Online Learning with Continuous Variations Dynamic Regret and Reductions

Ching-An Cheng* Georgia Tech

Jonathan Lee* UC Berkeley

Ken Goldberg UC Berkeley

Byron Boots Georgia Tech

Abstract

Online learning is a powerful tool for analyzing iterative algorithms. However, the classic adversarial setup fails to capture regularity that can exist in practice. Motivated by this observation, we establish a new setup, called Continuous Online Learning (COL), where the gradient of online loss function changes continuously across rounds with respect to the learner's decisions. We show that COL appropriately describes many interesting applications, from general equilibrium problems (EPs) to optimization in episodic MDPs. Using this new setup, we revisit the difficulty of sublinear dynamic regret. We prove a fundamental equivalence between achieving sublinear dynamic regret in COL and solving certain EPs. With this insight, we offer conditions for efficient algorithms that achieve sublinear dynamic regret, even when the losses are chosen adaptively without any *a priori* variation budget. Furthermore, we show for COL a reduction from dynamic regret to both static regret and convergence in the associated EP, allowing us to analyze the dynamic regret of many existing algorithms.

1 INTRODUCTION

Online learning (Gordon, 1999; Zinkevich, 2003), which studies the interactions between a learner (i.e. an algorithm) and an opponent through regret minimization, has proved to be a powerful framework for analyzing and designing iterative algorithms. However, while classic setups focus on bounding the worst case, many applications are not naturally adversarial. In this work, we aim to bridge this reality gap by establishing a new online learning setup that better captures certain regularity that appears in practical problems.

Formally, an online learning problem repeats the following steps: in round n, the learner plays a decision x_n from a decision set $\mathcal{X}$, the opponent chooses a loss function $l_n : \mathcal{X} \to \mathbb{R}$ based on the decisions of the learner, and then information about l_n (e.g. $\nabla l_n(x_n)$) is revealed to the learner for making the next decision. This abstract setup (Shalev-Shwartz et al., 2012; Hazan et al., 2016) studies the *adversarial* setting where l_n can be almost arbitrarily chosen except for minor restrictions like convexity. Often the performance is measured relatively through *static regret*,

$$\text{Regret}_N^s \coloneqq \sum_{n=1}^N l_n(x_n) - \min_{x\in\mathcal{X}} \sum_{n=1}^N l_n(x). \quad (1)$$

Recently, interest has emerged in algorithms that make decisions that are nearly optimal at every round. The regret is therefore measured on-the-fly and suitably named *dynamic regret*,

$$\text{Regret}_N^d \coloneqq \sum_{n=1}^N l_n(x_n) - \sum_{n=1}^N l_n(x_n^*), \quad (2)$$

where $x_n^* \in \arg\min_{x\in\mathcal{X}} l_n(x)$. As dynamic regret by definition upper bounds static regret, minimizing dynamic regret is a more difficult problem.

While algorithms with sublinear static regret are well understood, the research on dynamic regret is relatively recent. As dynamic regret grows linearly in the adversarial setup, most papers (Zinkevich, 2003; Mokhtari et al., 2016; Yang et al., 2016; Dixit et al., 2019; Besbes et al., 2015; Jadbabaie et al., 2015; Zhang et al., 2017) focus on how dynamic regret depends on certain variations of the loss sequence across rounds (such as the path variation $V_N = \sum_{n=1}^{N-1} \|x_n^* - x_{n+1}^*\|$). Even if the algorithm does not require knowing the variation, the bound is still written in terms of it. While tight bounds have been established (Yang et al., 2016), their results do not always translate into conditions for achieving sublinear dynamic regret in practice, because the size (i.e. budget) of the variation can be difficult to verify beforehand. This is especially the

Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108.

case when the opponent is *adaptive*, responding to the learner's decisions at each round. In these situations, it is unknown if existing results become vacuous or yield sublinear dynamic regret.

Motivated by the use of online learning to analyze iterative algorithms in practice, we consider a new setup we call Continuous Online Learning (COL), which directly models regularity in losses as part of the problem definition, as opposed to the classic adversarial setup that adds ad-hoc budgets. As we will see, this minor modification changes how regret and feedback interact and makes the quest of seeking sublinear dynamic regret well-defined and interpretable, even for adaptive opponents, without imposing variation budgets.

1.1 Definition of COL

A COL problem is defined as follows. We suppose that the opponent possesses a bifunction $f : (x, x') \mapsto f_x(x') \in \mathbb{R}$, for $x, x' \in \mathcal{X}$, that is *unknown* to the learner. This bifunction is used by the opponent to determine the per-round losses: in round n, if the learner chooses x_n, then the opponent responds with

$$l_n(\cdot) = f_{x_n}(\cdot). \tag{3}$$

Finally, the learner suffers $l_n(x_n)$ and receives feedback about l_n. For $f_x(x')$, we treat x as the *query argument* that proposes a question (i.e. an optimization objective $f_x(\cdot)$), and treat x' as the *decision argument* whose performance is evaluated. This bifunction f generally can be defined online as queried, with only the limitation that the same loss function $f_x(\cdot)$ must be selected by the opponent whenever the learner plays the same decision x. Thus, the opponent can be adaptive, but in response to only the learner's current decision.

In addition to the restriction in (3), we impose regularity into f to relate l_n across rounds so that seeking sublinear dynamic regret becomes well defined.[1]

Definition 1. We say an online learning problem is *continuous* if l_n is set as in (3) by a bifunction f satisfying, $\forall x' \in \mathcal{X}$, $\nabla f_x(x')$ is a continuous map in x [2].

The continuity structure in Definition 1 and the constraint (3) in COL limit the degree that losses can vary, making it possible for the learner to partially infer future losses from the past experiences.

The continuity may appear to restrict COL to purely deterministic settings, but adversity such as stochasticity can be incorporated via an important nuance in the relationship between loss and feedback. In the classic online learning setting, the adversity is incorporated in the loss: the losses l_n and decisions x_n may themselves be generated adversarially or stochastically and then they directly determine the feedback, e.g., given as full information (receiving l_n or $\nabla l_n(x_n)$) or bandit (just $l_n(x_n)$). The (expected) regret is then measured with respect to these intrinsically adversarial losses l_n. By contrast, in COL, we always measure regret with respect to the true underlying bifunction $l_n = f_{x_n}$. However, we give the opponent the freedom to add an additional stochastic or adversarial component into the feedback; e.g., in first-order feedback, the learner could receive $g_n = \nabla l_n(x_n) + \xi_n$, where ξ_n is a probabilistically bounded and potentially adversarial vector, which can be used to model noise or bias in feedback. In other words, the COL setting models a true underlying loss with regularity, but allows the adversary to be modeled within the feedback. This addition is especially important for dynamic regret, as it allows us to always consider regret against the true f_{x_n} while incorporating the possibility of stochasticity.

