Journal of Dynamics and Games
doi:10.3934/jdg.2017008

A CONTINUOUS-TIME APPROACH TO ONLINE OPTIMIZATION

Joon Kwon

Centre de mathématiques appliquées
École polytechnique
Université Paris–Saclay
Palaiseau, France

Panayotis Mertikopoulos

CNRS (French National Center for Scientific Research)
Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP
LIG, F-38000, Grenoble, France

(Communicated by Michel Benaim)

Abstract. We consider a family of mirror descent strategies for online optimization in continuous-time and we show that they lead to no regret. From a more traditional, discrete-time viewpoint, this continuous-time approach allows us to derive the no-regret properties of a large class of discrete-time algorithms including as special cases the exponential weights algorithm, online mirror descent, smooth fictitious play and vanishingly smooth fictitious play. In so doing, we obtain a unified view of many classical regret bounds, and we show that they can be decomposed into a term stemming from continuous-time considerations and a term which measures the disparity between discrete and continuous time. This generalizes the continuous-time based analysis of the exponential weights algorithm from [29]. As a result, we obtain a general class of infinite horizon learning strategies that guarantee an $\mathcal{O}(n^{-1/2})$ regret bound without having to resort to a doubling trick.

1. **Introduction.** Online optimization focuses on decision-making in sequentially changing environments (the weather, the stock market, etc.). More precisely, at each stage of a repeated decision process, the agent/decision-maker obtains a payoff (or incurs a loss) based on the state of the environment and his decision, and his long-term objective is to maximize his cumulative payoff via the use of past observations.

The worst-case scenario for the agent – and one which has attracted considerable interest in the literature – is when he has no Bayesian-like prior belief on the environment. In this context, the cumulative payoff difference between a decision rule which prescribes an action based on knowledge of the future and a *learning strategy* (a rule which only relies on past observations) can become arbitrarily large, even in very simple problems. As a result, in the absence of absolute payoff guarantees, the most widely used online optimization criterion is that of *regret minimization*, a

2010 *Mathematics Subject Classification.* Primary: 68Q32, 68T05, 91A26; Secondary: 90C25.
Key words and phrases. Online optimization, regret minimization, mirror descent, gradient descent, continuous time, convex optimization.

notion which was first introduced by [14] and has since given rise to a vigorous literature at the interface of optimization, statistics and theoretical computer science – see e.g. [10, 28] for a survey. Specifically, the *cumulative regret* of a strategy compares the payoff obtained by an agent that follows it to the payoff that he would have obtained by constantly choosing one action. Accordingly, one of the main goals in online optimization is to devise strategies that lead to (vanishingly) small average regret against any fixed action, and irrespective of how the environment evolves over time.

In this paper, we take a continuous-time approach to online optimization and we consider a class of strategies that lead to no regret in continuous time. From a more traditional, discrete-time viewpoint, the importance of this approach lies in that it provides a unifying view of the regret properties of a broad class of well-known online optimization algorithms. In particular, the discrete-time version of our family of strategies is an extension of the general class of online mirror descent (OMD) algorithms (also known as "Following the Regularized Leader" (FtRL) in the case of linear payoffs; see e.g. [28, 7, 15]) with a time-varying parameter. As such, our analysis contains as special cases *a*) the exponential weights (EW) algorithm [19, 30] and its decreasing parameter variant [1]; *b*) smooth fictitious play (SFP) [13, 4] and vanishingly smooth fictitious play (VSFP) [3]; and *c*) the method of online gradient descent (OGD) introduced by [33] (the Euclidean predecessor of OMD).

With regards to the OMD/FtRL family of algorithms, the vanishing regret bounds that we derive by using a time-varying parameter are not particularly new. Bounds of the same order can be obtained by taking existing guarantees for learning with a finite horizon and then using the so-called "doubling trick" [9, 31].[1] That said, the introduction of a time-varying parameter has several advantages: *a*) it allows us to integrate SFP and VSFP into the fold and to derive explicit bounds for their regret; *b*) it provides a unified any-time analysis without needing to reboot the algorithm every so often (to the best of our knowledge, such an analysis only exists for a few special cases like the EW algorithm with a time-varying parameter [1]; the analysis of the algorithm from [33] could also be adapted); and *c*) in the case of ordinary convex optimization problems with an open-ended termination criterion (as opposed to a fixed number of steps), a variable parameter leads to more efficient value convergence bounds than a variable step-size.

Building on an idea that was introduced by [32] in the framework of convex optimization and by [29] in the study of the exponential weights algorithm, the key ingredient of our analysis is the descent from continuous to discrete time. More precisely, given an online optimization problem in discrete time, we construct a continuous-time interpolation where our continuous-time dynamics lead to no regret. Then, by comparing the agent's payoffs in discrete and continuous time, we are able to deduce a bound for the agent's regret in the original discrete-time framework.

One of the main contributions of this approach is that it leads to a unified derivation of several existing regret bounds with disparate proofs. Additionally, it allows us to decompose many classical bounds into two components, a term coming from continuous-time considerations and a comparison term which measures the disparity between discrete and continuous time (see also [20] for an alternative

[1] In a nutshell, the doubling trick amounts to breaking up the learning timeline in blocks of exponentially increasing horizon, and then resetting the algorithm at the start of each block with an optimal parameter for the block's (finite) horizon.

interpretation of such a decomposition). Each of these terms can be made arbitrarily small by itself, but their sum is coupled in a nontrivial way that induces a trade-off between continuous- and discrete-time considerations: in a sense, faster decay rates in continuous time lead to greater discrepancies in the discrete/continuous comparison – and hence, to slower regret decay bounds in discrete time.

Finally, we also give a brief account of how the derived regret bounds are related to classical convergence results for certain convex optimization and stochastic convex optimization algorithms – including the projected subgradient (PSG) method, mirror descent (MD), and their stochastic variants [23, 22], and we illustrate a (somewhat surprising) performance gap incurred by using an optimization algorithm with a decreasing parameter instead of a decreasing step-size.

1.1. **Paper outline.** In Section 2, we present some basics of online optimization to fix notation and terminology; subsequently, in Section 3, we define regularizer functions, choice maps and the class of variable-parameter OMD/FtRL strategies that we will focus on. The core of our paper consists of Sections 4 and 5: we first show that the corresponding class of continuous-time strategies leads to no regret in Section 4; this analysis is then translated to discrete time in Section 5 where we derive the no-regret properties of the class of algorithms under consideration. Finally, in Section 6, we review a number of existing online learning and convex optimization algorithms, and show how their properties can be derived as corollaries of our analysis.

1.2. **Notation and preliminaries.** Let d be a positive integer and let $V = \mathbb{R}^d$ be equipped with an arbitrary norm $\|\cdot\|$. The dual of V is denoted by V^* and the induced dual norm on V^* is given by the familiar expression:

$$\|y\|_* = \sup_{\|x\| \leqslant 1} |\langle y|x\rangle|\,, \tag{1}$$

where $\langle y|x\rangle$ denotes the canonical pairing between $y \in V^*$ and $x \in V$. For a nonempty subset $U \subset V$ we use the notation $\|U\| = \sup_{x\in U} \|x\|$.

In the rest of our paper, $\mathcal{C}$ denotes a nonempty compact convex subset of V. Moreover, given a convex function $f\colon V \to \mathbb{R}\cup\{+\infty\}$, its *effective domain* is defined as $\operatorname{dom} f = \{x \in V : f(x) < \infty\}$. For convenience, if $f\colon \mathcal{C} \to \mathbb{R}$ is convex, we will treat f as a convex function on V by setting $f(x) = +\infty$ for $x \in V \setminus \mathcal{C}$; conversely, if $f\colon V \to \mathbb{R}\cup\{+\infty\}$ has domain $\operatorname{dom} f = \mathcal{C}$, we will also treat f as a real-valued function on $\mathcal{C}$ (in all cases, the ambient space V will be clear from the context). We then say that $v \in V^*$ is *a subgradient of f at $x \in \mathcal{C}$* if $f(x') - f(x) \geqslant \langle v|x' - x\rangle$ for all $x' \in \mathcal{C}$. Likewise, the set $\partial f(x) = \{v \in V^* : v \text{ is a subgradient of } f \text{ at } x\}$ is called the *subdifferential of f at x* and f is called *subdifferentiable* if $\partial f(x)$ is nonempty for all $x \in \operatorname{dom} f$.

If $\mathcal{A} = \{a_1, \dots, a_d\}$ is a finite set, the set $\Delta(\mathcal{A})$ of probability measures on $\mathcal{A}$ will be identified with the standard $(d-1)$-dimensional simplex $\Delta_d = \{x \in \mathbb{R}^d_+ : \sum_{i=1}^d x_i = 1\}$ of $\mathbb{R}^d$. Also, the elements of $\mathcal{A}$ will be identified with the corresponding vertices of $\Delta(\mathcal{A})$, i.e. the canonical basis vectors $\{e_i\}_{i=1}^d$ of $\mathbb{R}^d$. Finally, for $x, y \in \mathbb{R}$, we let $\lfloor x \rfloor = \max\{k \in \mathbb{Z} : k \leqslant x\}$ and $\lceil x \rceil = \min\{k \in \mathbb{Z} : k \geqslant x\}$, and we write $x \vee y = \max\{x, y\}$ and $x \wedge y = \min\{x, y\}$.

2. The model.

Our basic model is as follows: at every discrete time instance $n \geqslant 1$, an agent (decision-maker) chooses an action from a nonempty convex action set $\mathcal{C} \subset V$ and gains a payoff (or incurs a loss) determined by some time-dependent

function. Information about this function is only revealed to the agent after he picks his action, and the agent's objective is to maximize his long-term payoff in an adaptive manner.

2.1. **The core model.** Let $\mathcal{C} \subset V$ denote the agent's action space. Then, at each stage $n \geqslant 1$, the process of play is as follows:

1. The agent chooses an action $x_n \in \mathcal{C}$.
2. Nature chooses and reveals the *payoff vector* $v_n \in V^*$ of the n-th stage and the agent receives a payoff of $\langle v_n | x_n \rangle$.[2]
3. The agent uses some decision rule to pick a new action $x_{n+1} \in \mathcal{C}$ and the process is repeated ad infinitum.

