REGRET MINIMIZATION IN STOCHASTIC NON-CONVEX LEARNING VIA A PROXIMAL-GRADIENT APPROACH

NADAV HALLAK[*,c], PANAYOTIS MERTIKOPOULOS[◇,♯], AND VOLKAN CEVHER[‡]

Abstract. Motivated by applications in machine learning and operations research, we study regret minimization with stochastic first-order oracle feedback in online constrained, and possibly non-smooth, non-convex problems. In this setting, the minimization of external regret is beyond reach, so we focus on a local regret measure defined via a proximal-gradient mapping. To achieve no (local) regret in this setting, we develop a prox-grad 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 min-max order-optimal, and we also establish a bound on the number of prox-grad 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 prox-grad 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, composite, etc., 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 and Seshadhri (2009) for variant regret measures, and Abernethy et al. (2008), Agarwal et al. (2010), and Bubeck and 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

[*] The Technion, 3200003, Haifa, Israel.
[c] Corresponding author.
[◇] Univ. Grenoble Alpes, CNRS, Inria, LIG, 38000, Grenoble, France.
[♯] Criteo AI Lab.
[‡] École Polytechnique Fédérale de Lausanne (EPFL).
E-mail addresses: ndvhllk@technion.ac.il, panayotis.mertikopoulos@imag.fr, volkan.cevher@epfl.ch.

The work of N. Hallak was conducted at EPFL, and 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 also grateful for financial support by the French National Research Agency (ANR) under grant no. ANR–16–CE33–0004–01 (ORACLESS). V. Cevher gratefully acknowledges the support of the Swiss National Science Foundation (SNSF) under grant № 200021–178865/1, the European Research Council (ERC) under the Horizon 2020 research and innovation programme (grant agreement № 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).

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.

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 search such as the hypercube. Krichene et al. (2015) and, more recently, Perkins et al. (2017) and Héliou et al. (2020), 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 and 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 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 $O(w^3)$. 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 $O(w^2)$ in the process.
- As a by-product, but of an independent interest and 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

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

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.,

$$\|\nabla f_t(\mathbf{x}) - \nabla f_t(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\| \qquad \forall \mathbf{x}, \mathbf{y} \in \operatorname{dom} g, \ \forall t \in [T].$$

(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, meaning that if $\operatorname{dom} g$ is bounded, then the assumptions on f_t 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., Agarwal et al. (2019), Hazan et al. (2017).

A particularly interesting instance of a two players non-convex game in which the feasible set is usually compact, and 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 and Gallager (1992) and Shakkottai and 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 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 $\mathrm{Reg}(T) = \max_{\mathbf{x}\in\mathrm{dom}\, g} \sum_{t=1}^{T} [\ell_t(\mathbf{x}_t) - \ell_t(\mathbf{x})]$: 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 $\mathrm{Reg}(T) = \Omega(T)$, so the notion of regret minimization must also be re-examined in this setting.

We address both of these problems by extending the *local regret minimization* framework of Hazan et al. (2017) to the composite problem (P). 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

$$T_\eta^g(\mathbf{x};\mathbf{d}) \equiv \mathrm{prox}_{\eta g}(\mathbf{x} - \eta\mathbf{d}) = \arg\min_{\mathbf{z}\in\mathbb{R}^n} \{\eta g(\mathbf{z}) + \tfrac{1}{2}\|\mathbf{x} - \eta\mathbf{d} - \mathbf{z}\|^2\}, \tag{1}$$

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

$$\mathcal{P}_\eta^g(\mathbf{x};\mathbf{d}) = \frac{1}{\eta}\left(\mathbf{x} - T_\eta^g(\mathbf{x};\mathbf{d})\right). \tag{2}$$

Remark 2.1. *We note that the purpose behind the use of a general vector* $\mathbf{d}$ *in Eq. (1) and Eq. (2) 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. (1) and Eq. (2) in the smooth unconstrained and constrained scenarios. If $g \equiv 0$, then Eq. (1) is the gradient descent operator and Eq. (2) reduces to $\mathcal{P}_\eta^g(\mathbf{x};\nabla f(\mathbf{x})) = \nabla f(\mathbf{x})$. 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. (1) and its corresponding projection residual $\mathcal{P}_\eta^g(\mathbf{x};\nabla f(\mathbf{x})) = \eta^{-1}(\mathbf{x} - \mathrm{proj}_{\mathcal{K}}(\mathbf{x} - \eta\nabla f(\mathbf{x})))$.

A fundamental result in optimization is that $\mathcal{P}_\eta^g(\mathbf{x};\nabla f(\mathbf{x})) = 0$ if and only if $\mathbf{x}$ is a stationary point of (P), making the residual quantity $\mathrm{Res}_\eta^g(\mathbf{x}) \equiv \|\mathcal{P}_\eta^g(\mathbf{x};\nabla f(\mathbf{x}))\|^2 \geq 0$ an efficient proxy for the first-order optimality condition (see also (Beck, 2017, Ch. 10)).

Motivated by this, it would seem natural to define the regret of an online policy $\mathbf{x}_t$ at time T as the classical measure in non-convex optimization

$$\mathrm{Reg}(T) \equiv \sum_{t=1}^{T} \mathrm{Res}_{\eta}^{g}(\mathbf{x}_t) = \sum_{t=1}^{T} \left\| \mathcal{P}_{\eta}^{g}(\mathbf{x}_t; \nabla f_t(\mathbf{x}_t)) \right\|^2 . \tag{3}$$

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. (3) seems impossible.

Because of this, 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. Formally, for all $w \in [T]$, consider the sliding average

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

with the convention $f_t \equiv 0$ for $t \leq 0$. 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

$$\mathrm{Reg}_w(T) = \sum_{t=1}^{T} \left\| \mathcal{P}_{\eta}^{g}\left(\mathbf{x}_t; \nabla S_{t,w}(\mathbf{x}_t)\right) \right\|^2 . \tag{4}$$

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. (4) is quantified by the minimax bound of Hazan et al. (2017) who showed that an informed adversary can impose $\mathrm{Reg}_w(T) = \Omega(T/w^2)$. 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. (4) generalizes standard measures in simpler models:

- In the offline case $f_t \equiv f$, we immediately recover the classical measure of Eq. (3).
- If $g \equiv 0$, we readily obtain $\mathrm{Reg}_w(T) = (1/w^2) \sum_{t=1}^{T} \| \sum_{i=t-w+1}^{t} \nabla f_i(\mathbf{x}_t) \|^2$, 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 and ω_t is an i.i.d sequence of random seeds, then $\mathbb{E}\left(\mathrm{Reg}_w(T)\right)/T \geq \sum_{t=1}^{T} \|\nabla f(\mathbf{x}_t)\|^2$, 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 $\mathbb{E}\left[\left\| \mathcal{P}_{\eta}^{g}\left(\mathbf{x}_{t_*}; \nabla f(\mathbf{x}_{t_*})\right) \right\|^2 \right] \leq \mathbb{E}\left(\mathrm{Reg}_w(T)\right)/T$,

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*

$$V_w[T] = \sup_{\mathbf{x}\in\operatorname{dom} g} \left\{ \sum_{i=1}^{T} \|\nabla f_i(\mathbf{x}) - \nabla f_{i-w}(\mathbf{x})\|^2 \right\}. \tag{5}$$

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 $t = 1, \dots, T$ do:

(1) $f_t : \mathbb{R}^n \to \mathbb{R}$ is determined;

(2) Set $\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t$;

(3) While $\left\| \mathcal{P}_\eta^g \left(\mathbf{x}_{t+1}; \nabla S_{t,w}(\mathbf{x}_{t+1})\right) \right\| > \delta/w$ 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 $O(w^2)$ prox-grad operations. We note that the bound $O(w^2)$ on the number of prox-grad operations improves the bound $O(Tw)$ established for the simplified case of $g \equiv 0$ by Hazan et al. (2017).