1.2 Examples

At this point, the setup of COL may sound abstract, but this setting is in fact motivated by a general class of problems and iterative algorithms used in practice, some of which have been previously analyzed in the online learning setting. Generally, COL describes the trial-and-error principle, which attempts to achieve a difficult objective $f_x(x)$ through iteratively constructing a sequence of simplified and related subproblems $f_{x_n}(x)$, similar to majorize-minimize (MM) algorithms. Our first application of this kind is the use of iterative algorithms in solving (stochastic) equilibrium problems (EPs) (Bianchi and Schaible, 1996). EPs are a well-studied subject in mathematical programming, which includes optimization, saddle-point problems, variational inequality (VI) (Facchinei and Pang, 2007), fixed-point problems (FP), etc. Except for toy cases, these problems usually rely on using iterative algorithms to generate ϵ-approximate solutions; interestingly, these algorithms often resemble known algorithms in online learning, such as mirror descent or Follow-the-Leader (FTL). In Sections 4 and 5, we will show how the residual function of these problems renders a natural choice of bifunction f in COL and how the regret of COL relates to its solution quality. In this example, it is particularly important to classify the adversary (e.g. due to bias or stochasticity) as feedback rather than as a loss function, to properly incorporate the continuity in the source problem.

Another class of interesting COL problems comes from optimization in episodic Markov decision processes (MDPs). In online imitation learning (IL) (Ross et al., 2011), the learner optimizes a policy to mimic an expert policy $\pi^\star$. In round n, the loss is $l_n(\pi) = \mathbb{E}_{s\sim d_{\pi_n}}[c(s, \pi; \pi^\star)]$, where d_{π_n} is the state distribu-

[1] Otherwise the opponent can define $f_x(\cdot)$ pointwise for each x to make $l_n(x_n) - l_n(x_n^*)$ constant.

[2] We define $\nabla f_x(x')$ as the derivative with respect to x'.

tion visited by running the learner's policy π_n in the MDP, and $c(s, \pi; \pi^\star)$ is a cost that measures the difference between a policy π and the expert $\pi^\star$. This is a bifunction form where continuity exists due to expectation and feedback is noisy about l_n (allowed by our feedback model). In fact, online IL is the main inspiration behind this research. An early analysis of IL was framed using the adversarial, static regret setup (Ross et al., 2011). Recently, results were refined through the use of continuity in the bifunction and dynamic regret (Cheng and Boots, 2018; Lee et al., 2018; Cheng et al., 2019b). This problem again highlights the importance of treating stochasticity as the feedback. We wish to measure regret with respect to the expected cost $l_n(\pi)$ which admits a continuous structure, but feedback only arrives via stochastic samples from the MDP. Structural prediction and system identification can be framed similarly (Ross and Bagnell, 2012; Venkatraman et al., 2015). Details, including new insights into the IL, can be found in Appendix F.

Lastly, we note that the classic fitted Q-iteration (Gordon, 1995; Riedmiller, 2005) for reinforcement learning also uses a similar setup. In the nth round, the loss can be written as $l_n(Q) = \mathbb{E}_{s,a\sim\mu_{\pi(Q_n)}}\mathbb{E}_{s'\sim\mathcal{P}(s,a)}[(Q(s,a) - r(s,a) - \gamma \max_{a'} Q_n(s',a'))^2]$, where $\mu_{\pi(Q_n)}$ is the state-action distribution[3] induced by running a policy $\pi(Q_n)$ based on the Q-function Q_n of the learner, and $\mathcal{P}$ is the transition dynamics, r is the reward, and γ is the discount factor. Again this is a COL problem.

1.3 Main Results

The goal of this paper is to establish COL and to study, particularly, conditions and efficient algorithms for achieving sublinear dynamic regret. We choose not to pursue algorithms with fast static regret rates in COL, as there have been studies on how algorithms can systematically leverage continuity in COL to accelerate learning (Cheng et al., 2019b,a) although they are framed as online IL research. Knowledge of dynamic regret is less well-known, with the exception of Cheng and Boots (2018); Lee et al. (2018) (both also framed as online IL), which study the convergence of FTL and mirror descent, respectively.

Our first result shows that achieving sublinear dynamic regret in COL is equivalent to solving certain EP, VI, and FP problems that are known to be PPAD-complete[4] (Daskalakis et al., 2009). In other words, we show that achieving sublinear dynamic regret that is polynomial in the dimension of the decision set can be extremely difficult.

Nevertheless, based on the solution concept of EP, VI, and FP, we show a reduction from monotone EPs to COL, and we present necessary conditions and sufficient conditions for achieving sublinear dynamic regret with polynomial dependency. Particularly, we show a *reduction* from sublinear dynamic regret to static regret and convergence to the solution of the EP/VI/FP. This reduction allows us to quickly derive non-asymptotic dynamic regret bounds of popular online learning algorithms based on their known static regret rates. Finally, we extend COL to consider partially adversarial loss and discuss open questions.

2 RELATED WORK

Much work in dynamic regret has focused on improving rates with respect to various measures of the loss sequence's variation. Zinkevich (2003); Mokhtari et al. (2016) showed the dynamic regret of gradient descent in terms of the path variation. Other measures of variation such as functional variation (Besbes et al., 2015) and squared path variation (Zhang et al., 2017) have also been studied. While these algorithms may not need to know the variation size beforehand, their guarantees are still stated in terms of these variations. Therefore, these results can be difficult to interpret when the losses can be chosen adaptively.

To illustrate, consider the online IL problem. It is impossible to know the variation budget *a priori* because the loss observed at each round of IL is a function of the policy selected by the algorithm. This budget could easily be linear, if an algorithm selects very disparate policies, or it could be zero if the algorithm always naively returns the same policy. Thus, existing budget-based results cannot describe the convergence of an IL algorithm.

Our work is also closely related to that of Rakhlin and Sridharan (2013); Hall and Willett (2013), which consider *predictable* loss sequences, i.e. sequences that are presumed to be non-adversarial and admit improved regret rates. The former considers static regret for both full and partial information cases, and the latter considers a similar problem setting but for the dynamic regret case. These analyses, however, still require a known variation quantity in order to be interpretable.