More precisely, define a *strategy* to be a sequence of maps $\sigma_n \colon (V^*)^{n-1} \to \mathcal{C}$, $n \geqslant 1$, such that σ_{n+1} determines the player's action at stage $n+1$ in terms of the payoff vectors $v_1, \dots, v_n \in V^*$ that have been revealed up to stage n (in a slight abuse of notation, σ_1 will be regarded as an element of $\mathcal{C}$). Then, given a sequence of payoff vectors $u = (v_n)_{n \geqslant 1}$ in V^*, the *sequence of actions generated by* σ will be

$$x_{n+1} \equiv \sigma_{n+1}(v_1, \dots, v_n), \tag{2}$$

and the agent's *cumulative regret* with respect to $x \in \mathcal{C}$ is defined as:

$$\begin{aligned} \operatorname{Reg}_n^{\sigma,v}(x) &= \sum_{k=1}^{n} \langle v_k | x \rangle - \sum_{k=1}^{n} \langle v_k | x_k \rangle \\ &= \sum_{k=1}^{n} \langle v_k | x \rangle - \sum_{k=1}^{n} \langle v_k | \sigma_k(v_1, \dots, v_{k-1}) \rangle . \end{aligned} \tag{3}$$

In what follows, we focus on strategies that lead to *no* (or, at worst, *small*) *regret*:

Definition 2.1. A strategy σ *leads to* ε*-regret* ($\varepsilon \geqslant 0$) if, for every sequence of payoff vectors $(v_n)_{n \geqslant 1}$ in V^* such that $\|v_n\|_* \leqslant 1$:

$$\limsup_{n \to \infty} \frac{1}{n} \max_{x \in \mathcal{C}} \operatorname{Reg}_n^{\sigma,v}(x) \leqslant \varepsilon. \tag{4}$$

In particular, if (4) holds with $\varepsilon = 0$, we say that σ *leads to no regret*.

Remark 1. The definition of an ε-regret strategy depends on the dual norm $\|\cdot\|_*$ of V^* (and hence, on the original norm $\|\cdot\|$ on V); on the other hand, the definition of "no regret" is independent of the norm.

Remark 2. In our framework, a strategy leading to ε-regret against "any sequence" is equivalent to leading to ε-regret against "any strategy of nature". However, this may not be true in the randomized setting we present in the following paragraph.

Despite its simplicity, this online linear optimization model may be used to analyze more general online optimization models. In what follows, we summarize some examples of this kind.

[2] Nature may be adversarial, i.e. v_n may be chosen as a function of $x_1, \dots, x_n$.

2.2. **The case of the simplex and mixed actions.** Consider a discrete decision process where, at each stage $n \geqslant 1$, the agent chooses an action a_n from a finite set of *pure* actions $\mathcal{A} = \{1, \dots, d\}$. To do so, the agent draws a_n according to some probability distribution $x_n \in \Delta(\mathcal{A})$; then, once a_n is drawn, the payoff vector $v_n \in [-1, 1]^d$ which prescribes the payoff $v_{n,a}$ of each action $a \in \mathcal{A}$ is revealed and the agent receives the payoff v_{n,a_n} that corresponds to his choice of action. Moreover, we assume that Nature's choice of payoff vector v_n does not depend on pure action a_n.

In this setting, a strategy is still defined as in the core model of Section 2.1 with the agent's action set replaced by the set of *mixed actions* $\Delta(\mathcal{A})$.[3] The agent's *realized* regret with respect to a pure action $a \in \mathcal{A}$ will then be

$$\sum_{k=1}^{n} (v_{k,a} - v_{k,a_k}), \tag{5}$$

and we will say that a strategy σ leads to *ε-realized-regret* (resp. to *no realized regret* for $\varepsilon = 0$) if

$$\limsup_{n\to\infty} \frac{1}{n} \max_{a\in\mathcal{A}} \sum_{k=1}^{n} (v_{k,a} - v_{k,a_k}) \leqslant \varepsilon \quad \text{(a.s.)}, \tag{6}$$

for every strategy of Nature choosing payoff vectors $(v_n)_{n\geqslant 1}$ in $\mathbb{R}^d$ such that $\|v_n\|_\infty \leqslant 1$.[4] Besides, consider the filtration $(\mathcal{F}_n)_{n\geqslant 1}$ where $\mathcal{F}_n$ is generated by

$$(x_1, v_1, i_1, \dots, x_{n-1}, v_{n-1}, i_{n-1}, x_n, v_n).$$

Then, then conditional expectation $\mathbb{E}\left[v_{n,a_n} \,|\, \mathcal{F}_n\right]$ is equal to $\langle v_n | x_n \rangle$. Using a classical argument based on Hoeffding's inequality and the Borel–Cantelli lemma, the realized regret can be shown to be close with high probability to the regret as defined in Section 2.1. The minimization of (5) is then reduced to the core model of Section 2.1:

Proposition 1 ([10], Corollary 4.3)**.** *If a strategy σ leads to ε-regret with respect to the uniform norm on V^*, it also leads to ε-realized-regret.*

2.3. **Online convex optimization.** We briefly discuss here a more general online convex optimization model where losses are determined by a sequence of convex functions. Formally, the only change from Section 2.1 is that at each stage $n \geqslant 1$, the agent incurs a loss $\ell_n(x_n)$ determined by a subdifferentiable convex *loss function* $\ell_n \colon \mathcal{C} \to \mathbb{R}$. In this nonlinear setting, the information revealed to the agent after playing includes a (negative) subgradient $v_n \in -\partial \ell_n(x_n) \subset V^*$ of ℓ_n at x_n, so the incurred cumulative regret with respect to a fixed action $x \in \mathcal{C}$ is:

$$\sum_{k=1}^{n} \ell_k(x_k) - \sum_{k=1}^{n} \ell_k(x). \tag{7}$$

As demonstrated by [18, 8], this problem can be reduced to the core model from Section 2.1 using the so-called *gradient trick*: by convexity, $\ell_k(x') - \ell_k(x) \leqslant \langle v | x' - x \rangle$

[3] In a more general setting, the choice at each stage might depend not only on the past payoff vectors, but also on the agent's realized actions $a_1, \dots, a_n$.

[4] This condition is also called external ε-consistency [13, 4].

for all $v \in \partial\ell_k(x')$ and for all $x \in \mathcal{C}$; in this way, (7) readily yields:

$$\sum_{k=1}^{n} \ell_k(x_k) - \sum_{k=1}^{n} \ell_k(x) \leqslant -\sum_{k=1}^{n} \langle v_k | x_k - x \rangle = \sum_{k=1}^{n} \langle v_k | x \rangle - \sum_{k=1}^{n} \langle v_k | x_k \rangle \tag{8}$$

where $v_k \in -\partial\ell_k(x_k)$. This last expression can obviously be interpreted as the regret incurred by an agent facing a sequence of payoff vectors $v_n \in V^*$, so a strategy which guarantees a bound on the right-hand side of (8) will guarantee the same for (7). Consequently, when the loss functions ℓ_n are uniformly Lipschitz continuous, results for the core model can be directly translated into this one.

3. Regularizer functions, choice maps and learning strategies.

3.1. Regularizer functions and choice maps.

We begin with the concept of *regularizer functions*:

Definition 3.1. A convex function $h\colon V \to \mathbb{R} \cup \{+\infty\}$ will be called a *regularizer* (or *penalty*) function on $\mathcal{C}$ if $\operatorname{dom} h = \mathcal{C}$ and $h|_{\mathcal{C}}$ is strictly convex and continuous.

For a given regularizer function on $\mathcal{C}$, we write

$$h_{\max} = \max_{x \in \mathcal{C}} h(x) \quad \text{and} \quad h_{\min} = \min_{x \in \mathcal{C}} h(x).$$

Remark 3. This definition is intimately related to the notion of a Legendre-type function (see e.g. [25, Section 26]); however, as was recently noted by [27] (and in contrast to the analysis of e.g. [3, 7, 4]), we will not require any differentiability or steepness assumptions.

A key tool in our analysis will be the *convex conjugate* $h^*\colon V^* \to \mathbb{R} \cup \{+\infty\}$ of h defined as

$$h^*(y) = \sup_{x \in V} \{\langle y | x \rangle - h(x)\}. \tag{9}$$

Since h is equal to $+\infty$ on $V \setminus \{\mathcal{C}\}$ and $h|_{\mathcal{C}}$ is continuous and strictly convex, the supremum in (9) will be attained at a *unique* point in $\mathcal{C}$. This unique maximizer then defines our choice map as follows:

Definition 3.2. The *choice map* associated to a regularizer function h on $\mathcal{C}$ will be the map $Q_h\colon V^* \to \mathcal{C}$ defined as

$$Q_h(y) = \arg\max_{x \in \mathcal{C}} \{\langle y | x \rangle - h(x)\}, \quad y \in V^*. \tag{10}$$

Example 3.3 (Entropy and logit choice)**.** In the case of the simplex ($\mathcal{C} = \Delta_d$),[5] a classical example of a choice map is generated by the entropy function

$$h(x) = \begin{cases} \sum_{i=1}^{d} x_i \log x_i & \text{if } x \in \Delta_d, \\ +\infty & \text{otherwise.} \end{cases} \tag{11}$$

A standard calculation then yields the so-called *logit choice map*:

$$Q_h(y) = \frac{1}{\sum_{j=1}^{d} e^{y_j}} \left(e^{y_1}, \dots, e^{y_d}\right). \tag{12}$$

This map is used to define the exponential weights algorithm (cf. Section 6), and its importance stems from the well known fact that it leads to the optimal regret bound for $\mathcal{C} = \Delta_d$ [10, Theorems 2.2 and 3.7].

[5] In this setting, choice maps are more commonly known as *smooth best reply maps* [12, 16, 4, 3].

Example 3.4 (Euclidean projection)**.** Another important example arises by taking the squared Euclidean distance as a regularizer function; more precisely, we define the *Euclidean regularizer* on $\mathcal{C}$ as

$$h(x) = \begin{cases} \frac{1}{2}\|x\|_2^2 & \text{if } x \in \mathcal{C}, \\ +\infty & \text{otherwise.} \end{cases} \tag{13}$$

The associated choice map $Q_h \colon \mathbb{R}^N \to \mathcal{C}$ corresponds to taking the orthogonal projection with respect to $\mathcal{C}$:

$$\begin{aligned} Q_h(y) &= \arg\max_{x\in\mathcal{C}} \left\{ \langle y|x\rangle - \tfrac{1}{2}\|x\|_2^2 \right\} \\ &= \arg\min_{x\in\mathcal{C}} \left\{ \tfrac{1}{2}\|x\|_2^2 - \langle y|x\rangle + \tfrac{1}{2}\|y\|_2^2 \right\} = \arg\min_{x\in\mathcal{C}} \|y-x\|_2^2. \end{aligned} \tag{14}$$

Example 3.5 (Bregman projections)**.** The Euclidean example above is a special case of a class of projection mappings known as *Bregman projections* [5].

