Dear reviewers, we are deeply grateful for all your helpful feedback. Your reviews will be invaluable in revising the manuscript, regardless of its acceptance. We first address common concerns, and then particular reviewers. **Organization & Accessibility:** All reviewers expressed concerns about the accessibility of the paper to non-experts, and pointed out that the structure contributed to the obfuscation. The intent of the "our techniques" section was to explain, at an intuitive level, the main technical challenges we overcome. However, to the reader unfamiliar with the control-to-policy-regret reduction [3,12,24] , we now recognize this may be confusing. Moreover, the paper requires background from both the control theory community *and* the online learning community; and whereas are main results pertain to the former, are techniques and technical novelty relate to the latter. To better modularize the paper, we plan to restructure Section 2 to describe *only* the problem of attaining low policy regret for online convex optimization with memory. This requires limited notational overheard compared to the control setting. The section will both state and analyze our online algorithm optimization algorithm Semi-ONS. The "our techniques" section can then be woven into a complete proof, as the proof of logarithmic regret of Semi-ONS is quite direct. The full proof of $\sqrt{T}$ regret for approximate system knowledge will remain deferred to the appendix, but the sketch will be lengthened. Notably, this section will be accessible to online learning theorists without control knowledge, and will highlight the novelty of our techniques. Section 3 will apply Section 2 to the control problem. Having seen the OCO-with-memory formalism, the control-oriented reader will now understand the motivation for expressing the problem in the way we do. With the additional page of space, we will state formal black-box reductions from control-to-OCO-with-memory from the Appendix. The will allow readers unfamiliar with past work to understand how our innovations in online learning translate to control, even if they wish to skip the details of those regret bounds. The reader familiar with relevant prior work can skip the reduction. **Clarity, Typos, Experiments:** We apologize for the numerous grammatical errors, and mis-spliced sentences. The manuscript went through many rounds of restructuring, and we may not have caught all errors that arose as a consequence. We will be sure to adress all typos in the final version. Regarding experiments: past work on non-stochastic control does not include experiments, and we followed this convention. Nevertheless, we will attempt to include a simple demonstration in the final manuscript comparing Newton to Gradient-Based methods. Note that we propose Newton as an *learning procedure*, not simply an optimization subroutine. **Reviewer 1:** *(a)* See above. *(b)* We agree that, from a perspective of optimal control, online non-stochastic control is much harder than stochastic , as the tracking problem and known lower bounds elucidate. We initially had a sentence to that effect, which must have been mistakenly removed. We will be sure to clarify this point in further revision. However, for the narrow definition of non-stochastic control introduced by [3], where regret to a fixed benchmark of LTI controllers is defined as the objective, our paper does indeed demonstrate that that the regret rates for this problem coincide with stochastic. Secondly, at multiple points throughout the paper, we stress that while our algorithm uses a static K for the parametrization, our benchmark are linear *dynamic* controls with internal state, which may fare well in many tracking tasks (e.g. targets generated by an LDS). The arXiv of [24] describes numerous other classes of control policies compatible with the DRC formalism, which may be better suited to various tasks. *(c)* This is addressed briefly in lines 110-114, but can be expanded upon in the appendix. *(d)* the algorithm admits an efficient implementation of maintaining the matrix inverse via the Woodbury identity, we can discuss this the revision. **Reviewer 2:** *Typos:* This manuscript underwent several revisions, and we apologize for the numerous typos which remain. We will address all in the subsequent revision. *Motivation:* This paper is best categorized as at the intersection of reinforcement learning theory, control, and online learning, and in the absence of clear RL theory categories, we decided to list control as our subject area. We agree that the setting could be better motivated to a broader control; while connections to robustness are described in prior work, we shall be sure to reiterate them in the introduction to motivate our setting. *Contribution:* We dispute the claim that the contribution is an incremental improvement over [24]. Our main techniques resolve a standing question in online learning with memory, open since 2015: whether "fast" policy regret is obtainable for non-strongly convex unary losses, as in possible in standard online learning (see discussion at 139). Moreover, our bounds demonstrate that the same results attained by a long list of online LQR papers [1,9,10,20,22] are attainable with adversarial noise. *Modeling Assumptions:* The modeling assumptions were stated formally in Section 3. We will include signposts to these conditions in the revision, and in the restructing, these assumptions will be placed at the beginning of the newly proposed control section. *Crucial Identity* We will be sure to expound upon the "crucial indenity in future revions". *Conclusion:* Appendix B.3 contains detailed concluding remarks. We will signpost to this more carefully from the main text. **Reviewer 3:** (1) See discussion at 139 for by GD fails (2) yes, good catch, (3) replace i with n, (4) see Alg 3 for definition of how see $\mathcal{C}$ is set (the reference in the Alg 2 should be to Allg 3). This is shown to be efficient in [24], and can be made more practical with a slight relaxation. We shall discuss/clarify this further **Reviewer 4:** Should be $R_{\mathcal{M}} > 0$; For optimal parameter dependences, problem parameters must be known a-prior. However, the performance will degrade gracefully when parameters are misspecified; Unknown horizon can be adressed by the doubling trick (we can discuss this and resilience to parameter inaccuracies further in the appendix); adaptive exploration is challenging due to biases from closed loop control, but can be adressed by alternating rounds of explore and commit; yes, for certain partially observed systems, further assumptions may be required. However, one can show that if the assumptions of [24] hold (as in standard LQG), our algorithm naturally adapts of the induced strong convexity from semi-stochastic noise. Thus, we obtain the best of both worlds. We can sketch this as well.