Rev 1. We thank the reviewer for the positive feedback. (1) On local linear models: One possible approach is to apply RARL to each local linear model and then use the idea in guided policy search to piece together all local controllers. One issue is that there will not be a "global" robustness guarantee. But our results still hold locally. (2) Model-based or model-free?: With model-knowledge, other methods from classical robust-control, such as LMIs or Riccati equations can be used. However, these methods can hardly be made "model-free", and are less scalable (to high dim. systems) than our PG methods (which can also be made model-free via zeroth-order optimization methods). We assume here that the model is known, since our focus is on the fundamental issues regarding "optimization landscape" and "stability" in LQ RARL: model-based update already illustrates the landscape well; while sample-based update will only worsen the stability issue we've identified. We have mentioned this in lines 137-140, and will emphasize it in revision. (3) We will include the references. (4) The reviewer's suggestions on improvements are helpful. We will revise accordingly.

Rev 2. We thank the reviewer for the positive feedback. With an extra page allowed, we will be able to add conclusion and more details on related work. Regarding the restrictions of the state-feedback case, we agree that the output feedback case is important. The "optimization landscape" might become more challenging for the output-feedback case, and there has been little theoretical work even on "PG for output-feedback (non-robust) LQG". We leave this as an important future direction. Regarding the control requirements, in the robust control context we have studied, the frequency domain requirement on *"$\mathcal{H}_\infty$-norm"* is equivalent to some LMI or Riccati conditions in the time domain (see our Lemmas 3.4 and 5.2), and can thus be imposed. Imposing *other* control requirements in RARL is left for future research. On (A, B, C), we chose the matrices randomly, and made sure that they satisfy our assumptions.

Rev 3. We appreciate the detailed and positive comments, and hope that our response below addresses your concerns and help improve the scoring. (1) *Novelty:* We respectfully disagree that our contributions are incremental: (a) RARL in [30] is a highly-cited approach, and the stability issue of RARL that we've identified has been overlooked in (RA)RL; (b) the convergence theory and proof techniques are different from either [44] or [10]. Our assumptions are also different from [10] and [44], and align much better with the common assumptions in robust control. We note that even in the zero-sum LQ game context, our proofs are the first correct ones that carry rigorous *robust control* implications; (c) Our algorithms are not "minor extensions" of those in [44]. "(Robust) stability" has been handled in [44] through a rough "projection" step, which requires *model-knowledge*, and has no robust control implications; we had new and non-trivial techniques to remove the projection, enabling model-free algorithm-design; (d) empirically, the study of other descent-ascent PG methods, and how the joint effect of "initialization" and "update-rule" affects the convergence is new, while [10] did not provide any empirical results or any study on descent-ascent methods; (e) the "robust control implication" of the "good initializations" in LQ RARL (satisfying certain $\mathcal{H}_\infty$-norm constraint), and the "zeroth-order optimization-based" robustification algorithm are both novel, and cannot be found in either the (RA)RL or control literature (including [10,30,44]). (2) Convergence rates: Yes, only sublinear rates were established. For nonconvex optimization without additional problem structure, e.g., the gradient domination property of the objective for the inner-loop LQR, this *global sublinear* rate is something one can hardly improve in general. But note that in our simulations (Figure 4), convergence of this double-loop algorithm is not that bad (sublinear only in the beginning). We will add the runtime discussions. Also, note that the discussion in Sec. B.4 about faster local linear rates is not only for *Gauss-Newton (G-N)*, but also for *natural PG*, which can be made model-free using zeroth-order methods. Finally, we would like to clarify that when saying "G-N cannot be made model-free", we meant that the "zeroth-order optimization"-based methods cannot be used for G-N. It is still possible to apply other methods. For the (non-robust) LQR problem, the G-N method can be implemented in a model-free manner using the approximate policy iteration method where the policy evaluation step uses LSTD-Q. The RARL case is similar. We shouldn't have made it sound like a dead-end. We will add clarifications. (3) On stability issues in Sec. 3.2: Example 3.5 is a "best-response" update (with large enough N_K, N_L), which was discussed in the paragraph before it. We will add more details like lines 718-720. (4) Suggestions on clarity: We will revise accordingly. (5) Ref.: [30] is indeed highly-cited and we will add evidence. Thanks for mentioning the other references. We will include them. [R2] is very helpful for robustification.

Rev 4. Thanks for the comments. We hope our response will help for re-evaluating our work. (1) "Limited work": We respectfully disagree that our work is "limited". It is well-acknowledged that, LQ is the most fundamental and common setting in (robust) control, covering also scenarios where nonlinear, norm-bounded perturbations are allowed around a nominal linear system. To our knowledge, "PG for LQ setting" has only been well-studied for "non-robust LQR" problems, *but not* for zero-sum LQ games, or robust control. The most related work [10,44] has already been discussed in detail. See reply (1) to **Rev. 3**. We have novel technical improvement over [10,44]. (2) "No empirical study in the main paper": Figure 1 is an empirical result, and more results are available in the appendix, with clear pointers in the main paper. (3) "Limited related work": Could the reviewer specify which references? (4) Reproducibility: We have provided all the code and experiment details, and all other reviewers see our work as reproducible. Could the reviewer specify it is not reproducible in what sense? (5)*Assumptions:* Robust stability is not an assumption. Guaranteeing it in fact makes the analysis harder. Assump. A.1 on the existence of the solution to GARE is standard in robust control (cf. [2,5,37]), and is weaker than the direct assumption on A, B, C, see [5, Chapter 3]. Robust stability is a significant property in robust control, and is essential in our LQ RARL. It is not some "necessary condition".