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

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

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*

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

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 $\omega_1, \omega_2, \ldots$, 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 ω, and a smooth function $h\colon \mathbb{R}^n \to \mathbb{R}$ satisfies:*

(1) $\mathcal{S}_\sigma(\mathbf{x}; \omega, h)$ is unbiased relative to $\nabla h(\mathbf{x})$: $\mathbb{E}\left(\mathcal{S}_\sigma(\mathbf{x}; \omega, h) - \nabla h(\mathbf{x})\right) = 0$;

(2) $\mathcal{S}_\sigma(\mathbf{x}; \omega, h)$ has variance bounded by $\sigma > 0$: $\mathbb{E}\left(\|\mathcal{S}_\sigma(\mathbf{x}; \omega, h) - \nabla h(\mathbf{x})\|^2\right) \leq \sigma^2$.

With all this 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. In detail, the algorithm is presented in pseudocode form below:

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 and 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: 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 $t = 1, 2, \dots, T$ do:

(1) Function is updated to $f_t : \mathbb{R}^n \to \mathbb{R}$;
(2) Sample $\tilde{\nabla} f_t(\mathbf{x}_t) \leftarrow \mathcal{S}_{\sigma/w}(\mathbf{x}_t; \omega, f_t)$;
(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 $\tilde{\nabla} f_i(\mathbf{y}_t^{k+1}) \leftarrow \mathcal{S}_{\sigma/w}(\mathbf{y}_t^{k+1}; \omega, f_i)$ for any $i = t - w + 1, \dots, t$;
  (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 $\tilde{\nabla} S_t(\mathbf{x}_{t+1}) = G_t^k$.

Algorithm 2 takes a different approach in this context by, instead of stating this connection in the analysis, it explicitly links the batch-size (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.

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. 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*

$$\delta^2 > \frac{2\sigma^2}{\eta\,(1 - \eta L)}. \tag{6}$$

Then τ_t and τ are almost surely finite, and

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

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*

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

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

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 and 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 $O(w^2)$ prox operations and $O(w^3)$ 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*

$$\tau = \sum_{t=1}^{T} \tau_t \leq \frac{2w^2(g(\mathbf{x}_1) + 2M)}{(1-\eta(L+1))\eta\delta^2 - \sigma^2} = O(w^2). \tag{7}$$

Remark 4.1. *Under the conditions of Theorem 4.3, both Theorem 4.1 and Theorem 4.3 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. (Beck, 2017, Ch. 2)), 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 and 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 guaranteess 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 (e.g., Wang et al. (2019)), as the former facilitate 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 if 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 $\{w, w+1, \dots, T\}$. Suppose that $V_w[T] \leq cT/6$ for some $c > 0$. Then $\mathbb{E}\left(\|\mathcal{P}(\mathbf{x}_{t_*}; \nabla f(\mathbf{x}_{t_*}))\|^2\right) \leq \frac{2T\left(\delta^2+7\sigma^2+c\right)}{(T-w)w^2}$.*

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 $\{w, w+1, \dots, T\}$. 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 was to develop an online prox-grad methodology for stochastic non-convex online optimization problems with constraints and regularization (possibly non-smooth). In this regard, the proposed framework achieves the min-max optimal bounds for 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. We defer these questions to future research.

A. Motivating examples

A.1. **A conceptual approach for non-convex games.** We extend here the solution concept for non-convex m-player games with *smoothed local equilibrium* proposed by Hazan et al. (2017) to be valid in our *stochastic composite* game setup. We emphasize that the guarantees we present in this section are also valid for when each player only has access to a stochastic first-order oracle, making it closer to practical use.

To model the multi-player setting, consider m problems of the form (P) corresponding to each of the players, where every player i observes her online part of her objective function

$$f_t^i(\mathbf{z}) := f(\mathbf{x}_t^1, \dots, \mathbf{x}_t^{i-1}, \mathbf{z}, \dots, \mathbf{x}_t^m), \tag{8}$$

and then decides on $\mathbf{x}_{t+1}^i$.

It is sometimes desirable to induce specific properties in the game, this is fully supported by our model (P). For example: (i) to incur risk-aversion, the regularizer of each player g^i can be chosen accordingly, e.g., L^1-norm; (ii) to ensure a meaningful solution, such as the

global minimax point condition defined by Jin et al. (2019), restriction of the decision set to a compact convex set can be applied.

In our non-convex setting, obtaining the global measure of Nash equilibrium is beyond reach, and may not exist at all (Jin et al., 2019, Prop. 6). Thus, a different, local, measure for equilibrium is essential. This topic is already receiving much attention in the literature, for example, for a multi-player non-convex games, Pang and Scutari (2011) proposes the local *quasi-Nash equilibrium* measure defined using KKT conditions. In the case of a (two-players) minmax game (e.g., GANs) for example, local measure is defined as the stationarity (first-order condition) of both players in the very recent Jin et al. (2019), Nouiehed et al. (2019). For additional details, we refer to the works alluded above.

We follow the smoothed local equilibrium approach (Hazan et al., 2017, Sec. 6) , and extend it here to our composite model. This approach comes naturally from assuming that the players take into account the behavior history of the other players. Other than that, it allows for a tractable notion of equilibrium.

The *smoothed local equilibrium* is defined for the joint cost function (8) as follows, where $S^i_{t,w}(\mathbf{x}) = \frac{1}{w}\sum_{j=t-w+1}^{t} f^i_j(\mathbf{x})$.

Definition A.1 (smoothed local equilibrium)**.** *Let $\eta > 0, w \geq 1$. For an m-player iterative game with cost functions as in (8), a joint strategy at iteration $t > 0$, $(\mathbf{x}^1_t, \ldots, \mathbf{x}^{i-1}_t, \mathbf{x}^i_t, \ldots, \mathbf{x}^m_t)$, is an ε-(η, w) smoothed local equilibrium with respect to the history of w-iterates if:*

$$\left\| \mathcal{P}^{g^i}_{\eta}(\mathbf{x}^i_t; \nabla S^i_{t,w}(\mathbf{x}^i_t)) \right\|^2 \leq \varepsilon \qquad \forall i \in [m]. \tag{9}$$

Denote by $\mathrm{Reg}^i_w(T)$ the local regret (cf. Eq. (4)) of the i-th player. We first derive a guarantee for when each player has access to a perfect first-order oracle (using Theorem 3.1).

Theorem A.1 (Equilibrium with perfect oracle)**.** *Let the sequence $(\mathbf{x}^1_t, \ldots, \mathbf{x}^{i-1}_t, \mathbf{x}^i_t, \ldots, \mathbf{x}^m_t)$, $t = 1, \ldots, T$ be generated by running Algorithm 1 for all players simultaneously with input $\eta > 0$ and $w = \lceil 2k(\delta^2 + c)\varepsilon^{-1/2} \rceil$, given that the online function is determined by (8). Suppose that $V_w[T] \leq cT$ for some $c > 0$. Then there exists $t^* \geq w$ such that (9) holds true.*

Proof. There exists a $t^* \geq w$ such that

$$\begin{aligned}\sum_{i=1}^{k} \left\| \mathcal{P}^{g^i}_{\eta}(\mathbf{x}^i_{t^*}; \nabla f^i_{t^*}(\mathbf{x}^i_{t^*})) \right\|^2 &\leq \frac{1}{T-w} \sum_{i=1}^{k} \sum_{t=w}^{T} \left\| \mathcal{P}^{g^i}_{\eta}(\mathbf{x}^i_t; \nabla f^i_t(\mathbf{x}^i_t)) \right\|^2 \\ &\leq \frac{1}{T-w} \sum_{i=1}^{k} \mathrm{Reg}^i_w(T).\end{aligned}$$

Thus, if each player has access to a perfect first-order oracle and $V_w[T] \leq cT$, then by Theorem 3.1

$$\sum_{i=1}^{k} \left\| \mathcal{P}^{g^i}_{\eta}(\mathbf{x}^i_{t^*}; \nabla f^i_{t^*}(\mathbf{x}^i_{t^*})) \right\|^2 \leq \frac{1}{T-w} \sum_{i=1}^{k} \frac{2}{w^2} \left(T\delta^2 + V_w[T] \right) \leq \frac{2kT(\delta^2 + c)}{(T-w)w^2}.$$

Consequently, by setting $T = w^2$ and $w = \lceil 2k(\delta^2 + c)\varepsilon^{-1/2} \rceil$ we obtain

$$\sum_{i=1}^{k} \left\| \mathcal{P}^{g^i}_{\eta}(\mathbf{x}^i_{t^*}; \nabla f^i_{t^*}(\mathbf{x}^i_{t^*})) \right\|^2 \leq \frac{2k(\delta^2 + c)}{(w-1)w} \leq \varepsilon,$$

as desired. □

By similar arguments, we derive the guarantees for when players have access via a stochastic first-order oracle, only now we utilize Theorem 4.2; we implicitly assume here that all the conditions of Theorem 4.2 are satisfied.