By contrast, we leverage extra structures of COL to provide interpretable dynamic regret rates, without *a priori* constraints on the variation. That is, our rates are internally governed by the algorithms, rather than externally dictated by a variation budget. This problem setup is in some sense more difficult, as achieving sublinear dynamic regret requires that both the per-round losses and the loss variation, as a function of the learner's decisions, be *simultaneously* small. Nonetheless, we can show conditions for sublinear dynamic regret using the bifunction structure in COL.

[3] Or some fixed distribution with sufficient excitation.

[4] In short, they are NP problems whose solutions are known to exist, but it is open as to if they belong to P.

3 PRELIMINARIES

We review background, in particular VIs and EPs, for completeness (Facchinei and Pang, 2007; Bianchi and Schaible, 1996; Konnov and Laitinen, 2002).

Notation Throughout the paper, we reserve the notation f to denote the bifunction that defines COL problems, and we assume $\mathcal{X} \subset \mathbb{R}^d$ is compact and convex, where $d \in \mathbb{N}_+$ is finite. We equip $\mathcal{X}$ with norm $\|\cdot\|$, which is not necessarily Euclidean, and write $\|\cdot\|_*$ to denote its dual norm. We denote its diameter by $D_{\mathcal{X}} \coloneqq \max_{x,x' \in \mathcal{X}} \|x - x'\|$.

As in the usual online learning, we are particularly interested in the case where $f_x(\cdot)$ is convex and continuous. For simplicity, we will assume all functions are continuously differentiable, except for $f_x(x')$ as a function over the querying argument x, where $x' \in \mathcal{X}$. We will use ∇ to denote gradients. In particular, for the bifunction f, we use ∇f to denote $\nabla f : x \mapsto \nabla f_x(x)$ and we recall, in the context of f, ∇ is always with respect to the decision argument. Likewise, given $x \in \mathcal{X}$, we use ∇f_x to denote $\nabla f_x(\cdot)$. Note that the continuous differentiability of $f_{x'}(\cdot)$ together with the continuity of $\nabla f_\cdot(x)$ implies ∇f is continuous; the analyses below can be extended to the case where $\nabla f_{x'}(\cdot)$ is a subdifferential.[5] Finally, we assume, $\forall x \in \mathcal{X}$, $\|\nabla f_x(x)\|_* \le G$ for some $G < \infty$.

Convexity For $\mu \ge 0$, a function $h : \mathcal{X} \to \mathbb{R}$ is called μ-strongly convex if it satisfies, for all $x, x' \in \mathcal{X}$, $h(x') \ge h(x) + \langle \nabla h(x), x' - x \rangle + \frac{\mu}{2}\|x - x'\|^2$. If h satisfies above with $\mu = 0$, it is called convex. A function h is called pseudo-convex if $\langle \nabla h(x), x' - x \rangle \ge 0$ implies $h(x') \ge h(x)$. These definitions have a natural inclusion: strongly convex functions are convex; convex functions are pseudo-convex. We say h is L-smooth if ∇h is L-Lipschitz continuous, i.e., there is $L \in [0, \infty)$ such that $\|\nabla h(x) - \nabla h(x')\|_* \le L\|x - x'\|$ for all $x, x' \in \mathcal{X}$. Finally, we will use Bregman divergence $B_R(x'||x) \coloneqq R(x') - R(x) - \langle \nabla R(x), x' - x \rangle$ to measure the difference between $x, x' \in \mathcal{X}$, where $R : \mathcal{X} \to \mathbb{R}$ is a μ-strongly convex function with $\mu > 0$; by definition $B_R(\cdot||x)$ is also μ-strongly convex.

Fixed-Point Problems Let $T : \mathcal{X} \to 2^{\mathcal{X}}$ be a point-to-set map, where $2^{\mathcal{X}}$ denotes the power set of $\mathcal{X}$. A fixed-point problem $\mathrm{FP}(\mathcal{X}, T)$ aims to find a point $x^\star \in \mathcal{X}$ such that $x^\star \in T(x^\star)$. Suppose T is λ-Lipschitz. It is called non-expansive if $\lambda = 1$ and λ-contractive if $\lambda < 1$.

Variational Inequalities VIs study equilibriums defined by vector-valued maps. Let $F : \mathcal{X} \to \mathbb{R}^d$ be a point-to-point map. The problems $\mathrm{VI}(\mathcal{X}, F)$ and $\mathrm{DVI}(\mathcal{X}, F)$ aim to find $x^\star \in \mathcal{X}$ and $x_\star \in \mathcal{X}$, respectively, such that the following conditions are satisfied:

$$\mathrm{VI} : \langle F(x^\star), x - x^\star \rangle \ge 0, \qquad \forall x \in \mathcal{X}$$
$$\mathrm{DVI} : \langle F(x), x - x_\star \rangle \ge 0, \qquad \forall x \in \mathcal{X}$$

VIs and DVIs are also known as Stampacchia and Minty VIs, respectively (Facchinei and Pang, 2007). The difficulty of solving VIs depends on the property of F. For $\mu \ge 0$, F is called μ-strongly monotone if $\forall x, x' \in \mathcal{X}$. $\langle F(x) - F(x'), x - x' \rangle \ge \mu\|x - x'\|^2$. If F satisfies the above with $\mu = 0$, F is called monotone. F is called pseudo-monotone if $\langle F(x'), x - x' \rangle \ge 0$ implies $\langle F(x), x - x' \rangle \ge 0$ for $x, x' \in \mathcal{X}$. It is known that the gradient of a (strongly/pseudo) convex function is (strongly/pseudo) monotone.

VIs are generalizations of FPs. For a point-to-point map $T : \mathcal{X} \to \mathcal{X}$, $\mathrm{FP}(\mathcal{X}, T)$ is equivalent to $\mathrm{VI}(\mathcal{X}, I - T)$, where I is the identity map. If T is λ-contractive, then F is $(1 - \lambda)$-strongly monotone.

Equilibrium Problems EPs further generalize VIs. Let $\Phi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a bifunction such that $\Phi(x, x) \ge 0$. The problems $\mathrm{EP}(\mathcal{X}, \Phi)$ and $\mathrm{DEP}(\mathcal{X}, \Phi)$ aim to find $x^\star, x_\star \in \mathcal{X}$, respectively, such that

$$\mathrm{EP} : \Phi(x^\star, x) \ge 0, \qquad \forall x \in \mathcal{X}$$
$$\mathrm{DEP} : \Phi(x, x_\star) \le 0, \qquad \forall x \in \mathcal{X}.$$

By definition, we have $\mathrm{VI}(\mathcal{X}, F) = \mathrm{EP}(\mathcal{X}, \Phi)$ if we define $\Phi(x, x') = \langle F(x), x' - x \rangle$.

