Regret Minimization in Stochastic Non-Convex Learning via a Proximal-Gradient Approach

Nadav Hallak [1] **Panayotis Mertikopoulos** [2] **Volkan Cevher** [3]

Abstract

This paper develops a methodology for regret minimization with stochastic first-order oracle feedback in online, constrained, non-smooth, non-convex problems. In this setting, the minimization of external regret is beyond reach for first-order methods, and there are no gradient-based algorithmic frameworks capable of providing a solution. On that account, we focus on a *local regret* measure defined via a proximal-gradient mapping, that also encompasses the original notion proposed by Hazan et al. (2017). To achieve no local regret in this setting, we develop a proximal-gradient method based on stochastic first-order feedback, and a simpler method for when access to a perfect first-order oracle is possible. Both methods are order-optimal (in the min-max sense), and we also establish a bound on the number of proximal-gradient queries these methods require. As an important application of our results, we also obtain a link between online and offline non-convex stochastic optimization manifested as a new proximal-gradient scheme with complexity guarantees matching those obtained via variance reduction techniques.

1. Introduction

First-order methods have proven to be extremely flexible and efficient in online convex optimization: They enjoy tight performance guarantees in a wide range of relevant settings such as convex, strongly convex and/or composite problems, and they can adapt to different measures of regret under different oracle feedback assumptions, e.g., perfect/stochastic gradients or bandit feedback. For example, see Abernethy et al. (2008), Hazan (2016), Hazan et al. (2007) and Xiao (2010) for applications to different convex settings, Besbes et al. (2015), Cesa-Bianchi et al. (2012), and Hazan & Seshadhri (2009) for variant regret measures, and Abernethy et al. (2008), Agarwal et al. (2010), and Bubeck & Eldan (2016, 2017) for a range of feedback assumptions.

On the other hand, many contemporary problems, especially in machine learning, involve highly multi-modal *non-convex* functions. In this case, the results obtained in the above framework do not – in fact, *cannot* – apply, and new analytical tools and algorithms are needed. Nevertheless, and somewhat surprisingly at that, online non-convex optimization problems are not as well explored, and significantly less is known about the performance of first-order methods in this context.

The key difficulties encountered in the online non-convex setting are twofold: First, the standard regret comparator of a "best action in hindsight" (fixed or otherwise) is too ambitious because, in general, even *offline* non-convex optimization problems are intractable. Second, compared to problems with a convex structure, non-convex problems have no local-to-global guarantees, so the adversary has a near-insurmountable advantage (in analogy to non-convexified/non-randomized optimizers facing an adversarial bandit). Our paper seeks to address these challenges in a unified way in the setting of first-order methods.

Related work. One approach to treat online non-convex optimization is to regard the problem as an adversarial multi-armed bandit (MAB) with a *continuum* of arms. This approach was pioneered by Bubeck et al. (2011), Kleinberg (2004) and Kleinberg et al. (2008), who proposed a range of hierarchical search methods, with and without a doubling trick, that guarantee no regret in problems with a geometry that is amenable to local

[1]Faculty of Industrial Engineering and Management, The Technion, Haifa, Israel [2]Univ. Grenoble Alpes, CNRS, Inria, LIG, Grenoble, France, & Criteo AI Lab [3]École Polytechnique Fédérale de Lausanne (EPFL). Correspondence to: Nadav Hallak <ndvhllk@technion.ac.il>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

search such as the hypercube. Krichene et al. (2015) and, more recently, Perkins et al. (2017) and Héliou et al. (2020, 2021), took an approach based on a suitable adaptation of the Hedge/EXP3 algorithms to bandits with a continuum of arms and established the method's no-regret properties under relatively mild regularity conditions. However, in full generality, sampling from continuous Gibbs distributions can be quite challenging, so it is not a-priori clear how to implement these methods without a sampling oracle in place.

Another approach, manifesting in the recent works of Agarwal et al. (2019) and Suggala & Netrapalli (2019), is the classical Follow-the-Perturbed-Leader algorithm with access to an *offline non-convex optimization oracle*, which was shown to enjoy a polynomial regret bound. Simplifying assumptions that render a non-convex problem tractable, were also considered in the literature in more particular cases such as the principal component analysis model; see Garber (2019) and references therein for additional examples.

Complementing this literature in an orthogonal direction, Hazan et al. (2017) took a more direct, "pure-strategy", approach based on a "smoothed" inner-loop / outer-loop version of projected gradient descent. In this general framework, a straightforward extension of Cover's impossibility result shows that the minimization of standard regret measures is unattainable. On account of this, Hazan et al. (2017) considered instead a *local regret* measure based on a sliding evaluation window and a suitable measure of stationarity (as opposed to *optimality*). When faced with a stream of Lipschitz smooth functions, the algorithm of Hazan et al. (2017) enjoys a local regret bound that scales with the horizon T of the process and the size w of the sliding window as $O(T/w^2)$, with projection calls complexity $O(Tw)$; as a result, sublinear (local) regret *is* possible as long as $w = \omega(1)$. Importantly, Hazan et al. (2017) also showed that the local regret bound is unimprovable from a min-max perspective, so the proposed algorithm is optimal in this regard. For *unconstrained* problems with stochastic gradient observations, Hazan et al. (2017) further showed that a suitable variant of their method achieves similar guarantees in expectation.

Our contributions. Our goals in this paper are twofold: First, we seek to treat online problems that are potentially *non-smooth*, covering e.g., the case of L^1-regularization. Second, in line with the above, we also wish to account for problems with *stochastic* oracle feedback, simultaneously with constraints and regularization, thus including problems subjected to both random and seasonal fluctuations. To achieve the desiderata, we consider a general *composite* non-convex online framework in which each loss function encountered consists of a smooth and non-smooth part; this study is the first to provide methods with theoretical guarantees to address this scenario. Concisely, our main contributions are

- Assuming access to only a stochastic first-order oracle, we introduce a smoothed *prox-grad* method to handle *stochastic, constrained, non-smooth, non-convex* online optimization problems with tight regret guarantees of $O(T/w^2)$ in expectation and stochastic first-order oracle calls bound of $O(Tw^2)$. This represents a significant step forward relative to the literature, mainly, compared to the online stochastic method proposed by (Hazan et al., 2017), as the latter can only address the basic *smooth unconstrained* case.
- Relaxing the feedback assumptions to a perfect first-order oracle, we also present a simpler method that can simultaneously tackle online non-convex optimization problems with both constraints and regularization, and obtain tight regret guarantees $O(T/w^2)$ with prox-grad calls complexity $O(Tw)$ in the process.
- As a by-product, but of an independent interest and a contribution of its own, we derive from our methods new schemes for stochastic offline optimization under the online framework assumptions with the best known guarantees, achievable only via variance reduction techniques – see Arjevani et al. (2019) and references therein.