Theorem A.2 (Equilibrium with stochastic first-order oracle)**.** *Suppose that the sequence* $(\mathbf{x}_t^1, \dots, \mathbf{x}_t^{i-1}, \mathbf{x}_t^i, \dots, \mathbf{x}_t^m)$, $t = 1, \dots, T$ *is generated by running Algorithm 1 for all players simultaneously with input* $\eta > 0$ *and* $w = \lceil \frac{2k\left(\delta^2+7\sigma^2+6c\right)}{\sqrt{\varepsilon}} \rceil$*, given that the online function is determined by (8). Suppose that* $V_w[T] \leq cT$ *for some* $c > 0$*. Then there exists* $t^* \geq w$ *such that (9) holds true in expectation.*

Proof. There exists a $t^* \geq w$ such that

$$\begin{aligned}\sum_{i=1}^{k} \left\| \mathcal{P}_\eta^{g^i}(\mathbf{x}_{t^*}^i; \nabla f_{t^*}^i(\mathbf{x}_{t^*}^i)) \right\|^2 &\leq \frac{1}{T-w} \sum_{i=1}^{k} \sum_{t=w}^{T} \left\| \mathcal{P}_\eta^{g^i}(\mathbf{x}_t^i; \nabla f_t^i(\mathbf{x}_t^i)) \right\|^2 \\ &\leq \frac{1}{T-w} \sum_{i=1}^{k} \mathrm{Reg}_w^i(T).\end{aligned}$$

Thus, by taking expectation and using the fact that $V_w[T] \leq cT$, we obtain from Theorem 4.2 that

$$\begin{aligned}\sum_{i=1}^{k} \mathbb{E} \left\| \mathcal{P}_\eta^{g^i}(\mathbf{x}_{t^*}^i; \nabla f_{t^*}^i(\mathbf{x}_{t^*}^i)) \right\|^2 &\leq \frac{1}{T-w} \sum_{i=1}^{k} 2 \left(\left(\frac{T}{w^2} \right) \left(\delta^2 + 7\sigma^2 \right) + \frac{6}{w^2} V_w[T] \right) \\ &= \frac{2kT \left(\delta^2 + 7\sigma^2 + 6c \right)}{(T-w)w^2}.\end{aligned}$$

Consequently, by setting $T = w^2$ and $w = \lceil \frac{2k\left(\delta^2+7\sigma^2+6c\right)}{\sqrt{\varepsilon}} \rceil$ we obtain

$$\sum_{i=1}^{k} \mathbb{E} \left\| \mathcal{P}_\eta^{g^i}(\mathbf{x}_{t^*}^i; \nabla f_{t^*}^i(\mathbf{x}_{t^*}^i)) \right\|^2 \leq \frac{2k \left(\delta^2 + 7\sigma^2 + 6c \right)}{(w-1)w} \leq \varepsilon,$$

as desired. □

A.2. **The online traffic assignment problem.** Referring to Bertsekas and Gallager (1992) and Shakkottai and 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. To state this precisely, consider a directed multi-graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$. Embedded in this network is a set of origin-destination (O/D) pairs $(o_i, d_i) \in \mathcal{V} \times \mathcal{V}$, $i \in \mathcal{N} = \{1, 2, \dots, N\}$, each routing a (possibly random) quantity of traffic from o_i to d_i via a set of paths $\mathcal{P}_i$ in $\mathcal{G}$. Writing $\mathcal{K}_i = \Delta(\mathcal{P}_i)$ for the simplex spanned by $\mathcal{P}_i$, a *trafic allocation vector* for the i-th O/D pair is defined to be a vector $\mathbf{x}_i = (x_{i,p_i})_{p_i \in \mathcal{P}_i} \in \mathcal{K}_i$ with each x_{i,p_i} denoting the fraction of the traffic of the i-th O/D pair that is routed via p_i. Then, collectively, a *traffic allocation profile* is an ensemble $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ of such vectors belonging to the product space $\mathcal{K} = \prod_i \mathcal{K}_i$.

In this general context, the cost (delay, latency, etc.) of routing a certain amount of traffic via a given path p_i is a function $\ell_{p_i}(\mathbf{x}; \boldsymbol{\lambda})$ of the chosen allocation profile $\mathbf{x} \in \mathcal{K}$ and the set of *traffic demands* $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$ of each O/D pair.These demands are typically assumed to follow a non-stationary probability distribution (e.g., accounting for diurnal variations in

an urban traffic network), leading to the *online traffic assignment problem* (OnTAP) stated below:

$$\begin{aligned} \text{minimize} \quad & \ell_t(\mathbf{x}) = \sum_{i\in\mathcal{N}} \sum_{p_i\in\mathcal{P}_i} x_{i,p_i} \ell_{p_i}(\mathbf{x};\boldsymbol{\lambda}_t) + \mu\|\mathbf{x}\|_1 \\ \text{subject to} \quad & \mathbf{x}\in\mathcal{K}. \end{aligned} \tag{OnTAP}$$

In the above formulation, the sparsity-inducing L^1 term is intended to "robustify" solutions by minimizing the overall number of paths employed. The cost functions ℓ_{p_i} are sums of positive polynomials (described below), so they are smooth over $\mathcal{K}$ but may otherwise be non-convex. As such, (OnTAP) can be cast in the framework of (P) by taking $g = \delta_{\mathcal{K}} + \mu\|\cdot\|_1$ with $\delta_{\mathcal{K}}$ denoting the convex indicator of $\mathcal{K}$.

Let us now detail the definition of the cost functions ℓ_{p_i} for (OnTAP). For simplicity, we will suppress the O/D index $i\in\mathcal{N}$, i.e., we will treat the problem as a single-O/D one; this doesn't play a major role in the sequel and only serves to make the notation ligther.

To begin, given a traffic allocation vector $\mathbf{x}\in\mathcal{K}$ and an inflow rate λ, the *traffic load* carried by edge $e\in\mathcal{E}$ is defined to be the total traffic routed via the edge in question, i.e.,

$$y_e \equiv y_e(\mathbf{x};\lambda) = \lambda \sum_{p:p\ni e} x_p, \tag{10}$$

and we write $\mathbf{y} = (y_e)_{e\in\mathcal{E}}$ for the corresponding *load profile* on the network. Given all this, the cost (delay, latency, etc.) experienced by an infinitesimal traffic element traversing edge e is given by a non-decreasing continuous *cost function* $\ell_e\colon \mathbb{R}_+ \to \mathbb{R}_+$; more precisely, if $\mathbf{y}\equiv\mathbf{y}(\mathbf{x};\lambda)$ is the load profile induced by a traffic allocation profile $\mathbf{x}\in\mathcal{K}$ and a traffic demand λ, the incurred cost on edge $e\in\mathcal{E}$ is simply $\ell_e(y_e)$. Hence, the associated cost for path $p\in\mathcal{P}$ will be

$$\ell_p(\mathbf{x};\lambda) \equiv \sum_{e\in p} \ell_e(y_e(\mathbf{x};\lambda)) = \sum_{e\in p} \ell_e\left(\lambda \sum_{p':p'\ni e} x_{p'}\right). \tag{11}$$

In urban traffic networks, the cost functions ℓ_e are typically non-decreasing positive polynomials fitted to appropriate statistical data; a common choice is the so-called "quartic BPR" model $\ell_e(y_e) = a_e + b_e y_e^4$ of the US Bureau of Public Roads (BPR), but this is beyond our scope.