Let $F : V \longrightarrow \mathbb{R} \cup \{+\infty\}$ be a proper convex function, differentiable on its domain. Let us denote $\mathcal{D} = \operatorname{dom} F$ and for $x, x' \in \mathcal{D}$, the *Bregman divergence* $D_F : \mathcal{D} \times \mathcal{D} \longrightarrow \mathbb{R}_+$ is defined as

$$D_F(x, x') = F(x) - F(x') - \langle \nabla F(x') | x - x' \rangle \,. \tag{15}$$

Hence, given a compact set $\mathcal{C} \subset \mathcal{D}$, the associated *Bregman projection* of a point $x_0 \in \mathcal{D}$ onto $\mathcal{C}$ is given by

$$\operatorname{pr}_F^{\mathcal{C}}(x_0) = \arg\min_{x\in\mathcal{C}} D_F(x, x_0). \tag{16}$$

Now assume that F^* is also differentiable on its domain which we will denote $\mathcal{D}^*$. It is easy to check that for $y \in \mathcal{D}^*$, $\nabla F^*(y) \in \mathcal{D}$ and $\nabla F(\nabla F^*(y)) = y$. Then, the process of mapping $y \in \mathcal{D}^*$ to $\nabla F^*(y)$ and then projecting to $\mathcal{C}$ can be written as a choice map in the sense of (10):

$$\begin{aligned} \operatorname{pr}_F^{\mathcal{C}} \nabla F^*(y) &= \arg\min_{x\in\mathcal{C}} \{F(x) - F(\nabla F^*(y)) - \langle \nabla F(\nabla F^*(y)) | x - \nabla F^*(y)\rangle\} \\ &= \arg\min_{x\in\mathcal{C}} \{F(x) - \langle y|x\rangle\} = \arg\max_{x\in\mathbb{R}^d} \{\langle y|x\rangle - h(x)\} = Q_h(y), \end{aligned} \tag{17}$$

where $h|_{\mathcal{C}} = F|_{\mathcal{C}}$ and $h(x) = +\infty$ for $x \in \mathbb{R}^d \setminus \{\mathcal{C}\}$.

3.2. **Strategies generated by regularizer functions.** The class of strategies that we will consider in the rest of this paper is a variable-parameter extension of the so-called online mirror descent (OMD) method—itself equivalent to the family of algorithms known as Follow the Regularized Leader (FtRL) in the case of linear payoffs (see e.g. [28, 15]).

In a nutshell, this class of strategies may be described as follows: the agent aggregates his payoffs over time into a score vector $y \in V^*$ and then uses a choice map to turn these scores into actions and continue playing. Formally, if h is a regularizer function on the agent's action space $\mathcal{C}$ and $(\eta_n)_{n\geqslant 1}$ is a positive nonincreasing sequence, *the strategy* $\sigma \equiv \left(\sigma_n^{h,\eta_n}\right)_{n\geqslant 1}$ *generated by* h *with parameter* η_n is defined as

$$\sigma_{n+1}(v_1, \ldots, v_n) = Q_h\left(\eta_n \sum_{k=1}^{n} v_k\right), \tag{18}$$

with $\sigma_1 = Q_h(0)$. The corresponding sequence of play $x_{n+1} = \sigma_{n+1}(v_1, \dots, v_n)$ will then be given by the recursion:

$$\begin{aligned} U_n &= U_{n-1} + v_n, \\ x_{n+1} &= Q_h(\eta_n U_n). \end{aligned}$$

In addition to the standard variants of OMD/FtRL, a list of examples of strategies and algorithms that can be expressed in this general form is given in Table 1. A more detailed analysis (including the regret properties of each algorithm) will also be provided in Section 6; we only mention here that the variability of η_n will be key for the no-regret properties of σ: when η_n is constant, known techniques do not guarantee a sublinear regret bound for strategy (18) (see e.g. [28, 7]).

3.3. **Regularity of the choice map and the role of strong convexity.** In this section, we derive some regularity properties of the choice map Q_h that will be needed in the analysis of the subsequent sections. We begin by showing that Q_h is continuous and equal to the gradient of h^*:

Proposition 2. *Let h be a regularizer function on $\mathcal{C}$. Then h^* is continuously differentiable on $\mathcal{C}$ and $\nabla h^*(y) = Q_h(y)$ for all $y \in V^*$.*

Proof. For $y \in V^*$, we have

$$x \in \partial h^*(y) \iff y \in \partial h(x) \iff x \in \arg\max_{x' \in \mathcal{C}} \{\langle y | x' \rangle - h(x')\}, \tag{19}$$

i.e. $\partial h^*(y) = \arg\max_{x' \in \mathcal{C}} \{\langle y|x'\rangle - h(x')\}$. However, since the latter set only consists of $Q_h(y)$, h^* will be differentiable with $\nabla h^*(y) = Q_h(y)$ for all $y \in V^*$. The continuity of ∇h^* then follows from [25, Corollary 25.5.1]. □

In the discrete-time analysis of Section 5, (18) will be shown to guarantee a regret bound of a simple form when Q_h is Lipschitz continuous. This last requirement is equivalent to h being *strongly convex*:

Definition 3.6. Let $f \colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a convex function, let $\|\cdot\|$ be a norm on $\mathbb{R}^d$, and let $K > 0$.

(1) f is *K-strongly convex w.r.t. $\|\cdot\|$* if, for all $w_1, w_2 \in \mathbb{R}^d$ and for all $\lambda \in [0,1]$:

$$f(\lambda w_1 + (1-\lambda) w_2) \leqslant \lambda f(w_1) + (1-\lambda) f(w_2) - \tfrac{1}{2} K \lambda (1-\lambda) \|w_2 - w_1\|^2. \tag{20}$$

(2) f is *K-strongly smooth w.r.t. $\|\cdot\|$* if it is differentiable and, for all $w_1, w_2 \in \mathbb{R}^d$:

$$f(w_2) \leqslant f(w_1) + \langle \nabla f(w_1) | w_2 - w_1 \rangle + \tfrac{1}{2} K \|w_2 - w_1\|^2. \tag{21}$$

Strong convexity of a function was shown in [17] to be equivalent to strong smoothness of its conjugate. In turn, this equivalence yields the following characterization of Lipschitz continuity:

Proposition 3. *Let $f \colon V \to \mathbb{R} \cup \{+\infty\}$ be proper and lower semi-continuous. Then, for $K > 0$, the following are equivalent:*

(i) f is K-strongly convex with respect to $\|\cdot\|$.
(ii) f^* is differentiable and ∇f^* is $1/K$-Lipschitz.
(iii) f^* is $1/K$-strongly smooth with respect to $\|\cdot\|_*$.

Hence, given that regularizer functions are proper and lower semi-continuous by definition, Proposition 3 leads to the following characterization:

Corollary 1. *Let h be a regularizer function $\mathcal{C}$ and $K > 0$. The associated choice map Q_h is K-Lipschitz continuous if and only if h is K-strongly convex with respect to $\|\cdot\|$.*

This characterization of the Lipschitz continuity of ∇f^* (which will be of particular interest to us) is a classical result in the case of the Euclidean norm—see e.g. [26, Proposition 12.60]. On the other hand, the implication (ii) $\Longrightarrow$ (iii) appears to be new in the case of an arbitrary norm (though the proof technique is fairly standard).

Proof of Proposition 3. We will show that (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii) $\Longrightarrow$ (i).
(i) $\Longrightarrow$ (ii). See e.g. [2, Proposition 3.1], [24, Lemma 1] or [27, Lemma 15].

(ii) $\Longrightarrow$ (iii). Fix $y_1, y_2 \in V^*$, let $z = y_2 - y_1$, and set $\phi(t) = f^*(y_1 + tz)$, $t \in [0,1]$. Identifying V with V^{**} and $\|\cdot\|_{**}$ with $\|\cdot\|$, we have:

$$\begin{aligned}\phi'(t) - \phi'(0) &= \langle \nabla f^*(y_1 + tz) - \nabla f^*(y_1) | z \rangle \\ &\leqslant \|z\|_* \|\nabla f^*(y_1 + tz) - \nabla f^*(y_1)\| \leqslant \frac{t}{K}\|z\|_*^2, \end{aligned} \tag{22}$$

where the first inequality follows from the definition of the dual norm and the second from the assumed Lipschitz continuity of f^*. By integrating, we then get:

$$\phi(t) - \phi(0) \leqslant \phi'(0)t + \frac{1}{2K}t^2\|z\|_*^2, \tag{23}$$

and hence, for $t = 1$:

$$f^*(y_2) - f^*(y_1) \leqslant \langle \nabla f^*(y_1) | y_2 - y_1 \rangle + \frac{1}{2K}\|y_2 - y_1\|_*^2, \tag{24}$$

which shows that f^* is $1/K$-strongly smooth.

(iii) $\Longrightarrow$ (i). Since f is proper and lower semi-continuous, it will also be closed. Our assertion then follows from e.g. [17, Theorem 3]. □

We close this section by stating the strong convexity properties of the regularizer functions of Examples 3.3 and 3.4 (which thus imply the Lipschitz continuity of the corresponding choice maps):

Proposition 4. *With notation as in Examples 3.3 and 3.4, we have:*
(i) The entropy $h\colon \Delta_d \to \mathbb{R}$ of (11) is 1-strongly convex w.r.t. $\|\cdot\|_1$.
(ii) The Euclidean regularizer $h\colon \mathcal{C} \to \mathbb{R}$ of (13) is 1-strongly convex w.r.t. $\|\cdot\|_2$.

Proof. The strong convexity of the Euclidean regularizer is trivial; for the strong convexity of the entropy with respect to $\|\cdot\|_1$, see e.g. [2, Proposition 5.1]. □

4. **The continuous-time analysis.** Motivated by a technique introduced by [29] for the study of the exponential weights (EW) algorithm, we present in this section a continuous-time version of the class of strategies of Section 3 and we derive a bound for the induced regret in continuous time. This will then enable us to bound the resulting discrete-time regret by comparing the continuous- and discrete-time variants of this and the previous section respectively.