2. Problem setup

2.1. Statement of the problem and blanket assumptions

We consider the class of online non-convex, nonsmooth, composite problems over a finite and discrete time horizon $T \geq 1$ of the form

$$\min\{\ell_t(\mathbf{x}) = f_t(\mathbf{x}) + g(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\}, \qquad t \in [T], \quad \text{(P)}$$

where

1. $g : \mathbb{R}^n \to \mathbb{R}_+ \cup \{\infty\}$ is a proper, convex, lower semicontinuous (l.s.c) function.
2. For any $t \in [T]$, the function $f_t \colon \mathbb{R}^n \to \mathbb{R}$ is L-smooth ($L > 0$) over $\operatorname{dom} g$, i.e.,

$$\|\nabla f_t(\mathbf{x}) - \nabla f_t(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\| \quad \forall \mathbf{x}, \mathbf{y} \in \operatorname{dom} g.$$

3. There exists $M > 0$ such that for any $\mathbf{x} \in \operatorname{dom} g$ and $t \in [T]$, it holds that $|f_t(\mathbf{x})| \leq M$.

Our blanket assumptions are fundamental in the study of online learning, even when the objective function is convex (see e.g., Hazan, 2016). We also note that f_t is assumed to be L-smooth and bounded only over the domain of g. Thus, if $\operatorname{dom} g$ is bounded and f_t is continuously differentiable over $\operatorname{dom} g$, then the model's assumptions trivially hold true.

2.2. Motivating applications

Examples of (P) are ubiquitous in theoretical computer science, operations research, and many other fields where online decision-making is the norm. For concreteness, we shortly describe next a few conceptual examples; further details are provided in the supplement.

- **Non-convex games:** A multi-player non-convex game can be modeled by simultaneously optimizing several copies of (P), where all share the same function f_t, and (un-shared) penalty functions may be utilized to induce stability (e.g., risk aversion) in the choices of each of the players independently; see e.g., Hazan et al. (2017) and Agarwal et al. (2019).

  A particularly interesting instance of a two-player non-convex game in which the objective function is accessible through a stochastic oracle, is the *generative adversarial network (GAN)* model; GANs were already considered via an online framework by Grnarova et al. (2017) and Agarwal et al. (2019) for example.

- **Online path planning with splittable traffic demands:** The online traffic assignment problem is a hallmark path planning problem that requires the full capacity of our model, and whose formulation further applies to learning perfect matchings, multitask bandits, spanning tree exploration, etc. Referring to Bertsekas & Gallager (1992) and Shakkottai & Srikant (2008) for an introduction to the topic, the key objective in traffic assignment problems is the optimal allocation of traffic over a given network with variable traffic inflows. The feasibe set here is compact, the cost functions are smooth yet non-convex, and a sparsity-inducing L^1 term is typically included to "robustify" solutions by minimizing the overall number of paths employed; we provide a fully detailed formulation in the supplement.

- **Stochastic (offline) optimization:** Stochastic optimization, which follows naturally from online optimization by restricting the adversarial behavior accordingly, plays a prominent role in modern applications, such as the training of neural networks.

2.3. Local regret minimization

In the online non-convex framework of (P), there are two key issues with the standard definition of the regret as $\operatorname{Reg}(T) = \max_{\mathbf{x}\in\operatorname{dom} g} \sum_{t=1}^{T} [\ell_t(\mathbf{x}_t) - \ell_t(\mathbf{x})]$: First, the global minimization of a non-convex objective is intractable in general, so using the best fixed action in hindsight as a comparator is too ambitious. Second, as we explain below, even if one uses a proxy for stationarity in lieu of a global minimizer, an informed adversary can still impose $\operatorname{Reg}(T) = \Omega(T)$, so the notion of regret minimization must also be re-examined in this setting.

We address both of these issues by a combined approach, leveraging optimality criteria and measures from (offline) non-convex analysis, together with smoothing of the online part of the objective function. This generalizes the proposed framework of Hazan et al. (2017) from a gradient projection scheme into a complete methodology that can be applied "off-the-shelf" to possibly non-smooth / non-convex problems.

Due to the of impossibility of finding a global optimizer for a non-convex problem in polynomial time (see also Kleinberg, 2004, on a similar result for multi-armed bandit algorithms), optimization schemes are analyzed and tested with different optimality measures and figures of merit. Typically, optimality measures are *scheme-dependent*, see, e.g., Beck (2017, Section 10.3 on the proximal gradient, and Section 13.2 on the conditional gradient). Complementing the latter, *smoothing* (i.e., averaging) is a common practice when dealing with uncertainty and fluctuations, e.g., in the fields of statistics, stochastic optimization, finance, and others. Indeed, intuitively, in the face of a time-varying non-convex objective function with possibly no behavioral pattern, a rational approach for the decision maker is to stabilize the decision-making protocol, so that, on average, the best possible decisions will be made.

In light of the above, and in the spirit of Hazan et al. (2017), define for all $w \in [T]$ the sliding average of the smooth part of the objective function, namely

$$S_{t,w}(\mathbf{x}) = \frac{1}{w} \sum_{i=t-w+1}^{t} f_i(\mathbf{x}),$$

with the convention $f_t \equiv 0$ for $t \leq 0$. To stabilize the optimization protocol's behavior, the decision maker determines her actions based on a smoothed objective function $S_{t,w}(x)+g(x)$ instead of the real objective $f_t(x)+g(x)$. This way, the adversary's ability to manipulate the decision maker's decisions by increasing the variability of her choices is reduced. This was formally established by Hazan et al. (2017) for the projected-gradient framework.

The characteristics of optimality measures in non-convex optimization can be summarized as *a*) dependency in the method at hand; *b*) positive-definiteness; and *c*) zero value only at points that satisfy the first-order optimality conditions (stationarity) This reflects the fact that a reasonable and tractable objective for the optimizer in non-convex learning is to find a point having zero optimality measure. Now, denote by $\mathcal{S}^{\text{alg}}(x; S_{t,w}, g)$ an optimality measure associated with an underlying method *alg*, at the point x, for $\min_z S_{t,w}(z) + g(z)$. Then, the notion of *local regret* of a policy x_t up to time T with window length w is defined via $\mathcal{S}^{\text{alg}}$ as

$$\operatorname{Reg}_w(T) = \sum_{t=1}^{T} \mathcal{S}^{\text{alg}}(x_t; S_{t,w}, g). \tag{1}$$

Hence, for the decision maker to achieve no-regret, $\frac{1}{T}\operatorname{Reg}_w(T)$ must go to zero as the time horizon T goes to infinity; this is in line with the original notion of regret (Hazan, 2016), and with the definition of local regret as pioneered by Hazan et al. (2017). Indeed, in the offline case, the requirement that $\frac{1}{T}\operatorname{Reg}_w(T) \to 0$, is translated to convergence of the optimality measure to zero, which is the desired goal.

