE Journal of Selected Topics in Signal Processing*, 4(9):647–662, 2015.

Cuize Han and Ming Yuan. Information based complexity for high dimensional sparse functions. *Journal of Complexity*, 57:101443, 2020.

Botao Hao, Tor Lattimore, Csaba Szepesvári, and Mengdi Wang. Online sparse reinforcement learning. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 316–324. PMLR, 2021.

Elad Hazan and Comandur Seshadhri. Efficient learning algorithms for changing environments. In *ICML*, pages 393–400, 2009.

Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

Elad Hazan, Karan Singh, and Cyril Zhang. Efficient regret minimization in non-convex games. In *International Conference on Machine Learning (ICML)*, pages 1433–1441. JMLR. org, 2017.

X. Hu, L. A. Prashanth, A. György, and C. Szepesvari. (Bandit) Convex Optimization with Biased Noisy Gradient Oracles. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 3420–3428, 2016.

Ali Jadbabaie, Alexander Rakhlin, Shahin Shahrampour, and Karthik Sridharan. Online optimization: Competing with dynamic comparators. In *Artificial Intelligence and Statistics*, pages 398–406, 2015.

Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M Kakade, and Michael I Jordan. How to escape saddle points efficiently. In *International Conference on Machine Learning (ICML)*, pages 1724–1732. PMLR, 2017.

Kenji Kawaguchi. Deep learning without poor local minima. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 586–594, 2016.

Kenji Kawaguchi, Jiaoyang Huang, and Leslie Pack Kaelbling. Effect of depth and width on local minima in deep learning. *Neural computation*, 31(7):1462–1498, 2019.

Gi-Soo Kim and Myunghee Cho Paik. Doubly-robust lasso bandit. *Advances in Neural Information Processing Systems*, 32:5877–5887, 2019.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 681–690, 2008.

Robert D Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems*, pages 697–704, 2005.

Jeffrey Larson, Matt Menickelly, and Stefan M Wild. Derivative-free optimization methods. *Acta Numerica*, 28:287–404, 2019.

Tor Lattimore. Improved regret for zeroth-order adversarial bandit convex optimisation. *arXiv:2006.00475*, 2020.

Joonseok Lee, Seungyeon Kim, Guy Lebanon, Yoram Singer, and Samy Bengio. Llorma: Local low-rank matrix approximation. *Journal of Machine Learning Research*, 17(15): 1–24, 2016.

Yuxi Li. Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*, 2017.

Sijia Liu, Jie Chen, Pin-Yu Chen, and Alfred O. Hero. Zeroth-order online alternating direction method of multipliers: Convergence analysis and applications, 2017.

Haipeng Luo and Robert E Schapire. Achieving all with no parameters: Adanormalhedge. In *Conference on Learning Theory*, pages 1286–1304, 2015.

Katta G Murty and Santosh N Kabadi. Some np-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39:117–129, 1987.

Arkadii Semenovich Nemirovsky and David Borisovich Yudin. Problem complexity and method efficiency in optimization. 1983.

Yurii Nesterov. *Lectures on convex optimization*, volume 137. Springer, 2018.

Yurii Nesterov and Boris T Polyak. Cubic regularization of newton method and its global performance. *Mathematical Programming*, 108(1):177–205, 2006.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017.

Gergely Neu, Andras Antos, András György, and Csaba Szepesvári. Online markov decision processes under bandit feedback. In *Advances in Neural Information Processing Systems*, pages 1804–1812, 2010.

Quynh Nguyen and Matthias Hein. The loss surface of deep and wide neural networks. In *International Conference on Machine Learning (ICML)*, pages 2603–2612. PMLR, 2017.

Erich Novak. *Deterministic and stochastic error bounds in numerical analysis*. Springer, 2006.

Erich Novak and Henryk Woźniakowski. *Tractability of Multivariate Problems: Standard information for functionals*, volume 2. European Mathematical Society, 2008.

Samet Oymak and Mahdi Soltanolkotabi. Overparameterized nonlinear learning: Gradient descent takes the shortest path? *arXiv preprint arXiv:1812.10004*, 2018.

Garvesh Raskutti, Martin J Wainwright, and Bin Yu. Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *The Journal of Machine Learning Research*, 13(1):389–427, 2012.

