Reviewer 1: Thank you for your constructive feedback. We hope our response below addresses your concerns:

• Indeed the constant α in the CR bound could exponentially increase as p increases, especially for marginally-stable systems and highly unstable systems. But we believe it is mainly from the system's property, instead of our algorithm's limitation. In other words, marginally stable or highly unstable A matrices are inherently harder to achieve small CR. As you mentioned, any online policy will suffer α^2 in the competitive ratio (Appendix E). It is interesting to understand which system will make online control harder (we partially discussed that in Appendix B, focusing on 1-D systems), and we will add more discussion about this point in the main paper.

• In online learning, it is common to have $q_{\min}$ and $q_{\max}$ in the CR or regret bounds (e.g.,[23, 31]). We also want to point out that even if q is fixed, bounding competitive ratio is non-trivial, and often it is unbounded [7].

• We agree our control setting is a subset of general LQR. We emphasize that our goal in the control part is to show the possibility for a policy to be competitive to the true, dynamic optimal offline cost, something not shown before. The only prior result is [23], which uses invertible B and known w_t at time t. This goal is important because all other prior works focus on achieving small regret compared to the best linear policy. Given our example in Appendix B, one may wonder if it is even possible to match the offline optimal. We give a positive answer. Though our setting is not fully general, it strictly generalized the prior art. We see this as a significant step towards competitive control.

• The most significant difference between our approach and the classical robust control (e.g., H_∞) is that H_∞ is neither online nor adaptive, i.e., the policy will not change even if some w_t is not adversarial. Our framework is naturally adaptive from the estimation set W_t (i.e., better estimation leads to more aggressive policy). We will add this discussion.

• Being optimistic is a common and powerful heuristic in general online learning, especially in regret minimization (e.g., UCB in multi-armed bandit, efficient Q-learning). But our optimistic idea is quite novel in SOCO, naturally because previous settings focus on precise information cases and then there is no need for being optimistic.

Reviewer 2: Thank you for your constructive feedback. We hope our response below addresses your concerns:

•We feel our assumption on strongly convexity and smoothness is not a significant weakness because: (1) Assuming strongly convexity and smoothness are very common in the online learning and optimization community [30, 31]; (2) Strong smoothness is not needed in [24] because the agent has the perfect prediction of the next hitting cost function, but it is critical in our setting where the prediction is imperfect, because the competitive ratio can be unbounded otherwise (e.g., consider the case when the hitting cost is an indicator function).

• Our setting strictly generalizes [23], where B is invertible and w_t is perfectly known at step t. To address the clarity issues, we will add more background and make notations easier to follow (e.g., add a notation list).

• For technical questions: (1) In line 207, we need to make an additional assumption that α is large (strictly speaking, larger than a constant c such that $c > 1$.) In this case, λ must be in the order of $O(m)$. (2) In line 153, when m tends to zero, the competitive ratio is of order $O(1/\sqrt{m})$ when $\alpha = 1$, and $O(1/m)$ if $\alpha > 1$. Hence we report $O(1/m)$ which is more conservative. (3) In line 150, it is not necessary to assume $\alpha^2 < m + 1$, because all assumptions in line 150 hold if we let $\lambda_2 = 0$ and $\lambda_1 = 2m/(m + \alpha^2 - 1 + \sqrt{(m + \alpha^2 - 1)^2 + 4m})$. We will discuss them in the revision.

Reviewer 3: Thank you for your constructive feedback. We hope our response below addresses your concerns:

• Previous competitive ratio results in SOCO focus on the setting when both the geometry and minimizer of f_t are revealed before the agent picks y_t [16, 23, 24]. In contrast, the minimizer is unknown when picking y_t in our setting. Our setting generalizes the previous ones, and is practical in many cases. For example, in power systems the geometry is governed by network topology (usually revealed before decision making) and minimizer is decided by users (which could be revealed afterwards). When reduced to control, the geometry is from cost functions and the minimizer is from adversarial disturbance. It is why we need p steps of future costs, but don't need any future disturbances. We want to emphasize (1) the access to future cost functions is common in control if the focus is on disturbances in dynamics (e.g., in LQ tracking problem the cost functions are pregiven) and (2) the only existing competitive policy [23] needs both future costs and disturbances, and we show a competitive policy exists even if disturbance is unknown in advance.

• Note that the cost bound in our Theorem 2 is $C_1 \cdot \mathrm{cost}(\mathrm{opt}) + (a + b - d) \sum_t \|v_t - \tilde{v}_t\|^2$, and we can get a pure competitive ratio $C_2 \cdot \mathrm{cost}(\mathrm{opt})$ because we have a "$-d$" term before the path length $\sum_t \|v_t - \tilde{v}_t\|^2$. We name it "beyond worst case" because the first bound will be tighter if the estimation is more precise. But note that we can always get a constant-competitive result even if the estimation is totally off, and C_2 does not depend on $\sum_t \|v_t - \tilde{v}_t\|^2$. Our numerical examples in Appendix C provide the example: if w_t is smooth (i.e., $\|w_{t+1} - w_t\|$ is small), then the path length is small and the first bound gets tighter (Fig. 1(b,d)). However, if the estimation set is large, the path length may get large and then the second bound would be better.

• For technical questions, we think the definition of g_t on page 14 is correct because the switching cost has the coefficient of $1/2$ by definition (between line 109 and 110). In Eq. (7), we should change t to i and q to p.

Reviewer 4: Thank you for the feedback on paper organization and presentation. We would move our examples in the main body (we will have one more page in the revision), add a notion table and restate all claims before the proofs.