On Provably Robust Meta-Bayesian Optimization

Abstract

Bayesian optimization (BO) has become popular for sequential optimization of black-box functions. When BO is used to optimize a target function, we often have access to previous evaluations of potentially related functions. This begs the question as to whether we can leverage these previous experiences to accelerate the current BO task through *meta-learning* (meta-BO), while ensuring *robustness* against potentially harmful dissimilar tasks that could sabotage the convergence of BO. This paper introduces two scalable and provably robust meta-BO algorithms: *robust meta-Gaussian process-upper confidence bound* (RM-GP-UCB) and *RM-GP-Thompson sampling* (RM-GP-TS). We prove that both algorithms are asymptotically no-regret even when some or all previous tasks are dissimilar to the current task, and show that RM-GP-UCB enjoys a better theoretical robustness than RM-GP-TS. We also exploit the theoretical guarantees to optimize the weights assigned to individual previous tasks through regret minimization via online learning, which diminishes the impact of dissimilar tasks and hence further enhances the robustness. Empirical evaluations show that (a) RM-GP-UCB performs effectively and consistently across various applications, and (b) RM-GP-TS, despite being less robust than RM-GP-UCB both in theory and in practice, performs competitively in some scenarios with less dissimilar tasks and is more computationally efficient.

1 INTRODUCTION

Bayesian optimization (BO) has recently gained immense popularity as an efficient method to optimize black-box functions [Shahriari et al., 2016], and it has found success in a variety of applications such as automated *machine learning* (ML) [Snoek et al., 2012], *reinforcement learning* (RL) [Wilson et al., 2014], among others. BO uses a *Gaussian process* (GP) [Rasmussen and Williams, 2006] as a surrogate to represent the belief about the objective function and, in each iteration, queries the input parameters that maximize an *acquisition function*. In particular, the BO algorithms based on the *GP-upper confidence bound* (GP-UCB) [Srinivas et al., 2010] and *GP-Thompson sampling* (GP-TS) [Chowdhury and Gopalan, 2017] acquisition functions have been shown to be asymptotically *no-regret* and perform competitively in practice. When using BO to optimize a *target function*, we sometimes have access to a set of evaluations of some potentially related functions. For example, when using BO for hyperparameter optimization of an ML model trained on a target dataset, we often have access to some previously completed BO tasks using other potentially related datasets [Golovin et al., 2017]. These previous tasks, if similar to the current task, may be exploited to accelerate the current BO task. However, if some (or even all) previous tasks are in fact dissimilar to the current task, their use may turn out to incorporate harmful information and sabotage the convergence of BO [Feurer et al., 2018]. This begs the question as to whether we can leverage previous tasks to improve the efficiency of the current BO task, while ensuring *robustness* against harmful dissimilar tasks such that they do not affect the trademark *no-regret* convergence of BO.

Exploiting previous learning experiences to improve the efficiency of the current task is the goal of *meta-learning* [Vanschoren, 2018]. Meta-learning is a broad field with various applications in supervised learning [Finn et al., 2017], RL [Xu et al., 2018], active learning [Pang et al., 2018], among others. The major challenges in meta-learning include (a) the transfer of information from previous tasks to the current task, and (b) characterization of task similarity which is crucial for identifying harmful dissimilar tasks [Vanschoren, 2018]. The application of meta-learning to BO (or *meta-BO*) has been explored by previous studies which differ in how these two challenges are addressed.

Submitted to the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022). **To be used for reviewing only**.

Some works, such as multitask BO [Swersky et al., 2013], transfer the information from previous tasks by building a joint GP surrogate using the observations from all previous and current tasks, with the task similarity either represented by meta-features [Bardenet et al., 2013, Yogatama and Mann, 2014] or learned from observations [Swersky et al., 2013, Wang et al., 2018]. These works, however, are limited by the scalability of GP due to including all previous and current observations in a single GP [Feurer et al., 2018].[1] To this end, other recent works transfer information from previous tasks using a more scalable approach: They build a separate GP surrogate for each individual task and use a weighted combination of either the individual surrogate functions or acquisition functions for query selection [Feurer et al., 2018, Wistuba et al., 2016, 2018]. A more detailed review of related works is presented in Sec. 7. However, none of the previous works has provided a theoretical performance guarantee to ensure robust performances in the presence of harmful dissimilar tasks. A robust theoretical guarantee is important for guaranteeing the consistent performances of meta-BO algorithms in various real-world applications, which is crucial for their practical deployment.

To this end, this paper introduces two scalable and provably robust meta-BO algorithms: *robust meta-GP-upper confidence bound* (RM-GP-UCB) and *robust meta-GP-Thompson sampling* (RM-GP-TS). Both algorithms compute the acquisition function (GP-UCB or GP-TS) for each individual task and select the next query via either a weighted combination (RM-GP-UCB) or in a probabilistic way (RM-GP-TS) (Sec. 3). As a result, like the works of Feurer et al. [2018], Wistuba et al. [2016, 2018], a separate GP surrogate is built for each previous task, making our algorithms scale well in the number of meta-tasks and observations in each meta-task. Our major contributions include: **Firstly**, we prove robust theoretical convergence guarantees for both RM-GP-UCB and RM-GP-TS (Sec. 4). In particular, both algorithms are asymptotically *no-regret* for *any* given set of previous tasks, i.e., even if some or all previous tasks are dissimilar to the target task. Moreover, we show that RM-GP-UCB enjoys a superior robustness guarantee compared with RM-GP-TS (Sec. 4.2). **Secondly**, to further enhance our robustness against dissimilar tasks, we exploit the theoretical guarantees to learn the task similarity (and hence identify dissimilar tasks) in a principled way, by minimizing the regret upper bounds via a computationally cheap online learning algorithm known as *Follow-The-Regularized-Leader* (Sec. 5). **Lastly**, we use extensive empirical evaluations to show that: RM-GP-UCB performs effectively and consistently across a wide range of tasks; RM-GP-TS, despite under-performing in adverse scenarios (i.e., when a large number of previous tasks are dissimilar), performs competitively in some favorable cases with less dissimilar tasks and is much more computationally efficient. Of note, our theoretical and empirical comparisons between RM-GP-UCB and RM-GP-TS may provide useful insights for other meta-BO algorithms in general (and potentially for other related algorithms such as meta-RL) in terms of the relative strengths and weaknesses of UCB- and TS-based meta-learning algorithms.

[1] Some works such as Perrone et al. [2018] and Volpp et al. [2020] replace GP by other surrogate models such as neural networks for scalability, however, they lack the principled uncertainty estimate and theoretical guarantee offered by GP.

2 BACKGROUND AND PROBLEM FORMULATION

Bayesian Optimization. This work tackles the problem of sequentially maximizing an unknown function $f : \mathcal{D} \to \mathbb{R}$. In each iteration $t = 1, \ldots, T$, an input $\mathbf{x}_t \in \mathcal{D}$ (a $D \geq 1$-dimensional vector) is queried to yield $y_t \triangleq f(\mathbf{x}_t) + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is a Gaussian noise with variance σ^2. The performance of BO is typically measured by *cumulative regret*: $R_T \triangleq \sum_{t=1,\ldots,T}[f(\mathbf{x}^*) - f(\mathbf{x}_t)]$ where $\mathbf{x}^* \in \arg\max_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x})$ is a global maximizer of f. It is desirable for a BO algorithm to achieve *no regret* by making its R_T grow sublinearly s.t. its *simple regret* $S_T \triangleq \min_{t=1,\ldots,T}[f(\mathbf{x}^*) - f(\mathbf{x}_t)] \leq R_T/T$ goes to 0 asymptotically. During BO, we model the belief about f using a *Gaussian process* (GP) $\{f(\mathbf{x})\}_{\mathbf{x}\in\mathcal{D}}$. That is, any finite subset of $\{f(\mathbf{x})\}_{\mathbf{x}\in\mathcal{D}}$ follows a multivariate Gaussian distribution [Rasmussen and Williams, 2006]. A GP is fully specified by its prior mean $\mu(\mathbf{x})$ and kernel function $k(\mathbf{x}, \mathbf{x}')$, and we assume w.l.o.g. that $\mu(\mathbf{x}) = 0$ and $k(\mathbf{x}, \mathbf{x}') \leq 1 \; \forall \mathbf{x}, \mathbf{x}' \in \mathcal{D}$. We focus on the widely used Squared Exponential (SE) kernel. Given T noisy observations $\mathbf{y}_T \triangleq [y_t]^\top_{t=1,\ldots,T}$ at inputs $\mathbf{x}_1, \ldots, \mathbf{x}_T$, the posterior GP belief of f at input $\mathbf{x} \in \mathcal{D}$ is Gaussian with the following posterior mean and variance:

$$\begin{aligned} \mu_T(\mathbf{x}) &\triangleq \mathbf{k}_T(\mathbf{x})^\top(\mathbf{K}_T + \lambda I)^{-1}\mathbf{y}_T, \\ \sigma^2_T(\mathbf{x}) &\triangleq k(\mathbf{x}, \mathbf{x}) - \mathbf{k}_T(\mathbf{x})^\top(\mathbf{K}_T + \lambda I)^{-1}\mathbf{k}_T(\mathbf{x}), \end{aligned} \tag{1}$$

where $\mathbf{K}_T \triangleq [k(\mathbf{x}_t, \mathbf{x}_{t'})]_{t,t'=1,\ldots,T}$, $\mathbf{k}_T(\mathbf{x}) \triangleq [k(\mathbf{x}_t, \mathbf{x})]^\top_{t=1,\ldots,T}$, λ is a regularization parameter.

Meta-Bayesian Optimization. We refer to the function f being maximized as the *target function* and the functions f_i for $i = 1, \ldots, M$ of the M previous tasks as *meta-functions*. We use *target task/observations* and *meta-tasks/observations* in a similar manner. All functions are defined on the same domain $\mathcal{D}$ which is assumed to be discrete for simplicity, but the theoretical results can be easily generalized to continuous domains following the analysis of Srinivas et al. [2010]. We assume that f and all f_i's lie in the *reproducing kernel Hilbert space* (RKHS) associated with the kernel k such that their norm induced by the RKHS is bounded: $\|f\|_k \leq B, \|f_i\|_k \leq B, \forall i = 1, \ldots, M$. This assumption intuitively suggests that the target and meta-functions have the same degree of smoothness. Same as

the work of Wang et al. [2018] which has also performed theoretical analysis of a meta-learning algorithm for BO, we also assume that all meta- and target observations are corrupted by a Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$ with variance σ^2. The number of observations from meta-task i is a constant denoted as N_i, and $N \triangleq \max_{i=1,\ldots,M} N_i$. $\mathbf{x}_{i,j}$ and $y_{i,j}$ represent the j-th input and noisy output of meta-task i respectively. We define the *function gap* $d_i \triangleq \max_{j=1,\ldots,N_i} \left|f(\mathbf{x}_{i,j}) - f_i(\mathbf{x}_{i,j})\right| < \infty$ which represents the maximum difference between the function values of f and f_i at any corresponding input $\mathbf{x}_{i,j}$ of meta-task i. Note that for a given set of meta-observations for meta-task i, the function gap d_i is an unknown constant characterizing the similarity between meta-task i and the target task: a smaller function gap implies a stronger similarity.

3 ROBUST META-BAYESIAN OPTIMIZATION

The acquisition function (2) adopted by RM-GP-UCB in iteration t is a weighted combination of $M+1$ individual GP-UCB acquisition functions [Srinivas et al., 2010] for the target task and the M meta-tasks, each of which is calculated using the observations from a particular task:

$$\overline{\zeta}_t(\mathbf{x}) \triangleq \nu_t \Big[\sum\nolimits_{i=1}^{M} \omega_i \left[\overline{\mu}_i(\mathbf{x}) + \tau\overline{\sigma}_i(\mathbf{x})\right]\Big] + (1-\nu_t)\left[\mu_{t-1}(\mathbf{x}) + \beta_t\sigma_{t-1}(\mathbf{x})\right]. \tag{2}$$

In (2), $\mu_{t-1}(\mathbf{x})$ and $\sigma_{t-1}(\mathbf{x})$ represent, respectively, the GP posterior mean and standard deviation (1) at $\mathbf{x}$ calculated using the target observations from iterations 1 to $t-1$. $\overline{\mu}_i(\mathbf{x})$ and $\overline{\sigma}_i(\mathbf{x})$ are computed using all meta-observations from meta-task i. $\beta_t > 0$ and $\tau > 0$ will be defined in Sec. 4. $\nu_t \in [0,1]$ can be interpreted as the overall weight given to all meta-tasks in iteration t and should be chosen to be non-increasing in t, which enforces the impact of meta-tasks in (2) to be non-increasing. The *meta-weights* ω_i's can be understood as the weights assigned to individual meta-tasks. Note that since the dataset used to calculate $\overline{\mu}_i(\mathbf{x})$ and $\overline{\sigma}_i(\mathbf{x})$ is fixed with size N_i, the matrix inversion in (1) (i.e., computational bottleneck for GP) can be pre-computed. So, after T iterations, RM-GP-UCB incurs $\mathcal{O}(T^3)$ time for covariance matrix inversion (since only the target covariance matrix of size $T \times T$ needs to be inverted) and $\mathcal{O}(MN^2 + T^2)$ time during predictive inference, which are less than the respective $\mathcal{O}((MN+T)^3)$ and $\mathcal{O}((MN+T)^2)$ time when all observations are included in a single GP. Hence, RM-GP-UCB is scalable in the number of meta-tasks and observations in each meta-task.

The acquisition function of RM-GP-TS is defined as:

$$\overline{\zeta}_t(\mathbf{x}) \triangleq \begin{cases} f^t(\mathbf{x}) & \text{with probability } 1-\nu_t\,, \\ \sum_{i=1}^{M}\omega_i \overline{f}_i^t(\mathbf{x}) & \text{with probability } \nu_t, \end{cases} \tag{3}$$

in which f^t is a function sampled from the GP posterior of the target task: $f^t \sim \mathcal{GP}\left(\mu_{t-1}(\cdot), \beta_t^2\sigma_{t-1}^2(\cdot)\right)$, and f_i^t is sampled from the GP posterior of meta-task i: $\overline{f}_i^t \sim \mathcal{GP}\left(\overline{\mu}_i(\cdot), \tau^2\overline{\sigma}_i^2(\cdot)\right)$. Using approximation techniques such as random Fourier features (RFF) approximation [Rahimi and Recht, 2008] (which we use in all our experiments), the functions f^t and f_i^t's can be sampled efficiently, hence making RM-GP-TS computationally efficient (as we will demonstrate in Sec. 6). Moreover, since the meta-observations of every meta-task is fixed, the use of approximation techniques such as RFF allows the functions $\overline{f}_i^t$'s to be sampled beforehand before the algorithm starts. Refer to Appendix E.5 for more details on RM-GP-TS.

