We greatly appreciate the feedback of the reviewers. We discuss the specific concerns of the reviewers below.

Reviewer 1:

- A theoretical upper-bound of the regret of Approx-Zooming-With-No-Arm-Similarity is stated in [7] as $O(KT^{\frac{d+1}{d+2}})$, where d is the dimension of the context space. Note that this regret bound is linear in the number of arms. We will include this discussion into the paper.
- We will include empirical results of a gaussian process-based bandit in the final paper. As one needs to specify the covariance/kernel matrix, we could either fit a Gaussian process to each arm separately when the metric is unknown, or we can construct the covariance matrix given oracle access to a metric among the arms.
- Generalizing to higher dimension: Yes, indeed our algorithm and analysis does extend to the general d-dimensional setting, and we will include that into the final paper. The only change required algorithmically is in the subpartitioning/clustering step. Let us define $C_d(q)$ to be the number of balls of radius r/q needed to cover a ball of radius r, which scales exponentially in the dimension d, e.g. q^d. Since we are now estimating the reward function f over a d-dimensional context space, the number of sub-regions of the context space that need to be clustered will be $C_d(2)$, and the number of samples needed to guarantee that the k-nearest neighbor samples are within distance $\frac{1}{16}$radius, will be equal to $\tilde{O}(kC_d(32))$. To compute $\hat{\mathcal{D}}$, we will instead have a d-dimensional summation over the subset of the context space. Once $\hat{\mathcal{D}}$ is computed, then the clustering of arms will have the same computational cost, i.e. linear in number of arms to be clustered. The analysis can be modified to account for the d-dimensional setting, and the final regret bound will look like

$$O\Big(C_d(2)C_d(32)\sigma^2L^{-2}K\ln(TK) + \min_{i_{\max}\in\mathbb{Z}_+}\big(LT2^{-i_{\max}} + \sum_{i=1}^{i_{max}-1} C_d(2)C_d(32)\sigma^2L^{-1}M_i2^i\ln(TK)\big)\Big),$$

  where M_i instead sums over an ϵ-net of the context space for $\epsilon = 2^{-i}$, and thus we may expect M_i to grow exponentially in $i \times d$, although modified with respect to the distribution of the reward function and the finite arms. The growth of M_i will dominate the regret bound with respect to the dependence on the dimension d.

Reviewer 2:

- Line 160: We will include an example, in fact the setup we chose for the simulation illustrates this as it is a periodic function that could be rearranged to be simply linear rather than the depicted zigzag.
- Adapting to the smoothness parameter of the reward function: Indeed our algorithm can be generalized to Holder continuous reward functions. If the smoothness parameter is known, then modifying the algorithm is straightforward. We will look into the techniques of Qian and Yang (2016) for adaptivity to the smoothness.
- Generalizing to higher dimension: see response to reviewer 1 above.

Reviewer 3:

- Regret bound scaling with Lipschitz constant: There are two terms in the final regret bound. The first term comes from the cost of the initial clustering, and as you pointed out it scales inversely with L^2. However, the first term scales as $\tilde{O}(K)$, logarithmic in T, whereas the second term has a polynomial dependence on T, e.g. $\tilde{O}(\sqrt{KT})$ in the given examples. For large T and $K = o(T)$, the second term will dominate, which does not scale inversely with the Lipschitz constant. To give more inutition though, the inverse dependence on L in the first term (only logarithmic scaling wrt T) is due to the fact that the clusters are constructed to satisfy that the bias in the reward due to different arms in the same cluster is controlled to be on the same order as the bias in reward from different contexts in the same set. For small Lipschitz constant, the reward varies less across the same size context width, and thus the algorithm requires that arm distances are measured more precisely to guarantee that the bias of the new ball is bounded by L times the context width. This initial clustering of the algorithm could be modified to allow for looser clusters, where the bias due to different arms in a cluster is larger than the context width. This would remove the inverse dependence on L in the first term, but would add more phases of subpartitioning that would contribute regret to the summation in the second term.
- The regret bound obtained in the contextual zooming algorithm when the metric among the arms is available is $O(T^{\frac{d+1}{d+2}})$, where d is the dimension of the joint context-arm space, i.e. $d = d_c + d_a$ where d_c is the dimension of the context space and d_a is the dimension of the arm space. We will include this into the final paper.

We will also address the minor typos/comments in the revision as well. Thank you for your detailed feedback!