

1 We thank the reviewers for their thoughtful feedback. We are pleased that they found the online learning with dynamics
2 problem interesting and appreciated the generality of our analysis. Below, we provide clarifications to their concerns.

3 **Reviewer R1. – (Sufficient conditions for learnability)** As noted on Line 72, an online learning with dynamics
4 problem is said to be learnable if the policy regret $\text{Reg}_T^{\text{pol}} = o(T)$. Our main results in Theorem 1 and Proposition 2
5 show that whenever sequential Rademacher complexity of the underlying policy class Π as well as the dynamic stability
6 parameters are $o(T)$, the corresponding problem instance is learnable. These comprise the sufficient conditions for
7 learnability for the online learning with dynamics problem; we shall make this connection clearer in our updated draft.

8 – **(Lower bounds)** While the lower bound in Theorem 2 shows the existence of a particular instance, the results of
9 Proposition 3 hold for a broad range of problem instances – given any choice of decision functions \mathcal{F} , adversary
10 instance \mathcal{Z} and loss ℓ , there exists an instance with policy class $\Pi_{\mathcal{F}}$ and dynamics Φ such that the upper bounds on
11 regret are tight. These comprise the broad class of instances for which our learnability characterization is tight.

12 – **(Conditions for Minimax theorem)** Our application of the minimax theorem requires the space of probability
13 distributions \mathcal{P} and \mathcal{Q} to be weakly-compact (see [1, Appendix A] for a detailed discussion). Such an assumption is
14 fairly general – A simple corollary of Prokhorov’s theorem establishes that for *any* compact metric space (\mathcal{X}, d) , the
15 metric space $(\mathcal{P}(\mathcal{X}), d_P)$ is compact where d_P is the Prokhorov metric. We will elaborate further in our revised draft.

16 – **(Comparison with [22])** We would like to highlight that both our problem definition as well as our techniques for
17 obtaining the upper bounds generalize those of [22]. The presence of an underlying dynamics adds an additional
18 complexity to the problem which requires the dual analysis to balance loss minimization and dynamic stability. This
19 creates a strong coupling between the strategies at any time t with the future strategies - such a coupling and the
20 associated difficulties are absent in the dual analysis for the classical online learning framework. Indeed, the greedy dual
21 algorithm used in the analysis by [22] is dynamically unstable and would lead to $O(T)$ policy regret in our framework.

22 – **(Dynamics function)** We would like to highlight that the learner having access to the function Φ does not imply
23 that it knows the underlying dynamics. Indeed, one can encode all the aspects of the dynamics in the variable ζ_t and
24 Φ then acts as an applicator of these dynamics. Furthermore, recall that the learner gets to observe ζ_t only at time
25 t . For example, consider linear dynamical systems parameterized by matrices (A, B) . By letting $\zeta_t = (A_t, B_t)$, our
26 framework can capture time varying linear dynamical systems with the learner observing these matrices at time t .

27 – **(Examples)** We chose to highlight several different examples in Section 5 to demonstrate the generality of our results
28 and show that it recovers regret bounds for a wide class of problems, each of which is an individual research paper.

29 **Reviewer R2. – (Regularization term)** The boundedness for a large class of regularizers follows from the boundedness
30 of the underlying policy class – for e.g., policy classes with parameter belonging to a compact domain. Further, several
31 well-studied regularization functions (e.g. entropic penalty) are themselves bounded and do not affect learnability.

32 – **($T^{\frac{1}{2}}$ -regret for examples)** There are definitely example instances in our framework wherein the regret is not $T^{\frac{1}{2}}$; for
33 instance, our lower bound example construction (Line 295) has regret scaling as $T^{\frac{2}{3}}$. However, since our focus in the
34 examples section is on recovering regret bounds for classical problems, we recover rates scaling as $T^{\frac{1}{2}}$. One reason
35 why such rates are common in online learning is the connection of the sequential Rademacher complexity with uniform
36 convergence of martingale difference sequences in the corresponding Banach space (see [2] for details).

37 – **(Policy Def.)** A policy is a mapping from state space \mathcal{X} to a distribution over actions; we will include this in our draft.

38 **Reviewer R3. – (Lower bounds)** We would like to highlight that our lower bounds in Thm 2 (eq. (7a) and (7b)) can
39 actually be combined to obtain one which matches the upper bound of Thm 1 – we will update this in our draft. Given
40 this, we believe that obtaining such a lower bound in the general setup of Prop. 3(b) is an interesting future direction.

41 **Reviewer R4. – (Clarification regarding dual game)** We would like to clarify that the RERM we consider in the
42 proof of Theorem 1 is indeed for the dual game – it is straightforward to observe from equation 3 (Line 217) that the
43 policy at time t involves an expectation with respect to $z_t \sim p_t$, which is clearly not available in the primal game.
44 Further, it is a well-known fact that primal ERM (which is effectively a follow-the-leader algorithm) does not guarantee
45 sub-linear regret in the classical online learning setup with no dynamics and our paper definitely does not propose it as
46 a solution to the online learning with dynamics problem.

47 – **(Non-algorithmic framework)** We disagree with the assessment that a non-algorithmic complexity theoretic approach
48 to studying such a fundamental problem is not a sufficient contribution. There have been decades of work in the
49 field of learning theory and statistical machine learning which have studied fundamental problems (like classification,
50 regression) from a complexity perspective leading to characterization of learnability in the form of VC dimension,
51 Littlestone dimension, Rademacher complexity amongst others. We believe that the study of fundamental questions
52 concerning learnability is the first step towards systematic algorithmic approaches for these classes of problems and is
53 therefore an important contribution in the learning community.

54 [1] A. Rakhlin, K. Sridharan, and A. Tewari. Online learning via sequential complexities. JMLR 2015.

55 [2] A. Rakhlin, K. Sridharan, and A. Tewari. Sequential complexities and uniform martingale laws of large numbers. PTRF 2015.