In continuous time, instead of a sequence of payoff vectors $(v_n)_{n\geqslant 1}$ in V^*, the agent will be facing a measurable and locally integrable stream of payoff vectors

$(v_t)_{t\in\mathbb{R}_+}$ in V^*. Hence, extending (18) to continuous time, we will consider the process

$$x_t^c = Q_h\left(\eta_t \int_0^t v_s\,ds\right), \tag{25}$$

where $(\eta_t)_{t\in\mathbb{R}_+}$ is a positive, nonincreasing and continuous parameter, while $x_t^c \in \mathcal{C}$ denotes the agent's action at time t given the history of payoff vectors v_s, $0 \leqslant s < t$.[6]

Our main result for (25) is as follows:

Theorem 4.1. *If h is a regularizer function on $\mathcal{C}$ and $(\eta_t)_{t\in\mathbb{R}_+}$ is a positive, nonincreasing and piecewise continuous parameter, then, for every locally integrable payoff stream $(v_t)_{t\in\mathbb{R}_+}$ in V^*, we have:*

$$\max_{x\in\mathcal{C}} \int_0^t \langle v_s|x\rangle\,ds - \int_0^t \langle v_s|x_s^c\rangle\,ds \leqslant \frac{h_{\max}-h_{\min}}{\eta_t}. \tag{26}$$

Proof. Assume first that η_t is of class C^1 and let $y_t = \eta_t \int_0^t v_s\,ds$. Then, for all $x\in\mathcal{C}$ and for all $t \geqslant 0$, Fenchel's inequality gives:

$$\int_0^t \langle v_s|x\rangle\,ds = \frac{\langle y_t|x\rangle}{\eta_t} \leqslant \frac{h^*(y_t)+h(x)}{\eta_t} \leqslant \frac{h^*(y_t)}{\eta_t} + \frac{h_{\max}}{\eta_t}. \tag{27}$$

On the other hand, with $x_t^c = Q_h(y_t)$, we also have

$$\frac{h^*(y_t)}{\eta_t} = \frac{\langle y_t|x_t^c\rangle - h(x_t^c)}{\eta_t} = \int_0^t \langle v_s|x_t^c\rangle\,ds - \frac{h(x_t^c)}{\eta_t}. \tag{28}$$

Consider now the function $\phi\colon (x,t)\mapsto \int_0^t \langle v_s|x\rangle\,ds - h(x)/\eta_t$. For fixed $t\geqslant 0$, one can check that x_t^c maximizes $\phi(x,t)$, so we can apply the envelope theorem (see e.g. [21, Theorem M.L.1]) to write

$$\frac{d}{dt}\frac{h^*(y_t)}{\eta_t} = \frac{\partial\phi}{\partial t}(x_t^c,t) = \langle v_t|x_t^c\rangle + \frac{\dot\eta_t}{\eta_t^2}h(x_t^c) \leqslant \langle v_t|x_t^c\rangle + h_{\min}\frac{\dot\eta_t}{\eta_t^2}, \tag{29}$$

where we used the fact that, by assumption, $\dot\eta \leqslant 0$. Integrating (29) then yields

$$\frac{h^*(y_t)}{\eta_t} \leqslant \frac{h^*(y_0)}{\eta_0} + \int_0^t \langle v_s|x_s^c\rangle\,ds + h_{\min}\int_0^t \frac{\dot\eta_s}{\eta_s^2}\,ds = \int_0^t \langle v_s|x_s^c\rangle\,ds - \frac{h_{\min}}{\eta_t}, \tag{30}$$

where we have used the fact that $h^*(y_0) = h^*(0) = -h_{\min}$ in the second step. Hence, by combining this last equation with (27), we finally obtain:

$$\int_0^t \langle v_s|x\rangle\,ds \leqslant \int_0^t \langle v_s|x_s^c\rangle\,ds - \frac{h_{\min}}{\eta_t} + \frac{h_{\max}}{\eta_t}, \tag{31}$$

and our claim follows by taking the maximum of the left-hand side over $x\in\mathcal{C}$.

If η_t is not smooth, let η_t^m, $m=1,2\dots$, be a sequence of positive and nonincreasing parameters of class C^1 that converges pointwise to η_t. If we let $y_t^m = \eta_t^m\int_0^t v_s\,ds$ and $x_t^m = Q_h(y_t^m)$, we will also have $x_s^m \to x_s^c$ pointwise for all $s\in[0,t]$ by the continuity of Q_h. By the dominated convergence theorem, this implies that $\int_0^t \langle v_s|x_s^m\rangle\,ds \to \int_0^t \langle v_s|x_s^c\rangle\,ds$ and our assertion follows by the bound (31) for smoothly varying η. □

[6] In the rest of the paper, we will consistently use n and k for discrete indices and $s,t,\dots$ for continuous ones.

Remark 4. We should note here that the quantity $\delta_h = h_{\max} - h_{\min}$ in (26) can be taken arbitrarily small so there is no "optimal" regret bound in continuous time. However, as we show in the next section, smaller values of δ_h result in a greater gap between continuous and discrete time, thus leading to a trade-off for the regret in discrete time.

5. **Regret minimization in discrete time.** In this section, our aim is to provide a bound for the regret incurred by the discrete-time strategy (18). To that end, our approach will be as follows: first, given a positive nonincreasing parameter $(\eta_n)_{n\geqslant 1}$ and a sequence of payoff vectors $(v_n)_{n\geqslant 1}$, we construct a continuous-time counterpart by setting

$$v_t = v_{\lceil t \rceil} \tag{32a}$$

and

$$\eta_t = \eta_{\lfloor t \rfloor \vee 1} \tag{32b}$$

for all $t \in \mathbb{R}_+$ (i.e. $\eta_t = \eta_{\lfloor t \rfloor}$ if $t \geqslant 1$ and $\eta_t = \eta_1$ otherwise). Then, given a regularizer $h\colon \mathcal{C} \to \mathbb{R}$, we will compare the cumulative payoffs of the processes $(x_n)_{n\geqslant 1}$ and $(x_t^c)_{t\in\mathbb{R}_+}$ that are generated by (18) and (25) in discrete and continuous time respectively. In this way, the derived regret bound will consist of two terms: one coming from the continuous-time bound (26), and a term coming from the discrete/continuous comparison. Formally, we have:

Theorem 5.1. *Let h be a K-strongly convex regularizer on $\mathcal{C}$ and let $(\eta_n)_{n\geqslant 1}$ be a positive nonincreasing parameter. Then, for every sequence of payoff vectors $(v_n)_{n\geqslant 1}$ in V^*, the sequence of play*

$$x_{n+1} = Q_h\left(\eta_n \sum_{k=1}^{n} v_k\right) \tag{33}$$

generated by the strategy $\sigma = (\sigma_n^{h,\eta_n})_{n\geqslant 1}$ of (18) enjoys the bound

$$\max_{x\in\mathcal{C}} \operatorname{Reg}_n^{\sigma,v}(x) \leqslant \frac{h_{\max} - h_{\min}}{\eta_n} + \frac{1}{2K}\sum_{k=1}^{n} \eta_{k-1}\|v_k\|_*^2, \tag{34}$$

where we have set $\eta_0 = \eta_1$. In particular, if $\|v_n\|_* \leqslant M$ for some $M > 0$, then:

$$\max_{x\in\mathcal{C}} \operatorname{Reg}_n^{\sigma,v}(x) \leqslant \frac{h_{\max} - h_{\min}}{\eta_n} + \frac{M^2}{2K}\sum_{k=1}^{n} \eta_{k-1}. \tag{35}$$

Proof. Define the continuous-time interpolations of v_n and η_n as in (32) and let $y_t = \eta_t \int_0^t v_s\, ds$. Then, for the continuous-time process $x_t^c = Q_h(y_t)$ generated by (25), we have:

$$x_n = Q_h\left(\eta_{n-1}\sum_{k=1}^{n-1} v_k\right) = x_{n-1}^c, \tag{36}$$

and hence, for $k \geqslant 1$ and $t \in (k-1, k)$, the payoffs corresponding to x_t^c and x_k will differ by at most

$$\begin{aligned} |\langle v_t | x_t^c\rangle - \langle v_k | x_k\rangle| &= |\langle v_k | x_t^c - x_{k-1}^c\rangle| \\ &\leqslant \|v_k\|_* \|Q_h(y_t) - Q_h(y_{k-1})\| \leqslant \frac{1}{K}\|v_k\|_*\, \|y_t - y_{k-1}\|_*, \end{aligned} \tag{37}$$

where the last inequality follows from the $1/K$-Lipschitz continuity of Q_h (Corollary 1). On the other hand, the definition of y_t gives

$$\|y_t - y_{k-1}\|_* = \left\|\eta_{k-1}\int_{k-1}^{t} v_s\,ds\right\|_* \leqslant \eta_{k-1}\|v_k\|_*(t-k+1), \tag{38}$$

which leads to the estimate:

$$\begin{aligned}\left|\int_0^n \langle v_t|x_t^c\rangle - \sum_{k=1}^n \langle v_k|x_k\rangle\right| &\leqslant \sum_{k=1}^n \int_{k-1}^k |\,\langle v_t|x_t^c\rangle - \langle v_k|x_k\rangle\,|\,dt\\ &\leqslant \frac{1}{K}\sum_{k=1}^n \eta_{k-1}\|v_k\|_*^2 \int_{k-1}^k (t-k+1)\,dt\\ &= \frac{1}{2K}\sum_{k=1}^n \eta_{k-1}\|v_k\|_*^2. \end{aligned}\tag{39}$$

In view of this discrete/continuous comparison, we thus obtain:

$$\begin{aligned}\max_{x\in\mathcal{C}} \sum_{k=1}^n \langle v_k|x\rangle &= \max_{x\in\mathcal{C}} \int_0^t \langle v_t|x\rangle\;dt\\ &\leqslant \int_0^n \langle v_t|x_t^c\rangle\;dt + \frac{h_{\max}-h_{\min}}{\eta_n}\\ &\leqslant \sum_{k=1}^n \langle v_k|x_k\rangle + \frac{1}{2K}\sum_{k=1}^n \eta_{k-1}\|v_k\|_*^2 + \frac{h_{\max}-h_{\min}}{\eta_n}, \end{aligned}\tag{40}$$

where the first inequality follows from Theorem 4.1 and the last one from (39). The bounds (34) and (35) are then immediate. □

To get the optimal dependence of the bound (35) in n, both terms should scale as $\sqrt{n}$ (otherwise, one would be slower than the other). In this case, we get a bound for the average regret which vanishes as $\mathcal{O}(n^{-1/2})$:

Corollary 2. *Let $(v_n)_{n\geqslant 1}$ be a sequence of payoff vectors in V^*. Then, with notation as in Theorem 5.1, the sequence of play*

$$x_{n+1} = Q_h\left(\sqrt{\frac{K(h_{\max}-h_{\min})}{M^2 n}}\sum_{k=1}^n v_k\right) \tag{41}$$

enjoys the regret bound:

$$\max_{x\in\mathcal{C}} \mathrm{Reg}_n^{\sigma,v}(x) \leqslant 2M\sqrt{\frac{(h_{\max}-h_{\min})n}{K}}. \tag{42}$$

Proof. Set $\delta_h = h_{\max} - h_{\min}$ and $\eta_n = \eta/\sqrt{n}$ with $\eta = M^{-1}\sqrt{K\delta_h}$. Then, since we set $\eta_0 = \eta_1$,

$$\sum_{k=1}^n \eta_{k-1} = \eta\left(2 + \sum_{k=2}^n \frac{1}{\sqrt{k-1}}\right) \leqslant \eta\left(2 + \int_1^{n-1}\frac{1}{\sqrt{t}}dt\right) = \eta\int_0^{n-1}\frac{1}{\sqrt{t}}dt \leqslant 2\sqrt{n},$$

so the bound (35) becomes:

$$\frac{\delta_h}{\eta_n} + \frac{M^2}{2K}\sum_{k=1}^n \eta_{k-1} \leqslant \frac{\delta_h}{\eta}\sqrt{n} + \frac{M^2\eta}{K}\sqrt{n} = 2M\sqrt{\frac{\delta_h n}{K}}.$$

□

Remark 5. Regret guarantees of the same order as (42) can be obtained for the OMD/FtRL family of algorithms by optimizing the choice of η over a finite learning horizon and then using a doubling trick to restart the algorithm ever so often [9, 31].