This paper focuses on the proximal-gradient framework, and thus we use the corresponding *prox residual* (also called *the gradient mapping*) optimality measure, see Beck (2017, Section 10.3). To do so, we begin by defining the *proximal mapping* of g along the search direction $\mathbf{d} \in \mathbb{R}^n$ with step-size $\eta > 0$ as

$$\begin{aligned} T^g_\eta(\mathbf{x}; \mathbf{d}) &\equiv \operatorname{prox}_{\eta g}(\mathbf{x} - \eta \mathbf{d}) \\ &= \arg\min_{\mathbf{z}\in\mathbb{R}^n} \{\eta g(\mathbf{z}) + \tfrac{1}{2}\|\mathbf{x} - \eta\mathbf{d} - \mathbf{z}\|^2\}, \end{aligned} \tag{2}$$

where $\|\cdot\|$ stands for the Euclidean norm, and the corresponding *prox residual* as

$$\mathcal{P}^g_\eta(\mathbf{x}; \mathbf{d}) = \frac{1}{\eta}\left(\mathbf{x} - T^g_\eta(\mathbf{x}; \mathbf{d})\right). \tag{3}$$

Remark 2.1. *We note that the purpose behind the use of a general vector* $\mathbf{d}$ *in Eq. (2) and Eq. (3) is to be able to accommodate for* stochastic *gradients later on in Section 4.*

As an illustration, let us set $\mathbf{d} = \nabla f(\mathbf{x})$ and examine Eq. (2) and Eq. (3) in the smooth unconstrained and constrained scenarios. If $g \equiv 0$, then Eq. (2) is the gradient descent operator and Eq. (3) reduces to $\mathcal{P}^g_\eta(\mathbf{x}; \nabla f(\mathbf{x})) = \nabla f(\mathbf{x})$. Likewise, if $g \equiv \delta_{\mathcal{K}}$ for some closed convex subset $\mathcal{K}$ of $\mathbb{R}^n$, we get the projected gradient descent in Eq. (2) and its corresponding projection residual $\mathcal{P}^g_\eta(\mathbf{x}; \nabla f(\mathbf{x})) = \eta^{-1}(\mathbf{x} - \operatorname{proj}_{\mathcal{K}}(\mathbf{x} - \eta\nabla f(\mathbf{x})))$.

The norm of the prox residual $\operatorname{Res}^g_\eta(\mathbf{x}) \equiv \|\mathcal{P}^g_\eta(\mathbf{x}; \nabla f(\mathbf{x}))\|^2 \geq 0$ is the standard optimality measure for the analysis of proximal gradient-based schemes for non-convex optimization: $\|\mathcal{P}^g_\eta(\mathbf{x}; \nabla f(\mathbf{x}))\| = 0$ if and only if $\mathbf{x}$ is a stationary point of (P). This makes the following definition of regret the most natural choice to quantify the regret of an online policy $\mathbf{x}_t$ at time T,

$$\operatorname{Reg}(T) \equiv \sum_{t=1}^{T} \operatorname{Res}^g_\eta(\mathbf{x}_t) = \sum_{t=1}^{T} \left\|\mathcal{P}^g_\eta(\mathbf{x}_t; \nabla f_t(\mathbf{x}_t)\right\|^2. \tag{4}$$

However, as was shown by Hazan et al. (2017), it is not difficult for the adversary to impose linear regret by providing a sequence of "spiked" non-convex loss functions with large $\|\nabla f_t(\mathbf{x}_t)\|$ and small gradient away from each $\mathbf{x}_t$ (for completeness, we provide a simple example in the supplement). Perhaps more intuitively, one may consider a dynamical system with a time varying function that is only accessible via a stochastic oracle (e.g. GAN as a two-players game), in which case, attaining stationarity through the classical use of Eq. (4) seems impossible.

Because of this, as we informally stated before, it is more reasonable to consider a *smoothed*, *local* version of the regret that averages the sequence of loss functions encountered over a sliding window of w consecutive time periods. Building on the notion of regret proposed by (Hazan et al., 2017), the *local regret* of a policy $\mathbf{x}_t$ up to time T with window leghth w is then defined as

$$\operatorname{Reg}_w(T) = \sum_{t=1}^{T} \left\|\mathcal{P}^g_\eta\left(\mathbf{x}_t; \nabla S_{t,w}(\mathbf{x}_t)\right)\right\|^2. \tag{5}$$

In the above, the sliding window w can be seen as an "effective time unit": essentially, instead of working with the stream of (potentially volatile) loss functions f_t directly, we work with the average loss over a window of length w. In practice, the sliding window w acts as a "stabilizer" controlling the effects of the noise and variability of the function on the decision making of the optimization protocol; this will become apparent in the sequel.

In the non-composite case, when g is the indicator of a closed convex set, the local regret measure Eq. (5) is quantified by the minimax bound of Hazan et al.

(2017) who showed that an informed adversary can impose $\text{Reg}_w(T) = \Omega(T/w^2)$. This bound becomes sublinear in T if $w = \omega(1)$, so this definition provides the required flexibility for a tractable measure of regret.

To further substantiate the motivation for our smoothing approach, we provide four prototypical scenarios in which Eq. (5) generalizes standard measures in simpler models:

- In the offline case $f_t \equiv f$, we immediately recover the classical measure of Eq. (4).
- If $g \equiv 0$, we readily obtain $\text{Reg}_w(T) = (1/w^2)\sum_{t=1}^{T} \| \sum_{i=t-w+1}^{t} \nabla f_i(\mathbf{x}_t)\|^2$, i.e., the original definition of Hazan et al. (2017) for unconstrained online non-convex problems.
- If additionally $f_t = F(\cdot, \omega_t)$ where F is a stochastic objective with $\mathbb{E}(F) = f$, and ω_t is an i.i.d sequence of random seeds, then choosing an output iteration t uniformly leads to $\mathbb{E}(\text{Reg}_w(T))/T \geq \mathbb{E}_t \|\nabla f(\mathbf{x}_t)\|^2$, meaning that local regret minimization leads to stationarity in expectation in unconstrained stochastic models; we will return to this example in Section 3.
- More generally, as discussed in detail in Section 4.2, if each f_t is drawn from an underlying stationary distribution with expectation f, and a stopping time t_* is selected uniformly at random from $[T]$, we will have $\mathbb{E}\left[\left\| \mathcal{P}_\eta^g(\mathbf{x}_{t_*}; \nabla f(\mathbf{x}_{t_*}))\right\|^2 \right] \leq \mathbb{E}(\text{Reg}_w(T))/T$, i.e., local regret minimization implies average stationarity in composite (offline) stochastic problems.

We close this section by introducing a measure of variation of the loss functions encountered by the optimizer, and which will be particularly useful in the sequel:

Definition 2.1 (Sliding window variation)**.** *The sliding window variation of a sequence of loss functions f_t is*

$$V_w[T] = \sup_{\mathbf{x}\in \text{dom} g} \left\{ \sum_{i=1}^{T} \|\nabla f_i(\mathbf{x}) - \nabla f_{i-w}(\mathbf{x})\|^2 \right\}. \quad (6)$$

An immediate observation is that if the gradients of the functions are bounded (e.g., if f_t is Lipschitz continuous), we automatically have $V_w[T] = O(T)$; as such, any regret guarantee stated in terms of $V_w[T]$ automatically translates to $O(T)$ in this context.

