Response to Reviewer TGdm

We sincerely thank the reviewer for the time and constructive comments. We find the comments all very helpful. We write in below responses to each of the comment. Each item starts with the original comment by the reviewer and follows with our response

1. **Comment:** Is there any assumption on the neural network $g(\theta)$? If there is no assumption, then it seems $g(\theta)$ can be any reward?

   Response: We thank the reviewer for the helpful comments. We require that $g(\theta)$ is continuous regarding θ to ensure that the kernel function with respect to θ adopts a Mercer decomposition. This assumption can be satisfied by selecting continuous activation functions of the neural network, for example, the ReLU function. In addition, in order to derive the theoretical upper bound of the cumulative regret in Theorem 4.2, we assume that the unknown mapping $g(\theta)$ is exactly captured by the neural network and known.

2. **Comment:** It seems the regret upper bound is sub-optimal, and there is not discussion on lower bound results.

   Response: We thank the reviewer for the helpful comments. Our work considers the scenarios when the reward function $f(\mathbf{x};\theta)$ regarding the decision variable $\mathbf{x}$ is approximated by the Gaussian process (GP) model. Based on this assumption, our work provides a bound of the regret $\tilde{\mathcal{O}}\left(\sqrt{T\gamma_T}\right)$ that is consistent with the relevant literature (Srinivas et al. 2010, Krause and Ong 2011); see Vakili et al. (2021) for a review. This upper bound depends on the maximal information gain γ_T, and we also provide that the information gain of NN-AGP is smaller than a composite GP, especially when the contextual variable is high-dimensional. This advantage is also supported by the experimental results in Section D.2.2.

   Before discussing the lower bound of cumulative regret, we first note that the reward function $f(\mathbf{x};\theta)$ in our work can be assumed as either 1) a sample of a GP or 2) a deterministic function that lives in Hilbert space, which is consistent with relevant literature (Srinivas et al. 2010). The difference between modeling the reward function as a GP sample or an element in an RKHS reflects the difference between Bayesianists and frequentists, as discussed in Srinivas et al. (2010). In our work, we adopt a Bayesian view since it helps us better understand the construction of the acquisition function.

In terms of the lower bound, to our knowledge, neither our work nor the existing literature has provided the lower bound when the objective function is regarded as a sample from a GP; see also Section A.5 of Cai and Scarlett (2021) as a summary. On the other hand, when taking a frequentist view that the reward function is from an RKHS, existing results focus on two specific commonly-selected kernels, squared exponential (SE) and Matern

$$k_{\mathrm{SE}}\left(\mathbf{x}, \mathbf{x}'\right) = \exp\left(-\frac{\left\|\mathbf{x}-\mathbf{x}'\right\|^2}{2l^2}\right)$$
$$k_{\text{Matérn}}\left(\mathbf{x}, \mathbf{x}'\right) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}\left\|\mathbf{x}-\mathbf{x}'\right\|}{l}\right)^{\nu} J_\nu\left(\frac{\sqrt{2\nu}\left\|\mathbf{x}-\mathbf{x}'\right\|}{l}\right),$$

where $l > 0$ denotes the length-scale, $\nu > 0$ is a smoothness parameter, and J_ν is the modified Bessel function. We note that existing results on the lower bound in Scarlett et al. (2017) can be employed to attain the lower bound of NN-AGP-UCB, since NN-AGP maintains a GP structure regarding the decision variable $\mathbf{x}$.

Specifically, we instead assume that the reward function lives in a reproducing kernel Hilbert space (RKHS). To simplify notation, we follow the model assumption in Section 4.3 that $\mathbf{p}(\mathbf{x}) = au(\mathbf{x})$, where $a \in \mathbb{R}^m$ and $u(\mathbf{x}) \in \mathcal{H}_k$. Here $\mathcal{H}_k$ represents an RKHS associated with the kernel function k. We assume that the RKHS norm of u is bounded, that is $\|u\|_k \leqslant B$. In this way, when $g(\theta)$ and a are exactly known (as assumed in **Theorem 4.2** in our work), we have that

$$\mathbb{E}\left[R_T\right] = \Omega\left(\sqrt{T\sigma_\epsilon^2\left(\log\frac{U^2B^2T}{\sigma_\epsilon^2}\right)^{d/2}}\right)$$

when the kernel function of $u(\mathbf{x})$ is the SE kernel as a representative. Here, R_T is the cumulative regret up to time T, the expectation is taken over the noise $\epsilon_t \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$, d denotes the dimension of the decision variable $\mathbf{x}$, and $U = \sup_{\theta\in\Theta} a^\top g(\theta)$. In summary, when the kernel function k of NN-AGP is assumed as SE,

$$\text{lowerbound} = \Omega\left(\sqrt{T(\log T)^{d/2}}\right)$$
$$\text{upperbound} = \tilde{\mathcal{O}}\left(\sqrt{T(\log T)^{d}}\right).$$

We will include the discussion on the frequentist setting and contain the details of this lower bound in the revised manuscript.

Again, we appreciate the valuable comments provided by the reviewer. We will make the necessary revisions to the manuscript based on your suggestions.

References

Cai, X. and Scarlett, J. (2021). On lower bounds for standard and robust gaussian process bandit optimization. In *International Conference on Machine Learning*, pages 1216–1226. PMLR.

Krause, A. and Ong, C. (2011). Contextual gaussian process bandit optimization. *Advances in neural information processing systems*, 24.

Scarlett, J., Bogunovic, I., and Cevher, V. (2017). Lower bounds on regret for noisy gaussian process bandit optimization. In *Conference on Learning Theory*, pages 1723–1742. PMLR.

Srinivas, N., Krause, A., Kakade, S., and Seeger, M. (2010). Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning*, number CONF. Omnipress.

Vakili, S., Khezeli, K., and Picheny, V. (2021). On information gain and regret bounds in gaussian process bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 82–90. PMLR.