Remark 6. The dependence of η on δ_h, K and M in (42) has been chosen precisely so as to minimize the expression $\big(\delta_h/\eta + M^2\eta/K\big)$ over all $\eta > 0$.

Remark 7 (On the dependence on K and the choice of optimal h)**.** The dependence of the bound (42) on K is clearly artificial: (42) remains invariant if h is rescaled by a positive constant, so it suffices to consider regularizer functions that are 1-strongly convex over $\mathcal{C}$. This then leads to the following question: *given a norm $\|\cdot\|$ on V and a compact convex subset $\mathcal{C} \subset V$, which 1-strongly convex function minimizes $h_{\max} - h_{\min}$?* With the exception of the Euclidean norm, this question does not seem to admit a trivial answer (cf. Section 7.1 for a more detailed discussion).

By expressing the cumulative payoff gap between discrete- and continuous-time *exactly*, Theorem 5.1 can be extended further to regularizer functions that are not strongly convex over $\mathcal{C}$. The only thing that changes in this case is that the comparison term of the bound (35) is replaced by a term involving the Bregman divergence associated with the convex conjugate h^* of h.

The following result is a variable-parameter extension of Theorem 5.6 in [6].

Theorem 5.2. *Let h be a regularizer function on $\mathcal{C}$. Then, with notation as in Theorem 5.1, the strategy $\sigma = (\sigma_n^{h,\eta_n})_{n\geqslant 1}$ of (18) enjoys the regret bound:*

$$\max_{x\in\mathcal{C}} \mathrm{Reg}_n^{\sigma,v}(x) \leqslant \frac{h_{\max} - h_{\min}}{\eta_n} + \sum_{k=1}^{n} \frac{1}{\eta_{k-1}} D_{h^*}(y_k^-, y_{k-1}^+), \tag{43}$$

where we have set $y_n^+ = \eta_n \sum_{k=1}^n v_k$, $y_n^- = \eta_{n-1}\sum_{k=1}^n v_k$ and $\eta_0 = \eta_1$.

Proof. With notation as in the proof of Theorem 5.1, the variables $y_n^\pm$ in the statement of the theorem may be expressed more concisely as:

$$y_n^\pm = \lim_{t\to n^\pm} y_t = \lim_{t\to n^\pm} \eta_t \int_0^t v_s\, ds, \tag{44}$$

and hence, since $t \longmapsto \eta_t$ is right-continuous, we get $x_n = Q_h(y_{n-1}) = Q_h(y_{n-1}^+)$. Accordingly, if $x_t^c = Q_h(y_t)$ denotes the continuous-time process generated by (25), then, for all $k \geqslant 1$ and for all $t \in (k-1, k)$, we will have:

$$\langle v_t | x_t^c\rangle - \langle v_k | x_k\rangle = \langle v_t | Q_h(y_t)\rangle - \big\langle v_k \big| Q_h(y_{k-1}^+)\big\rangle = \langle v_k | \nabla h^*(y_t)\rangle - \big\langle v_k \big| \nabla h^*(y_{k-1}^+)\big\rangle. \tag{45}$$

In this way, noting that $\langle v_t | \nabla h^*(y_t)\rangle$ is simply the derivative of $h^*(y_t)/\eta_{k-1}$ for $t \in (k-1, k)$, we obtain the following comparison over $(k-1, k)$:

$$\begin{aligned}
\int_{k-1}^{k} \langle v_t | x_t^c\rangle\, dt - \langle v_k | x_k\rangle &= \int_{k-1}^{k} \frac{1}{\eta_{k-1}} \frac{d}{dt}\left(h^*(y_t)\right) dt - \frac{1}{\eta_{k-1}} \big\langle \eta_{k-1} v_k \big| \nabla h^*(y_{k-1}^+)\big\rangle \\
&= \frac{1}{\eta_{k-1}} \left(h^*(y_k^-) - h^*(y_{k-1}^+) - \big\langle y_k^- - y_{k-1}^+ \big| \nabla h^*(y_{k-1}^+)\big\rangle\right) \\
&= \frac{1}{\eta_{k-1}} D_{h^*}\left(y_k^-, y_{k-1}^+\right).
\end{aligned} \tag{46}$$

In view of the above, the claim follows by summing this bound over $k = 1, \dots, n$ and plugging the resulting expression in the first inequality of (40)—which holds independently of any assumptions on h. □

6. **Links with existing results.** In this section, we discuss how certain existing results in online optimization and (stochastic) convex programming can be obtained as corollaries of the general analysis of the previous sections.

6.1. Links with known online optimization algorithms.

6.1.1. *The exponential weights algorithm.* The exponential weights (EW) algorithm was introduced independently by [19] and [30] as a learning strategy in discrete time. Motivated by the approach of [29] who used a continuous-time variant to retrieve the algorithm's classical regret bounds, we show here how the same bounds can be obtained directly from Theorem 5.1.

The framework of the EW algorithm is that of randomized action selection as in Section 2.2. Specifically, let $\mathcal{A} = \{1, \dots, d\}$ be a finite set of *pure* actions, and let the agent's action set be the unit simplex $\mathcal{C} = \Delta_d$ of $\mathbb{R}^d$ – the latter being endowed with the ℓ^1 norm $\|\cdot\|_1$. In this context, the EW algorithm is defined as:

$$\begin{aligned} U_n &= U_{n-1} + v_n, \\ x_{i,n+1} &= \frac{e^{\eta U_{i,n}}}{\sum_{j=1}^d e^{\eta U_{j,n}}} \end{aligned} \tag{EW}$$

where $\eta > 0$ is a (fixed) parameter and $(v_n)_{n\geqslant 1}$ is a sequence of payoff vectors in $[-1,1]^d$ (so that $\|v_n\|_\infty \leqslant 1$ in the induced dual norm).

Example 3.3 in Section 3.1 shows that (EW) corresponds to (18) with $\eta_n = \eta$ and $h(x) = \sum_{i=1}^d x_i \log x_i$. Since $h_{\max} - h_{\min} = \log d$ and h is 1-strongly convex with respect to $\|\cdot\|_1$ (cf. Proposition 4), Theorem 5.1 readily yields the bound

$$\max_{a\in\mathcal{A}} \operatorname{Reg}_n(a) \leqslant \frac{\log d}{\eta} + \frac{n\eta}{2}. \tag{47}$$

Additionally, if the time horizon n is known in advance, the optimal parameter choice $\eta = \sqrt{2\log d/n}$ leads to

$$\max_{a\in\mathcal{A}} \operatorname{Reg}_n(a) \leqslant \sqrt{2n\log d}, \tag{48}$$

which, as far as the dependence on d and n is concerned, is the best possible bound a strategy can guarantee in this framework – see e.g. [10, Theorem 3.7].

Remark 8. By taking $v_n \in [0,1]^d$ (as is often the case in the literature) and then shifting to $[-1/2, 1/2]^d$, Theorem 5.1 can be applied with $M = 1/2$. This yields a factor of $1/8$ in the second term of (47) and leads to the bound obtained by [8] and [10].

6.1.2. *The exponential weights algorithm with $\eta_n = 1/\sqrt{n}$.* [1] considered the following variant of (EW)

$$\begin{aligned} U_n &= U_{n-1} + v_n, \\ x_{i,n+1} &= \frac{e^{\eta U_{i,n}/\sqrt{n}}}{\sum_{j=1}^d e^{\eta U_{j,n}/\sqrt{n}}}. \end{aligned} \tag{EW$'$}$$

In our context, a direct application of Corollary 2 with $M = K = 1$ then gives

$$\max_{a\in\mathcal{A}} \operatorname{Reg}_n(a) \leqslant 2\sqrt{n\log d}, \tag{49}$$

a bound which, unlike (48), has the advantage of holding uniformly in time.

6.1.3. *Smooth fictitious play.* The smooth fictitious play (SFP) process was introduced by [11] (see also [12, 13]), and its regret properties were examined further by [4] using the theory of stochastic approximation – but without providing any quantitative bounds for the regret.

Just like the EW algorithm, SFP falls within the randomized actions framework of Section 2.2. In particular, SFP corresponds to the sequence of plays generated by (18) for an arbitrary regularizer on Δ_d and with parameter η/n for some $\eta > 0$; specifically:

$$x_{n+1} = Q_h\left(\frac{\eta}{n}\sum_{k=1}^{n} v_k\right). \tag{SFP}$$

With regards to the regret induced by (SFP), [4, Theorem 6.6] show that for every $\varepsilon > 0$, there exists some $\eta^* \equiv \eta^*(\varepsilon)$ such that the strategy (SFP) with parameter $\eta \geqslant \eta^*$ leads to ε-realized-regret. On the other hand, combining Proposition 1 with Theorem 5.1 yields the following more precise statement:

Proposition 5. *Let h be a K-strongly convex regularizer on the unit simplex $\Delta_d \subset \mathbb{R}^d$ endowed with the ℓ^1 norm. Then, for every sequence of payoff vectors $(v_n)_{n\geqslant 1}$ in $[-1,1]^d$, the strategy (SFP) with parameter $\eta > 0$ guarantees*

$$\max_{a\in\mathcal{A}} \mathrm{Reg}_n(e_a) \leqslant \frac{h_{\max} - h_{\min}}{\eta} n + \frac{\eta \log n}{2K} + \frac{\eta}{K}. \tag{50}$$

In particular, (SFP) with parameter η leads to $(h_{\max} - h_{\min})/\eta$ (realized) regret.

Proof. Simply combine the logarithmic growth estimate $\sum_{k=1}^{n} k^{-1} < 1 + \log n$ for the harmonic series and Theorem 5.1 with $\eta_n = \eta/n$; the claim for the realized regret then follows from Proposition 1. □

Remark 9. It should be noted here that the qualitative analysis of [4] does not require h to be strongly convex; that said, if h is strongly convex, Proposition 5 gives a quantitative bound on the regret.

6.1.4. *Vanishingly smooth fictitious play.* The variant of SFP known as vanishingly smooth fictitious play (VSFP) was introduced by [3], and its regret properties were established using sophisticated tools from the theory of differential inclusions and stochastic approximation – but, again, without providing explicit regret bounds.

Using the same notation as before, VSFP corresponds to the sequence of play

$$x_{n+1} = Q_h\left(\eta_n \sum_{k=1}^{n} v_k\right), \tag{VSFP}$$

where h is a K-strongly convex regularizer on Δ_d and the sequence η_n satisfies:

(A1) $\lim_{n\to\infty} n\eta_n = +\infty$,

(A2) $\eta_n = \mathcal{O}(n^{-\alpha})$ for some $\alpha > 0$.