The main reason that we introduce this variation measure instead of working with a more uniform hypothesis, such as the standard Lipschitz continuity of the objective function, is to account for cases where this quantity is naturally small. For example, in the routing problem mentioned in Section 2.2 and detailed in the supplemental, $V_w[T]$ corresponds to the variability of the encountered traffic demands at a time-scale of w. As such, if the sliding window w is attuned to the seasonal variability of the process (e.g., an hour, a day or a week, depending on granularity), $V_w[T]$ could be considerably smaller than T, so the obtained regret bounds would be considerably sharper as a result.

We should also note that, when $w = 1$, $V_w[T]$ boils down to the "gradual variation" measure of Chiang et al. (2012) – and, indirectly, to the variation budget of Besbes et al. (2015). The above suggests an interesting interplay between our analysis and regret minimization relative to a dynamic comparator; this is also part of the reason that we state our results in terms of $V_w[T]$ in the sequel.

3. The time-smoothed online prox-grad method

Assuming perfect first-order oracle, we introduce the *Time-Smoothed Online Prox-Grad Descent* method, cf. Algorithm 1, which generalizes the *time-smoothed online gradient descent* method of (Hazan et al., 2017).

Algorithm 1: Time-smoothed online prox-grad descent

Input. $\mathbf{x}_1 \in \mathbb{R}^n$, $\eta \in (0, 1/L)$, $w \in [T]$, $\delta > 0$.

General step. For any $t = 1, \ldots, T$ do:

1. $f_t : \mathbb{R}^n \to \mathbb{R}$ is determined;
2. Set $\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t$;
3. While $\left\|\mathcal{P}_\eta^g(\mathbf{x}_{t+1}; \nabla S_{t,w}(\mathbf{x}_{t+1}))\right\| > \delta/w$ do:
   (a) Update $\mathbf{x}_{t+1} \leftarrow \arg\min_{\mathbf{z}\in\mathbb{R}^n} g(\mathbf{z}) + \langle \nabla S_{t,w}(\mathbf{x}_{t+1}), \mathbf{z} - \mathbf{x}_{t+1}\rangle + \frac{1}{2\eta}\|\mathbf{z} - \mathbf{x}_{t+1}\|^2$;

As we show below, Algorithm 1 achieves an optimal regret bound of $O\left(\frac{T}{w^2}\right)$ when $V_w[T]$ is bounded by $O(T)$, and executes $O(Tw)$ prox-grad operations.

Theorem 3.1 (Local regret minimization)**.** *Algorithm 1 enjoys the local regret bound*

$$\text{Reg}_w(T) \leq \frac{2}{w^2}\left(T\delta^2 + V_w[T]\right).$$

Theorem 3.2 (Oracle queries)**.** *Let τ_t be the number of prox-grad operations at time $t \in [T]$. The total number of oracle queries $\tau = \sum_{t=1}^{T} \tau_t$ made by Algorithm 1 is bounded as*

$$\tau \leq \frac{2Tw(g(\mathbf{x}_1) + 3M)}{(2 - \eta L)\,\eta\delta^2} = O(Tw).$$

We conclude this section by examining the theoretical guarantees of Algorithm 1 when f_t is an unbiased stochastic approximation of f, so that, implicitly, ∇f_t is generated via an unbiased SFO. It should be noted that the SFO must satisfy that $V_w[T]$ is $O(T)$, which effectively bounds the variability of the stochastic gradient; this assumption is different than the standard variance bound in stochastic gradient analysis (cf. Definition 4.1).

Corollary 3.1. *Suppose that* $g \equiv 0$, $\mathbb{E}(\nabla f_t(\mathbf{x}) - \nabla f(\mathbf{x})) = 0$ *for any* $\mathbf{x} \in \mathbb{R}^n$*, and that* $V_w[T] \leq cT$ *for some* $c > 0$*. Let* $\varepsilon > 0$*, and* $t_* \in [T]$ *be chosen uniformly from* $\{w, w+1, \ldots, T\}$*. If* $T = 2w$ *and* $w = \left\lceil 2\sqrt{(\delta^2 + c)/\varepsilon} \right\rceil$*. Then Algorithm 1 achieves* $\mathbb{E}\left(\|\nabla f(\mathbf{x}_{t_*})\|^2\right) \leq \varepsilon$ *with at most* $O(\varepsilon^{-1})$ *prox-grad operations and* $O(\varepsilon^{-3/2})$ *SFO calls.*

Note that the complexities reported in Corollary 3.1 match those obtained for the state-of-the-art *Prox-SpiderBoost* method proposed by Wang et al. (2019), but under a different procedure using more stringent assumptions (boundedness of f and that $V_w[T]$ is $O(T)$). We stress that the Prox-SpiderBoost method is only applicable to stochastic problems, and as such, it has no online guarantees, unlike Algorithm 1.

The proofs of Theorems 3.1 and 3.2, and of Corollary 3.1, are deferred to the supplemental.

4. Stochastic time-smoothed online prox-grad method

4.1. Method and Analysis

Moving forward from the deterministic guarantees of Algorithm 1, we proceed to consider a more flexible framework that only posits access to a *stochastic first-order oracle* (SFO). Specifically, following Nemirovski et al. (2009), we assume that it is possible to generate an i.i.d. sequence of random seeds $\zeta_1, \zeta_2, \ldots$, that are concurrently used as input to an SFO as follows:

Definition 4.1 (Stochastic first-order oracle)**.** *A stochastic first-order oracle (SFO) is a function* $\mathcal{S}_\sigma$ *such that, given a point* $\mathbf{x} \in \mathcal{R}^n$*, a random seed* ζ*, and a smooth function* $h \colon \mathbb{R}^n \to \mathbb{R}$ *satisfies:*

1. $\mathcal{S}_\sigma(\mathbf{x}; \zeta, h)$ *is unbiased relative to* $\nabla h(\mathbf{x})$*:* $\mathbb{E}\left(\mathcal{S}_\sigma(\mathbf{x}; \zeta, h) - \nabla h(\mathbf{x})\right) = 0$*;*
2. $\mathcal{S}_\sigma(\mathbf{x}; \zeta, h)$ *has variance bounded by* $\sigma^2 > 0$*:* $\mathbb{E}\left(\|\mathcal{S}_\sigma(\mathbf{x}; \zeta, h) - \nabla h(\mathbf{x})\|^2\right) \leq \sigma^2$.