In iteration t of either RM-GP-UCB or RM-GP-TS (Algorithm 1), we first optimize the meta-weights and update ν_t (Sec. 5.2), which corresponds to line 2 of Algorithm 1. Next, the input $\mathbf{x}_t$ is selected by maximizing the acquisition function (2) (RM-GP-UCB) or (3) (RM-GP-TS), after which we query $\mathbf{x}_t$ and use the newly collected $(\mathbf{x}_t, y_t)$ to update the GP posterior belief (1).

Algorithm 1 RM-GP-UCB/RM-GP-TS

1: **for** $t = 1, 2, \ldots, T$ **do**
2: Update ω_i for $i = 1, \ldots, M$ via online meta-weight optimization and update ν_t (Sec. 5.2)
3: $\mathbf{x}_t \leftarrow \arg\max_{\mathbf{x}\in\mathcal{D}} \overline{\zeta}_t(\mathbf{x})$, with $\overline{\zeta}_t(\mathbf{x})$ given in (2) (for RM-GP-UCB) or (3) (for RM-GP-TS)
4: Query $\mathbf{x}_t$ to observe y_t, and update GP posterior belief (1) using $(\mathbf{x}_t, y_t)$
5: **end for**

4 THEORETICAL ANALYSIS

4.1 RM-GP-UCB

Theorem 1 presents an upper bound on the cumulative regret of RM-GP-UCB (proof in Appendix B).

Theorem 1 (RM-GP-UCB). *Let $\delta \in (0,1)$. Denote by γ_t the max information gain about f from observing any set of t observations. If RM-GP-UCB is run with: $\lambda = 1 + 2/T$, $\beta_t = B + \sigma\sqrt{2(\gamma_{t-1} + 1 + \log(4/\delta))}$, $\tau = B + \sigma\sqrt{2(\gamma_N + 1 + \log(4M/\delta))}$, $\nu_t \in [0,1]$ and $\nu_{t+1} \leq \nu_t$, $\omega_i \geq 0$ and $\sum_{i=1}^{M}\omega_i = 1$. Then, with probability of $\geq 1 - 3\delta/4$,*

$$R_T \leq 2(\alpha+\tau)\sum_{t=1}^{T}\nu_t + \beta_T\sqrt{C_1 T\gamma_T} = \widetilde{\mathcal{O}}\big(\big(\sum_{i=1}^{M} d_i\big)\sum_{t=1}^{T}\nu_t + \gamma_T\sqrt{T}\big), \tag{4}$$

where $C_1 \triangleq \frac{8}{1+\sigma^{-2}}$, and $\alpha \triangleq \sum_{i=1}^{M}\omega_i\frac{N_i}{\sigma^2}\big(2\sqrt{2\sigma^2\log\frac{8N_i}{\delta}} + d_i\big)$.

The second term $\gamma_T\sqrt{T}$ in the regret upper bound (4) grows sub-linearly for the SE kernel for which $\gamma_T = \mathcal{O}((\log T)^{D+1})$. Therefore, if ν_t is designed such that $\nu_t \to 0$ as $t \to \infty$, the first term also grows sub-linearly and hence RM-GP-UCB is asymptotically no-regret.

Theorem 1 holds for a given set of meta-tasks with fixed yet unknown d_i's. Note that we do not impose assumptions on the values of d_i's, i.e., the similarities between the meta- and target tasks. Therefore, Theorem 1 gives a robust regret upper bound which holds for *any* given set of meta-tasks. In other words, even in adverse scenarios where some or all meta-tasks are extremely dissimilar to the target task (i.e., when some or all d_i's are very large), RM-GP-UCB is still asymptotically no-regret, which indicates the robustness and generality of our algorithm. This provides an assurance about the *worst-case behavior* in any given scenario.[2] In our proof, the key step (Lemma 3 in Appendix B) is to upper bound (by α in Theorem 1) the overall error induced by the use of any given set of meta-observations, instead of the target observations at the same corresponding input locations, when calculating the acquisition function (2). These interpretations also explain the dependence of α, hence the regret bound, on d_i and N_i: Larger function gaps increase the error resulting from the use of the meta-observations, and a larger number of meta-observations also inflates the worst-case upper bound by accumulating the individual errors.

Meta-tasks Can Improve the Convergence by Accelerating Exploration. In addition to characterizing the worst-case behavior, we also exploit our theoretical analysis to illustrate how meta-tasks can help RM-GP-UCB converge faster than standard GP-UCB. As we have proved in Appendix B.3, at the early stage of the algorithm, the meta-tasks (if similar to the target task) can help RM-GP-UCB obtain a smaller regret upper bound than GP-UCB by *reducing the uncertainty at the selected input*. Equivalently, the additional information from the meta-tasks allows RM-GP-UCB to *reduce the degree of exploration at the early stage*. Since initial exploration of BO usually incurs large regrets, less exploration results in smaller regrets. At later stages when ν_t becomes close to 0, RM-GP-UCB converges to no regret at a similar rate to GP-UCB (i.e., the second term in Theorem 1 dominates).

4.2 RM-GP-TS

Theorem 2 gives an upper bound on the cumulative regret of RM-GP-TS (proof in Appendix C).

Theorem 2 (RM-GP-TS). *Define* $d'_i \triangleq \max_{\mathbf{x}\in\mathcal{D}} |f(\mathbf{x}) - f_i(\mathbf{x})|$. *With the same parameters as those defined in Theorem 1, we have that with probability of at least* $1 - 3\delta/4$,

$$R_T = \widetilde{\mathcal{O}}\big(\big(\sum_{i=1}^{M}\omega_i d'_i\big)\sum_{t=1}^{T}\nu_t + \sum_{t=1}^{T}\nu_t\sqrt{\gamma_t} + \gamma_T\sqrt{T}\big).$$

Note that by definition, we have that $d'_i \geq d_i, \forall i$. Similar to RM-GP-UCB, as long as ν_t is chosen such that $\nu_t \to 0$ as $t \to \infty$ and that $\nu_t = o(1/\sqrt{\gamma_t})$, all three terms in Theorem 2 are sub-linear (for the SE kernel). That is, RM-GP-TS is also asymptotically no-regret for any set of meta-tasks, even when some or all meta-tasks are dissimilar to the target task. Moreover, comparing the extra terms in the regret upper bounds resulting from the use of the meta-tasks for both RM-GP-UCB (i.e., the first term of equation (4) in Theorem 1) and RM-GP-TS (i.e., the first two terms of Theorem 2) reveals that compared with RM-GP-UCB, RM-GP-TS suffers from a worse extra dependence on T due to meta-tasks. Specifically, while the first terms of Theorems 1 and 2 have the same dependence on T, the second term of Theorem 2 introduces an extra dependence on T which dominates the first term. This suggests that in adverse scenarios with a large number of dissimilar tasks, RM-GP-TS may suffer from a worse convergence than RM-GP-UCB. In other words, RM-GP-UCB enjoys a better theoretically guaranteed robustness against dissimilar tasks.

4.3 PRACTICAL IMPLICATIONS

Besides the theoretical insights, Theorems 1 and 2 also provide two natural hints to the practical algorithmic design. Firstly, note that both Theorems hold for all choices of meta-weights ω_i's. Therefore, we have the flexibility to choose the optimal ω_i's (i.e., learn the task similarity) by minimizing the regret upper bounds in Theorems 1 and 2. Secondly, the first term in Theorem 1 suggests that we can lower the regret by making ν_t (i.e., the influence of the meta-tasks) decay faster if α in Theorem 1 (i.e., an upper bound on the error produced by using the meta-tasks) is larger. The same reasoning applies to Theorem 2, i.e., we can decay ν_t faster if $\sum_{i=1,\ldots,M}\omega_i d'_i$ in Theorem 2 is larger. Both design choices can further strengthen the robustness of our algorithms against dissimilar meta-tasks by lessening their impact. Unfortunately, they both require the values of the function gaps d_i's which are unavailable.[3] To this end, we devise a principled technique to estimate upper bounds on the function gaps, which is presented in the next section.

5 ONLINE META-WEIGHT OPTIMIZATION

In this section, we first introduce a principled technique for estimating high-probability upper bounds on the function

[2] This notion of robustness is in line with that of *robust optimization* (RO) [Beyer and Sendhoff, 2007] which also attempts to optimize the performance in the worst-case scenario. The difference is that RO optimizes an explicit objective, while we aim at preserving the no-regret property in the worst case.

[3] d_i can be used as an estimate of d'_i since $d'_i \geq d_i$ (Sec. 4.2).

gaps (Sec. 5.1) that, when combined with Theorems 1 and 2, naturally yields a principled method for optimizing the meta-weights through regret minimization via online learning.

5.1 ONLINE ESTIMATION OF FUNCTION GAPS

Inspired by the confidence region constructed by GP-UCB [Srinivas et al., 2010, Chowdhury and Gopalan, 2017] that contains the target function with high probability, after $t \geq 1$ target observations have been collected, define

$$\begin{aligned} U_{t,i,j} &\triangleq \mu_t(\mathbf{x}_{i,j}) + \beta_{t+1}\sigma_t(\mathbf{x}_{i,j})\,, \\ L_{t,i,j} &\triangleq \mu_t(\mathbf{x}_{i,j}) - \beta_{t+1}\sigma_t(\mathbf{x}_{i,j}), \end{aligned} \tag{5}$$

where $\mathbf{x}_{i,j}$ is the j-th input of meta-task i, β_{t+1} is previously defined in Theorem 1, and $U_{t,i,j}$ and $L_{t,i,j}$ can be interpreted, respectively, as the upper and lower confidence bounds of f at $\mathbf{x}_{i,j}$ after t iterations. Lemma 2 (Appendix B) implies that with probability of at least $1-\delta/4$ (δ is defined in Theorem 1): $L_{t,i,j} \leq f(\mathbf{x}_{i,j}) \leq U_{t,i,j}, \forall t, i, j$. Consequently, the following result gives high-probability upper bounds on the function gaps (proof in Appendix D.1):

Lemma 1. *With probability of at least* $1-\delta$*,*

$$d_i \leq \sqrt{2\sigma^2 \log\left[\left(8\sum\nolimits_{i=1}^{M} N_i\right)/\delta\right]} + \max_{j=1,\ldots,N_i}\left[\max\{|y_{i,j} - U_{t,i,j}|, |y_{i,j} - L_{t,i,j}|\}\right] \triangleq \overline{d}_{i,t},$$

for $t = 1, \ldots, T$ *and* $i = 1, \ldots, M$.

Unlike d_i, $\overline{d}_{i,t}$ can be efficiently calculated as its incurred time is linear in both M and N.

5.2 ONLINE META-WEIGHT OPTIMIZATION THROUGH REGRET MINIMIZATION

In this section, we focus on RM-GP-UCB since the analysis for RM-GP-TS (deferred to Appendix D.4) is similar and leads to the same update rules for ω_i's and ν_t. Combining Lemma 1 and Theorem 1 allows us to derive the following result for RM-GP-UCB (proof in Appendix D.2):

Proposition 1 (RM-GP-UCB)**.** *With probability of* $\geq 1-\delta$*,*

$$R_T \leq \frac{2}{\sigma^2}\left[\sum\nolimits_{t=1}^{T} \boldsymbol{\omega}^\top \boldsymbol{l}_t\right]\left[\sum\nolimits_{t=1}^{T} \nu_t\right] + 2\tau\sum\nolimits_{t=1}^{T} \nu_t + \beta_T\sqrt{C_1 T \gamma_T},$$

where $\boldsymbol{\omega} \triangleq [\omega_i]_{i=1,\ldots,M}$*,* $\boldsymbol{l}_t \triangleq [l_{i,t}]_{i=1,\ldots,M}$*, and* $l_{i,t} \triangleq N_i(2\sqrt{2\sigma^2\log(8N_i/\delta)} + \overline{d}_{i,t})$.

Note that $\boldsymbol{l}_t$ can be efficiently computed after the t-th observation is collected. The regret upper bound in Proposition 1 depends on ω_i's only through the term $\sum_{t=1}^{T} \boldsymbol{\omega}^\top \boldsymbol{l}_t$ which can be minimized to derive the optimal meta-weights. This constitutes an *online learning* problem with linear loss function and its solution $\boldsymbol{\omega}$ constrained to a probability simplex. An additional entropic regularization term is usually preferred so as to encourage a solution with a large entropy to stabilize it [Bubeck, 2011]. This corresponds to encouraging the meta-weights to spread across a large number of meta-tasks, in order to discover as many similar meta-tasks as possible. As a result, by using $1/\eta$ ($\eta > 0$) as the regularization parameter, the optimal $\boldsymbol{\omega}$ in iteration $t > 1$ is obtained by solving the following optimization problem:

$$\boldsymbol{\omega} \triangleq \arg\min_{\boldsymbol{\omega}'} \sum\nolimits_{s=1}^{t-1} \boldsymbol{\omega}'^\top \boldsymbol{l}_s + \eta^{-1}\sum\nolimits_{i=1}^{M} \omega'_i \log \omega'_i, \tag{6}$$

subject to the constraints: $\omega'_i \geq 0, \forall i$ and $\sum_{i=1}^{M} \omega'_i = 1$. When $t = 1$, the optimal $\boldsymbol{\omega}$ follows from optimizing only the entropic regularization term, thus naturally entailing the uniform distribution $\omega_i = 1/M, \forall i$. Consequently, (6) corresponds exactly to the online learning algorithm called *Follow-The-Regularized-Leader* with an entropic regularizer [Bubeck, 2011] where η represents the learning rate. Its optimal solution in iteration t can be derived via Lagrange multiplier (Appendix D.3) as

$$\omega_i = \frac{e^{-\eta\sum_{s=1}^{t-1} l_{i,s}}}{\sum_{j=1}^{M} e^{-\eta\sum_{s=1}^{t-1} l_{j,s}}} \overset{(a)}{\approx} \frac{e^{-\eta N^{\frac{3}{2}}\sum_{s=1}^{t-1} \overline{d}_{i,s}}}{\sum_{j=1}^{M} e^{-\eta N^{\frac{3}{2}}\sum_{s=1}^{t-1} \overline{d}_{j,s}}}, \tag{7}$$

for $i = 1, \ldots, M$ where (a) follows from assuming that all N_i's are close to N for simplicity. With this simplification, the first (noise-correction) term in the expression of $\overline{d}_{i,t}$ from Lemma 1 also cancels out, thus leading to a neat and elegant update rule for ω_i which we use in all our experiments. As is evident from (7), the update of ω_i's in each iteration only involves computing $\overline{d}_{i,t}$'s (incurring $\mathcal{O}(MN)$ time), adding one term to the summation on the exponent ($\mathcal{O}(M)$ time), and a normalization step ($\mathcal{O}(M)$ time), all of which are computationally cheap. Intuitively, (7) assigns small weights to meta-tasks with a large cumulative estimated function gap which implies a less similar meta-task.