Under these assumptions, the main result of [3] is that (VSFP) leads to no realized regret; in our framework, this follows directly from Proposition 1 and Theorem 5.1 (which also gives a quantitative regret guarantee):

Proposition 6. *With notation as in Proposition 5, and against payoff vectors in $[-1,1]^d$, the strategy (VSFP) with η_n satisfying assumptions (A1) and (A2) guarantees the regret bound*

$$\max_{a\in\mathcal{A}} \frac{1}{n}\mathrm{Reg}_n(e_a) \leqslant \frac{h_{\max} - h_{\min}}{n\eta_n} + \frac{1}{2nK}\sum_{k=1}^{n} \eta_{k-1}, \tag{51}$$

and thus leads to no regret. In particular, if $\eta_n = \eta n^{-\alpha}$ *for some* $\alpha \in (0,1)$*, then:*

$$\max_{a\in\mathcal{A}} \frac{1}{n}\operatorname{Reg}_n(e_a) \leqslant \frac{h_{\max} - h_{\min}}{\eta n^{1-\alpha}} + \frac{\eta n^{-\alpha}}{2(1-\alpha)K} + \frac{\eta}{2Kn}. \tag{52}$$

Proof. The bound (51) is an immediate corollary of Theorem 5.1; the no-regret property then follows from Assumptions (A1) and (A2). Finally, if $\eta_n = \eta n^{-\alpha}$, we get

$$\frac{1}{\eta}\sum_{k=1}^{n} \eta_{k-1} = 1 + \sum_{k=1}^{n-1} k^{-\alpha} \leqslant 1 + \int_0^{n-1} t^{-\alpha}\,dt = 1 + \frac{n^{1-\alpha}}{1-\alpha}, \tag{53}$$

and (52) follows by substituting the above in (51). □

Remark 10. If we take $h(x) = \sum_{i=1}^{d} x_i \log x_i$ and $\alpha = 1/2$, (VSFP) boils down to (EW′); the bound (49) then also follows from (52).

6.1.5. *Online gradient descent.* The online gradient descent (OGD) algorithm was introduced by [33] in the context of online convex optimization that we described in Section 2.3 – see also [7, Section 4.1]. Here, we focus on a so-called *lazy* variant [28, p. 144] defined by means of the recursion

$$\begin{aligned} U_n &\in U_{n-1} - \eta\,\partial\ell_n(x_n),\\ x_{n+1} &= \arg\min_{x\in\mathcal{C}} \|x - U_n\|^2, \end{aligned} \tag{OGD-L}$$

where $\ell_n\colon \mathcal{C} \to \mathbb{R}$ is a sequence of M-Lipschitz loss functions, $\eta > 0$ is a constant parameter, and the algorithm is initialized with $U_0 = 0$.

In view of Example 3.4, (OGD-L) corresponds to the strategy $\sigma = (\sigma_n^{h,\eta})_{n\geqslant 1}$ generated by the Euclidean regularizer h on $\mathcal{C}$ – defined itself as in (13). Theorem 5.1 thus yields the regret bound:

$$\max_{x\in\mathcal{C}} \frac{1}{n}\operatorname{Reg}_n(x) \leqslant \frac{\delta_{\mathcal{C}}^2}{2n\eta} + \frac{\eta M^2}{2} \tag{54}$$

with $\delta_{\mathcal{C}}^2 = \max_{x\in\mathcal{C}} \|x\|_2^2 - \min_{x\in\mathcal{C}} \|x\|_2^2$. Accordingly, if the time horizon n is known in advance, the optimal choice for η is $\eta = \delta_{\mathcal{C}}/(M\sqrt{n})$, leading to a cumulative regret guarantee of $M\delta_{\mathcal{C}}\sqrt{n}$, which is essentially the bound derived by [28, Corollary. 2.7] (see also [7, Theorem 3.1] for the greedy variant).[7]

6.1.6. *Online mirror descent.* The family of (lazy) online mirror descent (OMD) algorithms studied by Shalev-Shwartz [27, 28] is the most general family of strategies that we discuss in this section (see also [7] for a greedy version). In particular, the OMD class of strategies contains EW and OGD as special cases, and it is also equivalent to the family of Follow the Regularized Leader (FtRL) algorithms in the case of linear payoffs [28, 15].

Following [28] (and with notation as in Section 2.3), let $\ell_n\colon \mathcal{C} \to \mathbb{R}$ be a sequence of convex functions which are M-Lipschitz with respect to some norm $\|\cdot\|$ on $\mathbb{R}^d$. Then, given a regularizer function h on $\mathcal{C}$, the lazy OMD algorithm is defined by means of the recursion:

$$\begin{aligned} U_n &\in U_{n-1} - \eta\,\partial\ell_n(x_n),\\ x_{n+1} &= Q_h(U_n), \end{aligned} \tag{OMD-L}$$

[7] For the difference between lazy and greedy variants, see Section 7.2.

where $\eta > 0$ is a *fixed* parameter and the algorithm is initialized with $U_0 = 0$. As a result, if h is taken K-strongly convex with respect to $\|\cdot\|$, Theorem 5.1 immediately yields the known regret bound for OMD:

$$\max_{x\in\mathcal{C}} \operatorname{Reg}_n(x) \leqslant \frac{h_{\max} - h_{\min}}{\eta} + \frac{\eta M^2 n}{2K}. \tag{55}$$

6.2. **Links with convex optimization.** Ordinary convex programs can be seen as online optimization problems where the loss function remains constant over time and the agent seeks to attain its minimum value. In what follows, we outline how regret-minimizing strategies can be used for this purpose and we describe the performance gap incurred by using a method with a variable step-size instead of a variable parameter.

Let $f\colon \mathcal{C} \to \mathbb{R}$ be a convex real-valued function on $\mathcal{C}$ and let $(\gamma_n)_{n\geqslant 1}$ be a positive sequence (which we will later interpret as a sequence of step-sizes); also, given a sequence $(x_n)_{n\geqslant 1}$ in $\mathcal{C}$, let

$$x_n^{\min} \in \arg\min_{1\leqslant k\leqslant n} f(x_k), \qquad\qquad x_n^{\gamma} = \frac{\sum_{k=1}^n \gamma_k x_k}{\sum_{k=1}^n \gamma_k}. \tag{56}$$

If we use the notation $x_n' \in \{x_n^{\min}, x_n^{\gamma}\}$ to refer interchangeably to either $x_n^{\min}$ or x_n^{γ}, Jensen's inequality readily gives:

$$f(x_n') \leqslant \frac{\sum_{k=1}^n \gamma_k f(x_k)}{\sum_{k=1}^n \gamma_k}. \tag{57}$$

Now consider the algorithm:

$$\begin{aligned} U_n &\in U_{n-1} - \gamma_n \partial f(x_n), \\ x_{n+1} &= Q_h(\eta_n U_n), \end{aligned} \tag{58}$$

where $(\gamma_n)_{n\geqslant 1}$ is a sequence of step sizes and $(\eta_n)_{n\geqslant 1}$ a sequence of parameters. In the case of a constant parameter $\eta_n = 1$, (58) then becomes

$$\begin{aligned} U_n &\in U_{n-1} - \gamma_n \partial f(x_n), \\ x_{n+1} &= Q_h(U_n), \end{aligned} \tag{MD-L}$$

which is a lazy variant of the mirror descent (MD) algorithm [23]. In particular, if h is the Euclidean regularizer on $\mathcal{C}$, the algorithm boils down to a lazy version of the standard projected subgradient (PSG) method:

$$\begin{aligned} U_n &\in U_{n-1} - \gamma_n \partial f(x_n), \\ x_{n+1} &= \arg\min_{x\in\mathcal{C}} \|x - U_n\|_2. \end{aligned} \tag{PSG-L}$$

The following corollary shows that these lazy versions guarantee the same convergence bounds as the corresponding greedy variants — see e.g. [2, Theorem 4.1].

Corollary 3 (Constant parameter, variable step size)**.** *Let $f\colon \mathcal{C} \to \mathbb{R}$ be an M-Lipschitz convex function and let $(x_n)_{n\geqslant 1}$ be the sequence of plays generated by* (MD-L) *for some K-strongly convex regularizer h on $\mathcal{C}$. Then, the adjusted iterates $x_n' \in \{x_n^{\min}, x_n^{\gamma}\}$ of x_n satisfy:*

$$f(x_n') \leqslant f_{\min} + \frac{h_{\max} - h_{\min} + \frac{1}{2}M^2 K^{-1} \sum_{k=1}^n \gamma_k^2}{\sum_{k=1}^n \gamma_k}. \tag{59}$$

Proof. With $\sigma = (\sigma_n^{h,\eta_n})_{n\geqslant 1}$, $v_k \in -\gamma_k \partial f(x_k)$ and $x'_n \in \{x_n^{\min}, x_n^{\gamma}\}$, we have:

$$\mathrm{Reg}_n^{\sigma,v}(x) = \sum_{k=1}^{n} \langle v_k | x - x_k\rangle \geqslant -\sum_{k=1}^{n} \gamma_k \big(f(x) - f(x_k)\big) \geqslant \left(\sum_{k=1}^{n} \gamma_k\right) \big(f(x'_n) - f(x)\big), \tag{60}$$

where the last step follows from (57). By taking $x \in \arg\min f$, we then obtain:

$$f(x'_n) - f_{\min} \leqslant \frac{\mathrm{Reg}_n^{\sigma,v}(x)}{\sum_{k=1}^{n} \gamma_k}. \tag{61}$$

The result then follows by applying Theorem 5.1 and using the fact that $\|v_k\|_* \leqslant \|\gamma_k \partial f(x_k)\|_* \leqslant \gamma_k M$ (recall that f is M-Lipschitz continuous). □

One can see that the best convergence rate that we get with constant η and step-sizes of the form $\gamma_n \propto n^{-\alpha}$ is $\mathcal{O}(\log n/\sqrt{n})$ for $\alpha = 1/2$ (and there is no straighforward choice of γ_n leading to a better convergence rate). On the other hand, by taking a *constant* step-size $\gamma_n = 1$ and *varying the algorithm's parameter* $\eta_n \propto n^{-1/2}$, we do achieve an $\mathcal{O}(n^{-1/2})$ rate of convergence.

Corollary 4 (Constant step size, variable parameter)**.** *With notation as in Corollary 3, let $(x_n)_{n\geqslant 1}$ be the sequence of plays generated by (58) with*

$$\eta_n = \frac{1}{M}\sqrt{\frac{K(h_{\max} - h_{\min})}{n}}, \tag{62}$$

and constant $\gamma_n = 1$. Then, the adjusted iterates $x'_n \in \{x_n^{\min}, x_n^{\gamma}\}$ of x_n guarantee

$$f(x'_n) \leqslant f_{\min} + 2M\sqrt{\frac{h_{\max} - h_{\min}}{Kn}}. \tag{63}$$

Proof. Similar to the proof of Corollary 3. □

6.3. **Noisy observations and links with stochastic convex optimization.** Assume that at every stage $n = 1, 2, \dots$ of the decision process, the agent does not observe the actual payoff vector $v_n \in V^*$, but the realization of a random vector $\hat{v}_n$ satisfying $\mathbb{E}\left[\hat{v}_n | \mathcal{F}_n\right] = v_n$, where $\mathcal{F}_n$ is generated by

$$(\hat{x}_1, v_1, \hat{v}_1, i_1, \dots, \hat{x}_{n-1}, v_{n-1}, \hat{v}_{n-1}, i_{n-1}, \hat{x}_n, v_n). \tag{64}$$

In this case, a learning strategy σ can be used with the observed vectors $\hat{v}_n$, thus leading to a (random) sequence of play $\hat{x}_{n+1} = \sigma_{n+1}(\hat{v}_1, \dots, \hat{v}_n)$ – see e.g. [28, Section 4.1] for a model of this kind.