With all this in hand, the heuristics of the proposed stochastic prox-grad method are as follows: (i) f_t is determined; (ii) successive SFO queries generate a noisy descent process in an inner loop until a δ/w-stationary point is reached. The pseudocode of the algorithm is shown in Algorithm 2:

Algorithm 2: Time-smoothed online stochastic prox-grad method

Input. $\mathbf{x}_1 \in \mathbb{R}^n$, $\eta \in (0, 1/L)$, $w \in [T]$, $\delta > 0$.
Initialization. $\tilde{\nabla} S_{i,w}(\mathbf{x}_1) = \mathbf{0}$ for all $i \leq 0$.
General step. For any $t = 1, 2, \ldots, T$ do:

1. Function is updated to $f_t : \mathbb{R}^n \to \mathbb{R}$;
2. Sample $\tilde{\nabla} f_t(\mathbf{x}_t) \leftarrow \mathcal{S}_{\sigma/w}(\mathbf{x}_t; \zeta, f_t)$;
3. Set $\tilde{\nabla} S_{t,w}(\mathbf{x}_t) = \tilde{\nabla} S_{t-1,w}(\mathbf{x}_t) + \frac{1}{w}(\tilde{\nabla} f_t(\mathbf{x}_t) - \tilde{\nabla} f_{t-w}(\mathbf{x}_t))$;
4. Set $\mathbf{y}_t^1 = \mathbf{x}_t$, $G_t^1 = \tilde{\nabla} S_{t,w}(\mathbf{x}_t)$, $k = 1$;
5. While $\left\|\mathcal{P}_\eta^g\left(\mathbf{y}_t^k; G_t^k\right)\right\| > \delta/w$ do:
   (a) Update $\mathbf{y}_t^{k+1} = \arg\min_{\mathbf{z} \in \mathbb{R}^n} g(\mathbf{z}) + \langle G_t^k, \mathbf{z} - \mathbf{y}_t^k \rangle + \frac{1}{2\eta}\|\mathbf{z} - \mathbf{y}_t^k\|^2$;
   (b) Sample $\tilde{\nabla} f_i(\mathbf{y}_t^{k+1}) \leftarrow \mathcal{S}_{\sigma/w}(\mathbf{y}_t^{k+1}; \zeta, f_i)$ for any $i = t-w+1, \ldots, t$;
   (c) Set $G_t^{k+1} = \frac{1}{w}\sum_{i=t-w+1}^{t} \tilde{\nabla} f_i(\mathbf{y}_t^{k+1})$;
   (d) Set $k \leftarrow k + 1$;
6. Set $\mathbf{x}_{t+1} = \mathbf{y}_t^k$ and $\tilde{\nabla} S_t(\mathbf{x}_{t+1}) = G_t^k$.

The process of Algorithm 2 might be better understood by comparing it to offline stochastic variance reduction methods (SVR); see e.g., (Fang et al., 2018, Metel & Takeda, 2019, Wang et al., 2019, Yurtsever et al., 2019), and references therein. For these methods, which usually implement a non-diminishing step-size policy in the non-convex setting, a batch-size variance relation is required in order to achieve the methods' guarantees.

Algorithm 2 takes a different approach in this context by, instead of stating this connection in the analysis, it explicitly links the batch-size (i.e., w mimics the role of the batch-size) to the variance of the SFO in the scheme itself. The affinity of Algorithm 2 to SVR methods is further expressed when considering its guarantees in the offline scenario of $f_t \equiv f$. Then, Algorithm 2 achieves the best known SFO complexity as that obtained by SVR methods; see our Section 4.2 for additional details. We stress however that Algorithm 2 is stated as a general-form schematics, without any assumption on the sampling procedure itself, or on the variance reduction mechanism, and their particular implementation.

Before stating Algorithm 2's guarantees, let us first define the algorithm's natural filtration: For all $t \geq 1$, the filtration $\mathcal{F}_t$ includes all gradient feedback up to, but not including, the execution of step 2 at stage t. In particular, it includes $f_t, \mathbf{x}_t$ and $\tilde{\nabla} S_{t-1}(\mathbf{x}_t)$, but it does not include $\tilde{\nabla} f_t(\mathbf{x}_t)$.

With all this in hand, we now state our main results, accompanied by a summarized version of the proofs; detailed proofs are deferred to the supplementary. Denote by τ_t the number of times the condition in step 5 at t-th iteration is checked, that is the number of prox-grad operations at the t-th iteration, and let $\tau = \sum_{t\in[T]} \tau_t$. We begin by establishing that Algorithm 2 almost surely executes a finite number of prox-grad operations provided that δ is not too small.

Theorem 4.1 (Oracle queries)**.** *Let $t \in [T]$ and let the filtration $\mathcal{F}_t$ be given. Suppose that the inputs δ and η satisfy that*

$$\delta^2 > 2\sigma^2 \left(\eta \left(1 - \eta L\right)\right)^{-1}. \tag{7}$$

Then τ_t and τ are almost surely finite, and $\forall K \geq 1$

$$\mathbb{P}(\tau_t > K) \leq \frac{(h_t^1 + M)w^2}{2\left(\eta\left(1-\eta L\right)\delta^2 - 2\sigma^2\right)K} = O(1/K).$$

Proof sketch. Recall that $\mathbf{y}_t^1 = \mathbf{x}_t, \mathbf{y}_t^{\tau_t} = \mathbf{x}_{t+1}$, and for any $k \in [\tau_t - 1]$ we have that $\mathbf{y}_t^{k+1} = \arg\min_{\mathbf{z}\in\mathbb{R}^n} g(\mathbf{z}) + \langle G_t^k, \mathbf{z} - \mathbf{y}_t^k\rangle + \frac{1}{2\eta}\|\mathbf{z} - \mathbf{y}_t^k\|^2$. Denote $h_t^k := S_t(\mathbf{y}_t^k) + g(\mathbf{y}_t^k)$. By combining the descent lemma (cf. Lemma C.1 entailed in supplementary) and the stopping criteria of the inner loop, we have that for any $k \in [\tau_t - 1]$ (assuming that $\mathcal{F}_t$ is given), $h_t^k - h_t^{k+1} \geq \langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \mathbf{y}_t^{k+1} - \mathbf{y}_t^k\rangle + \frac{1}{2}\left(\eta - \eta^2 L\right)\delta^2 w^{-2}$. Applying expectation to the latter, using the law of total expectation, the technical result of Lemma E.2 in the supplementary, and relation (7), we obtain that for any $k \in [\tau_t - 1]$ it holds that $\mathbb{E}\left(h_t^k - h_t^{k+1}\right) \geq 2w^{-2}\left(\eta\left(1-\eta L\right)\delta^2 - 2\sigma^2\right) > 0$. Set $\alpha := 2\left(\eta\left(1-\eta L\right)\delta^2 - 2\sigma^2\right)/w^2 > 0$. From the former, for any $K \geq 1$ we have that

$$\begin{aligned} h_t^1 + M &\geq \mathbb{E}\left(h_t^1 - h_t^{K+1}\right) = \mathbb{E}\left(\sum_{k=1}^{K}(h_t^k - h_t^{k+1})\right) \\ &= \sum_{k=1}^{K}\mathbb{E}\left(h_t^k - h_t^{k+1} | \tau_t \geq k+1\right)\mathbb{P}(\tau_t \geq k+1) \\ &\geq \alpha\sum_{k=1}^{K}\mathbb{P}(\tau_t > K) = \alpha K \mathbb{P}(\tau_t > K). \end{aligned}$$

Consequently, we must have that τ_t is almost surely finite, which in turn implies that τ must be almost surely finite as it is the finite sum of almost surely finite variables. □

Next we provide a tight bound on the expected local regret in terms of $V_w[T]$; recall that under the standard assumptions of bounded feasible domain or Lipschitz continuity of f_t, $V_w[T]$ is bounded by $O(T)$, in which case we have that $\mathbb{E}\left[\mathrm{Reg}_w(T)\right]$ achieves the optimal local regret bound of $O\left(\frac{T}{w^2}\right)$.