In addition, $\overline{d}_{i,t}$ from Lemma 1 also allows for the estimation of an upper bound on α (Theorem 1) in each iteration (i.e., by simply replacing d_i with $\overline{d}_{i,t}$) and thus facilitates an adaptive selection of ν_t, as mentioned in Sec. 4. Specifically, we set $\nu_1 = 1$ and $\nu_t = \nu_{t-1} \times \min(r, (\sum_{i=1}^{M} \omega_i \overline{d}_{i,t})^{-\epsilon})$ for $t > 1$, in which we have dropped the constants independent of $\overline{d}_{i,t}$. $r \in (0, 1)$ represents the minimum decaying rate to ensure the monotonic decay of ν_t such that RM-GP-UCB is no-regret (Sec. 4.1). $\epsilon > 0$ controls the aggressiveness of the adaptive decay such that a larger ϵ results in a faster decay. With this scheme, when the overall estimated function gaps are larger (the meta-tasks are dissimilar), ν_t decays faster and thus the impact of the meta-tasks vanishes more quickly.

6 EXPERIMENTS AND DISCUSSION

We use extensive real-world experiments to compare our RM-GP-UCB and RM-GP-TS with *(1)* standard GP-UCB, two other GP-based scalable meta-BO algorithms: *(2) ranking-weighted Gaussian process ensemble* (RGPE) [Feurer et al., 2018] and *(3) transfer acquisition function* (TAF) [Wistuba et al., 2018] (Appendix A), *(4)* multitask BO (MTBO) [Swersky et al., 2013], and *(5)* the method from [Wang et al., 2018] named *point estimate meta-BO* (PEM-BO). Since MTBO is relatively not scalable (Sec. 1), we only apply it to those experiments with relatively small number of meta-tasks and observations for which MTBO is still computationally feasible. We compare with PEM-BO [Wang et al., 2018] in the experiment that is most favorable for this algorithm, i.e., with the largest number of meta-observations and a discrete domain (refer to Sec. 6.2 for more details). We set $\eta = 1/N^{3/2}$, $\epsilon = 0.7$ and $r = 0.7$ in all real-world experiments to demonstrate the robustness of our algorithm against the choice of these parameters. In practice, the upper bound on the function gap, $\overline{d}_{i,t}$, from Lemma 1 may be too conservative; so, we replace the outer $\max$ operator over $j = 1, ..., N_i$ with the empirical mean in our experiments.[4] Some details and results are deferred to Appendix E due to lack of space. All error bars represent standard errors.

6.1 SYNTHETIC EXPERIMENTS

We firstly explore the effectiveness of our online meta-weight optimization (Sec. 5) and the impact of different algorithmic parameters by optimizing synthetic functions drawn from GPs. For each objective function, we construct $M = 4$ meta-tasks with $N = N_i = 20$ meta-observations each. The function gaps are chosen as $d_1 = d_2 = 0.05$ and $d_3 = d_4 = 4.0$ such that the last 2 meta-tasks are dissimilar to the target task. Fig. 1a plots the simple regrets averaged over 20 randomly drawn synthetic functions, with $\eta N^{3/2} = 1.0$, $\epsilon = 0.7$, and $r = 0.7$. The figure shows that RM-GP-UCB with online meta-weight optimization significantly outperforms RM-GP-UCB with fixed meta-weights ($\omega_i = 1/4$ for all i). Fig. 1b plots the meta-weights optimized by RM-GP-UCB for the red curve in Fig. 1a, showing that the weights given to the last two meta-tasks which are dissimilar to the target task are rapidly reduced. These results verify the effectiveness of online meta-weight optimization in reducing the impact of dissimilar meta-tasks.

We also investigate the impact of η and ϵ. Fig. 1c shows the performances of different values of η, with fixed $\epsilon = 0.7$ and $r = 0.7$. The figure demonstrates that an excessively small η (purple curve) negatively impacts the performance, since RM-GP-UCB is unable to quickly reduce the weights of dissimilar meta-tasks (Fig. 4a in Appendix E.1). Moreover, an overly large η is also slightly detrimental (green curve) since it rapidly assigns a large weight to one of the two useful meta-tasks (Fig. 4c in Appendix E.1), thus failing to utilize the other useful meta-task. Fig. 1d illustrates the impact of ϵ when all function gaps are large: $d_i = 8.0$ for all i.[5] The figure shows that even when all meta-tasks are dissimilar, our adaptive selection of ν_t is able to diminish their negative impact and allow RM-GP-UCB to perform comparably to GP-UCB. Furthermore, in this adverse scenario, a faster decline of the impact of the meta-tasks (i.e., faster decay of η_t via larger ϵ) leads to slightly better performance.

[4]We explore the difference between them in Appendix E.3.

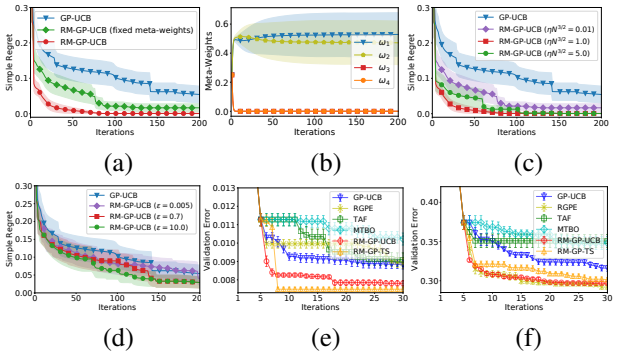


Figure 1: (a) The simple regret and (b) meta-weights optimized by RM-GP-UCB. The impact of (c) η and (d) ϵ. Best validation error of CNN for (e) MNIST and (f) CIFAR-10.

6.2 REAL-WORLD EXPERIMENTS

Hyperparameter Tuning for Convolutional Neural Networks (CNNs). We apply meta-BO to hyperparameter tuning of ML models with the previous tasks using other datasets as the meta-tasks. We tune 3 hyperparameters of CNNs using 4 widely used image datasets: MNIST, SVHN, CIFAR-10 and CIFAR-100. Specifically, in each experiment, one of the four datasets is selected to produce the target function f which maps a hyperparameter setting to a validation accuracy obtained using this dataset. The meta-observations are generated from 3 independent BO tasks (each with 50 iterations) using the other 3 datasets, i.e., $M = 3$ and $N_i = 50$ for $i = 1, 2, 3$ in all 4 experiments. The results for MNIST and CIFAR-10 are plotted in Figs. 1e and 1f while the remaining results are shown in Appendix E.2 (Fig. 5). The results show that RM-GP-UCB is the only method that consistently performs well in all tasks, and that RM-GP-TS performs much better than RM-GP-UCB (and other methods) for MNIST, yet worse in the other tasks. We have also adopted the Omniglot dataset [Lake et al., 2015] commonly used in meta-learning, for which RM-GP-UCB performs the best (Fig. 6, Appendix E.2).

[5]We use $\eta = 1/N^{3/2}$ and fix r at a large value (0.99) so that the decaying rate of ν_t is purely decided by ϵ.

Non-stationary Bayesian Optimization. Meta-BO can be naturally applied to non-stationary BO problems in which the unknown objective function evolves over time since the previous (outdated) observations can be treated as the meta-observations. We consider here automated ML for clinical diagnosis. As the data from new patients becomes available regularly, clinicians often need to periodically update the dataset and re-run hyperparameter optimization for the ML model used for clinical diagnosis. This stimulates the question as to whether the previous hyperparameter tuning tasks using the outdated patients data can help accelerate the current task. We consider the problem of diabetes prediction [Smith et al., 1988] with *logistic regression* (LR) and tune 3 LR hyperparameters. We create 5 progressively growing datasets (including the full dataset), treating (the hyperparameter tuning task using) the full dataset as the target task and the 4 smaller datasets as the meta-tasks. The results (Fig. 2a) show that RM-GP-TS outperforms all other methods in this task. Moreover, we also compare the runtime of different methods in Fig. 2b: RM-GP-TS is significantly more efficient than all other methods, and methods building separate GP surrogates for different tasks (i.e. RM-GP-UCB, RGPE and TAF) are more efficient than MTBO which includes all observations in a single GP (Sec. 1).

Hyperparameter Tuning for Support Vector Machines (SVMs). We also tune the hyperparameters of SVMs using a tabular benchmark dataset [Wistuba et al., 2015a] which has also been adopted by RGPE [Feurer et al., 2018]. The benchmark was constructed by evaluating a fixed grid of 288 SVM hyperparameter configurations using 50 *diverse* datasets (i.e., containing many dissimilar tasks). We follow the setting used by RGPE [Feurer et al., 2018]: In every trial, we fix one of the tasks as the target task, and the remaining $M = 49$ tasks as the meta-tasks; for every meta-task i, we randomly select $N_i = 50$ hyperparameter configurations as the meta-observations. The results in Fig. 2c show that our RM-GP-UCB performs comparably to RGPE, outperforming the other methods; RM-GP-TS performs unsatisfactorily in this experiment with diverse tasks. Of note, this experiment has the most favorable setting for PEM-BO [Wang et al., 2018] because (a) PEM-BO has been shown to require a massive set of meta-observations ($\geq$ 5000) to perform well [Wang et al., 2018], and this experiment has the largest number ($49 \times 50 = 2450$) of meta-observations among all experiments; (b) the domain here is discrete, which is much easier for the application of PEM-BO.

Human Activity Recognition (HAR). HAR using mobile devices has promising applications in various domains such as healthcare [Reyes-Ortiz et al., 2013]. When optimizing the configurations (hyperparameters) of the activity prediction model (ML model) for a subject, the previous optimization tasks for other subjects might be helpful. However, cross-subject transfer in HAR is challenging due to high *individual variability* [Soleimani and Nazerfard, 2019], which

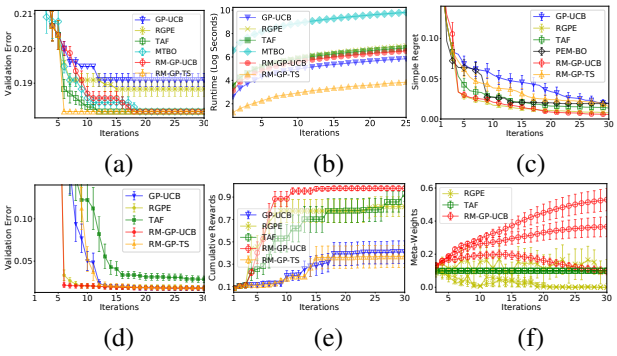


Figure 2: (a) Best validation error of LR for diabetes diagnosis. (b) Runtime in non-stationary BO experiment. (c) Simple regret on SVM benchmark. (d) Best validation errors for HAR. (e) Best cumulative rewards and (f) learned meta-weights for the 3 similar meta-tasks for the RL experiment.

makes HAR suitable for evaluating the robustness of a meta-BO algorithm against dissimilar meta-tasks. We use the data collected through mobile phones from 30 subjects performing 6 activities and use *support vector machines* (SVM) for activity prediction. Every task corresponds to tuning 2 SVM hyperparameters for a subject. We run a separate BO (30 iterations) for each of the 21 subjects to generate the meta-observations ($M = 21$, $N_i = 30$ for $i = 1, \ldots, 21$) and use the other 9 subjects for validation. The results are shown in Fig. 2d (averaged over the 9 subjects, each further averaged over 5 random initializations), in which RM-GP-UCB delivers the best performance, followed by RGPE; RM-GP-TS again fails to perform effectively, suggesting that it is less robust against the individual variability in HAR.

Policy Search for Reinforcement Learning (RL). When optimizing the RL policy of an agent in an environment, the agent's experience in other related environments may help to make learning more efficient [Duan et al., 2016, Wang et al., 2016]. We apply meta-BO to policy search in RL to maximize the cumulative rewards in an episode, using the Cart-Pole environment from OpenAI Gym [Brockman et al., 2016] with 8 policy parameters. We simulate different environments by setting the agent to different initial states. In particular, we choose $M = 10$ different initial states, among which the majority (i.e., 7) are randomly generated (i.e., dissimilar meta-tasks) and the other 3 are designed to be close to the initial state of the target task so that they are similar to the target task. An independent BO task with 50 iterations is run for every initial state, i.e., $N_i = 50$ for $i = 1, \ldots, 10$. Figs. 2e and 2f plot the (normalized) cumulative rewards of different algorithms and their learned meta-weights for the 3 similar meta-tasks. The results show that RM-GP-UCB achieves the best performance (Fig. 2e), and it is more effective than RGPE and TAF at identifying the 3 similar meta-tasks (Fig. 2f). RGPE and TAF fail to correctly identify similar meta-tasks because they learn the meta-weights based on how accurately each GP surrogate predicts the

pairwise ranking of the target observations (more details in Sec. 7). However, in the Cart-Pole environment, many target observations have equal values, which confuses the pairwise ranking and makes the learned meta-weights unreliable. RM-GP-TS again only performs comparably with standard GP-UCB (Fig. 2e).

6.3 EXPERIMENTAL DISCUSSION

In most experimental results (Figs. 1 and 2), the performance advantage of RM-GP-UCB is most evident at the initial stage. This is likely to corroborate our theoretical insights that the meta-tasks can help improve the convergence of RM-GP-UCB at the initial stage by reducing the degree of exploration (Sec. 4.1). A potential limitation of our online meta-weight optimization (Sec. 5) is that it does not account for the scenario where the meta-functions is shifted or scaled versions of the target function. However, note that in some scenarios, the scale of the meta-functions is informative about task similarity and thus should not be removed. For example, in our clinical diagnosis experiment, the more recently completed meta-tasks (with larger training set, smaller validation errors, and thus smaller function gaps) are expected to be more similar to the target task. Furthermore, as demonstrated by the green curve in Fig. 1a, in some cases, even though the meta-weights are not optimized, RM-GP-UCB still performs favorably. This implies its robustness against mis-specification of the meta-weights.