We can also define monotonicity properties for EPs. For $\mu \ge 0$, Φ is called μ-strongly monotone if for $\forall x, x' \in \mathcal{X}$, $\Phi(x, x') + \Phi(x', x) \le -\mu\|x - x'\|^2$. It is called monotone if it satisfies the above with $\mu = 0$. Similarly, Φ is called pseudo-monotone if $\Phi(x, x') \ge 0$ implies $\Phi(x', x) \le 0$ for $x, x' \in \mathcal{X}$. One can verify that these definitions are consistent with the ones for VIs.

Primal and Dual Solutions We establish some basics of the solution concepts of EPs. As VIs are a special case of EPs, these results can be applied to VIs too. First, we have a basic relationship between the solution sets, $X^\star$ of EP and $X_\star$ of DEP.

Proposition 1. (Bianchi and Schaible, 1996) *If Φ is pseudo-monotone, $X^\star \subseteq X_\star$. If $\Phi(\cdot, x)$ is continuous $\forall x \in \mathcal{X}$, $X_\star \subseteq X^\star$.*

The proposition states that a dual solution is always a primal solution when the problem is continuous, and a primal solution is a dual solution when the problem is pseudo-monotone. Intuitively, we can think of the primal solutions $X^\star$ as local solutions and the dual solutions $X_\star$ as global solutions. In particular for VIs, if F is a gradient of some, even nonconvex, function,

[5] Our proof can be extended to upper hemicontinuity for set-valued maps, such as subdifferentials.

any solution in $X_\star$ is a global minimum; any local minimum of a pseudo-convex function is a global minimum (Konnov and Laitinen, 2002).

We note, however, that Proposition 1 does not directly ensure that the solution sets are non-empty. The existence of primal solutions $X^\star$ has been extensively studied. Here we include a basic result that is sufficient for the scope of our online learning problems with compact and convex $\mathcal{X}$.

Proposition 2. (Bianchi and Schaible, 1996) *If $\Phi(x,\cdot)$ is convex and $\Phi(\cdot,x)$ is continuous $\forall x \in \mathcal{X}$, then $X^\star$ is non-empty.*

Analogous results have been established for VIs and FPs as well. If F and T are continuous then solutions exist for both $\mathrm{VI}(\mathcal{X}, F)$ and $\mathrm{FP}(\mathcal{X}, T)$, respectively (Facchinei and Pang, 2007). On the contrary, the existence of dual solutions $X_\star$ is mostly based on assumptions. For example, by Proposition 1, $X_\star$ is non-empty when the problem is pseudo-monotone. Uniqueness can be established with stronger conditions.

Proposition 3. (Konnov and Laitinen, 2002) *If the conditions of Proposition 2 are met and Φ is strongly monotone, then the solution to $EP(\mathcal{X}, \Phi)$ is unique.*

4 EQUIVALENCE AND HARDNESS

We first ask what extra information the COL formulation entails. We present this result as an equivalence between achieving sublinear dynamic in COL and solving several mathematical programming problems.

Theorem 1. *Let f be given in Definition 1. Suppose $f_x(\cdot)$ is convex and continuous. The following problems are equivalent:*

1. *Achieving sublinear dynamic regret w.r.t. f.*
2. *$VI(\mathcal{X}, F)$ where $F(x) = \nabla f_x(x)$.*
3. *$EP(\mathcal{X}, \Phi)$ where $\Phi(x, x') = f_x(x') - f_x(x)$.*
4. *$FP(\mathcal{X}, T)$ where $T(x) = \arg\min_{x' \in X} f_x(x')$.*

Therefore, if there is an algorithm that achieves sublinear dynamic regret that in poly(d), then it solves all PPAD problems in polynomial time.

Theorem 1 says that, because of the existence of a hidden bifunction, achieving sublinear dynamic regret is essentially equivalent to finding an equilibrium $x^\star \in X^\star$, in which $X^\star$ denotes the set of solutions of the EP/VI/FP problems in Theorem 1. Therefore, a *necessary* condition for sublinear dynamic regret is that $X^\star$ is non-empty. Fortunately, this is true for our problem definition by Proposition 2.

Moreover, it suggests that extra structure on COL is necessary for algorithms to achieve sublinear dynamic regret that depends polynomially on d (the dimension of $\mathcal{X}$). The requirement of polynomial dependency is important to properly define the problem. Without it, sublinear dynamic regret can be achieved already at least asymptotically, e.g. by simply discretizing $\mathcal{X}$ (as $\mathcal{X}$ is compact and ∇f is continuous) and grid-searching, albeit with an exponentially large constant.

Due to space limitation, we defer the proof of Theorem 1 to Appendix A, along with other proofs for this section. But we highlight the key idea is to prove that the gap function $\rho(x) \coloneqq f_x(x) - \min_{x' \in X} f_x(x')$ can be used as a residual function for the above EP/VI/FP in Theorem 1. In particular, we note that, for the Φ in Theorem 1, $\rho(x)$ is equivalent to a residual function $r_{ep}(x) \coloneqq \max_{x' \in \mathcal{X}} -\Phi(x, x')$ used in the EP literature.

Below we discuss sufficient conditions on f based on the equivalence between problems in Theorem 1, so that the EP/VI/FP in Theorem 1 becomes better structured and hence allows efficient algorithms.

4.1 EP and VI Perspectives

We first discuss some structures on f such that the VI/EP in Theorem 1 can be efficiently solved. From the literature, we learn that the existence of dual solutions is a common prerequisite to design efficient algorithms (Konnov, 2007; Dang and Lan, 2015; Burachik and Millán, 2016; Lin et al., 2018). For example, convergence guarantees on combined relaxation methods (Konnov, 2007) for VIs rely on the assumption that $X_\star$ is non-empty. Here we discuss some sufficient conditions for non-empty $X_\star$, which by Proposition 1 and Definition 1 is a subset of $X^\star$.

By Proposition 1 and 2, a sufficient condition for non-empty $X_\star$ is *pseudo-monotonicity* of F or Φ (which we recall is a consequence of monotonicity). For our problem, the dual solutions of the EP and VI are *different*, while their primal solutions X^* are the same.