Theorem 4.2 (Local regret minimization)**.** *Algorithm 2 enjoys the average local regret bound*

$$\mathbb{E}\left[\mathrm{Reg}_w(T)\right] \leq 2\left(\frac{T}{w^2}\right)\left(\delta^2 + 7\sigma^2\right) + \frac{6}{w^2}V_w[T].$$

Proof sketch. By simple algebra,

$$\begin{aligned} &\left\|\mathbf{x}_t - T_\eta^g(\mathbf{x}_t; \nabla S_t(\mathbf{x}_t))\right\|^2 \leq 2\left\|\mathbf{x}_t - T_\eta^g(\mathbf{x}_t; \tilde{\nabla} S_t(\mathbf{x}_t))\right\|^2 \\ &+ 2\left\|T_\eta^g(\mathbf{x}_t; \tilde{\nabla} S_t(\mathbf{x}_t)) - T_\eta^g(\mathbf{x}_t; \nabla S_t(\mathbf{x}_t))\right\|^2. \end{aligned} \tag{8}$$

Using the nonexpansivity of the prox operator (cf. Theorem 6.42 in (Beck, 2017)) we have that

$$\begin{aligned} &\left\|T_\eta^g(\mathbf{x}_t; \tilde{\nabla} S_t(\mathbf{x}_t)) - T_\eta^g(\mathbf{x}_t; \nabla S_t(\mathbf{x}_t))\right\|^2 \\ \leq &\eta^2\left\|\tilde{\nabla} S_t(\mathbf{x}_t) - \nabla S_t(\mathbf{x}_t)\right\|^2. \end{aligned}$$

Subsequently, we obtain the relation (cf. Lemma E.1)

$$\begin{aligned} &\mathbb{E}\left(\left\|T_\eta^g(\mathbf{x}_t; \tilde{\nabla} S_t(\mathbf{x}_t)) - T_\eta^g(\mathbf{x}_t; \nabla S_t(\mathbf{x}_t))\right\|^2\right) \\ \leq &\eta^2\mathbb{E}\left[\mathbb{E}\left(\left\|\tilde{\nabla} S_t(\mathbf{x}_t) - \nabla S_t(\mathbf{x}_t)\right\|^2 |\mathcal{F}_t\right)\right] \leq \eta^2\sigma^2 w^{-2}. \end{aligned}$$

Then, plugging the latter to the expected value of (8) yields

$$\begin{aligned} \mathbb{E}\left(\mathrm{Reg}_w(T)\right) \leq &2\sum_{t=1}^{T}\mathbb{E}\left(\left\|\mathcal{P}(\mathbf{x}_t; \tilde{\nabla} S_{t,w}(\mathbf{x}_t))\right\|^2\right) \\ &+ 2T\sigma^2 w^{-2}. \end{aligned} \tag{9}$$

Setting $G_1 = \tilde{\nabla} S_{t-1}(\mathbf{x}_t)$, $G_2 = \frac{1}{w}(\tilde{\nabla} f_t(\mathbf{x}_t) - \tilde{\nabla} f_{t-w}(\mathbf{x}_t))$, and applying Lemma C.3 together with the termination rule of the inner loop, then implies that

$$\left\|\mathcal{P}(\mathbf{x}_t; \tilde{\nabla} S_t(\mathbf{x}_t))\right\|^2 \leq \frac{2\delta^2}{w^2} + \frac{2}{w^2}\left\|\tilde{\nabla} f_t(\mathbf{x}_t) - \tilde{\nabla} f_{t-w}(\mathbf{x}_t)\right\|^2.$$

Using the triangle inequality, the relation $(a+b+c)^2 \leq 3(a^2 + b^2 + c^2)$, and by applying expectation and the law of total expectation together with Definition 4.1, we can derive that $\mathbb{E}\left(\left\|\tilde{\nabla} f_t(\mathbf{x}_t) - \tilde{\nabla} f_{t-w}(\mathbf{x}_t)\right\|^2\right) \leq \frac{6\sigma^2}{w^2} +$

$3\mathbb{E}\left(\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-w}(\mathbf{x}_t)\|^2\right)$, and consequently

$$\mathbb{E}\left(\left\|\mathcal{P}(\mathbf{x}_t; \tilde{\nabla} S_t(\mathbf{x}_t))\right\|^2\right) \leq \frac{2\delta^2}{w^2} + \frac{12\sigma^2}{w^4} + \frac{6}{w^2}\mathbb{E}\left(\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-w}(\mathbf{x}_t)\|^2\right).$$

Summing over $t \in [T]$, and plugging $V_w[T]$ defined in (6), then yields $\sum_{t=1}^{T} \mathbb{E}\left(\left\|\mathcal{P}(\mathbf{x}_t; \tilde{\nabla} S_t(\mathbf{x}_t))\right\|^2\right) \leq 2\left(\delta^2 + \frac{6\sigma^2}{w^2}\right)\left(\frac{T}{w^2}\right) + \frac{6}{w^2} V_w[T]$. Finally, plugging the latter into (9), and recalling that $w \geq 1$, results with the desired bound. □

The local regret bound established in Theorem 4.2, and the almost sure termination in finite time proved in Theorem 4.1, leave the question of the number of prox operation still unattended. To answer this nontrivial question, we require more control of the random processes originating from the SFO in the form of the following assumption on the noise.

Assumption 1. *Given any point $(\mathbf{x}, \zeta) \in \mathbb{R}^n \times \Omega$ and a function $h : \mathbb{R}^n \to \mathbb{R}$, the stochastic first-order oracle $\mathcal{S}_\sigma$ satisfies that $\|\mathcal{S}_\sigma(\mathbf{x}; \zeta, h) - \nabla h(\mathbf{x})\| \leq \sigma$;*

Assumption 1 is not uncommon in the stochastic setting, even in convex problems, see e.g., (Jain et al., 2019, Kavis et al., 2019, Li & Orabona, 2019), and references therein. We emphasize that Theorems 4.1 and 4.2 *do not* require, *nor* assume, that Assumption 1 holds true.

The next theorem states that Algorithm 2 executes $O(Tw)$ prox operations and $O(Tw^2)$ SFO calls.