B. Regretfulness when $w = 1$

For completeness, we provide a simple example for when the "standard" stationarity measure Eq. (3), obtained from the local regret when $w = 1$, fails. The bound $O(T/w^2)$ established in (Hazan et al., 2017, Thm. 2.7) is proved via a similar example.

Suppose that $g(x) = \delta_{[-1,1]}(x)$ is the indicator function for the set $[-1,1]$, and that

$$f_t(x) = \begin{cases} -x & \text{with probability } 0.5, \\ x & \text{with probability } 0.5. \end{cases}$$

Then

$$\mathbb{E}\,\mathrm{Reg}_1(T) = \mathbb{E}\sum_{t=1}^{T} \left\|\mathcal{P}_\eta^g(\mathbf{x}_t;\nabla f_t(\mathbf{x}_t))\right\|^2 \geq O(T).$$

C. Fundamental Properties

Throughout the analysis, we utilize fundamental properties of the prox operator for L-smooth functions. The descent lemma (see e.g., (Beck, 2017, Lem. 5.7)) and the sufficient decrease property of the prox-grad operator (cf. (Beck, 2017, Lem. 10.4)) are given as follows.

Lemma C.1 (Descent lemma)**.** *Let $f:\mathbb{R}^n\to(-\infty,\infty]$ be an L-smooth function ($L\geq 0$) over a convex set $C\subseteq\mathbb{R}^n$. Then for any $\mathbf{x},\mathbf{y}\in C$, $f(\mathbf{y})\leq f(\mathbf{x})+\langle\nabla f(\mathbf{x}),\mathbf{y}-\mathbf{x}\rangle+\frac{L}{2}\|\mathbf{x}-\mathbf{y}\|^2$.*

Lemma C.2 (Sufficient decrease property)**.** *Let $h:\mathbb{R}^n\to\mathbb{R}\cup\{\infty\}$ be a proper, convex, l.s.c function, and $f:\mathbb{R}^n\to(-\infty,\infty)$ be an L-smooth function ($L\geq 0$) over $\operatorname{dom}h$. Then for any $\mathbf{x}\in\int\operatorname{dom}h$ and $\eta\in(0,L/2)$ it holds for $\mathbf{x}^+=\operatorname{prox}_{\eta h}(\mathbf{x}-\eta\nabla f(\mathbf{x}))$ that*

$$h(\mathbf{x})+f(\mathbf{x})-h(\mathbf{x}^+)-f(\mathbf{x}^+)\geq\eta\left(1-\frac{\eta L}{2}\right)\left\|\frac{1}{\eta}\left(\mathbf{x}^+-\mathbf{x}\right)\right\|^2.$$

We also use a trivial, yet essential, property of the prox-grad mapping.

Lemma C.3. *For any $\mathbf{x},\mathbf{d}_1,\mathbf{d}_2\in\mathbb{R}^n$ and $\eta>0$ it holds that*

$$\left\|\mathcal{P}_\eta^g(\mathbf{x};\mathbf{d}_1+\mathbf{d}_2)\right\|\leq\left\|\mathcal{P}_\eta^g(\mathbf{x};\mathbf{d}_1)\right\|+\|\mathbf{d}_2\|.$$

Proof. By the triangle inequality and non-expensiveness of the prox operator (cf. (Beck, 2017, Theorem 6.42))

$$\begin{aligned}\left\|\mathcal{P}_\eta^g(\mathbf{x};\mathbf{d}_1+\mathbf{d}_2)\right\|-\left\|\mathcal{P}_\eta^g(\mathbf{x};\mathbf{d}_1)\right\|&\leq\left\|\mathcal{P}_\eta^g(\mathbf{x};\mathbf{d}_1+\mathbf{d}_2)-\mathcal{P}_\eta^g(\mathbf{x};\mathbf{d}_1)\right\|\\&\leq\frac{1}{\eta}\left\|(\mathbf{x}-\eta(\mathbf{d}_1+\mathbf{d}_2))-(\mathbf{x}-\eta\mathbf{d}_1)\right\|=\|\mathbf{d}_2\|.\end{aligned}$$

□

D. Proofs of Section 3

Proof of Theorem 3.1. Note that

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

Setting $h_1=S_{t-1}$, $h_2=\frac{1}{w}(f_t-f_{t-w})$, applying Lemma C.3 and the triangle inequality yields

$$\begin{aligned}\|\mathcal{P}(\mathbf{x}_t;\nabla S_t(\mathbf{x}_t))\|&=\|\mathcal{P}(\mathbf{x}_t;\nabla(h_1+h_2)(\mathbf{x}_t))\|\\&\leq\|\mathcal{P}(\mathbf{x}_t;\nabla S_{t-1}(\mathbf{x}_t))\|+\frac{1}{w}\|\nabla f_t(\mathbf{x}_t)-\nabla f_{t-w}(\mathbf{x}_t)\|.\end{aligned}$$

By the definition of the method, i.e. $\|\mathcal{P}(\mathbf{x}_t;\nabla S_{t-1}(\mathbf{x}_t))\|\leq\frac{\delta}{w}$, we thus have that

$$\|\mathcal{P}(\mathbf{x}_t;\nabla S_t(\mathbf{x}_t))\|\leq\frac{\delta}{w}+\frac{1}{w}\|\nabla f_t(\mathbf{x}_t)-\nabla f_{t-w}(\mathbf{x}_t)\|,\qquad\forall t\in[T],$$

and consequently, for any $t\in[T]$,

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

Summing over $t = 1, \ldots, T$, then results with

$$\mathrm{Reg}_w(T) = \sum_{t=1}^{T} \left\| \mathcal{P}(\mathbf{x}_t; \nabla S_t(\mathbf{x}_t)) \right\|^2 \leq \frac{2}{w^2} \left(T\delta^2 + V_w[T] \right).$$

□

To prove that Algorithm 1 executes $O(w^2)$ prox-grad calls, we require a sufficient decrease property that is given next.

Lemma D.1 (Sufficient decrease property)**.** *Let $t \in [T]$, and let τ_t be the number of times step 3 is executed at the t-th iteration. Then*

$$S_{t,w}(\mathbf{x}_t) + g(\mathbf{x}_t) - S_{t,w}(\mathbf{x}_{t+1}) - g(\mathbf{x}_{t+1}) \geq \tau_t \left(\eta - \frac{\eta^2 L}{2} \right) \frac{\delta^2}{w^2}, \qquad \forall t \in [T].$$

Proof. Denote the sequence generated in the inner loop at time $t \in [T]$ by

$$\mathbf{y}_t^0 = \mathbf{x}_t, \quad \mathbf{y}_t^{k+1} = \operatorname*{arg\,min}_{\mathbf{z} \in \mathbb{R}^n} g(\mathbf{z}) + \langle \nabla S_t(\mathbf{y}_t^k), \mathbf{z} - \mathbf{y}_t^k \rangle + \frac{1}{2\eta} \|\mathbf{z} - \mathbf{y}_t^k\|^2, \qquad k = 0, 1, \ldots, \tau_t - 1,$$

and note that $\mathbf{y}_t^{\tau_t} = \mathbf{x}_{t+1}$. By the sufficient decrease property of the prox-grad operator (cf. Lemma C.2), and the stopping criteria of the inner loop, we have that for all $k = 0, 1, \ldots, \tau_t - 1$

$$S_t(\mathbf{y}_t^k) + g(\mathbf{y}_t^k) - S_t(\mathbf{y}_t^{k+1}) - g(\mathbf{y}_t^{k+1}) \geq \left(\eta - \frac{\eta^2 L}{2} \right) \left\| \mathcal{P}(\mathbf{y}_t^k; \nabla S_t(\mathbf{y}_t^k)) \right\|^2 \geq \left(\eta - \frac{\eta^2 L}{2} \right) \frac{\delta^2}{w^2}. \tag{12}$$

Summing (12) over $k = 0, 1, \ldots, \tau_t - 1$, then yields

$$\begin{aligned} S_t(\mathbf{x}_t) + g(\mathbf{x}_t) - S_t(\mathbf{x}_{t+1}) - g(\mathbf{x}_{t+1}) &= S_t(\mathbf{y}_t^0) + g(\mathbf{y}_t^0) - S_t(\mathbf{y}_t^{\tau_t}) - g(\mathbf{y}_t^{\tau_t}) \\ &\geq \tau_t \left(\eta - \frac{\eta^2 L}{2} \right) \frac{\delta^2}{w^2} \end{aligned}$$

which completes our proof. □

We will now bound the number of prox-grad iterations executed by Algorithm 1.