Proposition 4. *Let $X_\star$ and $X_{\star\star}$ be the solutions to $DVI(\mathcal{X}, F)$ and $DEP(\mathcal{X}, \Phi)$, respectively, where F and Φ are defined in Theorem 1. Then $X_{\star\star} \subseteq X_\star$. The converse is true if $f_x(\cdot)$ is linear $\forall x \in \mathcal{X}$.*

Proposition 4 shows that, for our problem, pseudo-monotonicity of Φ is stronger than that of F. This is intuitive: as the pseudo-monotonicity of Φ implies that there is $x_\star$ such that $f_x(x_\star) \leq f_x(x)$, i.e. a decision argument that is consistently better than the querying argument under the latter's own question, whereas the pseudo-monotonicity of F merely requires the intersection of the half spaces of $\mathcal{X}$ cut by $\nabla f_x(x)$ to be non-empty. Another sufficient assumption for non-empty $X_\star$ of VIs is that $\mathcal{X}$ is sufficiently strongly convex. This condition has recently been used to show fast convergence of mirror descent and conditional gradient

descent (Garber and Hazan, 2015; Veliov and Vuong, 2017). We leave this discussion to Appendix B.

The above assumptions, however, are sometimes hard to verify for COL. Here we define a subclass of COL and provide constructive (but restrictive) conditions.

Definition 2. We say a COL problem with f is (α, β)-regular if for some $\alpha, \beta \in [0, \infty)$, $\forall x \in \mathcal{X}$,

1. $f_x(\cdot)$ is a α-strongly convex function.
2. $\nabla f_\cdot(x)$ is a β-Lipschitz continuous map.

We call β the *regularity* constant; for short, we will also say ∇f is *β-regular* and f is (α, β)-regular. We note that β is different from the Lipschitz constant of $\nabla f_x(\cdot)$. The constant β defines the degree of online components; in particular, when $\beta = 0$ the learning problem becomes offline. Based on (α, β)-regularity, we have a sufficient condition to monotonicity.

Proposition 5. *∇f is $(\alpha - \beta)$-strongly monotone.*

Proposition 5 shows if $\nabla f_x(\cdot)$ does not change too fast with x, then ∇f is strongly monotone in the sense of VI, implying $X^\star = X_\star$ is equal to a singleton (but not necessarily the existence of $X_{\star\star}$). Strong monotoncity also implies fast linear convergence is possible for deterministic feedback (Facchinei and Pang, 2007). When $\alpha = \beta$, it implies at least monotonicity, by which we know $X_\star$ is non-empty.

We emphasize that the condition $\alpha \geq \beta$ is not necessary for monotonicity. The monotonicity condition of ∇f more precisely results from the monotonicity of $\nabla f_\cdot(x')$ and $\nabla f_x(\cdot)$, as $\langle \nabla f_x(x) - \nabla f_{x'}(x'), x - x' \rangle = \langle \nabla f_x(x) - \nabla f_x(x'), x - x' \rangle + \langle \nabla f_x(x') - \nabla f_{x'}(x'), x - y \rangle$. From this decomposition, we can observe that as long as the sum of $\nabla f_\cdot(x')$ and $\nabla f_x(\cdot)$ is monotone for any $x, x' \in \mathcal{X}$, then ∇f is monotone. In the definition of (α, β)-regular problems, no condition is imposed on $\nabla f_\cdot(x)$, so we need $\alpha \geq \beta$ in Proposition 5.

4.2 Fixed-point Perspective

We can also study the feasibility of sublinear dynamic regret from the perspective of the FP in Theorem 1. Here again we consider (α, β)-regular problems.

Proposition 6. *Let $\alpha > 0$. If $\alpha > \beta$, then T is $\frac{\beta}{\alpha}$-contractive; if $\alpha = \beta$, T is non-expansive.*

We see again that the ratio $\frac{\beta}{\alpha}$ plays an important role in rating the difficulty of the problem. When $\alpha > \beta$, an efficient algorithm for obtaining the the fixed point solution is readily available (i.e. by contraction) An alternative interpretation is that x_n^* changes at a slower rate than x_n when $\alpha > \beta$ with respect to $\|\cdot\|$.

5 MONOTONE EP AS COL

After understanding the structures that determine the difficulty of COL, we describe a converse result of Theorem 1, which converts monotone EPs into COL.

Theorem 2. *Let EP$(\mathcal{X}, \Phi)$ be monotone with $\Phi(x, x) = 0$.[6] Consider COL with $f_x(x') = \Phi(x, x')$. Let $\{x_n\}_{n=1}^N$ be any sequence of decisions and define $\hat{x}_N \coloneqq \frac{1}{N}\sum_{n=1}^N x_n$ It holds that $r_{dep}(\hat{x}_N) \leq \frac{1}{N}$Regret$_N^s$, where $r_{dep}(x') \coloneqq \max_{x \in \mathcal{X}} \Phi(x, x')$ is the dual residual. The same holds for the best decision in $\{x_n\}_{n=1}^N$.*

Theorem 2 shows monotone EPs can be solved by achieving sublinear static regret in COL, at least in terms of the dual residual. Below we relate bounds on the dual residual back to the primal residual, which we recall is given as $r_{ep}(x) \coloneqq \max_{x' \in \mathcal{X}} -\Phi(x, x')$.

Theorem 3. *Suppose $\Phi(\cdot, x)$ is L-Lipschitz, $\forall x \in \mathcal{X}$. If Φ satisfies $\Phi(x, x') = -\Phi(x', x)$, i.e. Φ is skew-symmetric, then $r_{ep}(x) = r_{dep}(x)$. Otherwise,*

1. *For $x \in \mathcal{X}$ such that $r_{dep}(x) \leq 2LD_{\mathcal{X}}$, it holds $r_{ep}(x) \leq 2\sqrt{2LD_{\mathcal{X}}}\sqrt{r_{dep}(x)}$.*
2. *If $\Phi(x, \cdot)$ is in addition μ-strongly convex with $\mu > 0$, for $x \in \mathcal{X}$ such that $r_{dep}(x) \leq L^2/\mu$, it holds $r_{ep}(x) \leq 2.8(L^2/\mu)^{1/3} r_{dep}(x)^{2/3}$*

We can view the above results as a generalization of the classic reduction from convex optimization and Blackwell approachability to no-regret learning (Abernethy et al., 2011). Generally, the rate of primal residual converges slower than the dual residual. However, when the problem is skew-symmetric (which is true for EPs coming from optimization and saddle-point problems; see Appendix C), we recover the classic results. In this case, we can show $r_{ep}(\hat{x}_N) = r_{dep}(\hat{x}_N) \leq \frac{1}{N}\text{Regret}_N^s \leq \frac{1}{N}\text{Regret}_N^d = \frac{1}{N}\sum_{n=1}^N r_{ep}(x_n)$.