Theorem 4.3 (Iteration bound)**.** *Suppose that Assumption 1 holds true, and that $\eta \in (0, 1/(L+1))$, $\delta^2 > \sigma^2/\eta(1-\eta(L+1))$. Then the number of SFO calls is $O(w\tau)$ with*

$$\tau = \sum_{t=1}^{T} \tau_t \leq \frac{2Tw(g(\mathbf{x}_1) + 3M)}{(1-\eta(L+1))\eta\delta^2 - \sigma^2} = O(Tw). \quad (10)$$

Remark 4.1. *Under the conditions of Theorem 4.3, Theorem 4.1, Theorem 4.2 and Theorem 4.3, all hold true.*

4.2. Implications to Offline Stochastic Optimization

This section considers the reduction of our model to an offline stochastic non-convex composite optimization problem by examining our results when $f_t \equiv f$ for any $t \in [T]$. In this scenario, where the goal is to obtain an ε-stationary point $\mathbf{x}_* \in \mathbb{R}^n$ satisfying that $\|\mathcal{P}(\mathbf{x}_*; \nabla f(\mathbf{x}_*))\|^2 \leq \varepsilon$ (cf. Chapter 2 in (Beck, 2017)), our sliding average $S_{t,w}(\mathbf{x})$ is reduced to the objective function itself, and the local regret measure $\mathrm{Reg}_w(T)$ is reduced to the standard sum of prox-residuals in the consecutive points generated by the algorithm. Algorithm 2 itself takes the form of a stochastic prox-grad type method in which w calls to the SFO are used to approximate the gradient at each iteration. This resulting scheme bare some resembles to variance reduction techniques appearing in (Metel & Takeda, 2019, Wang et al., 2019, Yurtsever et al., 2019), where here, w seemingly takes the role of the batch-size, and the process of Algorithm 2 enforces the relation between the SFO's variance and w.

The connection between Algorithm 2 and SVR methods is further supported by the $O(M\sigma\varepsilon^{-3/2})$ SFO calls complexity guarantee for obtaining a ε-stationary point in expectation, which we will derive shortly. This complexity is currently the best known (sometimes written as $O(M\sigma\varepsilon^{-3})$ due to square-difference in the stationarity definition), and can only be obtained by SVR methods; see the already mentioned (Arjevani et al., 2019) for details.

Although obtained as a by-product, our offline-related result are of an independent interest and contribution, as, apart from providing a new connection between online learning and offline stochastic optimization, we also derive a new stochastic method with the best known guarantess under different model assumptions and procedure compared to the SVR literature.

It should be noted though that our assumptions, albeit standard in online optimization, are more restrictive compared to the related stochastic (offline) optimization literature (see e.g., Wang et al., 2019), as the former facilitates guarantees, first and foremost, for our *online* stochastic model. Indeed, methods for stochastic problems cannot address the adversarial online settings we study here. Notwithstanding, our complexity results suggest new scheme's design directions to explore in the development of (offline) stochastic methods, encouraging future study on the matter, that is unfortunately out of the scope of this paper.

Let us now derive the aforementioned guarantees, proofs are provided in the supplemental.

Theorem 4.4. *Let $\varepsilon > 0$, and t_* be chosen uniformly from $\{w, w+1, \ldots, T\}$. Suppose that $V_w[T] \leq cT/6$ for some $c > 0$. Then $\mathbb{E}\left(\|\mathcal{P}(\mathbf{x}_{t_*}; \nabla f(\mathbf{x}_{t_*}))\|^2\right) \leq \frac{2T(\delta^2 + 7\sigma^2 + c)}{(T-w)w^2}$.*

From Theorem 4.3 and Theorem 4.4 we obtain the desired guarantees.

Corollary 4.1. *Let $\varepsilon > 0$, and $t_* \in [T]$ be chosen* *uniformly from* $\{w, w+1, \ldots, T\}$*. Suppose that* $V_w[T] \leq cT/6$ *for some* $c > 0$*. If* $T = 2w$ *and* $w = \left\lceil 2\sqrt{(\delta^2 + 7\sigma^2 + c)/\varepsilon} \right\rceil$*. Then Algorithm 2 achieves* $\mathbb{E}\left(\|\mathcal{P}(\mathbf{x}_{t_*}; \nabla f(\mathbf{x}_{t_*}))\|^2\right) \leq \varepsilon$*. Additionally, under the conditions of Theorem 4.3 with* $\delta^2 = 2\eta\sigma^2/(1-\eta(L+1))$*, Algorithm 2 executes at most* $O(M\sigma\varepsilon^{-3/2})$ *SFO calls.*

5. Conclusions and future work

Our aim in this paper is to develop an online methodology for stochastic non-convex online optimization problems with constraints and regularization. Our focus on proximal-gradient schemes allows us to achieve min-max optimal bounds in terms of local regret minimization while at the same time bounding the number of overall operator queries. From a top-down perspective, this departure from standard notions of regret suggests various extensions based on different notions of local regret, ranging from measures of stationarity in offline non-convex analysis, to proxies for constraint qualification in problems with sufficient regularity. Additionally, our reductions to the offline stochastic setting suggest new and interesting schemes to address stochastic non-convex optimization problems.

Acknowledgements

The work of N. Hallak was partially conducted at EPFL, where it was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no 725594 - time-data). P. Mertikopoulos is grateful for financial support by the French National Research Agency (ANR) in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02), the LabEx PERSYVAL (ANR-11-LABX-0025-01), MIAI@Grenoble Alpes (ANR-19-P3IA-0003), and the grants ORACLESS (ANR-16-CE33-0004) and ALIAS (ANR-19-CE48-0018-01). V. Cevher gratefully acknowledges the support of the Swiss National Science Foundation (SNSF) under grant number 200021–178865/1, the European Research Council (ERC) under the Horizon 2020 research and innovation programme (grant agreement number 725594 - time-data), and 2019 Google Faculty Research Award. This research was also supported by the COST Action CA16228 "European Network for Game Theory" (GAMENET).

References

Abernethy, J., Bartlett, P. L., Rakhlin, A., and Tewari, A. Optimal strategies and minimax lower bounds for online convex games. In *COLT '08: Proceedings of the 21st Annual Conference on Learning Theory*, 2008.

Agarwal, A., Dekel, O., and Xiao, L. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *COLT '10: Proceedings of the 23rd Annual Conference on Learning Theory*, 2010.

Agarwal, N., Gonen, A., and Hazan, E. Learning in non-convex games with an optimization oracle. In Beygelzimer, A. and Hsu, D. (eds.), *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 18–29, Phoenix, USA, 25–28 Jun 2019. PMLR. URL http://proceedings.mlr.press/v99/agarwal19a.html.

Arjevani, Y., Carmon, Y., Duchi, J. C., Foster, D. J., Srebro, N., and Woodworth, B. Lower bounds for non-convex stochastic optimization. *arXiv preprint arXiv:1912.02365*, 2019.