RM-GP-UCB is the only method that consistently outperforms standard GP-UCB in *all* experiments (Figs. 1 and 2), whereas other methods perform either comparably with or worse than GP-UCB in some experiments (e.g., RGPE in Figs. 1e and 2a, TAF in Figs. 1e, 1f and 2d). This might be attributed to RM-GP-UCB's theoretically guaranteed robustness against dissimilar meta-tasks (Sec. 4) and its ability to diminish their impact in a principled way (Sec. 5). In particular, RM-GP-UCB performs significantly better than RM-GP-TS in those experiments with a large number of dissimilar meta-tasks (Figs. 2c-e), which may be explained by RM-GP-UCB's better theoretically guaranteed robustness against dissimilar meta-tasks than RM-GP-TS (Sec. 4.2). However, Figs. 1e-f and Fig. 2a show that RM-GP-TS performs competitively in some experiments with more favorable settings (i.e., less dissimilar meta-tasks), which might result from the repeatedly observed empirical effectiveness of TS-based algorithms [Chapelle and Li, 2011, Russo et al., 2017]. Moreover, the computational efficiency of RM-GP-TS is markedly superior to other methods (Fig. 2b). These theoretical and empirical comparisons between RM-GP-UCB and RM-GP-TS may provide useful insights for other meta-BO algorithms and potentially for a broader range of problems (e.g., meta-learning for multi-armed bandits and RL) in terms of the relative strengths and weaknesses of UCB- and TS-based algorithms.

7 RELATED WORKS

Some previous works on meta-BO, such as multitask BO [Swersky et al., 2013], build a joint GP surrogate using all previous and current observations. We defer more detailed discussions of these works, as well as other recent works on meta-BO, to Appendix A. Wang et al. [2018] have learned the GP prior from previous tasks and given theoretical guarantees. However, they have shown in both theory and practice that a large training set of meta-observations (≥ 5000) is required for their method to work well, while we focus on the more practical setting of meta-BO where the number of available meta-observations may be small. We have also verified that our algorithm outperforms the method from Wang et al. [2018] in the experiment that is most favorable for their method among all our experiments (more details in the third paragraph of Sec. 6.2).

Some works have aimed to improve the scalability of GP-based meta-BO algorithms by building a separate GP surrogate for each task [Feurer et al., 2018, Wistuba et al., 2016, 2018]. Wistuba et al. [2016] use a weighted combination of the posterior mean of each individual GP surrogate as the joint posterior mean while the posterior variance is derived using only the target observations. RGPE [Feurer et al., 2018] has extended the work of Wistuba et al. [2016] by estimating the joint objective function as a weighted combination of individual objective functions, such that the resulting joint surrogate remains a GP (unlike Wistuba et al. [2016]) and can thus be plugged into standard BO algorithms. Note that RGPE differs from our RM-GP-UCB algorithm in that RGPE uses a weighted combination of individual GP surrogates to derive a joint GP surrogate, whereas our RM-GP-UCB leverage a weighted combination of individual acquisition functions. Wistuba et al. [2018] have proposed TAF, which also uses a weighted combination of the acquisition functions (i.e., expected improvement) from the individual tasks for query selection. In these works, the weight of a previous task is heuristically chosen to be proportional to the accuracy of the *pairwise ranking of the target observations* produced by either (a) the posterior mean of the GP surrogate of the previous task (TAF) [Wistuba et al., 2018] or (b) functions sampled from the posterior GP surrogate (RGPE) [Feurer et al., 2018].

8 CONCLUSION

We have introduced RM-GP-UCB and RM-GP-TS, both of which are asymptotically no-regret even if all meta-tasks are dissimilar to the target task. We leverage the theoretical results to learn the task similarities in a principled way via online learning. Theoretical and empirical comparisons show that RM-GP-UCB is more robust against dissimilar tasks, whereas RM-GP-TS performs effectively in more favorable cases and is more computationally efficient.

References

Rémi Bardenet, Mátyás Brendel, Balázs Kégl, and Michele Sebag. Collaborative hyperparameter tuning. In *Proc. ICML*, pages 199–207, 2013.

Hans-Georg Beyer and Bernhard Sendhoff. Robust optimization–a comprehensive survey. *Computer methods in applied mechanics and engineering*, 196(33-34): 3190–3218, 2007.

Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. OpenAI Gym. arXiv:1606.01540, 2016.

Sébastien Bubeck. Introduction to online optimization. Lecture notes, 2011.

Olivier Chapelle and Lihong Li. An empirical evaluation of Thompson sampling. In *Proc. NeurIPS*, volume 24, pages 2249–2257. Citeseer, 2011.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *Proc. ICML*, pages 844–853, 2017.

Zhongxiang Dai, Kian Hsiang Low, and Patrick Jaillet. Federated Bayesian optimization via Thompson sampling. In *Proc. NeurIPS*, 2020.

Zhongxiang Dai, Bryan Kian Hsiang Low, and Patrick Jaillet. Differentially private federated Bayesian optimization with distributed exploration. In *Proc. NeurIPS*, volume 34, 2021.

Yan Duan, John Schulman, Xi Chen, Peter L Bartlett, Ilya Sutskever, and Pieter Abbeel. RL^2: Fast reinforcement learning via slow reinforcement learning. arXiv:1611.02779, 2016.

Matthias Feurer, Jost Tobias Springenberg, and Frank Hutter. Initializing Bayesian hyperparameter optimization via meta-learning. In *Proc. AAAI*, pages 1128–1135, 2015.

Matthias Feurer, Benjamin Letham, and Eytan Bakshy. Scalable meta-learning for Bayesian optimization using ranking-weighted Gaussian process ensembles. In *Proc. ICML Workshop on Automatic Machine Learning*, 2018.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proc. ICML*, pages 1126–1135, 2017.

Daniel Golovin, Benjamin Solnik, Subhodeep Moitra, Greg Kochanski, John Karro, and D Sculley. Google Vizier: A service for black-box optimization. In *Proc. ACM SIGKDD*, pages 1487–1495, 2017.

Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In *Proc. ICML*, volume 2. Lille, 2015.

Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338, 2015.

Kunkun Pang, Mingzhi Dong, Yang Wu, and Timothy Hospedales. Meta-learning transferable active learning policies by deep reinforcement learning. arXiv:1806.04798, 2018.

Valerio Perrone, Rodolphe Jenatton, Matthias W Seeger, and Cedric Archambeau. Scalable hyperparameter transfer learning. In *Proc. NIPS*, pages 6845–6855, 2018.

Valerio Perrone, Huibin Shen, Matthias W Seeger, Cédric Archambeau, and Rodolphe Jenatton. Learning search spaces for bayesian optimization: Another view of hyperparameter transfer learning. In *Proc. NeurIPS*, pages 12771–12781, 2019.

Matthias Poloczek, Jialei Wang, and Peter I. Frazier. Warm starting Bayesian optimization. In *Proc. WSC*, pages 770–781, 2016.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Proc. NeurIPS*, pages 1177–1184, 2008.

Anil Ramachandran, Sunil Gupta, Santu Rana, and Svetha Venkatesh. Information-theoretic transfer learning framework for Bayesian optimisation. In *Proc. ECML/PKDD*, pages 827–842. Springer, 2018.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Jorge Luis Reyes-Ortiz, Alessandro Ghio, Xavier Parra, Davide Anguita, Joan Cabestany, and Andreu Catala. Human activity and motion disorder recognition: Towards smarter interactive cognitive environments. In *Proc. ESANN*, 2013.

Daniel Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, and Zheng Wen. A tutorial on Thompson sampling. arxiv:1707.02038, 2017.

David Salinas, Huibin Shen, and Valerio Perrone. A quantile-based approach for hyperparameter transfer learning. In *Proc. ICML*, pages 8438–8448. PMLR, 2020.

Nicolas Schilling, Martin Wistuba, and Lars Schmidt-Thieme. Scalable hyperparameter optimization with products of Gaussian process experts. In *Proc. ECML/PKDD*, pages 33–48, 2016.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016.

Alistair Shilton, Sunil Gupta, Santu Rana, and Svetha Venkatesh. Regret bounds for transfer learning in Bayesian optimisation. In *Proc. AISTATS*, pages 1–9, 2017.

Jack W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, and R. S. Johannes. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In *Proc Annu. Symp. Comput. Appl. Med. Care*, pages 261–265, 1988.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical Bayesian optimization of machine learning algorithms. In *Proc. NeurIPS*, pages 2951–2959, 2012.

Elnaz Soleimani and Ehsan Nazerfard. Cross-subject transfer learning in human activity recognition systems using generative adversarial networks. arxiv:1903.12489, 2019.

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proc. ICML*, pages 1015–1022, 2010.

Kevin Swersky, Jasper Snoek, and Ryan P. Adams. Multi-task Bayesian optimization. In *Proc. NeurIPS*, pages 2004–2012, 2013.

Joaquin Vanschoren. Meta-learning: A survey. arXiv:1810.03548, 2018.

Michael Volpp, Lukas Froehlich, Kirsten Fischer, Andreas Doerr, Stefan Falkner, Frank Hutter, and Christian Daniel. Meta-learning acquisition functions for transfer learning in Bayesian optimization. In *Proc. ICLR*, 2020.

Jane X. Wang, Zeb Kurth-Nelson, Dhruva Tirumala, Hubert Soyer, Joel Z Leibo, Rémi Munos, Charles Blundell, Dharshan Kumaran, and Matt Botvinick. Learning to reinforcement learn. arXiv:1611.05763, 2016.

Zi Wang, Beomjoon Kim, and Leslie Pack Kaelbling. Regret bounds for meta Bayesian optimization with an unknown Gaussian process prior. In *Proc. NeurIPS*, pages 10477–10488, 2018.

Aaron Wilson, Alan Fern, and Prasad Tadepalli. Using trajectory data to improve Bayesian optimization for reinforcement learning. *Journal of Machine Learning Research*, 15(1):253–282, 2014.

Martin Wistuba, Nicolas Schilling, and Lars Schmidt-Thieme. Learning hyperparameter optimization initializations. In *Proc. DSAA*, pages 1–10. IEEE, 2015a.

Martin Wistuba, Nicolas Schilling, and Lars Schmidt-Thieme. Sequential model-free hyperparameter tuning. In *Proc. ICDM*, pages 1033–1038, 2015b.

Martin Wistuba, Nicolas Schilling, and Lars Schmidt-Thieme. Two-stage transfer surrogate model for automatic hyperparameter optimization. In *Proc. ECML/PKDD*, pages 199–214, 2016.

Martin Wistuba, Nicolas Schilling, and Lars Schmidt-Thieme. Scalable Gaussian process-based transfer surrogates for hyperparameter optimization. *Machine Learning*, 107(1):43–78, 2018.

Zhongwen Xu, Hado P van Hasselt, and David Silver. Meta-gradient reinforcement learning. In *Proc. NeurIPS*, pages 2396–2407, 2018.

Dani Yogatama and Gideon Mann. Efficient transfer learning method for automatic hyperparameter tuning. In *Proc. AISTATS*, pages 1077–1085, 2014.

A MORE RELATED WORKS

Some previous works on meta-BO build a joint GP surrogate using all previous and current observations, and represent task similarity through meta-features [Bardenet et al., 2013, Schilling et al., 2016, Yogatama and Mann, 2014]. However, these algorithms suffer from the requirement of handcrafted meta-features, which is avoided in other works that learn task similarity from the observations [Swersky et al., 2013, Shilton et al., 2017]. For example, multitask BO [Swersky et al., 2013] uses a multitask GP as a surrogate and models each task as an output of the GP. These works include all previous and current observations in a single GP surrogate and are thus limited by GP's scalability. There have also been other empirical works which replace GP with Bayesian linear regression for scalability [Perrone et al., 2018], tackle sequentially arriving tasks [Golovin et al., 2017, Poloczek et al., 2016], learn a set of good initializations [Feurer et al., 2015, Wistuba et al., 2015b], learn a reduced search space for BO from previous tasks [Perrone et al., 2019], handle the issue of different function scales using Gaussian Copulas [Salinas et al., 2020], learn the task similarities through the distance between the distributions of the optima from different tasks [Ramachandran et al., 2018], or use the meta-observations to learn the entire acquisition function through RL [Volpp et al., 2020].

B PROOF OF THEOREM 1

To begin with, we need the following lemma to give a high-probability confidence bound on the target function, which will be used in the theoretical analysis of both Theorems 1 and 2.

Lemma 2. *Let $\delta \in (0,1)$ and $\beta_t = B + \sigma\sqrt{2(\gamma_{t-1} + 1 + \log(4/\delta))}$, then*

$$|f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})| \leq \beta_t \sigma_{t-1}(\mathbf{x}) \qquad \forall \mathbf{x} \in \mathcal{D},\, t \geq 1$$

which holds with probability of $\geq 1 - \delta/4$.

Lemma B follows directly from Theorem 2 of [Chowdhury and Gopalan, 2017].

To facilitate the theoretical analysis of RM-GP-UCB, we introduce the following auxiliary term:

$$\widetilde{\zeta}_t(\mathbf{x}) = \nu_t \left[\sum_{i=1}^{M} \omega_i \left[\widetilde{\mu}_i(\mathbf{x}) + \tau\widetilde{\sigma}_i(\mathbf{x})\right]\right] + (1-\nu_t)\left[\mu_{t-1}(\mathbf{x}) + \beta_t\sigma_{t-1}(\mathbf{x})\right] \tag{8}$$

in which $\widetilde{\mu}_i(\mathbf{x})$ and $\widetilde{\sigma}_i(\mathbf{x})$ are obtained by replacing each noisy output of the meta-observations $y_{i,j}$ in the calculation of $\overline{\mu}_i(\mathbf{x})$ and $\overline{\sigma}_i(\mathbf{x})$ (2) by the (hypothetically available) noisy target function output observation at the corresponding input $\mathbf{x}_{i,j}$. Eq. (8) will serve as the bridge to connect the acquisition function of RM-GP-UCB (2) with the target function f in the subsequent theoretical analysis, which will be demonstrated in Appendix B.2. The next lemma shows that the difference between $\overline{\zeta}_t(\mathbf{x})$ (2) and $\widetilde{\zeta}_t(\mathbf{x})$ (8) is bounded $\forall \mathbf{x} \in \mathcal{D}$, whose proof is given in Appendix B.1.