Proof of Theorem 3.2. Recall that $S_0(\mathbf{x}_0) \equiv 0$, and $S_t(\mathbf{x}) = \frac{1}{w}(f_t(\mathbf{x}) - f_{t-w}(\mathbf{x})) + S_{t-1}(\mathbf{x})$. Thus,

$$\begin{aligned} S_T(\mathbf{x}_T) &= \sum_{t=1}^{T} \left(S_t(\mathbf{x}_t) - S_{t-1}(\mathbf{x}_{t-1}) \right) \\ &= \frac{1}{w} \sum_{t=1}^{T} \left(f_t(\mathbf{x}_t) - f_{t-w}(\mathbf{x}_t) \right) + \sum_{t=2}^{T} \left(S_{t-1}(\mathbf{x}_t) - S_{t-1}(\mathbf{x}_{t-1}) \right) \\ &= \frac{1}{w} \sum_{t=T-w+1}^{T} f_t(\mathbf{x}_t) + \sum_{t=2}^{T} \left(S_{t-1}(\mathbf{x}_t) - S_{t-1}(\mathbf{x}_{t-1}) \right) \\ &\leq M + \sum_{t=2}^{T} \left(S_{t-1}(\mathbf{x}_t) - S_{t-1}(\mathbf{x}_{t-1}) \right), \end{aligned}$$

where the last inequality follows from our blanket assumptions. Consequently, by Lemma D.1, we have that

$$S_T(\mathbf{x}_T) + g(\mathbf{x}_T) - g(\mathbf{x}_1) \leq M + \sum_{t=2}^{T} \left(S_{t-1}(\mathbf{x}_t) + g(\mathbf{x}_t) - S_{t-1}(\mathbf{x}_{t-1}) - g(\mathbf{x}_{t-1}) \right)$$

$$\leq M - \sum_{t=1}^{T-1} \tau_t \left(\eta - \frac{\eta^2 L}{2}\right) \frac{\delta^2}{w^2}$$
$$\leq M - \tau \left(\eta - \frac{\eta^2 L}{2}\right) \frac{\delta^2}{w^2},$$

where the last inequality uses $\tau = \sum_{t=1}^{T-1} \tau_t$. On the other hand, by our blanket assumptions,

$$S_T(\mathbf{x}_T) = \frac{1}{w} \sum_{i=T-w+1}^{T} f_i(\mathbf{x}_i) \geq -M.$$

By combining both sides we obtain that

$$-M \leq g(\mathbf{x}_1) - g(\mathbf{x}_T) + M - \tau \left(\eta - \frac{\eta^2 L}{2}\right) \frac{\delta^2}{w^2},$$

and the desired immediately follows from the nonnegativity of g:

$$\tau \leq \frac{g(\mathbf{x}_1) - g(\mathbf{x}_T) + 2M}{\left(\eta - \frac{\eta^2 L}{2}\right) \frac{\delta^2}{w^2}} \leq \frac{2w^2 (g(\mathbf{x}_1) + 2M)}{(2 - \eta L)\, \eta \delta^2}.$$ □

We conclude with the implication of our guarantees to the stochastic offline setting.

Proof of Corollary 3.1. From the choice of t_*, Jensen's inequality, and Theorem 3.1, we have that

$$\begin{aligned}
\mathbb{E}_{t_*} \left(\|\nabla f(\mathbf{x}_{t_*})\|^2 \right) &= \frac{1}{T-w} \sum_{t=w}^{T} \|\mathbb{E}\left(\nabla f_t(\mathbf{x}_t)\right)\|^2 \\
&= \frac{1}{T-w} \sum_{t=w}^{T} \left\| \mathbb{E}\left(\frac{1}{w} \sum_{i=t-w+1}^{t} \nabla f_i(\mathbf{x}_t) \right) \right\|^2 \\
&\leq \frac{1}{T-w} \sum_{t=w}^{T} \mathbb{E}\left(\left\| \frac{1}{w} \sum_{i=t-w+1}^{t} \nabla f_i(\mathbf{x}_t) \right\|^2 \right) \\
&\leq \frac{1}{T-w} \mathbb{E}\left(\mathrm{Reg}_w(T)\right) \\
&\leq \frac{2}{(T-w)w^2} \left(T\delta^2 + V_w[T] \right).
\end{aligned}$$

Plugging the parameters' values $T = 2w$, $w = \lceil \sqrt{\frac{2(\delta^2 + c)}{\varepsilon}} \rceil$, and $V_w[T] = cT$, we immediately obtain that

$$\mathbb{E}\left(\|\nabla f(\mathbf{x}_{t_*})\|^2 \right) \leq \frac{2}{(T-w)w^2} \left(\delta^2 T + V_w[T] \right) \leq \frac{4}{w^2} \left(\delta^2 + c \right) \leq \varepsilon.$$

Once again, by plugging the parameters' values we obtain from (7) in Theorem 3.2 that

$$\tau \leq \frac{2w^2 (g(\mathbf{x}_1) + 2M)}{(2 - \eta L)\, \eta \delta^2} \propto O(\varepsilon^{-1}).$$

Since for each prox-grad update the algorithm computes w gradient samples (for each function sampled in the time-window), the SFO complexity is

$$\tau w \propto O(\varepsilon^{-3/2}).$$ □

E. Proofs of Section 4

Before proceeding to the stochastic analysis, we make some notational conventions for the sake of readability: $S_t \equiv S_{t,w}$, $T(\mathbf{x};\mathbf{d}) \equiv T_\eta^{f,g}(\mathbf{x};\mathbf{d})$, and $\mathcal{P}(\mathbf{x};\mathbf{d}) \equiv \mathcal{P}_\eta^g(\mathbf{x};\mathbf{d})$. Additionally, we set $\mathbf{y}_t^k = \mathbf{y}_t^{\tau_t}$ for all $k \geq \tau_t$; this means that $\mathbf{y}_t^k = \mathbf{y}_t^{k+1}$ if and only if $k \geq \tau_t$.

The forthcoming analysis of Algorithm 2 requires delicate treatment of what is known, and what is not, at specific moments during the run. To avoid confusion, we state explicitly what is included in the algorithm's natural filtration at time $t \geq 1$ and at each inner iteration $k \geq 1$, thus extending on our original description.

Definition E.1 (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 $\tilde{\nabla} f_t(\mathbf{x}_t)$.*

For all $t \geq 1$ and all $k \geq 1$, the filtration $\mathcal{F}_{t,k}$ includes all gradient feedback up to, but not including, the execution of the k-th iteration of step 5(b) at time t. In particular, it contains $\mathcal{F}_t$, and includes $\mathbf{y}_t^k, G_t^k$, and $\mathbf{y}_t^{k+1}$, but it does not include $\{\tilde{\nabla} f_i(\mathbf{y}_t^{k+1})\}_{i=t-w}^t$, G_t^{k+1}.

We will utilize two trivial technical corollaries of Definition 4.1 given next.