Abhishek Roy, Krishnakumar Balasubramanian, Saeed Ghadimi, and Prasant Mohapatra. Escaping saddle-point faster under interpolation-like conditions. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Abhishek Roy, Lingqing Shen, Krishnakumar Balasubramanian, and Saeed Ghadimi. Stochastic zeroth-order discretizations of langevin diffusions for bayesian inference. *Bernoulli (forthcoming)*, 2021.

Ankan Saha and Ambuj Tewari. Improved regret guarantees for online smooth convex optimization with bandit feedback. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 636–642, 2011.

Iman Shames, Daniel Selvaratnam, and Jonathan H Manton. Online optimization using zeroth order oracles. *IEEE Control Systems Letters*, 4(1):31–36, 2019.

Ohad Shamir. On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on Learning Theory (COLT)*, pages 3–24, 2013.

Ohad Shamir. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *Journal of Machine Learning Research*, 18(52):1–11, 2017.

Jasper Snoek, Hugo Larochelle, and Ryan Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

James C Spall. An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins apl technical digest*, 19(4):482–492, 1998.

James C Spall. *Introduction to stochastic search and optimization: estimation, simulation, and control*, volume 65. John Wiley & Sons, 2005.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction.* MIT press, 2018.

Nilesh Tripuraneni, Mitchell Stern, Chi Jin, Jeffrey Regier, and Michael I Jordan. Stochastic cubic regularization for fast nonconvex optimization. In *Advances in neural information processing systems*, pages 2899–2908, 2018.

Joel A Tropp. The expected norm of a sum of independent random matrices: An elementary approach. In *High dimensional probability VII*, pages 173–202. Springer, 2016.

Hemant Tyagi and Bernd Gärtner. Continuum armed bandit problem of few variables in high dimensions. In *International Workshop on Approximation and Online Algorithms*, pages 108–119. Springer, 2013.

Hemant Tyagi, Anastasios Kyrillidis, Bernd Gärtner, and Andreas Krause. Algorithms for learning sparse additive models with interactions in high dimensions. *Information and Inference: A Journal of the IMA*, 7(2):183–249, 2018.

Yining Wang, Simon Du, Sivaraman Balakrishnan, and Aarti Singh. Stochastic zeroth-order optimization in high dimensions. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1356–1365, 2018.

Yining Wang, Yi Chen, Ethan X Fang, Zhaoran Wang, and Runze Li. Nearly dimension-independent sparse linear bandit over small action spaces via best subset selection. *arXiv preprint arXiv:2009.02003*, 2020.

Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256, 1992.

Przemysław Wojtaszczyk. Complexity of approximation of functions of few variables in high dimensions. *Journal of Complexity*, 27(2):141–150, 2011.

Liangbei Xu and Mark A Davenport. Dynamic matrix recovery from incomplete observations under an exact low-rank constraint. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 3592–3600, 2016.

Liangbei Xu and Mark A Davenport. Simultaneous recovery of a series of low-rank matrices by locally weighted matrix smoothing. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5. IEEE, 2017.

Tianbao Yang, Lijun Zhang, Rong Jin, and Jinfeng Yi. Tracking slowly moving clairvoyant: optimal dynamic regret of online learning with true and noisy gradient. In *International Conference on Machine Learning (ICML)*, pages 449–457. JMLR. org, 2016.

Deming Yuan, Alexandre Proutiere, and Guodong Shi. Distributed online linear regression. *arXiv preprint arXiv:1902.04774*, 2019.

Zelda B Zabinsky. *Stochastic adaptive search for global optimization*, volume 72. Springer Science & Business Media, 2013.

Haixiang Zhang, Yingjie Bi, and Javad Lavaei. General low-rank matrix optimization: Geometric analysis and sharper bounds. *arXiv preprint arXiv:2104.10356*, 2021.

Richard Y Zhang. Sharp global guarantees for nonconvex low-rank matrix recovery in the overparameterized regime. *arXiv preprint arXiv:2104.10790*, 2021.

Zhihui Zhu, Qiuwei Li, Gongguo Tang, and Michael B Wakin. The global optimization geometry of low-rank matrix optimization. *IEEE Transactions on Information Theory*, 67(2):1308–1331, 2021.