Lemma 3. *Let $\delta \in (0,1)$. Suppose the RM-GP-UCB algorithm is run with parameters $\nu_t \in [0,1]\ \forall t \geq 1$, and $\omega_i \geq 0$ for $i = 1, \ldots, M$ and $\sum_{i=1,\ldots,M} \omega_i = 1$. Then with probability of $\geq 1 - \delta/4$,*

$$\left|\overline{\zeta}_t(\mathbf{x}) - \widetilde{\zeta}_t(\mathbf{x})\right| \leq \nu_t \alpha \qquad \forall \mathbf{x} \in \mathcal{D}$$

in which

$$\alpha \triangleq \sum_{i=1}^{M} \omega_i \frac{N_i}{\sigma^2}\left(2\sqrt{2\sigma^2 \log\frac{8N_i}{\delta}} + d_i\right).$$

Next, because $\widetilde{\mu}_i(\mathbf{x})$ and $\widetilde{\sigma}_i(\mathbf{x})$ are calculated using the (hypothetically available) noisy observations of the target function (i.e., same as $\mu_{t-1}(\mathbf{x})$ and $\sigma_{t-1}(\mathbf{x})$), we can also get the following lemma on the concentration of the target function f which, similar to Lemma 2 above, also follows directly from Theorem 2 of [Chowdhury and Gopalan, 2017].

Lemma 4. *Let $\tau = B + \sigma\sqrt{2(\gamma_N + 1 + \log(4M/\delta))}$, we have that*

$$|f(\mathbf{x}) - \widetilde{\mu}_i(\mathbf{x})| \leq \tau\widetilde{\sigma}_i(\mathbf{x}) \qquad \forall \mathbf{x} \in \mathcal{D},\, i = 1, ..., M,$$

which also holds with probability $\geq 1 - \delta/4$.

B.1 PROOF OF LEMMA 3

Let $\mathbf{K}_i = [k(\mathbf{x}_{i,j}, \mathbf{x}_{i,j'})]_{j,j'=1,\ldots,N_i}$ represent the Gram matrix corresponding to the inputs of the meta-observations from meta-task i, and $\mathbf{k}_i = [k(\mathbf{x}_{i,j}, \mathbf{x})]^\top_{j=1,\ldots,N_i}$. Denote by $\lambda_j[\mathbf{A}]$ the j-th eigenvalue of matrix $\mathbf{A}$. Firstly, we need the following lemma proving an upper bound on matrix L_2 norm:

Lemma 5. *For all $i = 1, \ldots, M$, we have that*

$$\left\| \left(\mathbf{K}_i + \sigma^2 I \right)^{-1} \right\|_2 \leq \frac{1}{\sigma^2}.$$

Proof.

$$\begin{aligned}
\left\| \left(\mathbf{K}_i + \sigma^2 I \right)^{-1} \right\|_2 &= \sqrt{\max_{j=1,\ldots,N_i} \lambda_j \left[\left((\mathbf{K}_i + \sigma^2 I)^{-1} \right)^\top (\mathbf{K}_i + \sigma^2 I)^{-1} \right]} \\
&= \sqrt{\max_{j=1,\ldots,N_i} \lambda_j \left[(\mathbf{K}_i + \sigma^2 I)^{-1} \right]^2} \\
&\leq \frac{1}{\sigma^2}
\end{aligned}$$

□

Next, define $\overline{\mathbf{f}}_i = [f_i(\mathbf{x}_{i,j})]_{j=1,\ldots,N_i}$ (in which $f_i(\mathbf{x}_{i,j})$ represents the value of meta-function i at input $\mathbf{x}_{i,j}$), and $\widetilde{\mathbf{f}}_i = [f(\mathbf{x}_{i,j})]_{j=1,\ldots,N_i}$ (in which $f(\mathbf{x}_{i,j})$ represents the value of target function at input $\mathbf{x}_{i,j}$). Similarly, define $\overline{\mathbf{y}}_i = [y_{i,j}]_{j=1,\ldots,N_i}$ (in which $y_{i,j}$ represents the noisy output observation of meta-task i at input $\mathbf{x}_{i,j}$), and $\widetilde{\mathbf{y}}_i = [y(\mathbf{x}_{i,j})]_{j=1,\ldots,N_i}$ (in which $y(\mathbf{x}_{i,j})$ represents the hypothetically observed noisy output observation of the target function at input $\mathbf{x}_{i,j}$). With these definitions, the next lemma shows upper bounds on the distance between $\overline{\mathbf{y}}_i$ and $\overline{\mathbf{f}}_i$, as well as that distance between $\widetilde{\mathbf{y}}_i$ and $\widetilde{\mathbf{f}}_i$.

Lemma 6. *With probability $\geq 1 - \delta/4$,*

$$\begin{aligned}
\left\| \overline{\mathbf{y}}_i - \overline{\mathbf{f}}_i \right\|_2 &\leq \sqrt{N_i} \sqrt{2\sigma^2 \log \frac{8N_i}{\delta}}, \\
\left\| \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{f}}_i \right\|_2 &\leq \sqrt{N_i} \sqrt{2\sigma^2 \log \frac{8N_i}{\delta}}.
\end{aligned}$$

Proof. Following the same analysis as Lemma 5.1 of [Srinivas et al., 2010], we have that for the standard Gaussian random variable $z \sim \mathcal{N}(0, 1)$,

$$\mathbb{P}(|z| > c) \leq e^{-\frac{c^2}{2}}. \tag{9}$$

Since for each $j = 1, \ldots, N_i$, we have that $y_{i,j} - f_i(\mathbf{x}_{i,j}) \sim \mathcal{N}(0, \sigma^2)$ and that $y(\mathbf{x}_{i,j}) - f(\mathbf{x}_{i,j}) \sim \mathcal{N}(0, \sigma^2)$, which leads to the following,

$$\begin{aligned}
\mathbb{P}\left(\left| \frac{y_{i,j} - f_i(\mathbf{x}_{i,j})}{\sigma} \right| > \sqrt{2 \log \frac{8N_i}{\delta}} \right) &= \mathbb{P}\left(\left| y_{i,j} - f_i(\mathbf{x}_{i,j}) \right| > \sqrt{2\sigma^2 \log \frac{8N_i}{\delta}} \right) \leq \frac{\delta}{8N_i}, \\
\mathbb{P}\left(\left| \frac{y(\mathbf{x}_{i,j}) - f(\mathbf{x}_{i,j})}{\sigma} \right| > \sqrt{2 \log \frac{8N_i}{\delta}} \right) &= \mathbb{P}\left(\left| y(\mathbf{x}_{i,j}) - f(\mathbf{x}_{i,j}) \right| > \sqrt{2\sigma^2 \log \frac{8N_i}{\delta}} \right) \leq \frac{\delta}{8N_i}.
\end{aligned}$$

Taking a union bound over $j = 1, \ldots, N_i$ for each of the two equations above, we have that for *all* $j = 1, \ldots, N_i$,

$$\begin{aligned}
\left| y_{i,j} - f_i(\mathbf{x}_{i,j}) \right| &\leq \sqrt{2\sigma^2 \log \frac{8N_i}{\delta}}, \\
\left| y(\mathbf{x}_{i,j}) - f(\mathbf{x}_{i,j}) \right| &\leq \sqrt{2\sigma^2 \log \frac{8N_i}{\delta}},
\end{aligned}$$

both of which hold with probability $\geq 1-\delta/8$. Therefore, with probability $\geq 1-\delta/8$,

$$\left\|\overline{\mathbf{y}}_i - \overline{\mathbf{f}}_i\right\|_2 = \sqrt{\sum_{j=1}^{N_i}\left|y_{i,j} - f_i(\mathbf{x}_{i,j})\right|^2} \leq \sqrt{\sum_{j=1}^{N_i} 2\sigma^2 \log\frac{8N_i}{\delta}} \leq \sqrt{N_i}\sqrt{2\sigma^2\log\frac{8N_i}{\delta}}. \tag{10}$$

Repeating the procedure above leads to

$$\left\|\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{f}}_i\right\|_2 \leq \sqrt{N_i}\sqrt{2\sigma^2\log\frac{8N_i}{\delta}} \tag{11}$$

which also holds with probability $\geq 1-\delta/8$. Taking a union bound over equations (10) and (11) completes the proof. □

With these supporting lemmas, Lemma 3 can be proved as follows:

$$\begin{aligned}
\left|\overline{\zeta}_t(\mathbf{x}) - \widetilde{\zeta}_t(\mathbf{x})\right| &= \left|\nu_t\left[\sum_{i=1}^{M}\omega_i[\overline{\mu}_i(\mathbf{x}) + \sqrt{\tau}\overline{\sigma}_i(\mathbf{x})]\right] - \nu_t\left[\sum_{i=1}^{M}\omega_i[\widetilde{\mu}_i(\mathbf{x}) + \sqrt{\tau}\widetilde{\sigma}_i(\mathbf{x})]\right]\right| \\
&\overset{\text{(a)}}{=} \left|\nu_t\sum_{i=1}^{M}\omega_i[\overline{\mu}_i(\mathbf{x}) - \widetilde{\mu}_i(\mathbf{x})]\right| \\
&\leq \nu_t\sum_{i=1}^{M}\omega_i\left|\overline{\mu}_i(\mathbf{x}) - \widetilde{\mu}_i(\mathbf{x})\right| \\
&\leq \nu_t\sum_{i=1}^{M}\omega_i\left|\mathbf{k}_i(\mathbf{x})^\top(\mathbf{K}_i + \sigma^2 I)^{-1}(\overline{\mathbf{y}}_i - \widetilde{\mathbf{y}}_i)\right| \\
&\overset{\text{(b)}}{\leq} \nu_t\sum_{i=1}^{M}\omega_i\left\|\mathbf{k}_i(\mathbf{x})\right\|_2\left\|(\mathbf{K}_i + \sigma^2 I)^{-1}\right\|_2\|\overline{\mathbf{y}}_i - \widetilde{\mathbf{y}}_i\|_2 \\
&\overset{\text{(c)}}{\leq} \nu_t\sum_{i=1}^{M}\omega_i\left\|\mathbf{k}_i(\mathbf{x})\right\|_2\frac{1}{\sigma^2}\|\overline{\mathbf{y}}_i - \widetilde{\mathbf{y}}_i\|_2 \\
&\overset{\text{(d)}}{\leq} \nu_t\sum_{i=1}^{M}\omega_i\sqrt{N_i}\frac{1}{\sigma^2}\|\overline{\mathbf{y}}_i - \widetilde{\mathbf{y}}_i\|_2 \\
&\leq \nu_t\sum_{i=1}^{M}\omega_i\frac{\sqrt{N_i}}{\sigma^2}\left\|\overline{\mathbf{y}}_i - \overline{\mathbf{f}}_i + \overline{\mathbf{f}}_i - \widetilde{\mathbf{f}}_i + \widetilde{\mathbf{f}}_i - \widetilde{\mathbf{y}}_i\right\|_2 \\
&\leq \nu_t\sum_{i=1}^{M}\omega_i\frac{\sqrt{N_i}}{\sigma^2}\left[\left\|\overline{\mathbf{y}}_i - \overline{\mathbf{f}}_i\right\|_2 + \left\|\overline{\mathbf{f}}_i - \widetilde{\mathbf{f}}_i\right\|_2 + \left\|\widetilde{\mathbf{f}}_i - \widetilde{\mathbf{y}}_i\right\|_2\right] \\
&\overset{\text{(e)}}{\leq} \nu_t\sum_{i=1}^{M}\omega_i\frac{\sqrt{N_i}}{\sigma^2}\left(2\sqrt{N_i}\sqrt{2\sigma^2\log\frac{8N_i}{\delta}} + \left\|\overline{\mathbf{f}}_i - \widetilde{\mathbf{f}}_i\right\|_2\right) \\
&= \nu_t\sum_{i=1}^{M}\omega_i\frac{\sqrt{N_i}}{\sigma^2}\left(2\sqrt{N_i}\sqrt{2\sigma^2\log\frac{8N_i}{\delta}} + \sqrt{\sum_{j=1}^{N_i}\left(f_i(\mathbf{x}_{i,j}) - f(\mathbf{x}_{i,j})\right)^2}\right) \\
&\overset{\text{(f)}}{\leq} \nu_t\sum_{i=1}^{M}\omega_i\frac{\sqrt{N_i}}{\sigma^2}\left(2\sqrt{N_i}\sqrt{2\sigma^2\log\frac{8N_i}{\delta}} + d_i\sqrt{N_i}\right) \\
&= \nu_t\sum_{i=1}^{M}\omega_i\frac{N_i}{\sigma^2}\left(2\sqrt{2\sigma^2\log\frac{8N_i}{\delta}} + d_i\right) \\
&\triangleq \nu_t\alpha
\end{aligned} \tag{12}$$

which holds with probability $\geq 1-\delta/4$. (a) holds because $\overline{\sigma}_i(\mathbf{x}) = \widetilde{\sigma}_i(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{D}$, because the posterior standard deviation only depends on the input locations and is independent of the corresponding output responses; (b) follows from Cauchy-Schwarz inequality, (c) follows from Lemma 5, (d) results from the assumption w.l.o.g. that $k\left(\mathbf{x}, \mathbf{x}'\right) \leq 1$ for all $\mathbf{x}, \mathbf{x}' \in \mathcal{D}$, (e) follows from Lemma 6, (f) is obtained from the definition of the function gap: $d_i \triangleq \max_{j=1,\ldots,N_i} \left|f(\mathbf{x}_{i,j}) - f_i(\mathbf{x}_{i,j})\right|$ for $i = 1, \ldots, M$. This completes the proof of Lemma 3.

B.2 PROOF OF THEOREM 1

To begin with, we need the following lemma showing a high-probability upper bound on the global maximum of the target function.