Corollary E.1. *Let $\mathbf{x} \in \mathbb{R}^n$, then*

$$\mathbb{E}\left(\|\mathcal{S}_\sigma(\mathbf{x};\omega,h) - \nabla h(\mathbf{x})\|\right)^2 \leq \mathbb{E}\left(\|\mathcal{S}_\sigma(\mathbf{x};\omega,h) - \nabla h(\mathbf{x})\|^2\right) \leq \sigma^2. \tag{13}$$

Lemma E.1. *Let $\mathbf{x} \in \mathbb{R}^n$ and $h_i : \mathbb{R}^n \to \mathbb{R}$ for any $i = 1, 2, \ldots, w$. Then*

$$\mathbb{E}\left(\left\|\frac{1}{w}\sum_{i=1}^w \mathcal{S}_\sigma(\mathbf{x};\omega,h_i) - \frac{1}{w}\sum_{i=1}^w \nabla h_i(\mathbf{x})\right\|^2\right) \leq \sigma^2.$$

Proof. Follows from Jensen's inequality. □

The following technical lemma is of key importance in the analysis ahead.

Lemma E.2. *Let $t \in [T]$ and $k \geq 2$. It holds that*

$$\mathbb{E}\left(\langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \mathbf{y}_t^{k+1} - \mathbf{y}_t^k\rangle | \mathcal{F}_{t,k-1}\right) \geq -\frac{\eta\sigma^2}{w^2}.$$

Proof. Define the full gradient prox-grad by $\hat{\mathbf{y}}_t^k = T_\eta^g(\mathbf{y}_t^k; \nabla S_{t,w}(\mathbf{y}_t^k))$, and note that

$$\begin{aligned}\langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \mathbf{y}_t^{k+1} - \mathbf{y}_t^k\rangle &= \langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \mathbf{y}_t^{k+1} - \hat{\mathbf{y}}_t^k\rangle + \langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \hat{\mathbf{y}}_t^k - \mathbf{y}_t^k\rangle \\ &\geq -\|G_t^k - \nabla S_t(\mathbf{y}_t^k)\|\|\mathbf{y}_t^{k+1} - \hat{\mathbf{y}}_t^k\| + \langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \hat{\mathbf{y}}_t^k - \mathbf{y}_t^k\rangle,\end{aligned} \tag{14}$$

where the last inequality follows from Cauchy-Schwartz inequality. By the nonexpansivity of the prox operator (Beck, 2017, Theorem 6.42) we have that

$$\|\mathbf{y}_t^{k+1} - \hat{\mathbf{y}}_t^k\| \leq \|\mathbf{y}_t^k - \eta G_t^k - \mathbf{y}_t^k + \eta\nabla S_t(\mathbf{y}_t^k)\| = \eta\|G_t^k - \nabla S_t(\mathbf{y}_t^k)\|,$$

meaning that

$$-\|G_t^k - \nabla S_t(\mathbf{y}_t^k)\|\|\mathbf{y}_t^{k+1} - \hat{\mathbf{y}}_t^k\| \geq -\eta\|G_t^k - \nabla S_t(\mathbf{y}_t^k)\|^2. \tag{15}$$

Plugging (15) to (14) then implies that

$$\langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \mathbf{y}_t^{k+1} - \mathbf{y}_t^k\rangle \geq -\eta\|G_t^k - \nabla S_t(\mathbf{y}_t^k)\|^2 + \langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \hat{\mathbf{y}}_t^k - \mathbf{y}_t^k\rangle. \tag{16}$$

Noting that by Definition 4.1

$$\mathbb{E}\left(\langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \hat{\mathbf{y}}_t^k - \mathbf{y}_t^k\rangle | \mathcal{F}_{t,k-1}\right) = 0,$$

we obtain, from taking expectation on (16) and using Lemma E.1, that

$$\mathbb{E}\left(\langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \mathbf{y}_t^{k+1} - \mathbf{y}_t^k\rangle | \mathcal{F}_{t,k-1}\right) \geq -\frac{\eta\sigma^2}{w^2}. \qquad \square$$

We can now embark on proving our claims stated in Section 4.

Proof of Theorem 4.1. Recall that $\mathbf{y}_t^1 = \mathbf{x}_t, \mathbf{y}_t^{\tau_t} = \mathbf{x}_{t+1}$, and

$$\mathbf{y}_t^{k+1} = \arg\min_{\mathbf{z}\in\mathbb{R}^n} g(\mathbf{z}) + \langle G_t^k, \mathbf{z} - \mathbf{y}_t^k\rangle + \frac{1}{2\eta}\|\mathbf{z} - \mathbf{y}_t^k\|^2, \qquad k \in [\tau_t - 1].$$

Denote $h_t^k := S_t(\mathbf{y}_t^k) + g(\mathbf{y}_t^k)$. By combining the descent lemma (cf. Lemma C.1), the definition of $\mathbf{y}_t^{k+1}$, 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),

$$\begin{aligned} h_t^k - h_t^{k+1} &\geq \langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \mathbf{y}_t^{k+1} - \mathbf{y}_t^k\rangle + \frac{1}{2}\left(\eta - \eta^2 L\right)\left\|\mathcal{P}(\mathbf{y}_t^k; G_t^k)\right\|^2 \\ &\geq \langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \mathbf{y}_t^{k+1} - \mathbf{y}_t^k\rangle + \frac{1}{2}\left(\eta - \eta^2 L\right)\frac{\delta^2}{w^2}. \end{aligned}$$

Applying expectation to the latter, using the law of total expection (tower rule), and invoking Lemma E.2 and relation (6), we obtain that for any $k \in [\tau_t - 1]$ it holds that

$$\begin{aligned} \mathbb{E}\left(h_t^k - h_t^{k+1}\right) &\geq \mathbb{E}\left(\langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \mathbf{y}_t^{k+1} - \mathbf{y}_t^k\rangle\right) + \frac{1}{2}\left(\eta - \eta^2 L\right)\frac{\delta^2}{w^2} \\ &\geq \frac{2}{w^2}\left(\eta\left(1 - \eta L\right)\delta^2 - 2\sigma^2\right) > 0. \end{aligned}$$

Set $\alpha := 2\left(\eta\left(1 - \eta L\right)\delta^2 - 2\sigma^2\right)/w^2 > 0$. From the former, by using the law of total expectation, for any $K \geq 1$ we have that

$$\begin{aligned} h_t^1 + M \geq \mathbb{E}\left(h_t^1 - h_t^{K+1}\right) &= \mathbb{E}\left(\sum_{k=1}^{K}(h_t^k - h_t^{k+1})\right) \\ &= \sum_{k=1}^{K}\mathbb{E}\left(h_t^k - h_t^{k+1}\right) \\ &= \sum_{k=1}^{K}\left(\mathbb{E}\left(h_t^k - h_t^{k+1} | \tau_t \geq k+1\right)\mathbb{P}(\tau_t \geq k+1) + 0\cdot\mathbb{P}(\tau_t \leq k)\right) \\ &\geq \alpha\sum_{k=1}^{K}\mathbb{P}(\tau_t > k) \\ &\geq \alpha\sum_{k=1}^{K}\mathbb{P}(\tau_t > K) = \alpha K\mathbb{P}(\tau_t > K). \end{aligned}$$

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. $\square$

Let us now establish the local regret bound stated in Theorem 4.2.