These results complement the discussion in Section 4.1, as monotonicity implies the dual solution set $X_{\star\star}$ is non-empty. Namely, these monotone EPs constitute a class of source problems of COL for which efficient algorithms are available. Proofs and further discussions of this reduction are given in Appendix C.

6 REDUCTION BY REGULARITY

Inspired by Theorem 1, we present a reduction from minimizing dynamic regret to minimizing static regret and convergence to $X^\star$. Intuitively, this is possible, because Theorem 1 suggests achieving sublinear dynamic regret should not be harder than finding $x^\star \in X^\star$. Define $\text{Regret}_N^s(x^\star) \coloneqq \sum_{n=1}^N l_n(x_n) - l_n(x^\star) \leq \text{Regret}_N^s$.

Theorem 4. *Let $x^\star \in X^\star$ and $\Delta_n \coloneqq \|x_n - x^\star\|$. If f is (α, β)-regular for $\alpha, \beta \in [0, \infty)$, then for all N,*

[6] $\Phi(x, x) = 0$ is not a restriction; see Appendix C.

$$\begin{aligned}\mathit{Regret}_N^d \le \min\{G\textstyle\sum_{n=1}^N \Delta_n, \mathit{Regret}_N^s(x^\star)\} \\ + \textstyle\sum_{n=1}^N \min\{\beta D_{\mathcal{X}} \Delta_n, \frac{\beta^2}{2\alpha}\Delta_n^2\}\end{aligned}$$

If further $X_{\star\star}$ of the dual EP is non-empty, $\mathit{Regret}_N^d \ge \frac{\alpha}{2}\sum_{n=1}^N \|x_n^* - x_\star\|^2$, where $x_\star \in X_{\star\star} \subseteq X^\star$.

Theorem 4 roughly shows that when $x^\star$ exists (e.g. given by the sufficient conditions in the previous section), it provides a stabilizing effect to the problem, so the dynamic regret behaves almost like the static regret when the decisions are around $x^\star$.

This relationship can be used as a powerful tool for understanding the dynamic regret of existing algorithms designed for EPs, VIs, and FPs. These include, e.g., mirror descent (Beck and Teboulle, 2003), mirror-prox (Nemirovski, 2004; Juditsky et al., 2011), conditional gradient descent (Jaggi, 2013), Mann iteration (Mann, 1953), etc. Interestingly, many of those are also standard tools in online learning, with static regret bounds that are well known (Hazan et al., 2016).

We can apply Theorem 4 in different ways, depending on the known convergence of an algorithm. For algorithms whose convergence rate of Δ_n to zero is known, Theorem 4 essentially shows that their dynamic regret is at most $O(\sum_{n=1}^N \Delta_n)$. For the algorithms with only known static regret bounds, we can use a corollary.

Corollary 1. *If f is (α,β)-regular and $\alpha > \beta$, it holds that $\mathit{Regret}_N^d \le \mathit{Regret}_N^s(x^\star) + \frac{\beta^2 \widetilde{\mathit{Regret}}_N^s(x^\star)}{2\alpha(\alpha-\beta)}$, where $\widetilde{\mathit{Regret}}_N^s(x^\star)$ denotes the static regret of the linear online learning problem with $l_n(x) = \langle \nabla f_n(x_n), x\rangle$.*

The purpose of Corollary 1 is not to give a tight bound, but to show that for nicer problems with $\alpha > \beta$, achieving sublinear dynamic regret is not harder than achieving sublinear static regret. For tighter bounds, we still refer to Theorem 4 to leverage the equilibrium convergence. We note that the results in Section 5 and here concern different classes of COL in general, because $\alpha > \beta$ does not necessarily imply the $\mathrm{EP}(\mathcal{X}, \Phi)$ is monotone, but only $\mathrm{VI}(\mathcal{X}, F)$ unless $f_x(\cdot)$ is linear.

Finally, we remark Theorem 4 is directly applicable to expected dynamic regret (the right-hand side of the inequality will be replaced by its expectation) when the learner only has access to stochastic feedback, because the COL setup in non-anticipating. Similarly, high-probability bounds can be obtained based on martingale convergence theorems, as in (Cesa-Bianchi et al., 2004). In these cases, we note that the regret is defined with respect to l_n in COL, *not* the sampled losses.

6.1 Example Algorithms

We showcase applications of Theorem 4. These bounds are *non-asymptotic* and depend polynomially on d. Also, these algorithms do not need to know α and β, except to set the stepsize upper bound for first-order methods. Please refer to Appendix D for the proofs.

6.1.1 Functional Feedback

We first consider the simple greedy update, which sets $x_{n+1} = \arg\min_{x\in X} l_n(x)$. By Proposition 6 and Theorem 4, we see that if $\alpha > \beta$, it has $\mathrm{Regret}_N^d = O(1)$. For $\alpha = \beta$, we can use algorithms for non-expansive fixed-point problems (Mann, 1953).

Proposition 7. *For $\alpha = \beta$, there is an algorithm that achieves sublinear dynamic regret in poly(d).*

6.1.2 Exact First-order Feedback

Next we use the reduction in Theorem 4 to derive dynamic regret bounds for mirror descent, under deterministic first-order feedback. We recall that mirror descent with step size $\eta_n > 0$ follows

$$x_{n+1} = \arg\min_{x\in\mathcal{X}} \langle \eta_n g_n, x\rangle + B_R(x\|x_n). \tag{4}$$

where g_n is feedback direction, B_R is a Bregman divergence with respect to some 1-strongly convex function R. Here we assume additionally that $f_x(\cdot)$ is γ-smooth with $\gamma > 0$ for all $x \in \mathcal{X}$.

Proposition 8. *Let f be (α,β)-regular and $f_x(\cdot)$ be γ-smooth, $\forall x \in \mathcal{X}$. Let R be 1-strongly convex and L-smooth. If $\alpha > \beta$, $g_n = \nabla l_n(x_n)$, and $\eta_n < \frac{2(\alpha-\beta)}{L(\gamma+\beta)^2}$, then, for some $0 < \nu < 1$, $\mathit{Regret}_N^d \le (G + \beta D_{\mathcal{X}})\sqrt{2B_R(x^\star\|x_1)}\sum_{n=1}^N \nu^{n-1} = O(1)$ for (4).*

