We thank the reviewers for their constructive comments.

———-REVIEWER 1———-

 # What if $|\sqrt{\beta_t}\theta - \min_{x\in\mathbb{R}^d} \alpha_{UCB}(x;\mathcal{D}_{t-1})| = 0$?

If $|\sqrt{\beta_t}\theta - \min_{x\in\mathbb{R}^d} \alpha_{UCB}(x;\mathcal{D}_{t-1})| = 0$, then $\forall\epsilon < |\sqrt{\beta_t}\theta - \max_{x\in\mathbb{R}^d} \alpha_{UCB}(x;\mathcal{D}_{t-1})|$ or $< |\sqrt{\beta_t}\theta - \max_{x\in\mathcal{D}_{t-1}} \alpha_{UCB}(x;\mathcal{D}_{t-1})|$, the bound in Theorem 4.1 remains valid. As in this case, the GP-UCB argmax is at a finite location and its acquisition function value $> \sqrt{\beta_t}\theta$, thus our arguments in Case 2 of Section 1.3 in the Supplementary material hold. However, it is worth noting that the scenario $|\sqrt{\beta_t}\theta - \min_{x\in\mathbb{R}^d} \alpha_{UCB}(x;\mathcal{D}_{t-1})| = 0$ is very rare in practice. With Assumption 4.1, most of the time, $\min_{x\in\mathbb{R}^d} \alpha_{UCB}(x;\mathcal{D}_{t-1}) \leq 0$, hence $|\sqrt{\beta_t}\theta - \min_{x\in\mathbb{R}^d} \alpha_{UCB}(x;\mathcal{D}_{t-1})| > 0$. Only when the noise is large, there is a very small chance $|\sqrt{\beta_t}\theta - \min_{x\in\mathbb{R}^d} \alpha_{UCB}(x;\mathcal{D}_{t-1})| = 0$ can happen.

 # Derivation of Eq. (14)

The two terms on the RHS of Eq. (13) are monotone increasing functions, and γ is smaller than both numbers in the set, hence each term in (13) is smaller than the value of the corresponding function operating on each number.

 # A) Derivation of Eq. (15). B) Did you consider the probability for $\max_{x\in\mathcal{D}_t} \alpha_{LCB} \leq \max_{x\in\mathcal{D}_t} f(x)$ to hold?

A) There is a typo in (15), the 1st inequality should be $\max_{x\in\mathcal{S}_k} f(x) - \max_{x\in\mathcal{D}_t} f(x) \leq \mu_{t-1}(x_t) + \beta_t^{1/2}\sigma_{t-1}(x_t) + 1/t^2 - \max_{x\in\mathcal{D}_t} f(x)$. This follows Lemmas 5.7 and 5.8's proof in Srinivas et al [19]. For the 2nd inequality, with the chosen β_t, following Lemma 5.5 in Srinivas et al [19], $\max_{x\in\mathcal{D}_t} \alpha_{LCB}(x;\mathcal{D}_{t-1}) \leq \max_{x\in\mathcal{D}_t} f(x)$. B) Yes. The chosen β_t ensures $\alpha_{LCB} \leq f(x) \leq \alpha_{UCB}$ $(x \in \mathcal{D}_t)$ with probability $\geq 1-\delta$ (Lemmas 5.5, 5.7 in Srinivas et al [19]).

 # The values of a_k and b_k in our experimental settings

We set $a_k = 1$ and derive an expression for b_k based on the kernel hyper-parameters and the search space size. The exact formula is in lines 141-145 of our script BO_unknown_searchspace_good.py or other GP-UCB based scripts.

 # The relation of iteration T and ϵ

Using our Lemma 5.1's proof, Lemma 5.8 and Theorem 2's proof in Srinivas et al [19], the regret bound $r_b(T) \leq \sqrt{C_1\beta_T\gamma_T}/T + 2/T$ with probability $\geq 1-\delta$. Thus, for any $T \geq (\sqrt{C_1\beta_T\gamma_T} + 2)/\epsilon$, the ϵ-*accuracy* is satisfied in each search space. Hence, we can see that when ϵ is smaller, T needs to be larger to guarantee the ϵ-*accuracy* condition.

———-REVIEWER 2———-

 # Clarity and organization must be improved according to detailed comments

Yes, we will, using all the comments and suggestions from all the reviewers.

 # Instead of being an expansion of UCB, if it was general, it would have ranked higher

GP-UCB is chosen as it has ability to analyze convergence, which is very important in the unknown search space setting. Note that it is still possible to use GP-UCB to define the expanded search space, then other acquisition functions can be subsequently used in this expanded search space to find optimal. However, in this case, the ϵ-*accuracy* cannot be guaranteed (e.g. EI convergence can be shown only in noiseless setting, PI/ES/PES do not have convergence proof yet).

 # The importance of our work

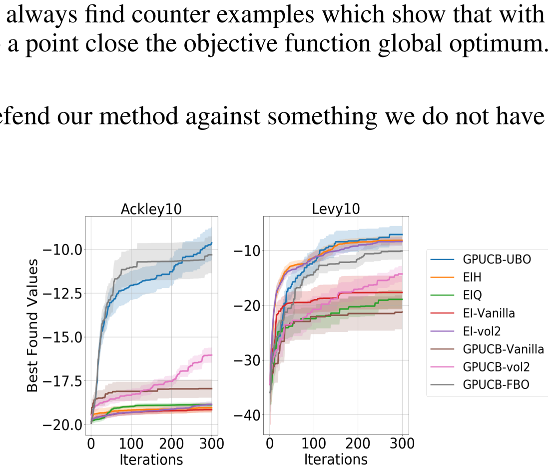
Our major contributions are: 1) formalizing the convergence analysis of Bayesian optimization when search space is unknown; and 2) proposing an effective algorithm that guarantees ϵ-*accuracy* convergence. To the best of our knowledge, there is no previous work that guarantees its solution converges to a point close to the objective function global optimum when the search space is unknown. In fact, we can always find counter examples which show that with high probability, the solutions of previous works do not converge to a point close the objective function global optimum.

 # Comparison with other NeurIPS submitted papers this year

This comparison disadvantages us as we are not in a position to defend our method against something we do not have access to. Thank you for the compliment though.

———-REVIEWER 3———-

 # More experiments with higher dimensional function (d=10)



We have conducted more experiments with the 10-dimensional functions Ackley10 and Levy10. The setups are same as in the paper. For Ackley10, #experiments=30 whilst for Levy10, #experiments=10 as GPUCB-FBO computation time for Levy10 is so expensive that we can only get 10 experiments during the whole rebuttal time (GPUCB-FBO's average runtime is 239.83 seconds/iteration while our method average runtime is 21.15 seconds/iteration). For Ackley10, our proposed method outperforms other 6 methods by a high margin and is better than GPUCB-FBO, and, note that GPUCB-FBO computation time is at least 5-6 times slower than our method. For Levy10, our proposed method is slightly better than EIH, EI-vol2 whilst outperforming other baselines significantly.