Beck, A. *First-Order Methods in Optimization*, volume 25. SIAM, 2017.

Bertsekas, D. P. and Gallager, R. *Data Networks*. Prentice Hall, Englewood Cliffs, NJ, 2 edition, 1992.

Besbes, O., Gur, Y., and Zeevi, A. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, October 2015.

Bubeck, S. and Eldan, R. Multi-scale exploration of convex functions and bandit convex optimization. In *COLT '16: Proceedings of the 29th Annual Conference on Learning Theory*, 2016.

Bubeck, S. and Eldan, R. Kernel-based methods for bandit convex optimization. In *STOC '17: Proceedings of the 49th annual ACM SIGACT symposium on the Theory of Computing*, 2017.

Bubeck, S., Munos, R., Stoltz, G., and Szepesvári, C. $\mathcal{X}$-armed bandits. *Journal of Machine Learning Research*, 12:1655–1695, 2011.

Cesa-Bianchi, N., Gaillard, P., Lugosi, G., and Stoltz, G. Mirror descent meets fixed share (and feels no regret). In 989-997 (ed.), *Advances in Neural Information Processing Systems*, volume 25, 2012.

Chiang, C.-K., Yang, T., Lee, C.-J., Mahdavi, M., Lu, C.-J., Jin, R., and Zhu, S. Online optimization with gradual variations. In Mannor, S., Srebro, N., and Williamson, R. C. (eds.), *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pp. 6.1–6.20, Edinburgh, Scotland, 25–27 Jun 2012. PMLR. URL http://proceedings.mlr.press/v23/chiang12.html.

Fang, C., Li, C. J., Lin, Z., and Zhang, T. Spider: Near-optimal non-convex optimization via stochastic path-integrated differential estimator. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 31*, pp. 689–699. Curran Associates, Inc., 2018.

Garber, D. On the regret minimization of nonconvex online gradient ascent for online pca. In Beygelzimer, A. and Hsu, D. (eds.), *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine*

Learning Research, pp. 1349–1373, Phoenix, USA, 25–28 Jun 2019. PMLR. URL http://proceedings.mlr.press/v99/garber19a.html.

Grnarova, P., Levy, K. Y., Lucchi, A., Hofmann, T., and Krause, A. An online learning approach to generative adversarial networks. *arXiv preprint arXiv:1706.03269*, 2017.

Hazan, E. *Introduction to online convex optimization*. 2016. ISBN 978-1-68083-171-9. OCLC: 1102388146.

Hazan, E. and Seshadhri, C. Efficient learning algorithms for changing environments. In *ICML '09: Proceedings of the 26th International Conference on Machine Learning*, 2009.

Hazan, E., Agarwal, A., and Kale, S. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, December 2007.

Hazan, E., Singh, K., and Zhang, C. Efficient regret minimization in non-convex games. In Precup, D. and Teh, Y. W. (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1433–1441, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR. URL http://proceedings.mlr.press/v70/hazan17a.html.

Héliou, A., Martin, M., Mertikopoulos, P., and Rahier, T. Online non-convex optimization with inexact models. In *NeurIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.

Héliou, A., Martin, M., Mertikopoulos, P., and Rahier, T. Zeroth-order non-convex learning via hierarchical dual averaging. In *ICML '21: Proceedings of the 38th International Conference on Machine Learning*, 2021.

Jain, P., Nagaraj, D., and Netrapalli, P. Making the last iterate of sgd information theoretically optimal. In Beygelzimer, A. and Hsu, D. (eds.), *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 1752–1755, Phoenix, USA, 25–28 Jun 2019. PMLR. URL http://proceedings.mlr.press/v99/jain19a.html.

Jin, C., Netrapalli, P., and Jordan, M. I. What is local optimality in nonconvex-nonconcave minimax optimization? *arXiv preprint arXiv:1902.00618*, 2019.

Kavis, A., Levy, K. Y., Bach, F., and Cevher, V. Unixgrad: A universal, adaptive algorithm with optimal guarantees for constrained optimization. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alche Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 32*, pp. 6257–6266. Curran Associates, Inc., 2019.

Kleinberg, R. D. Nearly tight bounds for the continuum-armed bandit problem. In *NIPS' 04: Proceedings of the 18th Annual Conference on Neural Information Processing Systems*, 2004.

Kleinberg, R. D., Slivkins, A., and Upfal, E. Multi-armed bandits in metric spaces. In *STOC '08: Proceedings of the 40th annual ACM symposium on the Theory of Computing*, 2008.

Krichene, W., Balandat, M., Tomlin, C., and Bayen, A. The Hedge algorithm on a continuum. In *ICML '15: Proceedings of the 32nd International Conference on Machine Learning*, 2015.

Li, X. and Orabona, F. On the convergence of stochastic gradient descent with adaptive stepsizes. In Chaudhuri, K. and Sugiyama, M. (eds.), *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pp. 983–992. PMLR, 16–18 Apr 2019. URL http://proceedings.mlr.press/v89/li19c.html.

Metel, M. and Takeda, A. Simple stochastic gradient methods for non-smooth non-convex regularized optimization. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4537–4545, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL http://proceedings.mlr.press/v97/metel19a.html.

Nemirovski, A., Juditsky, A., Lan, G., and Shapiro, A. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on optimization*, 19(4):1574–1609, 2009.

Nouiehed, M., Sanjabi, M., Huang, T., Lee, J. D., and Razaviyayn, M. Solving a class of non-convex min-max games using iterative first order methods. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 32*, pp. 14934–14942. Curran Associates, Inc., 2019.

Pang, J.-S. and Scutari, G. Nonconvex games with side constraints. *SIAM Journal on Optimization*, 21(4):1491–1522, 2011. doi: 10.1137/100811787.

Perkins, S., Mertikopoulos, P., and Leslie, D. S. Mixed-strategy learning with continuous action sets. *IEEE Transactions on Automatic Control*, 62(1):379–384, January 2017.

Shakkottai, S. and Srikant, R. Network optimization and control. *Foundations and Trends in Networking*, 2(3):271–379, 2008.

Suggala, A. S. and Netrapalli, P. Online Non-Convex Learning: Following the Perturbed Leader is Optimal. *arXiv:1903.08110 [cs, math, stat]*, March 2019. URL http://arxiv.org/abs/1903.08110. arXiv: 1903.08110.

Wang, Z., Ji, K., Zhou, Y., Liang, Y., and Tarokh, V. Spiderboost and momentum: Faster variance reduction algorithms. In *Advances in Neural Information Processing Systems*, pp. 2403–2413, 2019.

Xiao, L. Dual averaging methods for regularized stochastic learning and online optimization. *Journal of Machine Learning Research*, 11:2543–2596, October 2010.

Yurtsever, A., Sra, S., and Cevher, V. Conditional gradient methods via stochastic path-integrated differential estimator. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 7282–7291, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL http://proceedings.mlr.press/v97/yurtsever19b.html.