Proof of Theorem 4.2. Recall that

$$\mathrm{Reg}_w(T) = \sum_{t=1}^{T} \|\mathcal{P}(\mathbf{x}_t; \nabla S_t(\mathbf{x}_t))\|^2 = \sum_{t=1}^{T} \frac{1}{\eta^2} \|\mathbf{x}_t - T(\mathbf{x}_t; \nabla S_t(\mathbf{x}_t))\|^2 . \tag{17}$$

By simple algebra,

$$\|\mathbf{x}_t - T(\mathbf{x}_t; \nabla S_t(\mathbf{x}_t))\|^2 \leq 2 \left\|\mathbf{x}_t - T(\mathbf{x}_t; \tilde{\nabla} S_t(\mathbf{x}_t))\right\|^2 + 2 \left\|T(\mathbf{x}_t; \tilde{\nabla} S_t(\mathbf{x}_t)) - T(\mathbf{x}_t; \nabla S_t(\mathbf{x}_t))\right\|^2 . \tag{18}$$

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

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

Subsequently, using the law of total expectation and Lemma E.1, we obtain the relation

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

Then, plugging the latter to the expected value of (18) yields

$$\mathbb{E}\left(\|\mathbf{x}_t - T(\mathbf{x}_t; \nabla S_t(\mathbf{x}_t))\|^2\right) \leq 2\eta^2 \mathbb{E}\left(\left\|\mathcal{P}(\mathbf{x}_t; \tilde{\nabla} S_t(\mathbf{x}_t))\right\|^2\right) + \frac{2\eta^2 \sigma^2}{w^2}.$$

Thus,

$$\mathbb{E}\left(\mathrm{Reg}_w(T)\right) \leq 2 \sum_{t=1}^{T} \left[\mathbb{E}\left(\left\|\mathcal{P}(\mathbf{x}_t; \tilde{\nabla} S_{t,w}(\mathbf{x}_t))\right\|^2\right) + \frac{\sigma^2}{w^2}\right]. \tag{19}$$

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 yields

$$\begin{aligned}\left\|\mathcal{P}(\mathbf{x}_t; \tilde{\nabla} S_t(\mathbf{x}_t))\right\| = \|\mathcal{P}(\mathbf{x}_t; G_1 + G_2)\| &\leq \left\|\mathcal{P}(\mathbf{x}_t; \tilde{\nabla} S_{t-1}(\mathbf{x}_t))\right\| + \frac{1}{w} \left\|\tilde{\nabla} f_t(\mathbf{x}_t) - \tilde{\nabla} f_{t-w}(\mathbf{x}_t)\right\| \\ &\leq \frac{\delta}{w} + \frac{1}{w} \left\|\tilde{\nabla} f_t(\mathbf{x}_t) - \tilde{\nabla} f_{t-w}(\mathbf{x}_t)\right\|, \end{aligned} \tag{20}$$

where the last inequality follows from the termination rule of the inner loop. Therefore,

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

Using the triangle inequality and the relation $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, yields that

$$\begin{aligned}&\left\|\tilde{\nabla} f_t(\mathbf{x}_t) - \tilde{\nabla} f_{t-w}(\mathbf{x}_t)\right\|^2 \leq \\ &3 \left\|\tilde{\nabla} f_t(\mathbf{x}_t) - \nabla f_t(\mathbf{x}_t)\right\|^2 + 3 \left\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-w}(\mathbf{x}_t)\right\|^2 + 3 \left\|\nabla f_{t-w}(\mathbf{x}_t) - \tilde{\nabla} f_{t-w}(\mathbf{x}_t)\right\|^2 .\end{aligned}$$

Applying expectation, from the law of total expectation together with Definition 4.1, we obtain that

$$\begin{aligned}\mathbb{E}\left[\left\|\tilde{\nabla} f_t(\mathbf{x}_t) - \nabla f_t(\mathbf{x}_t)\right\|^2\right] &= \mathbb{E}\left[\mathbb{E}\left(\left\|\tilde{\nabla} f_t(\mathbf{x}_t) - \nabla f_t(\mathbf{x}_t)\right\|^2 |\mathcal{F}_t\right)\right] &\leq \frac{\sigma^2}{w^2}, \\ \mathbb{E}\left[\left\|\nabla f_{t-w}(\mathbf{x}_t) - \tilde{\nabla} f_{t-w}(\mathbf{x}_t)\right\|^2\right] &= \mathbb{E}\left[\mathbb{E}\left(\left\|\nabla f_{t-w}(\mathbf{x}_t) - \tilde{\nabla} f_{t-w}(\mathbf{x}_t)\right\|^2 |\mathcal{F}_{t-w}, \mathbf{x}_t\right)\right] &\leq \frac{\sigma^2}{w^2}.\end{aligned}$$

Thus, $\mathbb{E}\left(\left\|\tilde{\nabla} f_t(\mathbf{x}_t) - \tilde{\nabla} f_{t-w}(\mathbf{x}_t)\right\|^2\right) \leq \frac{6\sigma^2}{w^2} + 3\mathbb{E}\left(\left\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-w}(\mathbf{x}_t)\right\|^2\right)$, and consequently

$$\begin{aligned}\mathbb{E}\left(\left\|\mathcal{P}(\mathbf{x}_t; \tilde{\nabla} S_t(\mathbf{x}_t))\right\|^2\right) &\leq \frac{2}{w^2}\left(\delta^2 + \mathbb{E}\left(\left\|\tilde{\nabla} f_t(\mathbf{x}_t) - \tilde{\nabla} f_{t-w}(\mathbf{x}_t)\right\|^2\right)\right) \\ &\leq \frac{2}{w^2}\left(\delta^2 + \frac{6\sigma^2}{w^2} + 3\mathbb{E}\left(\left\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-w}(\mathbf{x}_t)\right\|^2\right)\right).\end{aligned}$$

Summing over $t \in [T]$ and plugging $V_w[T]$ defined in (5) 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 (19), and recalling that $w \geq 1$, results with the desired bound. □

Finally, we prove the bound on the number of SFO calls, as stated by Theorem 4.3.

Proof of Theorem 4.3. Denote $h_t^k := S_t(\mathbf{y}_t^k) + g(\mathbf{y}_t^k)$. By combining the descent lemma (cf. Lemma C.1), the definition of the sequence $\{\mathbf{y}_t^k\}_{k\geq 1}$, Young's inequality, and the stopping criteria of the inner loop, we have that for any $K \geq 1$ (assuming that $\mathcal{F}_t$ is given)

$$\begin{aligned}h_t^1 - h_t^{K+1} = \sum_{k=1}^{K}(h_t^k - h_t^{k+1}) &\geq \sum_{k=1}^{\min\{K,\tau_t\}}\left(\langle G_t^k - \nabla S_t(\mathbf{y}_t^k), \mathbf{y}_t^{k+1} - \mathbf{y}_t^k\rangle + \frac{1-\eta L}{2\eta}\left\|\mathbf{y}_t^{k+1} - \mathbf{y}_t^k\right\|^2\right) \\ &\geq \frac{1}{2}\sum_{k=1}^{\min\{K,\tau_t\}}\left(-\left\|G_t^k - \nabla S_t(\mathbf{y}_t^k)\right\|^2 - \left\|\mathbf{y}_t^{k+1} - \mathbf{y}_t^k\right\|^2 + \frac{1-\eta L}{\eta}\left\|\mathbf{y}_t^{k+1} - \mathbf{y}_t^k\right\|^2\right).\end{aligned}$$

Hence, by Assumption 1 and the stopping condition of the inner loop, we obtain

$$h_t^1 - h_t^{K+1} \geq \frac{1}{2w^2}\sum_{k=1}^{\min\{K,\tau_t\}}\left(-\sigma^2 + (1-\eta(L+1))\eta\delta^2\right) = \frac{(1-\eta(L+1))\eta\delta^2 - \sigma^2}{2w^2}\min\{K,\tau_t\} > 0.$$

Recall that $S_{0,w}(\mathbf{x}_0) \equiv 0$, and $S_t(\mathbf{x}) = \frac{1}{w}(f_t(\mathbf{x}) - f_{t-w}(\mathbf{x})) + S_{t-1}(\mathbf{x})$. Using the previous derivations for $t-1$ (setting $K = \tau_{t-1}$ and noting that $h_{t-1}^{\tau_{t-1}+1} = h_{t-1}^{\tau_{t-1}}$), we have that

$$S_{t-1}(\mathbf{x}_t) + g(\mathbf{x}_t) - S_{t-1}(\mathbf{x}_{t-1}) - g(\mathbf{x}_{t-1}) = h_{t-1}^{\tau_{t-1}} - h_{t-1}^1 \leq -\tau_{t-1}\frac{(1-\eta(L+1))\eta\delta^2 - \sigma^2}{2w^2}. \tag{21}$$