Lemma 7. *Given $\delta \in (0,1)$. Let $\mathbf{x}^*$ denote a global maximizer of the target function f, and α be as defined in Lemma 3. Suppose the RM-GP-UCB algorithm is run with the parameter $\nu_t \in [0,1]$ for all $t \geq 1$. Then, with probability $\geq 1-3\delta/4$,*

$$f(\mathbf{x}^*) \leq \overline{\zeta}_t(\mathbf{x}_t) + \nu_t \alpha \qquad \forall t \geq 1.$$

Proof. Firstly, as a result of Lemma 2 and Lemma 4 (both hold with probability of $\geq 1-\delta/4$), at any iteration $t \geq 1$ and for all $\mathbf{x} \in \mathcal{D}$, we have that with probability $\geq 1-\delta/4-\delta/4$, $\widetilde{\zeta}_t(\mathbf{x})$ is an upper bound on $f(\mathbf{x})$:

$$\begin{aligned} \widetilde{\zeta}_t(\mathbf{x}) - f(\mathbf{x}) &= \widetilde{\zeta}_t(\mathbf{x}) - \left[\nu_t \sum_{i=1}^{M} \omega_i f(\mathbf{x}) + (1-\eta_t) f(\mathbf{x})\right] \\ &= \nu_t \sum_{i=1}^{M} \omega_i \left[\widetilde{\mu}_i(\mathbf{x}) + \sqrt{\tau}\widetilde{\sigma}_i(\mathbf{x}) - f(\mathbf{x})\right] + (1-\nu_t)\left[\mu_{t-1}(\mathbf{x}) + \sqrt{\beta_t}\sigma_{t-1}(\mathbf{x}) - f(\mathbf{x})\right] \geq 0. \end{aligned} \tag{13}$$

Therefore, with probability $\geq 1-\delta/4-\delta/4-\delta/4$,

$$f(\mathbf{x}^*) \overset{\text{(a)}}{\leq} \widetilde{\zeta}_t(\mathbf{x}^*) \overset{\text{(b)}}{\leq} \overline{\zeta}_t(\mathbf{x}^*) + \nu_t \alpha \overset{\text{(c)}}{\leq} \overline{\zeta}_t(\mathbf{x}_t) + \nu_t \alpha \tag{14}$$

in which (a) results from (13), (b) is obtained via Lemma 3 which holds with probability of $\geq 1-\delta/4$, and (c) follows from the policy for selecting $\mathbf{x}_t$, i.e., by maximizing (2). This completes the proof. □

Subsequently, we can show a high-probability upper bound on the instantaneous regret with the following lemma .

Lemma 8. *Given $\delta \in (0,1)$. Let α be as defined in Lemma 3. Suppose the RM-GP-UCB algorithm is run with the parameters β_t, τ and ν_t. Then, with probability $\geq 1-3\delta/4$, $\forall t \geq 1$,*

$$r_t \leq 2\nu_t(\alpha+\tau) + 2(1-\nu_t)\beta_t \sigma_{t-1}(\mathbf{x}_t).$$

Proof. The instantaneous regret can be upper-bounded by

$$
\begin{aligned}
r_t &= f(\mathbf{x}^*) - f(\mathbf{x}_t) \overset{\text{(a)}}{\leq} \overline{\zeta}_t(\mathbf{x}_t) + \nu_t\alpha - f(\mathbf{x}_t) \\
&\leq \underline{\overline{\zeta}_t(\mathbf{x}_t) - \widetilde{\zeta}_t(\mathbf{x}_t)} + \widetilde{\zeta}_t(\mathbf{x}_t) - f(\mathbf{x}_t) + \nu_t\alpha \\
&\overset{\text{(b)}}{\leq} \nu_t\alpha + \nu_t \sum_{i=1}^{M} \omega_i \left[\widetilde{u}_i(\mathbf{x}_t) + \tau\widetilde{\sigma}_i(\mathbf{x}_t)\right] + (1-\nu_t)\left[u_{t-1}(\mathbf{x}_t) + \beta_t\sigma_{t-1}(\mathbf{x}_t)\right] \\
&\qquad - f(\mathbf{x}_t) + \nu_t\alpha \\
&= \nu_t\alpha + \nu_t \sum_{i=1}^{M} \omega_i \left[\widetilde{u}_i(\mathbf{x}_t) + \tau\widetilde{\sigma}_i(\mathbf{x}_t)\right] + (1-\nu_t)\left[u_{t-1}(\mathbf{x}_t) + \beta_t\sigma_{t-1}(\mathbf{x}_t)\right] \\
&\qquad - \left[\nu_t \sum_{i=1}^{M} \omega_i f(\mathbf{x}_t) + (1-\nu_t) f(\mathbf{x}_t)\right] + \nu_t\alpha \qquad (15) \\
&\leq \nu_t\alpha + \nu_t \sum_{i=1}^{M} \omega_i \left[\widetilde{u}_i(\mathbf{x}_t) - f(\mathbf{x}_t)\right] + \nu_t \sum_{i=1}^{M} \omega_i \tau\widetilde{\sigma}_i(\mathbf{x}_t) \\
&\qquad + (1-\nu_t)\left[u_{t-1}(\mathbf{x}_t) - f(\mathbf{x}_t)\right] + (1-\nu_t)\beta_t\sigma_{t-1}(\mathbf{x}_t) + \nu_t\alpha \\
&\overset{\text{(c)}}{\leq} 2\nu_t\alpha + 2\nu_t \sum_{i=1}^{M} \omega_i \tau\widetilde{\sigma}_i(\mathbf{x}_t) + 2(1-\nu_t)\beta_t\sigma_{t-1}(\mathbf{x}_t) \\
&\overset{\text{(d)}}{\leq} 2\nu_t\alpha + 2\nu_t\tau + 2(1-\nu_t)\beta_t\sigma_{t-1}(\mathbf{x}_t) \\
&\leq 2\nu_t(\alpha+\tau) + 2(1-\nu_t)\beta_t\sigma_{t-1}(\mathbf{x}_t)
\end{aligned}
$$

which holds with probability $\geq 1 - 3\delta/4$. (a) follows from Lemma 7 which holds with probability of $\geq 1 - 3\delta/4$, (b) results from Lemma 3 as well as the definition of $\widetilde{\zeta}_t(\mathbf{x}_t)$ (8), (c) is a result of Lemma 2 and Lemma 4, and (d) follows because $\widetilde{\sigma}_i(\mathbf{x}_t) \leq 1$ for all $\mathbf{x}_t \in \mathcal{D}$, which can be easily verified using the formula of the GP posterior variance (1) and the assumption that $k(\mathbf{x}, \mathbf{x}') \leq 1$ for all $\mathbf{x}, \mathbf{x}' \in \mathcal{D}$. The error probabilities $3\delta/4 = \delta/4 + \delta/4 + \delta/4$ result from Lemmas 2, 3 and 4. □

Next, we need to connect the second term from Lemma 8 with the information gain. The following lemma, which is Lemma 5.3 of [Srinivas et al., 2010], defines the information gain on the target function from any set of observations.

Lemma 9. *Let $\mathbf{f}_T$ and $\mathbf{y}_T$ denote the set of function values and noisy observations of the target function respectively after T iterations. Then, the information gain about f from the first T observations can be expressed as*

$$
I(\mathbf{y}_T; \mathbf{f}_T) = \frac{1}{2} \sum_{t=1}^{T} \log\left[1 + \sigma^{-2}\sigma_{t-1}^2(\mathbf{x}_t)\right].
$$

Subsequently, we can upper bound the second term from Lemma 8 (summed from iterations 1 to T) by the maximum information gain via the following lemma.

Lemma 10. *Suppose the RM-GP-UCB algorithm is run with the parameters β_t $\forall t \geq 1$ and a non-increasing sequence $\nu_t \in [0,1]$ $\forall t \geq 1$. Define the maximum information gain as $\gamma_T = \max_{A \in \mathcal{D}, |A| = T} I(\mathbf{y}_A; \mathbf{f}_A)$ in which $\mathbf{f}_A$ and $\mathbf{y}_A$ represent the function values and noisy observations from a set A of inputs of size T. Then,*

$$
\sum_{t=1}^{T} \left[2(1-\nu_t)\beta_t\sigma_{t-1}(\mathbf{x}_t)\right]^2 \leq (1-\nu_T)^2 C_1 \beta_T^2 \gamma_T
$$

in which $C_1 \triangleq \frac{8}{\log(1+\sigma^{-2})}$.

Proof. Each term inside the summation can be upper-bounded by

$$\begin{aligned}
4(1-\nu_t)^2\beta_t^2\sigma_{t-1}^2(\mathbf{x}_t) &\overset{\text{(a)}}{\leq} 4(1-\nu_T)^2\beta_T^2\sigma^2\left(\sigma^{-2}\sigma_{t-1}^2(\mathbf{x}_t)\right) \\
&\overset{\text{(b)}}{\leq} 4(1-\nu_T)^2\beta_T^2\sigma^2\left(\frac{\sigma^{-2}}{\log(1+\sigma^{-2})}\log\left(1+\sigma^{-2}\sigma_{t-1}^2(\mathbf{x}_t)\right)\right) \\
&= (1-\nu_T)^2\beta_T^2\frac{8}{\log(1+\sigma^{-2})}\left[\frac{1}{2}\log\left(1+\sigma^{-2}\sigma_{t-1}^2(\mathbf{x}_t)\right)\right]
\end{aligned} \tag{16}$$

in which (a) follows since β_t is non-decreasing in t and ν_t is non-increasing in t, (b) follows since $\sigma^{-2}x \leq \frac{\sigma^{-2}}{\log(1+\sigma^{-2})}\log(1+\sigma^{-2}x)$ for all $x \in (0,1]$ and $\sigma_{t-1}^2(\mathbf{x}_t) \in (0,1]$.

As a result, the summation can be decomposed as

$$\begin{aligned}
\sum_{t=1}^{T}\left[2(1-\nu_t)\beta_t\sigma_{t-1}(\mathbf{x}_t)\right]^2 &\overset{\text{(a)}}{\leq} (1-\nu_T)^2\beta_T^2\frac{8}{\log(1+\sigma^{-2})}\sum_{t=1}^{T}\left[\frac{1}{2}\log\left(1+\sigma^{-2}\sigma_{t-1}^2(\mathbf{x}_t)\right)\right] \\
&\overset{\text{(b)}}{=} (1-\nu_T)^2\beta_T^2\frac{8}{\log(1+\sigma^{-2})}I(\mathbf{y}_T;\mathbf{f}_T) \\
&\overset{\text{(c)}}{\leq} (1-\nu_T)^2 C_1\beta_T^2\gamma_T
\end{aligned}$$

in which (a) results from (16), (b) follows from Lemma 9, and (c) is obtained by making use of the definition of C_1 and γ_T. □

Finally, an upper bound on the cumulative regret follows from combining these supporting lemmas:

$$\begin{aligned}
R_T = \sum_{t=1}^{T} r_t &\overset{\text{(a)}}{\leq} \sum_{t=1}^{T}\left[2\nu_t(\alpha+\tau) + 2\left(1-\nu_t\right)\beta_t\sigma_{t-1}(\mathbf{x}_t)\right] \\
&= 2(\alpha+\tau)\sum_{t=1}^{T}\nu_t + \sum_{t=1}^{T}2(1-\nu_t)\beta_t\sigma_{t-1}(\mathbf{x}_t) \\
&\overset{\text{(b)}}{\leq} 2(\alpha+\tau)\sum_{t=1}^{T}\nu_t + \sqrt{T}\sqrt{\sum_{t=1}^{T}\left[2(1-\nu_t)\beta_t\sigma_{t-1}(\mathbf{x}_t)\right]^2} \\
&\overset{\text{(c)}}{\leq} 2(\alpha+\tau)\sum_{t=1}^{T}\nu_t + \sqrt{C_1 T(1-\nu_T)^2\beta_T^2\gamma_T} \\
&\overset{\text{(d)}}{\leq} 2(\alpha+\tau)\sum_{t=1}^{T}\nu_t + \beta_T\sqrt{C_1 T\gamma_T}
\end{aligned} \tag{17}$$

which holds with probability $\geq 1-3\delta/4$. (a) is a result of Lemma 8, (b) follows from Cauchy-Schwarz inequality, (c) is obtained using Lemma 10, and (d) follows since $1-\nu_T \leq 1$. This completes the proof.

If the meta-weights ω_i's are allowed to change with t (i.e., when our online meta-weight optimization is used), then the proof here only needs to be modified to let α depend on t: $R_T \leq 2\tau\sum_{t=1}^{T}\nu_t + 2\sum_{t=1}^{T}\nu_t\alpha_t + \beta_T\sqrt{C_1 T\gamma_T}$. In this case, the no-regret convergence guarantee of RM-GP-UCB (Sec. 4.1) is still preserved since in this case, we can simply upper-bound every $\omega_{i,t}$ by 1. That is $R_T \leq 2(\alpha'+\tau)\sum_{t=1}^{T}\nu_t + \beta_T\sqrt{C_1 T\gamma_T}$, with $\alpha' \triangleq \sum_{i=1}^{M}\frac{N_i}{\sigma^2}(2\sqrt{2\sigma^2\log\frac{8N_i}{\delta}} + d_i)$.

B.3 META-TASKS CAN IMPROVE THE CONVERGENCE BY ACCELERATING EXPLORATION

Here, we utilize the analysis in Appendix B.2 to illustrate how the meta-tasks (if similar to the target task) can help RM-GP-UCB obtain a better regret bound than standard GP-UCB in the early stage of the algorithm. For simplicity, we focus on the most favorable scenario where all meta-functions have equal values to the target function at their corresponding input

locations, i.e., all function gaps are 0: $d_i = \max_{j=1,\ldots,N_i} \left| f(\mathbf{x}_{i,j}) - f_i(\mathbf{x}_{i,j}) \right| = 0, \forall i = 1, \ldots, M$. Although not realistic, this scenario is useful for illustrating how the meta-tasks help our RM-GP-UCB algorithm achieve a better convergence at the initial stage.

In this case, according to the definition of $\widetilde{\zeta}_t$ (8) and $\overline{\zeta}_t$ (2), we have that $\widetilde{\zeta}_t(\mathbf{x}) = \overline{\zeta}_t(\mathbf{x}), \forall \mathbf{x} \in \mathcal{D}, t \geq 1$. As a result, the analysis of (14) in the proof of Lemma 7 can be similarly applied, yielding:

$$f(\mathbf{x}^*) \leq \widetilde{\zeta}_t(\mathbf{x}^*) = \overline{\zeta}_t(\mathbf{x}^*) \leq \overline{\zeta}_t(\mathbf{x}_t). \tag{18}$$

Next, we can re-analyze the instantaneous regret following similar steps to (15):

$$\begin{aligned} r_t &= f(\mathbf{x}^*) - f(\mathbf{x}_t) \leq \overline{\zeta}_t(\mathbf{x}_t) - f(\mathbf{x}_t) \\ &\leq 2\nu_t \sum_{i=1}^{M} \omega_i \tau \overline{\sigma}_i(\mathbf{x}_t) + 2(1-\nu_t)\beta_t \sigma_{t-1}(\mathbf{x}_t) \\ &= \underbrace{2\nu_t \left(\sum_{i=1}^{M} \omega_i \tau \overline{\sigma}_i(\mathbf{x}_t) - \beta_t \sigma_{t-1}(\mathbf{x}_t) \right)}_{A1} + \underbrace{2\beta_t \sigma_{t-1}(\mathbf{x}_t)}_{A2}, \end{aligned} \tag{19}$$

in which some intermediate steps that are identical to those used in (15) have been omitted for simplicity. Note that term A_2 in (19) is identical to the upper bound on the instantaneous regret for the standard GP-UCB algorithm [Srinivas et al., 2010]. Therefore, the meta-tasks affect the upper bound on the instantaneous regret through the term A_1.