6.1.3 Stochastic & Adversarial Feedback

We now consider stochastic and adversarial cases in COL. As discussed, these are directly handled in the feedback, while the (expected) regret is still measured against the true underlying bifunction. Importantly, we make the subtle assumption that bifunction f is fixed before learning. We consider mirror descent in (4) with additive stochastic and adversarial feedback given as $g_n = \nabla l_n(x_n) + \epsilon_n + \xi_n$, where $\epsilon_n \in \mathbb{R}^d$ is zero-mean noise with $\mathbb{E}\left[\|\epsilon_n\|_*^2\right] < \infty$ and $\xi_n \in \mathbb{R}^d$ is a bounded adversarial bias. The component ϵ_n can come from observing a stochastic loss $l_n(x;\zeta_n)$ with random variable ζ_n, when the true loss is $l_n(x) = \mathbb{E}_{\zeta_n}[l_n(x;\zeta_n)]$ (i.e. $\nabla l_n(x_n;\zeta_n) = \nabla l_n(x_n) + \epsilon_n$). On the other hand the adversarial component ξ_n can describe extra bias in computation. We consider the expected dynamic regret $\mathbb{E}[\mathrm{Regret}_N^d] = \mathbb{E}[\sum_{n=1}^N l_n(x_n) - \min_{x\in\mathcal{X}} l_n(x)]$, where the expectation is over ϵ_n. Define $\Xi \coloneqq \sum_{n=1}^N \|\xi_n\|_*$. By reduction to static regret in Corollary 1, we have the following proposition.

Proposition 9. *If f is fixed before learning, $\alpha > \beta$ and $\eta_n = \frac{1}{\sqrt{n}}$, then mirror descent with $g_n = \nabla l_n(x_n) + \epsilon_n + \xi_n$ has $\mathbb{E}[\mathit{Regret}_N^d] = O(\sqrt{N} + \Xi)$.*

6.2 Remark

Essentially, our finding indicates that the feasibility of sublinear dynamic regret is related to a problem's properties. For example, the difficulty of the problem depends largely on the ratio $\frac{\beta}{\alpha}$ when there is no other directional information about $\nabla f_\cdot(x)$, such as monotonicity. When $\beta \leq \alpha$, we have shown efficient algorithms are possible. But, for $\beta > \alpha$, we are not aware of any efficient algorithm. If one exists, it would solve all (α, β)-regular problems, which, in turn, would efficiently solve all EP/VI/FP problems as we can formulate them into the problem of solving COL problems with sublinear dynamic regret by Theorem 1.

7 EXTENSIONS

The COL framework reveals some core properties of dynamic regret. However, while we allow feedback to be adversarial, we still assume that the same loss function $f_x(\cdot)$ must be returned by the bifunction for the same query argument $x \in \mathcal{X}$. Therefore, COL does not capture time-varying situations where the opponent's strategy can change across rounds. Also, this constraint allows the learner to potentially enumerate the opponent. Here we relax (3) and define a generalization of COL. The proofs of this section are included in Appendix E.

Definition 3. We say an online learning problem is (α, β)-*predictable* with $\alpha, \beta \in [0, \infty)$ if $\forall x \in \mathcal{X}$,

1. $l_n(\cdot)$ is a α-strongly convex function.
2. $\|\nabla l_n(x) - \nabla l_{n-1}(x)\|_* \leq \beta \|x_n - x_{n-1}\| + a_n$, where $a_n \in [0, \infty)$ and $\sum_{n=1}^N a_n = A_N = o(N)$.

This problem generalizse COL along two directions: 1) it makes the problem non-stationary; 2) it allows adversarial components within a sublinear budget inside the loss function. We note that the second condition above is different from having adversarial feedback, e.g., in Section 6.1.3, because the regret now is measured with respect to the adversarial loss as opposed to those generated by a fixed bifunction. This new condition can make achieving sublinear dynamic regret considerably harder.

Let us further discuss the relationship between (α, β)-predictable and (α, β)-regular problems. First, a contraction property like Proposition 6 still holds.

Proposition 10. *For (α, β)-predictable problems with $\alpha > 0$, $\|x_n^* - x_{n-1}^*\| \leq \frac{\beta}{\alpha}\|x_n - x_{n-1}\| + \frac{a_n}{\alpha}$.*

Proposition 10 shows that when functional feedback is available and $\frac{\beta}{\alpha} < 1$, sublinear dynamic regret can be achieved, e.g., by a greedy update. However, one fundamental difference between predictable problems and COL problems is the lack of equilibria X^*, which is the foundation of the reduction in Theorem 4. This makes achieving sublinear dynamic regret much harder when functional feedback is unavailable or when $\alpha = \beta$. Using Proposition 10, we establish some preliminary results below.

Theorem 5. *Let $\frac{\beta}{\alpha} < \frac{\alpha}{2L^2\gamma}$. For (α, β)-predictable problems, if $l_n(\cdot)$ is γ-smooth and R is 1-strongly convex and L-smooth, then mirror descent with deterministic feedback and step size $\eta = \frac{\alpha}{2L\gamma^2}$ achieves $\mathit{Regret}_N^d = O(1 + A_N + \sqrt{NA_N})$.*

We find that, in Theorem 5, mirror descent must maintain a sufficiently large step size in predictable problems, unlike COL problems which allow for decaying step size. When $\alpha = \beta$, we can show that sublinear dynamic regret is possible under functional feedback.

Theorem 6. *For $\alpha = \beta$, if $A_\infty < \infty$ and $\|\cdot\|$ is the Euclidean norm, then there is an algorithm with functional feedback achieving sublinear dynamic regret. For $d = 1$ and $a_n = 0$ for all n, sublinear dynamic regret is possible regardless of α, β.*

We do not know, however, whether sublinear dynamic regret is feasible when $\alpha = \beta$ and $A_\infty = \infty$. We conjecture this is infeasible when the feedback is only first-order, as mirror descent is insufficient to solve monotone problems using the last iterate (Facchinei and Pang, 2007) which contain COL with $\alpha = \beta$ (a simpler case than predictable online learning with $\alpha = \beta$).

8 CONCLUSION

We present COL, a new class of online problems where the gradient varies continuously across rounds with respect to the learner's decisions. We show that this setting can be equated with certain equilibrium problems (EPs). Leveraging this insight, we present a reduction from monotone EPs to COL, and show necessary conditions and sufficient conditions for achieving sublinear dynamic regret. Furthermore, we show a reduction from dynamic regret to static regret and the convergence to equilibrium points.

There are several directions for future research on this topic. Our current analyses focus on classical algorithms in online learning. We suspect that the use of adaptive or optimistic methods can accelerate convergence to equilibria, if some coarse model can be estimated. In addition, although we present some preliminary results showing the possibility for interpretable dynamic regret rates in predictable online learning, further refinement and understanding the corresponding lower bounds remain important future work. Finally, while the current formulations restrict the loss to be determined solely by the learner's current decision, extending the discussion to history-dependent bifunctions is an interesting topic.

