We thank the reviewer for the detailed discussion and insights, which have helped us improve the presentation of our work.

Regarding point 1 mentioned by the reviewer, we will add $T < N$ to Algorithm 1 and the statement of Theorem 4.4 in the revised version of our manuscript. We remark that this is essentially not a limitation of the theorem, and our result is meaningful because it provides interesting insights for large enough N which is practically feasible to assume in most applications.

Regarding point 2, we would like to remark that the current interpretation of the reviewer is misleading and could lead to wrong conclusions about the usefulness of our final results in Theorem 4.4. We try to elaborate on this point as follows by showing the importance of our bounds, and explicitly showing the existence of T.

We start by considering equation (13) in our paper which says that we need to find T such that

$$\max_{\mathbf{x}\in M_T} \underbrace{\beta_T^{1/2}}_{:=M_1} \cdot \underbrace{\sigma_{A_T}(\mathbf{x})}_{:=M_2} \leq \frac{\epsilon}{4} \tag{13}$$

holds, where we note the product of two terms M_1 and M_2. It is important to note that M_1 increases with respect to T and M_2 decreases with respect to T, hence we need to find a T such that the product in the left hand side is less than $\frac{\epsilon}{4}$ (ignoring the max operation for simplicity). In order to see the existence of such T for (13), we consider $M_1 = \mathcal{O}(\log(T))$ (which holds in practice) and assume we sample such that $M_2 = \mathcal{O}(\frac{1}{T})$. We plot the rates individually (by dotted lines) and in product (by solid lines) in Figure 1. We can see that choosing any threshold $\epsilon > 0$ on the y-axis would give our T on the x axis.

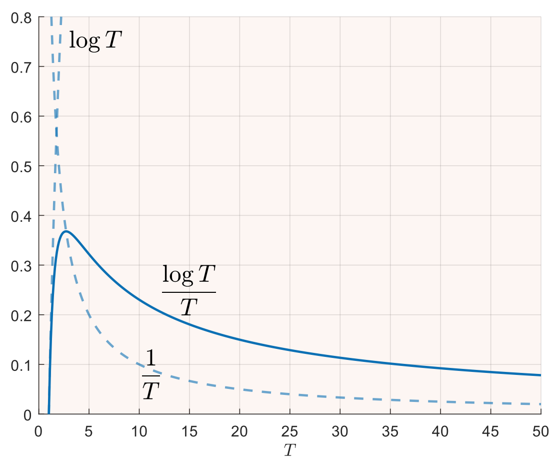

Figure 1. This figure provides a pictorial representation of the bounds presented in Equation (13).

Our Proof and Upper Bound: In order to provide an explicit estimate of T from Equation (13), we proceed by upper bounding the term $\sigma^2_{A_T}(\mathbf{x})$ in the left hand side of (13) first as follows (equation (16) in the manuscript)

$$\sigma^2_{A_T}(\mathbf{x}) \leq 1 - \frac{k((2Rb)T^{-1/d})^2}{1+\sigma_n^2}, \tag{16}$$

giving us the sufficient condition in equation (18),

$$\underbrace{\beta_{T+1}^{1/2}}_{:=M_1} \cdot \underbrace{\sqrt{1 - \frac{k(2RbT^{-1/d})^2}{1+\sigma_n^2}}}_{:=M_3} \leq \frac{\epsilon}{4}. \tag{18}$$

Tradeoff of the Upper Bound we Use: Although the above upper bound achieves a straightforward estimate of T, the drawback of the upper bound $1 - \frac{k((2Rb)t^{-1/d})^2}{1+\sigma_n^2}$ (labeled as M_3) is that it approaches $\frac{\sigma_n^2}{1+\sigma_n^2}$ as $T \to \infty$ (assuming $N \to \infty$

for the sake of argument), whereas $\sigma^2_{A_T}(\mathbf{x})$ approaches 0 as $T \to \infty$. Hence, M_3 introduces a lower bound on the accuracy ϵ one could achieve for a strictly positive $\sigma_n^2 > 0$. However, before it reaches a saturation point, $1 - \frac{k((2Rb)t^{-1/d})^2}{1+\sigma_n^2}$ decays faster than the rate at which β_T increases, which again is $\mathcal{O}(\log T)$. We plot the left hand side of Equation (18) for two practical kernels as follows, providing useful insights in showing the set of valid accuracies one can choose.

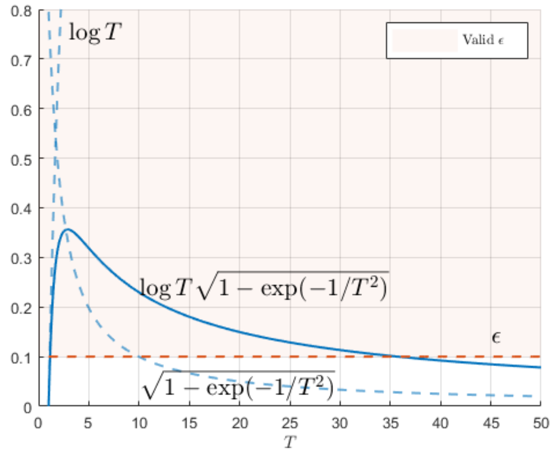

Figure 2. Plotting the left hand side of Equation (18) for the isotropic squared exponential kernel, which is $\mathcal{O}(\exp(-1/T^2))$. The dashed blue lines are the growth rates for M_1, M_3 in (18). The solid blue line represents the decay rate of $M_1 \cdot M_3$ in (18). The dashed red line represents the lower bound on ϵ. The area above this line is the set of valid accuracy levels one can choose. Note this lower bound approaches 0 as the noise variance $\sigma_n^2 \to 0$.

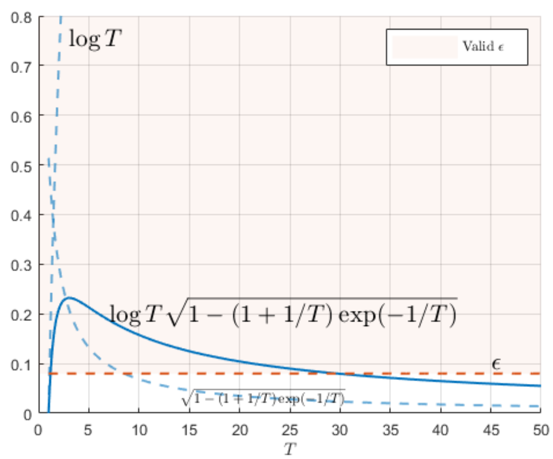

Figure 3. Plotting the left hand side of Equation (18) for the isotropic Matérn kernel, which has the rate $\mathcal{O}((1+1/T)\exp(-1/T))$.