Recall Theorem 1 has told us that we should choose $\nu_t \to 0$ as $t \to \infty$. In the initial stage of the algorithm when ν_t is large, the impact of A_1 on the regret of the algorithm is large. In this case, the meta-tasks improve the upper bound on the instantaneous regret (compared with standard GP-UCB) if $A_1 < 0$, that is:

$$\sum_{i=1}^{M} \omega_i \overline{\sigma}_i(\mathbf{x}_t) < \frac{\beta_t}{\tau} \sigma_{t-1}(\mathbf{x}_t). \tag{20}$$

In other words, RM-GP-UCB converges faster than standard GP-UCB in the initial stage if the (weighted combination of) meta-tasks have smaller uncertainty (i.e., posterior standard deviation) at $\mathbf{x}_t$ compared with the target task (scaled by β_t/τ). Fortunately, in the early stage of the algorithm, this condition is highly likely to be satisfied: When the number of observations of the target task is small, the posterior standard deviation of the target GP posterior (i.e., RHS of Equation (20)) is usually large; therefore, Equation (20) is highly likely to be satisfied. This insight turns out to have an intuitive and elegant interpretation as well. In the initial stage of the standard GP-UCB algorithm, due to the lack of observations, the algorithm *has large uncertainty* regarding the objective function and hence tends to *explore*; however, the meta-tasks (assuming that they are similar to the target task) provides additional information for the algorithm, which *reduces the uncertainty* about the objective function and hence *decreases the requirement for initial exploration*. To summarize, in the initial stage, the meta-tasks, if similar to the target task, help RM-GP-UCB achieve smaller regret upper bound (hence converge faster) than GP-UCB by reducing the degree of exploration. In less favorable scenarios where the function gaps are nonzero (i.e., the meta-functions are not exactly equal to the target function), some amount of errors will be introduced to the upper bound on the instantaneous regret (19). As a results, a positive error term will be added to the LHS of (20), making the theoretical condition for a faster convergence (20) harder to satisfy. At later stages where ν_t is already small and close to 0, the impact of the term A_1 is significantly diminished, thus allowing our RM-GP-UCB algorithm to converge to no regret at a similar rate to standard GP-UCB.

C PROOF OF THEOREM 2

Our theoretical analysis of RM-GP-TS shares similarity with the works of [Dai et al., 2020, 2021] but has important differences, e.g., unlike the works of [Dai et al., 2020, 2021], RM-GP-TS does not suffer from the error introduced by random Fourier features approximation since we do not need to consider the issues of retaining (hence not transmitting) the raw data and communication efficiency.

Based on the acquisition function for RM-GP-TS (3), define $\mathcal{E}_t^1$ as the event that $\overline{\zeta}_t(\mathbf{x}) = f^t(\mathbf{x})$ which happens with probability $1 - \nu_t$, and define $\mathcal{E}_t^2$ as the event that $\overline{\zeta}_t(\mathbf{x}) = \sum_{i=1}^{M} \omega_i \left[\overline{f}_i^t(\mathbf{x}) \right]$ which happens with probability ν_t. Define $\mathcal{F}_{t-1}$ as the filtration containing the history of input-output pairs of the target task up to and including iteration $t-1$.

Lemma 11. *With τ defined in Lemma 4, we have that*

$$|f_i(\mathbf{x}) - \overline{\mu}_i(\mathbf{x})| \leq \tau \overline{\sigma}_i(\mathbf{x}) \qquad \forall \mathbf{x} \in \mathcal{D},\ i = 1, ..., M$$

which holds with probability $\geq 1 - \delta/4$.

Similar to Lemma 2 and Lemma 4, Lemma 11 also follows from Theorem 2 of [Chowdhury and Gopalan, 2017]. Next, we also need the following lemma showing the concentration of functions sampled from the GP posterior around the posterior mean, for both the target function and the meta-functions.

Lemma 12. *With β_t defined in Lemma 2 and τ defined in Lemma 4, we have that*

$$|f^t(\mathbf{x}) - \mu_{t-1}(\mathbf{x})| \leq \beta_t \sqrt{2\log(\frac{|\mathcal{D}|t^2 2\pi^2}{\delta})}\sigma_{t-1}(\mathbf{x}), \qquad \forall \mathbf{x} \in \mathcal{D},\ t \geq 1,$$

which holds with probability $\geq 1 - \delta/12$, and that

$$|f_i^t(\mathbf{x}) - \overline{\mu}_i(\mathbf{x})| \leq \tau \sqrt{2\log(\frac{M|\mathcal{D}|t^2 2\pi^2}{\delta})}\overline{\sigma}_i(\mathbf{x}), \qquad \forall \mathbf{x} \in \mathcal{D},\ t \geq 1,\ i = 1, ..., M$$

which holds with probability $\geq 1 - \delta/12$.

The proof of Lemma 12 follows straightforwardly from Lemma 5 of [Chowdhury and Gopalan, 2017], together with a union bound over all $\mathbf{x} \in \mathcal{D}$ and over all $t \geq 1$, as well as an additional union bound over all M meta-tasks for the second inequality.

Lemma 13. *Define $d_i' \triangleq \max_{\mathbf{x}\in\mathcal{D}} |f(\mathbf{x}) - f_i(\mathbf{x})|$. Define $c_t \triangleq \beta_t \left(1 + \sqrt{2\log(\frac{|\mathcal{D}|t^2 2\pi^2}{\delta})}\right)$, and $c_t' \triangleq \tau \left(1 + \sqrt{2\log(\frac{M|\mathcal{D}|t^2 2\pi^2}{\delta})}\right)$. With probability $\geq 1 - \delta/4 - \delta/4 - \delta/12 - \delta/12 = 1 - 2\delta/3$, we have that*

$$|f^t(\mathbf{x}) - \sum_{i=1}^{M} \omega_i f_i^t(\mathbf{x})| \leq c_t + c_t' + \sum_{i=1}^{M} \omega_i d_i', \qquad \forall \mathbf{x} \in \mathcal{D}, t \geq 1.$$

Proof. Firstly, we can bound the difference between the target function and a sampled function from its GP posterior.

$$\begin{aligned} |f^t(\mathbf{x}) - f(\mathbf{x})| &\leq |f^t(\mathbf{x}) - \mu_{t-1}(\mathbf{x})| + |\mu_{t-1}(\mathbf{x}) - f(\mathbf{x})| \\ &\overset{(a)}{\leq} \beta_t \sqrt{2\log(\frac{|\mathcal{D}|t^2 2\pi^2}{\delta})}\sigma_{t-1}(\mathbf{x}) + \beta_t \sigma_{t-1}(\mathbf{x}) \\ &= c_t \sigma_{t-1}(\mathbf{x}), \end{aligned} \tag{21}$$

where (a) results from Lemma 12 and Lemma 2, and hence holds with probability of $\geq 1 - \delta/12 - \delta/4$. Next, we do the same for all meta-functions $i = 1, \ldots, M$.

$$\begin{aligned} |f_i^t(\mathbf{x}) - f_i(\mathbf{x})| &\leq |f_i^t(\mathbf{x}) - \overline{\mu}_i(\mathbf{x})| + |\overline{\mu}_i(\mathbf{x}) - f_i(\mathbf{x})| \\ &\overset{(a)}{\leq} \tau \overline{\sigma}_i(\mathbf{x}) + \tau \sqrt{2\log(\frac{M|\mathcal{D}|t^2 2\pi^2}{\delta})}\overline{\sigma}_i(\mathbf{x}) \\ &= c_t' \overline{\sigma}_i(\mathbf{x}), \end{aligned} \tag{22}$$

where (a) results from Lemma 12 and Lemma 11, and hence also holds with probability of $\geq 1 - \delta/12 - \delta/4$. Therefore, combining the above two inequalities gives us:

$$\begin{aligned} |f^t(\mathbf{x}) - f_i^t(\mathbf{x})| &\leq |f^t(\mathbf{x}) - f(\mathbf{x})| + |f(\mathbf{x}) - f_i(\mathbf{x})| + |f_i(\mathbf{x}) - f_i^t(\mathbf{x})| \\ &\leq c_t \sigma_{t-1}(\mathbf{x}) + c_t' \overline{\sigma}_i(\mathbf{x}) + d_i' \\ &\leq c_t \sigma_{t-1}(\mathbf{x}) + c_t' + d_i', \end{aligned} \tag{23}$$

in which the last inequality follows since $\overline{\sigma}_i(\mathbf{x}) \leq 1$. Finally, the lemma can be proved as:

$$
\begin{aligned}
|f^t(\mathbf{x}) - \sum_{i=1}^{M} \omega_i f_i^t(\mathbf{x})| &\leq \sum_{i=1}^{M} \omega_i |f^t(\mathbf{x}) - f_i^t(\mathbf{x})| \\
&\leq \sum_{i=1}^{M} \omega_i \left(c_t \sigma_{t-1}(\mathbf{x}) + c'_t + d'_i \right) \\
&\leq c_t + c'_t + \sum_{i=1}^{M} \omega_i d'_i.
\end{aligned} \tag{24}
$$

□

Next, we define the set of "saturated points" in an iteration t, which are those inputs which incur large regrets in iteration t.

Definition 1. *At iteration t, define the set of saturated points as*

$$S_t \triangleq \{\mathbf{x} \in \mathcal{D} | \Delta(\mathbf{x}) > c_t \sigma_{t-1}(\mathbf{x})\},$$

where $\Delta(\mathbf{x}) \triangleq f(\mathbf{x}^*) - f(\mathbf{x})$.

The next lemma will be useful in proving that the input we query in iteration t is unsaturated (i.e., in proving Lemma 15), and its proof makes use of Gaussian anti-concentration inequality.

Lemma 14. *With probability of $\geq 1 - \delta/4$,*

$$\mathbb{P}\left(f^t(\mathbf{x}) > f(\mathbf{x}) | \mathcal{F}_{t-1}, \mathcal{E}_t^1 \right) \geq p, \qquad \forall t \geq 1.$$

where $p \triangleq \frac{e^{-1}}{4\sqrt{\pi}}$.

Proof. Define $\theta_t \triangleq \frac{|f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})|}{\beta_t \sigma_{t-1}(\mathbf{x})}$.

$$
\begin{aligned}
\mathbb{P}\left(f^t(\mathbf{x}) > f(\mathbf{x}) | \mathcal{F}_{t-1}, \mathcal{E}_t^1 \right) &= \mathbb{P}\left(\frac{f^t(\mathbf{x}) - \mu_{t-1}(\mathbf{x})}{\beta_t \sigma_{t-1}(\mathbf{x})} > \frac{f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})}{\beta_t \sigma_{t-1}(\mathbf{x})} | \mathcal{F}_{t-1}, \mathcal{E}_t^1 \right) \\
&\geq \mathbb{P}\left(\frac{f^t(\mathbf{x}) - \mu_{t-1}(\mathbf{x})}{\beta_t \sigma_{t-1}(\mathbf{x})} > \frac{|f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})|}{\beta_t \sigma_{t-1}(\mathbf{x})} | \mathcal{F}_{t-1}, \mathcal{E}_t^1 \right) \\
&= \mathbb{P}\left(\frac{f^t(\mathbf{x}) - \mu_{t-1}(\mathbf{x})}{\beta_t \sigma_{t-1}(\mathbf{x})} > \theta_t | \mathcal{F}_{t-1}, \mathcal{E}_t^1 \right) \\
&\overset{(a)}{\geq} \frac{e^{-\theta_t^2}}{4\sqrt{\pi}\theta_t} \overset{(b)}{\geq} \frac{e^{-1}}{4\sqrt{\pi}}.
\end{aligned} \tag{25}
$$

Note that due to the way in which the function f^t is sampled from the GP posterior, i.e., $f^t \sim \mathcal{GP}\left(\mu_{t-1}(\cdot), \beta_t^2 \sigma_{t-1}^2(\cdot)\right)$ (Sec. 3), we have that $\frac{f^t(\mathbf{x}) - \mu_{t-1}(\mathbf{x})}{\beta_t \sigma_{t-1}(\mathbf{x})}$ follows a standard Gaussian distribution. Therefore, step (a) above results from the Gaussian anti-concentration inequality: denote by Z the standard Gaussian distribution $\mathcal{N}(0,1)$, then $\mathbb{P}(Z > \theta_t) \geq \frac{e^{-\theta_t^2}}{4\sqrt{\pi}\theta_t}$. Step (b) follows from Lemma 2 (i.e., $\theta_t \leq 1$) and hence holds with probability of $\geq 1 - \delta/4$. □

The next lemma shows that in every iteration, the probability that we choose an unsaturated input is lower-bounded.