In this framework, the agent's (maximal) cumulative regret, which is the quantity of interest, is given by

$$\max_{x\in\mathcal{C}} \sum_{k=1}^{n} \langle v_k | x\rangle - \sum_{k=1}^{n} \langle v_k | \hat{x}_k\rangle. \tag{65}$$

On the other hand,

$$\max_{x\in\mathcal{C}} \sum_{k=1}^{n} \langle \hat{v}_k | x\rangle - \sum_{k=1}^{n} \langle \hat{v}_k | \hat{x}_k\rangle. \tag{66}$$

can be interpreted as the agent's cumulative regret against the observed payoff sequence $(\hat{v}_n)_{n\geqslant 1}$. The above two quantites can be related (in average) as follows.

We assume that $\|\hat{v}_k\|_* \leqslant M$ (a.s.). As for the first term involving the maximum,

$$\begin{aligned}\max_{x\in\mathcal{C}} \frac{1}{n}\sum_{k=1}^{n} \langle v_k|x\rangle &= \max_{x\in\mathcal{C}} \frac{1}{n}\left\langle \sum_{k=1}^{n}\hat{v}_k + \sum_{k=1}^{n}(v_k-\hat{v}_k)\middle| x\right\rangle \\ &\leqslant \max_{x\in\mathcal{C}} \frac{1}{n}\sum_{k=1}^{n}\langle \hat{v}_k|x\rangle + \left\|\frac{1}{n}\sum_{k=1}^{n}(v_k-\hat{v}_k)\right\|_* \|\mathcal{C}\|, \end{aligned} \tag{67}$$

where the last term is small with high probability: indeed, since $\mathbb{E}\left[\hat{v}_k - v_k \,|\, \mathcal{F}_k\right] = 0$, a classical argument based on bounded martingale differences can be used. We deal with the second sum similarly by noting that $\mathbb{E}\left[\langle \hat{v}_k|\hat{x}_k\rangle \,|\, \mathcal{F}_k\right] = \langle \mathbb{E}\left[\hat{v}_k \,|\, \mathcal{F}_k\right]|\hat{x}_k\rangle = \langle v_k|\hat{x}_k\rangle$ and that

$$\frac{1}{n}\sum_{k=1}^{n}\langle v_k|\hat{x}_k\rangle = \frac{1}{n}\sum_{k=1}^{n}\langle \hat{v}_k|\hat{x}_k\rangle + \frac{1}{n}\sum_{k=1}^{n}\langle v_k-\hat{v}_k|\hat{x}_k\rangle\,. \tag{68}$$

The guarantees of Theorem 5.1 therefore translate to the present framework with high probability.

The above can be adapted to the framework of stochastic convex optimization as follows: let $f\colon \mathcal{C}\to\mathbb{R}$ be a Lipschitz convex function on $\mathcal{C}$, let $(\gamma_n)_{n\geqslant 1}$ be a positive sequence of step sizes, and consider the strategy σ generated by (18) with $\eta = 1$ and h a K-strongly convex regularizer on $\mathcal{C}$. Then, the sequence of play

$$\hat{x}_{n+1} = \sigma_{n+1}(-\gamma_1\hat{g}_1,\dots,-\gamma_n\hat{g}_n) = Q_h\left(-\sum_{k=1}^{n}\gamma_k\hat{g}_k\right) \tag{69}$$

where $\hat{g}_n$ is a random vector with $\mathbb{E}\left[\hat{g}_n|\hat{g}_{n-1},\dots,\hat{g}_1\right] = g_n \in \partial f(\hat{x}_n)$ may be written recursively as:

$$\begin{aligned}\hat{U}_n &\in \hat{U}_{n-1} - \gamma_n\partial f(\hat{x}_n),\\ \hat{x}_{n+1} &= Q_h(\hat{U}_n).\end{aligned} \tag{RSA-L}$$

This algorithm may be seen as a lazy version of the so-called robust stochastic approximation (RSA) process of [22]; in particular, using the Euclidean regularizer leads to the lazy stochastic projected subgradient (SPSG) method:

$$\begin{aligned}\hat{U}_n &\in \hat{U}_{n-1} - \gamma_n\partial f(\hat{x}_n),\\ \hat{x}_{n+1} &= \operatorname*{arg\,min}_{x\in\mathcal{C}} \|x-\hat{U}_n\|_2.\end{aligned} \tag{SPSG-L}$$

ALGORITHM	$\mathcal{C}$	$h(x)$	η_n	INPUT	NORM
EW	Δ_d	$\sum_i x_i \log x_i$	CONSTANT	v_n	ℓ^1
EW′	Δ_d	$\sum_i x_i \log x_i$	$\eta/\sqrt{n}$	v_n	ℓ^1
SFP	Δ_d	ANY	η/n	v_n	ℓ^1
VSFP	Δ_d	ANY	$\eta n^{-\alpha}$ $(0<\alpha<1)$	v_n	ℓ^1
OGD-L	ANY	$\frac{1}{2}\|x\|_2^2$	CONSTANT	$-\nabla f_n(x_n)$	ℓ^2
OMD-L	ANY	ANY	CONSTANT	$-\nabla f_n(x_n)$	ANY
PSG-L	ANY	$\frac{1}{2}\|x\|_2^2$	1	$-\gamma_n \nabla f(x_n)$	ℓ^2
MD-L	ANY	ANY	1	$-\gamma_n \nabla f(x_n)$	ANY
RSA-L	ANY	ANY	1	$-\gamma_n(\nabla f(x_n)+\xi_n)$	ANY
SPSG-L	ANY	$\frac{1}{2}\|x\|_2^2$	1	$-\gamma_n(\nabla f(x_n)+\xi_n)$	ℓ^2

TABLE 1. Summary of the algorithms discussed in Section 6. The suffix "L" indicates a "lazy" variant; the INPUT column stands for the stream of payoff vectors which is used as input for the algorithm and the NORM column specifies the norm of the ambient space; finally, ξ_n represents a zero-mean stochastic process with values in $\mathbb{R}^d$.

Setting $v_n = -\gamma_n g_n$, $\hat{v}_n = -\gamma_n \hat{g}_n$ and taking $\hat{x}'_n \in \{\hat{x}_n^{\min}, \hat{x}_n^{\gamma}\}$ as before, we can adapt Corollary 3 to we get, for all $x \in \mathcal{C}$,

$$\mathbb{E}\left[f(\hat{x}'_n) - f(x)\right] \leqslant \mathbb{E}\left[\frac{1}{\sum_{k=1}^n \gamma_k}\sum_{k=1}^n \gamma_k (f(\hat{x}_k) - f(x))\right] \tag{70}$$

$$\leqslant \mathbb{E}\left[\frac{1}{\sum_{k=1}^n \gamma_k}\sum_{k=1}^n \langle v_k | x - \hat{x}_k\rangle\right] \tag{71}$$

$$= \mathbb{E}\left[\frac{1}{\sum_{k=1}^n \gamma_k}\sum_{k=1}^n \mathbb{E}\left[\langle \hat{v}_k | x - \hat{x}_k\rangle \,|\, \mathcal{F}_k\right]\right] \tag{72}$$

$$= \mathbb{E}\left[\frac{1}{\sum_{k=1}^n \gamma_k}\sum_{k=1}^n \langle hat v_k | x - \hat{x}_k\rangle\right] \tag{73}$$

$$\leqslant \frac{h_{\max} - h_{\min} + \frac{1}{2}M^2 K^{-1}\sum_{k=1}^n \gamma_k^2}{\sum_{k=1}^n \gamma_k}, \tag{74}$$

which is essentially the same value guarantee as that of greedy RSA [22, Eq. 2.41].

7. Discussion.

7.1. **On the optimal choice of h.** As mentioned in the discussion after Corollary 2, the following open question arises: *given a norm $\|\cdot\|$ on V and a compact, convex subset $\mathcal{C} \subset V$, which 1-strongly convex regularizer on $h\colon \mathcal{C} \to \mathbb{R}$ has minimal depth $\delta_h = h_{\max} - h_{\min}$?*

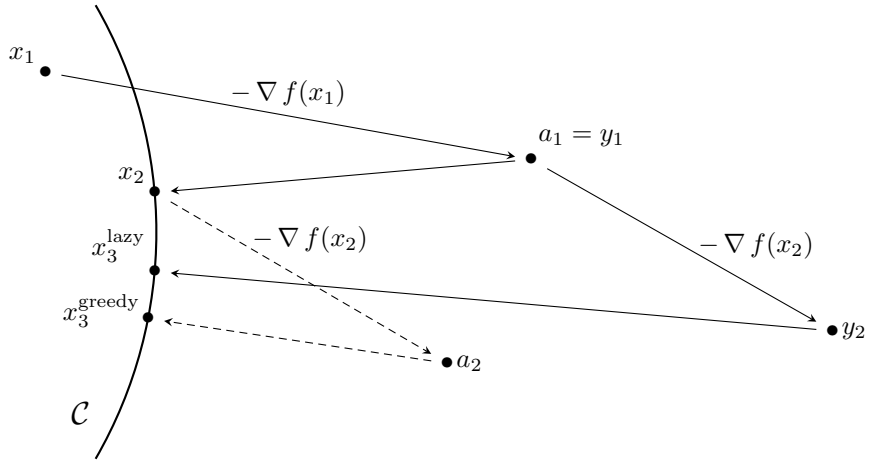


FIGURE 1. Graphical illustration of the greedy (dashed) and lazy (solid) branches of the projected subgradient (PSG) method.

As the following proposition shows, in the case of the Euclidean norm on V, this minimal depth is half the radius squared of the smallest enclosing sphere of $\mathcal{C}$:

Proposition 7. *Let $h\colon \mathcal{C} \to \mathbb{R}$ be a 1-strongly convex regularizer function on $\mathcal{C}$ with respect to the ℓ^2 norm $\|\cdot\|_2$ on V. Then:*

$$h_{\max} - h_{\min} \geqslant \frac{1}{2} \min_{x' \in \mathcal{C}} \max_{x \in \mathcal{C}} \|x' - x\|_2^2, \tag{75}$$

and equality is attained by taking

$$h(x) = \begin{cases} \frac{1}{2}\|x - x_0\|_2^2 & \text{if } x \in \mathcal{C}, \\ +\infty & \text{otherwise}, \end{cases} \tag{76}$$

where $x_0 \in \arg\min_{x' \in \mathcal{C}} \max_{x \in \mathcal{C}} \|x' - x\|_2^2$ is the center of the smallest enclosing sphere of $\mathcal{C}$.

