We thank the reviewers for providing valuable comments. Below are point-to-point responses to the important questions.

Reviewer 1: Q1: Non-i.i.d. issue can be alleviated by experience replay. Markovian setting is not that significant.
A: We respectfully disagree with the reviewer. We agree that experience replay can be used in the offline Markovian setting, in which the optimization problem involves finite samples and essentially falls into the i.i.d. setup. In comparison, we study VRTDC in the online Markovian setting, which covers many real-world RL applications that have online nature, e.g., traffic control, online portfolio optimization, etc. We also believe that studying policy evaluation algorithms in the online Markovian setting has become an important fundamental topic for the RL theory community.

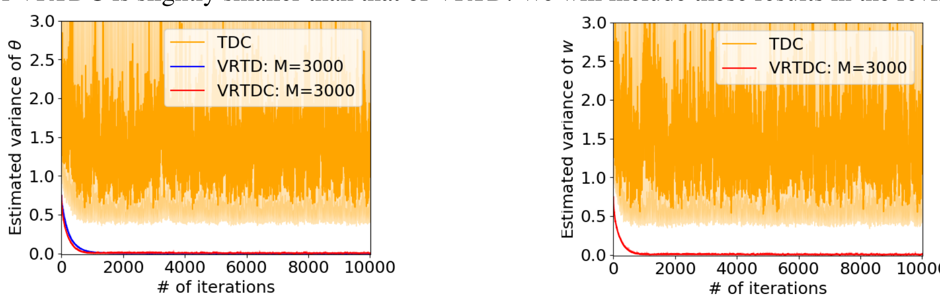
Q2: Keep problem's condition number in the complexity result.
A: Thanks for the suggestion. We will include all constants explicitly in the complexity result in the revision.

Q3: Summarize the bounds along with other existing algorithms' bounds in a table.
A: Thanks. We will add a table to compare our bounds with those of VRTD and TDC in i.i.d. and Markovian setting.

Reviewer 2: Q4: Intuition behind step-size choice $\alpha = O(\beta^{2/3})$?
A: For i.i.d. case, the inequality above Line 482 has the error term $c_1\beta + c_2(\alpha^2/\beta^2)$ for constants c_1, c_2. Minimizing it yields the desired learning rate $\alpha = O(\beta^{2/3})$. For the Markovian case, we obtain the same error term in Line 703.

Q5: Better to provide tight error bounds even though they do not lead an improved complexity.
A: Thanks for the suggestion and we totally agree. We will derive refined bounds for the Markovian setting and update them in the revision (it involves heavy computation that takes some time).

Q6: Empirically evaluate and compare the update variance.
A: In the following figure, we plot the estimated variance of the stochastic updates of θ (left) and w (right) for different algorithms. It can be seen that VRTDC significantly reduces the variance of TDC in both time-scales. Also, the variance of θ updates of VRTDC is slightly smaller than that of VRTD. We will include these results in the revision.



Reviewer 3: Q7: Emphasize importance of two time-scale, variance reduction and technical difficulties.
A: Thanks for the suggestions. We will emphasize and elaborate on these issues in the revision.

Reviewer 4: Q10: No strict sample complexity improvement in the Markovian case.
A: We agree with the reviewer, and our complexity result almost match the lower bound of linear two time-scale SA in Kaledin'20 (up to a logarithm factor).

Q11: The experimental results don't show a clear advantage.
A: In Figure 1 (left) and 2 (left), VRTDC achieves the highest solution accuracy, while TD and TDC cannot achieve a high accuracy. This is the desired effectiveness of variance reduction (i.e., can find high-accuracy solutions). From the above figure, it can be seen that VRTDC can significantly reduce the optimization variance of TDC in both time-scales.

Q12: Variance reduction idea is from CTD so there is no novel contribution.
A: VRTDC is the first variance reduction method for two time-scale Markovian TD learning. Our analysis requires to deal with the coupled θ and w, which leads to the new development of recursively refined error bounds to decouple these parameters and obtain the tight bounds.

Q13: Intuition behind the CTD and how it resulted in variance reduction should be added.
A: Thanks for the suggestion. We will discuss and elaborate on CTD with more details in the revision.

Q14: Discuss ETD. Can variance reduction be applied to ETD?
A: Thanks for pointing out the very interesting ETD method. Yes, one can still apply variance reduction to the ETD update. However, the main difference is that ETD is a one time-scale algorithm, and its update involves an emphasis factor F_t (i.e., the discounted interests of the states in history). The analysis will need to bound the variance and Markovian bias of ETD update in the presence of F_t. We expect that one needs to develop certain recursive bounds to address this issue. We will discuss ETD and cite related references in the revision.