We thank the reviewers for their thoughtful reviews; below we address their main concerns. While it only impacts our
LSVI analysis and our analysis of FRANCIS remains unchanged, we wish to note that in our own internal re-review we
found a small error in the proof of Lemma 23. We apologize for this. Fortunately we can address it by a relatively small
change: to define inherent Bellman error using the, perhaps more common, \(\infty\)-norm, i.e., if eq 2 (def 1) reads
\[
I_E = \max_{Q_{t+1} \in \mathcal{Q}_{t+1}} \min_{Q_t \in \mathcal{Q}_t} \max_{(s,a)} |[Q_t - T_t^\pi(Q_{t+1})](s,a)|.
\]
This allows us to express the misspecification error (e.g., eqn 37 in appendix) directly in every \((s,a)\) pair, as opposed to
in expectation, and the concentration argument in Lemma 23 (which is a concentration argument on deviation of the
encountered features wrt their expectation) is no longer needed. Note all the results of the paper continue to hold when
using the \(\infty\)-norm misspecification error in equation (\*) above.

Explorability (R1, R2, R4). We are sorry for the lack of clarity and we are happy to address our explorabil-
ity assumption and the relation to Chi et al’s bounds. Our analysis gives guarantees for the algorithm under
two sets of distinct assumptions, which we called implicit and explicit regularity in the paper (see def. 6 in app.).
1) Under implicit regularity, we do not put assumptions on the norm of reward parameter \(\|\theta^*\|_2\), but only a bound on
the expected value of the rewards under any policy; \(\mathbb{E}_{x_t \sim \pi_t(x_t)}[r(x_t)] \leq \frac{1}{H}\), (see line 762 in appendix). This
representation allows us to represent \textit{very high rewards} \((\gg 1)\) \textit{in hard-to-reach states}. It basically controls how big the
value function can get. This setting is more challenging for an agent to explore \textit{even in the tabular setting and even in
the case of a single reward function}. If a state is hard to reach, the reward there can be very high, and a policy that tries
to go there can \textit{still have high value}. Under this implicit regularity assumption, the explorability parameter would show
up for tabular algorithms as well (as minimum visit probability to any state under an appropriate policy): the classical
assumption that \(|r(s,a)| \leq 1\) would be replaced by \(|r(s,a)| \lesssim \frac{1}{\nu_{\min}}\), and the reward / transition noise would become
\(1/\nu_{\min}\) subgaussian; this would ultimately command a corresponding \(1/\nu_{\min}\) increase in sample complexity even for
tabular algorithms \textit{in the fixed reward setting as well as in the reward free setting}. Note that the results from Chi et al.
are derived under bounds on the reward parameter which do not satisfy this setting, and therefore our lower bounds are
not incompatible with their prior results for the tabular setting.
2) Under explicit regularity (Definition 6 in the appendix) we do make the classical assumption that bounds the
parameter norm \(\|\theta^*\|_2 \leq 1/H\) (line 767 in appendix). In this case, our lower bound no longer applies, but the proposed
algorithm still requires good “explorability” to proceed (in contrast to, e.g., LSVI-UCB as several reviewers have
noticed). We then completely agree with R1, R2 and R4 that the assumption should not be necessary in this case, and
removing it is certainly very important, but given the already existing challenges brought by the inherent Bellman
error setting we have to leave this as future work. Nonetheless, we would like to point out that explorability does not
make the exploration problem trivial. A random policy (i.e., \(c\)-greedy) can still take exponential time to learn; other
authors \cite{1,5} have made similar or even stronger assumptions (a bound on the minimum visit probability to any state)
to proceed in the context of function approximation.

Layer-by-layer learning (R4) As the reviewer suggests, an algorithm that is able to learn all layers together might be
more horizon efficient. Interestingly, with our approach we do already achieve the same horizon dependence as the state
of the art \cite{3} for tabular RL at the time of submission \(O(S^2AH^5/\epsilon^2)\) (see also table 1 in our manuscript).

Motivation (R4) With this work our aim was to try to weaken the assumptions the current literature makes in the
exploration setting with linear function approximation. Although the minimum set of assumptions for RL to work are
not well understood even in the linear case \cite{2}, our objective is to at least have algorithms that work under assumptions
typically made in the batch setting, i.e., when mainstream batch algorithms like least square value (or policy) iteration
work \cite{6,4} using already collected data. We consider this work as a first step in this direction.

Clarity (R2) We thank the reviewer for highlighting the clarity issue, and we do plan to use the extra page that is
allowed for the camera ready to expand the motivation behind the algorithm, to give a more detailed proof sketch and to
clarify the role of explorability, which is especially important as the algorithm style is different than those presented in
a number of recent theoretical papers.

References
\begin{itemize}
\end{itemize}