To R1, R2, R4: *About the proof sketch:* Due to the 8 pages limit, we had to put the proof sketch in Appendix A (Page 12-14) and use the main text to highlight the contribution. In the camera ready version, we can reorganize and move the proof sketch into the main text, and also add references/pointers to the specific section in the appendix.

To R1: *The paper is hard to follow...:* We apologize we did not provide enough background. We will provide more explanation to those terminologies in the revision.
Contribution compared with previous works: The main contributions of our work include: (1) We are the first to provide a finite-time analysis for a practical actor-critic algorithm introduced in [25]; (2) Our analysis does not require the unrealistic i.i.d. assumption and directly deals with the Markov decision process; (3) Our analysis provides a better sample complexity $\mathcal{O}(\epsilon^{-2.5})$ than the best-known result $\mathcal{O}(\epsilon^{-4})$ in previous work under strong assumptions [16,21].
1.Why decoupled actor-critic assumption? The decoupled actor-critic is not an assumption, but a different algorithm appearing in [16] and [21]. It is not a practical algorithm, but easier to analyze. Our two time-scale actor-critic algorithm is more realistic and more sample efficient.
2.What is τ_t? τ_t is the mixing time of the Markov chain, which characterizes the time it takes the ergodic Markov chain in Assumption 4.2 to converge to its stationary distribution.
3.What is Markovian noise? Both actor and critic are updated based on observation tuples $\{O_t = (s_t, a_t, s_{t+1})\}_{t=0,1,...}$. In previous work [16, 21], they assume that each tuple is sampled i.i.d. in order to simplify the analysis. However, this is obviously not true in practice. In this paper, we follow [3] and directly deal with the true data which are sampled from the Markov decision process. We refer to this setting as the "Markovian noise" setting.
4.How is "iterative refinement" used? The "iterative refinement" is not used in our paper. It is used in [36] and more details can be found therein.
5.Proof of Corollary 4.9: The proof is in Section C.4, line 667.
6.Proof of Lemma B.3: This is a typo. All a_i should be a_k.
7.Proof of Lemma C.2 The proof is as follows:

$$\begin{aligned}\big\|\Delta h(O,\eta,\boldsymbol{\omega},\boldsymbol{\theta})\big\|^2 &:= \big(\eta(\boldsymbol{\theta}) - \eta + (\boldsymbol{\phi}(s') - \boldsymbol{\phi}(s))^\top(\boldsymbol{\omega}-\boldsymbol{\omega}^*)\big)^2 \cdot \big\|\nabla \log \pi_{\boldsymbol{\theta}}(a|s)\big\|^2 \\ &\le \big[2(\eta(\boldsymbol{\theta})-\eta)^2 + 2\big((\boldsymbol{\phi}(s') - \boldsymbol{\phi}(s))^\top(\boldsymbol{\omega}-\boldsymbol{\omega}^*)\big)^2\big]B^2 \\ &\le \big[2(\eta(\boldsymbol{\theta})-\eta)^2 + 2\big\|\boldsymbol{\phi}(s') - \boldsymbol{\phi}(s)\big\|^2\|\boldsymbol{\omega}-\boldsymbol{\omega}^*\|^2\big]B^2 \le \big[2(\eta(\boldsymbol{\theta})-\eta)^2 + 2\cdot 4\cdot\|\boldsymbol{\omega}-\boldsymbol{\omega}^*\|^2\big]B^2,\end{aligned}$$

where the equality is by the definition of $\Delta h(O,\eta,\boldsymbol{\omega},\boldsymbol{\theta})$, the first inequality is by $(a+b)^2 \le 2a^2+2b^2$ and Assumption 4.3(a), the second inequality is by Cauchy-Schwartz, and the last inequality is by triangle inequality and $\|\phi(s)\| \le 1$.

To R2: *Technical novelty in Remark 4.8:* Take the estimation error $\mathbf{z}_t$ as an example, the inequality at line 642 involves bounding $\mathbb{E}\|\mathbf{z}_t\|^2$ with $\mathbb{E}\|\mathbf{z}_t\|$ at its right hand side. Directly unrolling the equation at line 642 yields a loose result and a complicated proof, as done in [36]. We find it is viable to postpone the unrolling and compute the average estimation error which can give a tighter bound. We will elaborate it in Remark 4.8 in the revision.
Technical difference with [Y]: Thanks for pointing out the related works which we were not aware of previously. We will compare with them in the revision. The problem settings of both works are very different. In specific, [Y] considers updates that are linear in the two parameters $\boldsymbol{\theta}_t$ and $\boldsymbol{\omega}_t$. In contrast, the actor-critic updates in our paper is not a linear function of $\boldsymbol{\theta}_t$ and $\boldsymbol{\omega}_t$, i.e., the policy gradient update for $\boldsymbol{\theta}_t$. So the analysis in [Y] is not directly applicable to our setting, and our analysis on the actor $\boldsymbol{\theta}_t$ update requires a very different approach.
Where the first term in (4.3) and (4.4) come from? The first term in (4.3) comes from the term I_1 at line 648, which is related to the estimation error of last iterate $\boldsymbol{\omega}_t$. Similarly, the term in (4.4) comes from I_1 at line 615, which is related to the estimation error of the last iterate η_t for the average reward.

To R3: *The ϵ-approximate stationary point:* We will explicitly acknowledge the existence of function approximation error to avoid any confusion.
Compatible function approximation: It is possible to use compatible function approximation instead of a fixed linear function approximation. The potential difficulty is to efficiently estimate the Q function for a given state-action pair, which might involve starting another sampling trajectory.
Discounted setting: It is possible to extend our analysis to discounted MDPs. As you suggested, we can discarding each transition w.p. $1-\gamma$ and restarting the episode. We will add a discussion in the future work section.
Q-function or Advantage function: Our analysis is applicable to both advantage function $\Delta(s,a)$ and Q-function, with a very minor change in the analysis. We use the advantage function just following the convention of practice.
Regularized critic: Thank you for pointing out the related work [ICML2020] and suggesting this very promising idea. We will comment on this work and study the regularized critic in our future work.

To R4: *The proofs are extending over more than 20 pages and they are marked as "sketches":* This is a misunderstanding. To clarify, we actually have both sketches of proofs (Appendix A, pp. 12-14) and detailed proofs (Appendix C). So it is not the sketch that is lengthy.
About Theorem 4.5 The first term is the linear approximation error; the second term is from upper bounding the performance function; the third term is due to the stochastic variance and Markovian noise; and the last term is the critic's error. More details can be found in the proof sketch, at line 446.
Joint loss: In Open AI's implementation of A2C, the gradients of the joint loss w.r.t. the actor and the critic can actually be separated, so our analysis still holds.