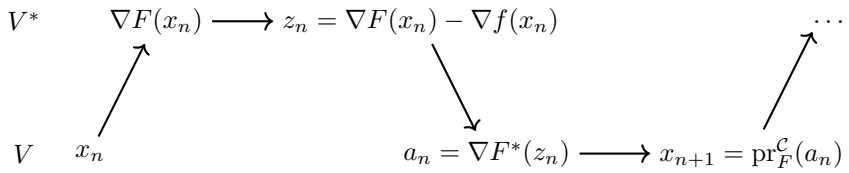
Proof. Letting $x_1 \in \arg\min_{x \in \mathcal{C}} h(x)$ and $x_2 \in \arg\max_{x \in \mathcal{C}} \|x - x_1\|_2^2$, we readily get:

$$\begin{aligned} h_{\max} - h_{\min} &\geqslant h(x_2) - h(x_1) \\ &\geqslant \frac{1}{2}\|x_2 - x_1\|_2^2 = \frac{1}{2} \max_{x \in \mathcal{C}} \|x - x_1\|_2^2 \geqslant \frac{1}{2} \min_{x' \in \mathcal{C}} \max_{x \in \mathcal{C}} \|x - x'\|_2^2, \end{aligned} \tag{77}$$

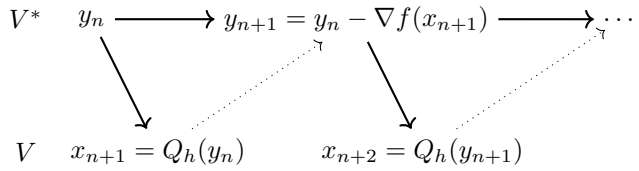
where the second inequality follows from the strong convexity of h and the fact that $\partial h(x_1) \ni 0$. That (76) attains the bound (75) is then a trivial consequence of its definition, as is its geometric characterization. □

Despite the simplicity of the bound (75), this analysis does not work for an arbitrary norm because $\frac{1}{2}\|x - x_0\|^2$ might fail to be 1-strongly convex with respect to $\|\cdot\|$ – for instance, $\|x - x_0\|_1^2$ is not even *strictly* convex.

7.2. **Greedy versus Lazy.** To illustrate the difference between *lazy* and *greedy* variants, we first focus on the PSG method run with constant step $\gamma = 1$ for a smooth function $f\colon \mathcal{C} \to \mathbb{R}$. The two variants may then be expressed by means of



(A) Greedy Mirror Descent



(B) Lazy Mirror Descent

FIGURE 2. Greedy and Lazy Mirror Descent with $\gamma_n = 1$.

the recursions:

$$\begin{aligned} a_n &= x_n - \nabla f(x_n) \\ x_{n+1} &= \arg\min_{x \in \mathcal{C}} \|x - a_n\|_2 \end{aligned} \tag{78a}$$

for the greedy version and:

$$\begin{aligned} y_n &= y_{n-1} - \nabla f(x_n) \\ x_{n+1} &= \arg\min_{x \in \mathcal{C}} \|x - y_n\|_2 \end{aligned} \tag{78b}$$

for the lazy one.

As can be seen in Fig. 1, the greedy variant is based on the classical idea of gradient descent, i.e. adding $-\nabla f(x_n)$ to x_n and projecting back to $\mathcal{C}$ if needed. On the other hand, in the lazy variant, the gradient term $-\nabla f(x_n)$ is *not* added to x_n, but to the "unprojected" iterate y_n; we only project to $\mathcal{C}$ in order to obtain the algorithm's next iterate. Owing to this modification, the lazy variant is thus driven by the sum $y_n = \sum_{k=1}^{n} \nabla f(x_n)$.

In the case of mirror descent with an arbitrary regularizer function h, the lazy version has an implementation advantage over its greedy counterpart. Specifically, given a proper convex function F such that $F = h$ on $\mathcal{C}$ (cf. Example 3.5), greedy mirror descent is defined as:

$$\begin{aligned} a_n &= \nabla F^* \left(\nabla F(x_n) - \nabla f(x_n) \right), \\ x_{n+1} &= \mathrm{pr}_F^{\mathcal{C}}(a_n), \end{aligned} \tag{79a}$$

where the Bregman projection $\mathrm{pr}_F^{\mathcal{C}}(a_n)$ is given by (16); on the other hand, lazy MD is defined as

$$\begin{aligned} y_n &= y_{n-1} - \nabla f(x_n), \\ x_{n+1} &= Q_h(y_n). \end{aligned} \tag{79b}$$

The computation steps for each variant are represented in Figure 2. The first step in the greedy version which consists in computing ∇F has no equivalent in the lazy version, which is thus computationally more lightweight.

Acknowledgments. The authors are greatly indebted to Vianney Perchet for his invaluable help in improving all aspects of this work, and to Rida Laraki, Gilles Stoltz and Sylvain Sorin for their insightful suggestions and careful reading of the manuscript. The authors would also like to express their gratitude to Guillaume Vigeral for many helpful discussions and remarks.

Part of this work was carried out during the authors' visit at the Hausdorff Research Institute for Mathematics at the University of Bonn in the framework of the Trimester Program "Stochastic Dynamics in Economics and Finance". This work was supported by the French National Research Agency (ANR) projects GAGA (grant no. ANR-13-JS01-0004-01), NETLEARN (grant no. ANR-13-INFR-004), and ORACLESS (grant no. ANR-16-CE33-0004-01).

REFERENCES

[1] P. Auer, N. Cesa-Bianchi and C. Gentile, Adaptive and self-confident on-line learning algorithms, *Journal of Computer and System Sciences*, **64** (2002), 48–75.

[2] A. Beck and M. Teboulle, Mirror descent and nonlinear projected subgradient methods for convex optimization, *Operations Research Letters*, **31** (2003), 167–175.

[3] M. Benaïm and M. Faure, Consistency of vanishingly smooth fictitious play, *Mathematics of Operations Research*, **38** (2013), 437–450.

[4] M. Benaïm, J. Hofbauer and S. Sorin, Stochastic approximations and differential inclusions, part II: Applications, *Mathematics of Operations Research*, **31** (2006), 673–695.

[5] L. M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, *USSR Computational Mathematics and Mathematical Physics*, **7** (1967), 200–217.

[6] S. Bubeck and N. Cesa-Bianchi, Regret analysis of stochastic and nonstochastic multi-armed bandit problems, *Foundations and trends in machine learning*, **5** (2012), 1–122.

[7] S. Bubeck, Introduction to online optimization, Lecture Notes, 2011.

[8] N. Cesa-Bianchi, Analysis of two gradient-based algorithms for on-line regression, in *COLT '97: Proceedings of the 10th Annual Conference on Computational Learning Theory*, 1997, 163–170.

[9] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire and M. K. Warmuth, How to use expert advice, *Journal of the ACM*, **44** (1997), 427–485.

[10] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*, Cambridge University Press, 2006.

[11] D. Fudenberg and D. K. Levine, Consistency and cautious fictitious play, *Journal of Economic Dynamics and Control*, **19** (1995), 1065–1089.

[12] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*, vol. 2 of Economic learning and social evolution, MIT Press, Cambridge, MA, 1998.

[13] D. Fudenberg and D. K. Levine, Conditional universal consistency, *Games and Economic Behavior*, **29** (1999), 104–130.

[14] J. Hannan, Approximation to Bayes risk in repeated play, in *Contributions to the Theory of Games, Volume III* (eds. M. Dresher, A. W. Tucker and P. Wolfe), vol. 39 of Annals of Mathematics Studies, Princeton University Press, Princeton, NJ, 1957, 97–139.

[15] E. Hazan, A survey: The convex optimization approach to regret minimization, in *Optimization for Machine Learning* (eds. S. N. Suvrit Spa and S. J. Wright), MIT Press, 2012, 287–304.

[16] J. Hofbauer and W. H. Sandholm, On the global convergence of stochastic fictitious play, *Econometrica*, **70** (2002), 2265–2294.

[17] S. M. Kakade, S. Shalev-Shwartz and A. Tewari, Regularization techniques for learning with matrices, *The Journal of Machine Learning Research*, **13** (2012), 1865–1890.

[18] J. Kivinen and M. K. Warmuth, Exponentiated gradient versus gradient descent for linear predictors, *Information and Computation*, **132** (1997), 1–63.

[19] N. Littlestone and M. K. Warmuth, The weighted majority algorithm, *Information and Computation*, **108** (1994), 212–261.

[20] S. Mannor and V. Perchet, Approchability, fast and slow, *Journal of Machine Learning Research: Workshop and Conference Proceedings*, **30** (2013), 1–16.

[21] A. Mas-Colell, M. D. Whinston and J. R. Green, *Microeconomic Theory*, Oxford University Press, New York, NY, USA, 1995.
[22] A. S. Nemirovski, A. Juditsky, G. G. Lan and A. Shapiro, Robust stochastic approximation approach to stochastic programming, *SIAM Journal on Optimization*, **19** (2009), 1574–1609.
[23] A. S. Nemirovski and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization*, Wiley, New York, NY, 1983.
[24] Y. Nesterov, Primal-dual subgradient methods for convex problems, *Mathematical Programming*, **120** (2009), 221–259.
[25] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[26] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, vol. 317 of A Series of Comprehensive Studies in Mathematics, Springer-Verlag, Berlin, 1998.
[27] S. Shalev-Shwartz, *Online Learning: Theory, Algorithms, and Applications*, PhD thesis, Hebrew University of Jerusalem, 2007.
[28] S. Shalev-Shwartz, Online learning and online convex optimization, *Foundations and Trends in Machine Learning*, **4** (2011), 107–194.
[29] S. Sorin, Exponential weight algorithm in continuous time, *Mathematical Programming*, **116** (2009), 513–528.
[30] V. G. Vovk, Aggregating strategies, in *COLT '90: Proceedings of the Third Workshop on Computational Learning Theory*, 1990, 371–383.
[31] V. G. Vovk, A game of prediction with expert advice, in *COLT '95: Proceedings of the 8th Annual Conference on Computational Learning Theory*, 1995, 51–60.
[32] M. K. Warmuth and A. K. Jagota, Continuous and discrete-time nonlinear gradient descent: Relative loss bounds and convergence, in *Electronic Proceedings of the 5th International Symposium on Artificial Intelligence and Mathematics*, 1997.
[33] M. Zinkevich, Online convex programming and generalized infinitesimal gradient ascent, in *ICML '03: Proceedings of the 20th International Conference on Machine Learning*, 2003.

Received January 2017; revised February 2017.

E-mail address: joon.kwon@ens-lyon.org
E-mail address: panayotis.mertikopoulos@imag.fr