Thus, since

$$\begin{aligned}S_T(\mathbf{x}_T) = \sum_{t=1}^{T}(S_t(\mathbf{x}_t) - S_{t-1}(\mathbf{x}_{t-1})) &= \sum_{t=1}^{T}\left(\frac{1}{w}(f_t(\mathbf{x}_t) - f_{t-w}(\mathbf{x}_t)) + S_{t-1}(\mathbf{x}_t) - S_{t-1}(\mathbf{x}_{t-1})\right) \\ &= \frac{1}{w}\sum_{t=T-w+1}^{T} f_t(\mathbf{x}_t) + \sum_{t=2}^{T}\left(S_{t-1}(\mathbf{x}_t) - S_{t-1}(\mathbf{x}_{t-1})\right),\end{aligned}$$

we have from our blanket assumptions and relation (21), that

$$S_T(\mathbf{x}_T) \leq g(\mathbf{x}_1) - g(\mathbf{x}_T) + M - \tau\frac{(1-\eta(L+1))\eta\delta^2 - \sigma^2}{2w^2}.$$

On the other hand, again by our blanket assumptions, $S_T(\mathbf{x}_T) = \frac{1}{w}\sum_{i=T-w+1}^{T} f_i(\mathbf{x}_i) \geq -M$. By combining both sides, we obtain that

$$-M \leq g(\mathbf{x}_1) - g(\mathbf{x}_T) + M - \tau\frac{(1-\eta(L+1))\eta\delta^2 - \sigma^2}{2w^2},$$

and the bound on τ immediately follows due to the nonnegativity of g. Finally, the desired bound on the SFO oracle calls follows from the fact that the inner loop makes $O(w)$ SFO calls per loop. □

E.1. **Implications to Offline Stochastic Optimization.** Next we establish our derivations in the offline scenario described in Section 4.2.

Proof of Theorem 4.4. Note that $f_t \equiv f$ for any $t \in [T]$ implies that $\nabla S_{t,w}(\mathbf{x}) \equiv \nabla f(\mathbf{x})$. From Theorem 4.2 and the choice of t_* we have that

$$\begin{aligned}\mathbb{E}\left(\left\|\mathcal{P}(\mathbf{x}_{t_*}; \nabla f(\mathbf{x}_{t_*}))\right\|^2\right) &= \frac{1}{T-w}\mathbb{E}\left(\sum_{t=w}^{T}\left\|\mathcal{P}(\mathbf{x}_t; \nabla f(\mathbf{x}_t))\right\|^2\right)\\ &\leq \frac{1}{T-w}\mathbb{E}\left(\mathrm{Reg}_w(T)\right)\\ &\leq \frac{2}{(T-w)w^2}\left(\left(\delta^2+7\sigma^2\right)T + 6V_w[T]\right).\end{aligned}$$

□

Proof of Corollary 4.1. From Theorem 4.4 we immediately obtain that

$$\frac{2}{(T-w)w^2}\left(\left(\delta^2+7\sigma^2\right)T + 6V_w[T]\right) = \frac{4w}{w^3}\left(\delta^2+7\sigma^2+c\right) \leq \varepsilon.$$

The bound $O(M\sigma\varepsilon^{-3/2})$ is obtained by plugging the assumed values of w, T, and δ^2, to (7) in Theorem 4.3:

$$w\tau \leq \frac{2\eta w^3(g(\mathbf{x}_1)+3M)}{(1-\eta(L+1))\delta^2 - \eta\sigma^2} = \frac{2w^3(g(\mathbf{x}_1)+3M)}{\sigma^2} \propto O(M\sigma\varepsilon^{-3/2}),$$

where we used the fact that w is $O(\sigma/\sqrt{\varepsilon})$. □

References

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

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

Naman Agarwal, Alon Gonen, and Elad Hazan. Learning in non-convex games with an optimization oracle. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 18–29, Phoenix, USA, 25–28 Jun 2019. PMLR. URL http://proceedings.mlr.press/v99/agarwal19a.html.

Yossi Arjevani, Yair Carmon, John C Duchi, Dylan J Foster, Nathan Srebro, and Blake Woodworth. Lower bounds for non-convex stochastic optimization. *arXiv preprint arXiv:1912.02365*, 2019.

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

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

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

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

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

Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. $\mathcal{X}$-armed bandits. *Journal of Machine Learning Research*, 12:1655–1695, 2011.

Nicolò Cesa-Bianchi, Pierre Gaillard, Gábor Lugosi, and Gilles Stoltz. Mirror descent meets fixed share (and feels no regret). In 989-997, editor, *Advances in Neural Information Processing Systems*, volume 25, 2012.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In Shie Mannor, Nathan Srebro, and Robert C. Williamson, editors, *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pages 6.1–6.20, Edinburgh, Scotland, 25–27 Jun 2012. PMLR. URL http://proceedings.mlr.press/v23/chiang12.html.

Cong Fang, Chris J Li, Zhouchen Lin, and Tong Zhang. Spider: Near-optimal non-convex optimization via stochastic path-integrated differential estimator. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 689–699. Curran Associates, Inc., 2018.

Dan Garber. On the regret minimization of nonconvex online gradient ascent for online pca. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1349–1373, Phoenix, USA, 25–28 Jun 2019. PMLR. URL http://proceedings.mlr.press/v99/garber19a.html.

Paulina Grnarova, Kfir Y Levy, Aurelien Lucchi, Thomas Hofmann, and Andreas Krause. An online learning approach to generative adversarial networks. *arXiv preprint arXiv:1706.03269*, 2017.

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

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

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

Elad Hazan, Karan Singh, and Cyril Zhang. Efficient regret minimization in non-convex games. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1433–1441, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR. URL http://proceedings.mlr.press/v70/hazan17a.html.

Amélie Héliou, Matthieu Martin, Panayotis Mertikopoulos, and Thibaud Rahier. Online non-convex optimization with inexact models. In *NeurIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.

Prateek Jain, Dheeraj Nagaraj, and Praneeth Netrapalli. Making the last iterate of sgd information theoretically optimal. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1752–1755, Phoenix, USA, 25–28 Jun 2019. PMLR. URL http://proceedings.mlr.press/v99/jain19a.html.

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

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

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

Robert David Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-armed bandits in metric spaces. In *STOC '08: Proceedings of the 40th annual ACM symposium on the Theory of Computing*, 2008.

Walid Krichene, Maximilian Balandat, Claire Tomlin, and Alexandre Bayen. The Hedge algorithm on a continuum. In *ICML '15: Proceedings of the 32nd International Conference on Machine Learning*, 2015.

Xiaoyu Li and Francesco Orabona. On the convergence of stochastic gradient descent with adaptive stepsizes. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 983–992. PMLR, 16–18 Apr 2019. URL http://proceedings.mlr.press/v89/li19c.html.

Michael Metel and Akiko Takeda. Simple stochastic gradient methods for non-smooth non-convex regularized optimization. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 4537–4545, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL http://proceedings.mlr.press/v97/metel19a.html.

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

Maher Nouiehed, Maziar Sanjabi, Tianjian Huang, Jason D Lee, and Meisam Razaviyayn. Solving a class of non-convex min-max games using iterative first order methods. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 14934–14942. Curran Associates, Inc., 2019.

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

Steven Perkins, Panayotis Mertikopoulos, and David S. Leslie. Mixed-strategy learning with continuous action sets. 62(1):379–384, January 2017.

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

Arun Sai Suggala and Praneeth Netrapalli. 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.

Zhe Wang, Kaiyi Ji, Yi Zhou, Yingbin Liang, and Vahid Tarokh. Spiderboost and momentum: Faster variance reduction algorithms. In *Advances in Neural Information Processing Systems*, pages 2403–2413, 2019.

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

Alp Yurtsever, Suvrit Sra, and Volkan Cevher. Conditional gradient methods via stochastic path-integrated differential estimator. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 7282–7291, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL http://proceedings.mlr.press/v97/yurtsever19b.html.