References

Abernethy, J., Bartlett, P. L., and Hazan, E. (2011). Blackwell approachability and no-regret learning are equivalent. In *Annual Conference on Learning Theory*, pages 27–46.

Alexander, S., Bishop, R., and Ghrist, R. (2006). Pursuit and evasion in non-convex domains of arbitrary dimensions. In *Robotics: Science and Systems*.

Beck, A. and Teboulle, M. (2003). Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175.

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

Bianchi, M. and Schaible, S. (1996). Generalized monotone bifunctions and equilibrium problems. *Journal of Optimization Theory and Applications*, 90(1):31–43.

Burachik, R. S. and Millán, R. D. (2016). A projection algorithm for non-monotone variational inequalities. *arXiv preprint arXiv:1609.09569*.

Cesa-Bianchi, N., Conconi, A., and Gentile, C. (2004). On the generalization ability of on-line learning algorithms. *IEEE Transactions on Information Theory*, 50(9):2050–2057.

Cheng, C.-A. and Boots, B. (2018). Convergence of value aggregation for imitation learning. In *International Conference on Artificial Intelligence and Statistics*, pages 1801–1809.

Cheng, C.-A., Yan, X., Ratliff, N., and Boots, B. (2019a). Predictor-corrector policy optimization. In *International Conference on Machine Learning*, pages 1151–1161.

Cheng, C.-A., Yan, X., Theodorou, E. A., and Boots, B. (2019b). Accelerating imitation learning with predictive models. In *International Conference on Artificial Intelligence and Statistics*.

Dang, C. D. and Lan, G. (2015). On the convergence properties of non-euclidean extragradient methods for variational inequalities with generalized monotone operators. *Computational Optimization and applications*, 60(2):277–310.

Daskalakis, C., Goldberg, P. W., and Papadimitriou, C. H. (2009). The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259.

Dixit, R., Bedi, A. S., Tripathi, R., and Rajawat, K. (2019). Online learning with inexact proximal online gradient descent algorithms. *IEEE Transactions on Signal Processing*, 67(5):1338–1352.

Facchinei, F. and Pang, J.-S. (2007). *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media.

Garber, D. and Hazan, E. (2015). Faster rates for the frank-wolfe method over strongly-convex sets. In *32nd International Conference on Machine Learning, ICML 2015*.

Gordon, G. J. (1995). Stable function approximation in dynamic programming. In *Machine Learning Proceedings 1995*, pages 261–268. Elsevier.

Gordon, G. J. (1999). Regret bounds for prediction problems. In *Conference on Learning Theory*, volume 99, pages 29–40.

Hall, E. and Willett, R. (2013). Dynamical models and tracking regret in online convex programming. In *International Conference on Machine Learning*, pages 579–587.

Hazan, E. et al. (2016). Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325.

Jadbabaie, A., Rakhlin, A., Shahrampour, S., and Sridharan, K. (2015). Online optimization: Competing with dynamic comparators. In *Artificial Intelligence and Statistics*, pages 398–406.

Jaggi, M. (2013). Revisiting frank-wolfe: Projection-free sparse convex optimization. In *ICML (1)*, pages 427–435.

Jofré, A. and Wets, R. J.-B. (2014). Variational convergence of bifunctions: motivating applications. *SIAM Journal on Optimization*, 24(4):1952–1979.

Journée, M., Nesterov, Y., Richtárik, P., and Sepulchre, R. (2010). Generalized power method for sparse principal component analysis. *Journal of Machine Learning Research*, 11(Feb):517–553.

Juditsky, A., Nemirovski, A., and Tauvel, C. (2011). Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems*, 1(1):17–58.

Konnov, I. and Schaible, S. (2000). Duality for equilibrium problems under generalized monotonicity. *Journal of Optimization Theory and Applications*, 104(2):395–408.

Konnov, I. V. (2007). Combined relaxation methods for generalized monotone variational inequalities. In *Generalized convexity and related topics*, pages 3–31. Springer.

Konnov, I. V. and Laitinen, E. (2002). *Theory and applications of variational inequalities*. University of Oulu, Department of Mathematical Sciences.

Lee, J., Laskey, M., Tanwani, A. K., Aswani, A., and Goldberg, K. (2018). A dynamic regret analysis and adaptive regularization algorithm for on-policy robot imitation learning. In *Workshop on the Algorithmic Foundations of Robotics*.

Lin, Q., Liu, M., Rafique, H., and Yang, T. (2018). Solving weakly-convex-weakly-concave saddle-point problems as weakly-monotone variational inequality. *arXiv preprint arXiv:1810.10207*.

Mann, W. R. (1953). Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4(3):506–510.

Mokhtari, A., Shahrampour, S., Jadbabaie, A., and Ribeiro, A. (2016). Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 7195–7201. IEEE.

Nemirovski, A. (2004). Prox-method with rate of convergence o (1/t) for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251.

Rakhlin, A. and Sridharan, K. (2013). Online learning with predictable sequences. In *Conference on Learning Theory*, pages 993–1019.

Riedmiller, M. (2005). Neural fitted q iteration–first experiences with a data efficient neural reinforcement learning method. In *European Conference on Machine Learning*, pages 317–328. Springer.

Ross, S. and Bagnell, J. A. (2012). Agnostic system identification for model-based reinforcement learning.

Ross, S., Gordon, G., and Bagnell, D. (2011). A reduction of imitation learning and structured prediction to no-regret online learning. In *International conference on artificial intelligence and statistics*, pages 627–635.

Shalev-Shwartz, S. et al. (2012). Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194.

Sun, W., Venkatraman, A., Gordon, G. J., Boots, B., and Bagnell, J. A. (2017). Deeply aggrevated: Differentiable imitation learning for sequential prediction. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 3309–3318. JMLR. org.

Veliov, V. and Vuong, P. T. (2017). Gradient methods on strongly convex feasible sets and optimal control of affine systems. *Applied Mathematics & Optimization*, pages 1–34.

Venkatraman, A., Hebert, M., and Bagnell, J. A. (2015). Improving multi-step prediction of learned time series models. In *Conference on Artificial Intelligence*.

Yang, T., Zhang, L., Jin, R., and Yi, J. (2016). Tracking slowly moving clairvoyant: optimal dynamic regret of online learning with true and noisy gradient. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48*, pages 449–457. JMLR. org.

Zhang, L., Yang, T., Yi, J., Rong, J., and Zhou, Z.-H. (2017). Improved dynamic regret for non-degenerate functions. In *Advances in Neural Information Processing Systems*, pages 732–741.

Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning*, pages 928–936.