Lemma 15. *With probability of $\geq 1 - \delta/4 - \delta/12 = 1 - \delta/3$,*

$$\mathbb{P}\left(\mathbf{x}_t \in \mathcal{D} \setminus S_t | \mathcal{F}_{t-1} \right) \geq (1 - \nu_t) p, \qquad \forall t \geq 1.$$

Proof. Firstly, we have that

$$\mathbb{P}\left(\mathbf{x}_t \in \mathcal{D} \setminus S_t | \mathcal{F}_{t-1}\right) \geq \mathbb{P}\left(\mathbf{x}_t \in \mathcal{D} \setminus S_t | \mathcal{F}_{t-1}, \mathcal{E}_t^1\right) \mathbb{P}\left(\mathcal{E}_t^1\right) = \mathbb{P}\left(\mathbf{x}_t \in \mathcal{D} \setminus S_t | \mathcal{F}_{t-1}, \mathcal{E}_t^1\right)(1-\nu_t). \tag{26}$$

Next, we attempt to lower-bound the term $\mathbb{P}\left(\mathbf{x}_t \in \mathcal{D} \setminus S_t | \mathcal{F}_{t-1}, \mathcal{E}_t^1\right)$.

$$\mathbb{P}\left(\mathbf{x}_t \in \mathcal{D} \setminus S_t | \mathcal{F}_{t-1}, \mathcal{E}_t^1\right) \geq \mathbb{P}\left(f^t(\mathbf{x}^*) > f^t(\mathbf{x}), \forall \mathbf{x} \in S_t | \mathcal{F}_{t-1}, \mathcal{E}_t^1\right). \tag{27}$$

The above inequality follows because $\mathbf{x}^*$ is always unsaturated: $\Delta(\mathbf{x}^*) = f(\mathbf{x}^*) - f(\mathbf{x}^*) = 0 \leq c_t\sigma_{t-1}(\mathbf{x}^*)$. As a result, if the event on the RHS of the above inequality holds (i.e., an unsaturated input has larger value of f^t than all saturated inputs), then the event on the LHS (i.e., $\mathbf{x}_t$ is unsaturated) also holds. Next, we also have that $\forall \mathbf{x} \in S_t$,

$$f^t(\mathbf{x}) \leq f(\mathbf{x}) + c_t\sigma_{t-1}(\mathbf{x}) \leq f(\mathbf{x}) + \Delta(\mathbf{x}) = f(\mathbf{x}) + f(\mathbf{x}^*) - f(\mathbf{x}) = f(\mathbf{x}^*), \tag{28}$$

in which the first inequality follows from (21) and hence holds with probability $\geq 1 - \delta/12 - \delta/4$, the second inequality is a result of the definition of saturated inputs (Definition 1). The above inequality implies that

$$\mathbb{P}\left(f^t(\mathbf{x}^*) > f^t(\mathbf{x}), \forall \mathbf{x} \in S_t | \mathcal{F}_{t-1}, \mathcal{E}_t^1\right) \geq \mathbb{P}\left(f^t(\mathbf{x}^*) > f(\mathbf{x}^*) | \mathcal{F}_{t-1}, \mathcal{E}_t^1\right). \tag{29}$$

Lastly, combining the above inequalities gives us

$$\mathbb{P}\left(\mathbf{x}_t \in \mathcal{D} \setminus S_t | \mathcal{F}_{t-1}, \mathcal{E}_t^1\right) \geq \mathbb{P}\left(f^t(\mathbf{x}^*) > f(\mathbf{x}^*) | \mathcal{F}_{t-1}, \mathcal{E}_t^1\right) \geq p, \tag{30}$$

where the last inequality follows from Lemma 14. This completes the proof. Note that the error probabilities for this lemma come from Lemma 12 ($\delta/12$) and Lemma 2 ($\delta/4$). □

Next, we prove an upper bound on the expected instantaneous regret $r_t = f(\mathbf{x}^*) - f(\mathbf{x}_t)$.

Lemma 16. *With probability of* $\geq 1 - \delta/4 - \delta/4 - \delta/12 - \delta/12 = 1 - 2\delta/3$,

$$\mathbb{E}[r_t | \mathcal{F}_{t-1}] \leq c_t\left(1 + \frac{2}{(1-\nu_1)p}\right)\mathbb{E}[\sigma_{t-1}(\mathbf{x}_t) | \mathcal{F}_{t-1}] + \psi_t,$$

where $\psi_t \triangleq 2\nu_t\left(c_t + c_t' + \sum_{i=1}^M \omega_i d_i'\right)$.

Proof. To begin with, define the unsaturated input with the smallest posterior standard deviation as

$$\overline{\mathbf{x}}_t \triangleq \arg\min_{\mathbf{x} \in \mathcal{D} \setminus S_t} \sigma_{t-1}(\mathbf{x}). \tag{31}$$

This allows us to obtain the following:

$$\mathbb{E}[\sigma_{t-1}(\mathbf{x}_t) | \mathcal{F}_{t-1}] \geq \mathbb{E}[\sigma_{t-1}(\mathbf{x}_t) | \mathcal{F}_{t-1}, \mathbf{x}_t \in \mathcal{D} \setminus S_t]\mathbb{P}\left(\mathbf{x}_t \in \mathcal{D} \setminus S_t\right) \overset{(a)}{\geq} \sigma_{t-1}(\overline{\mathbf{x}}_t)(1-\nu_t)p, \tag{32}$$

where (a) results from Lemma 15 and hence holds with probability $\geq 1 - \delta/12 - \delta/4$ (the error probabilities come from Lemma 12 and Lemma 2). Subsequently, the instataneous regret can be upper-bounded as

$$\begin{aligned} r_t &= \Delta(\mathbf{x}_t) = f(\mathbf{x}^*) - f(\overline{\mathbf{x}}_t) + f(\overline{\mathbf{x}}_t) - f(\mathbf{x}_t) \\ &\overset{(a)}{\leq} \Delta(\overline{\mathbf{x}}_t) + f^t(\overline{\mathbf{x}}_t) + c_t\sigma_{t-1}(\overline{\mathbf{x}}_t) - f^t(\mathbf{x}_t) + c_t\sigma_{t-1}(\mathbf{x}_t) \\ &\overset{(b)}{\leq} c_t\sigma_{t-1}(\overline{\mathbf{x}}_t) + c_t\sigma_{t-1}(\overline{\mathbf{x}}_t) + c_t\sigma_{t-1}(\mathbf{x}_t) + f^t(\overline{\mathbf{x}}_t) - f^t(\mathbf{x}_t) \\ &= c_t(2\sigma_{t-1}(\overline{\mathbf{x}}_t) + \sigma_{t-1}(\mathbf{x}_t)) + f^t(\overline{\mathbf{x}}_t) - f^t(\mathbf{x}_t), \end{aligned} \tag{33}$$

in which (a) follows from (21), and (b) results from the definition of saturated input (Definition 1) and that $\overline{\mathbf{x}}_t$ is unsaturated. Next, we attempt to upper-bound the expected value of the term $f^t(\overline{\mathbf{x}}_t) - f^t(\mathbf{x}_t)$ from the equation above:

$$\begin{aligned}
&\mathbb{E}\left[f^t(\overline{\mathbf{x}}_t) - f^t(\mathbf{x}_t)|\mathcal{F}_{t-1}\right] \\
&= \mathbb{P}(\mathcal{E}_t^1)\mathbb{E}\left[f^t(\overline{\mathbf{x}}_t) - f^t(\mathbf{x}_t)|\mathcal{F}_{t-1}, \mathcal{E}_t^1\right] + \mathbb{P}(\mathcal{E}_t^2)\mathbb{E}\left[f^t(\overline{\mathbf{x}}_t) - f^t(\mathbf{x}_t)|\mathcal{F}_{t-1}, \mathcal{E}_t^2\right] \\
&\overset{(a)}{\leq} \nu_t \mathbb{E}\left[f^t(\overline{\mathbf{x}}_t) - f^t(\mathbf{x}_t)|\mathcal{F}_{t-1}, \mathcal{E}_t^2\right] \\
&\overset{(b)}{\leq} \nu_t \mathbb{E}\Big[\sum_{i=1}^M \omega_i f_i^t(\overline{\mathbf{x}}_t) + c_t + c_t' + \sum_{i=1}^M \omega_i d_i' + c_t + c_t' + \sum_{i=1}^M \omega_i d_i' - \sum_{i=1}^M \omega_i f_i^t(\mathbf{x}_t)|\mathcal{F}_{t-1}, \mathcal{E}_t^2\Big] \\
&\overset{(c)}{\leq} 2\nu_t \left(c_t + c_t' + \sum_{i=1}^M \omega_i d_i'\right) \triangleq \psi_t.
\end{aligned} \tag{34}$$

Step (a) follows since conditioned on the event $\mathcal{E}_t^1$ ($\overline{\zeta}_t(\mathbf{x}) = f^t(\mathbf{x})$), we have that $f^t(\mathbf{x}) \leq f^t(\mathbf{x}_t), \forall \mathbf{x} \in \mathcal{D}$; step (b) results from Lemma 13; step (c) follows since conditioned on the event $\mathcal{E}_t^2$ (i.e., $\overline{\zeta}_t(\mathbf{x}) = \sum_{i=1}^M \omega_i \left[\overline{f}_i^t(\mathbf{x})\right]$), we have that $\sum_{i=1}^M \omega_i \left[\overline{f}_i^t(\mathbf{x})\right] \leq \sum_{i=1}^M \omega_i \left[\overline{f}_i^t(\mathbf{x}_t)\right], \forall \mathbf{x} \in \mathcal{D}$. Lastly,

$$\begin{aligned}
\mathbb{E}[r_t|\mathcal{F}_{t-1}] &\leq \mathbb{E}[c_t(2\sigma_{t-1}(\overline{\mathbf{x}}_t) + \sigma_{t-1}(\mathbf{x}_t)) + \psi_t|\mathcal{F}_{t-1}] \\
&\leq \mathbb{E}\left[c_t\left(\frac{2}{(1-\nu_t)p}\sigma_{t-1}(\mathbf{x}_t) + \sigma_{t-1}(\mathbf{x}_t)\right) + \psi_t|\mathcal{F}_{t-1}\right] \\
&\leq c_t\left(1 + \frac{2}{(1-\nu_1)p}\right)\mathbb{E}[\sigma_{t-1}(\mathbf{x}_t)|\mathcal{F}_{t-1}] + \psi_t,
\end{aligned} \tag{35}$$

in which the second inequality results from (32). Note that the error probabilities for this Lemma follow from Lemma 13. □

Subsequently, we make use of martingale concentration inequalities to bound the cumulative regret.

Definition 2. *Define $Y_0 = 0$, and for $t \geq 1$,*

$$X_t = r_t - c_t\left(1 + \frac{2}{(1-\nu_1)p}\right)\sigma_{t-1}(\mathbf{x}_t) - \psi_t,$$

$$Y_t = \sum_{s=1}^t X_s.$$

The next lemma shows that $\{Y_t\}_{t\geq 1}$ is a super-martingale.

Lemma 17. *With probability $\geq 1 - \delta/4 - \delta/4 - \delta/12 - \delta/12 = 1 - 2\delta/3$, $\{Y_t\}_{t\geq 1}$ is a super-martingale with respect to the filtration $\mathcal{F}_{t-1}$.*

Proof.

$$\begin{aligned}
\mathbb{E}[Y_t - Y_{t-1}|\mathcal{F}_{t-1}] &= \mathbb{E}[X_t|\mathcal{F}_{t-1}] \\
&= \mathbb{E}[r_t - c_t\left(1 + \frac{2}{(1-\nu_1)p}\right)\sigma_{t-1}(\mathbf{x}_t) + \psi_t|\mathcal{F}_{t-1}] \\
&= \mathbb{E}[r_t|\mathcal{F}_{t-1}] - \left[c_t\left(1 + \frac{2}{(1-\nu_1)p}\right)\mathbb{E}[\sigma_{t-1}(\mathbf{x}_t)|\mathcal{F}_{t-1}] + \psi_t\right] \leq 0,
\end{aligned} \tag{36}$$

where the last inequality follows from Lemma 16. □

Finally, we are ready to use martingale concentration inequalities to bound the cumulative regret.

Lemma 18. *With probability of* $\geq 1-\delta/4-\delta/4-\delta/12-\delta/12-\delta/12 = 1-3\delta/4$,

$$R_T \leq \left(2B + c_T\left(1+\frac{2}{(1-\nu_1)p}\right)+\psi_1\right)\sqrt{T(C_1\gamma_T + 2\log(12/\delta))} + 2\sum_{t=1}^{T}\nu_t(c_t + c_t' + \sum_{i=1}^{M}\omega_i d_i')$$

where $C_1 = 2/\log(1+\sigma^{-2})$.

Proof. To begin with, we have that

$$\begin{aligned} |Y_t - Y_{t-1}| = |X_t| &\leq |r_t| + |c_t\left(1+\frac{2}{(1-\nu_1)p}\right)\sigma_{t-1}(\mathbf{x}_t)| + |\psi_t| \\ &\leq 2B + c_t\left(1+\frac{2}{(1-\nu_1)p}\right)+\psi_t, \end{aligned} \tag{37}$$

where the last inequality follows since $|r_t| = |f(\mathbf{x}^*) - f(\mathbf{x}_t)| \leq 2B$ (because $\|f\|_k \leq B$ as we have assumed in Sec. 2, which immediately implies that $|f(\mathbf{x})| \leq B, \forall \mathbf{x} \in \mathcal{D}$), and $\sigma_{t-1}(\mathbf{x}) \leq 1, \forall \mathbf{x} \in \mathcal{D}$.

Next, we apply the Azuma-Hoeffding Inequality with an error probability of $\delta/12$ (first inequality):

$$\begin{aligned} \sum_{t=1}^{T} r_t &\leq \sum_{t=1}^{T} c_t\left(1+\frac{2}{(1-\nu_1)p}\right)\sigma_{t-1}(\mathbf{x}_t) + \sum_{t=1}^{T}\psi_t + \\ &\quad \sqrt{2\log\frac{10}{\delta}\sum_{t=1}^{T}\left(2B + c_t\left(1+\frac{2}{(1-\nu_1)p}\right)+\psi_t\right)^2} \\ &\leq c_T\left(1+\frac{2}{(1-\nu_1)p}\right)\sum_{t=1}^{T}\sigma_{t-1}(\mathbf{x}_t) + \sum_{t=1}^{T}\psi_t + \\ &\quad \left(2B + c_T\left(1+\frac{2}{(1-\nu_1)p}\right)+\psi_1\right)\sqrt{2T\log\frac{12}{\delta}} \\ &\leq c_T\left(1+\frac{2}{(1-\nu_1)p}\right)\sqrt{C_1'\gamma_T T} + \sum_{t=1}^{T}\psi_t + \\ &\quad \left(2B + c_T\left(1+\frac{2}{(1-\nu_1)p}\right)+\psi_1\right)\sqrt{2T\log\frac{12}{\delta}} \\ &\leq \left(2B + c_T\left(1+\frac{2}{(1-\nu_1)p}\right)+\psi_1\right)\sqrt{T(C_1'\gamma_T + 2\log(12/\delta))} + \\ &\quad 2\sum_{t=1}^{T}\nu_t(c_t + c_t' + \sum_{i=1}^{M}\omega_i d_i). \end{aligned} \tag{38}$$

The second last inequality makes use of Lemma 10 from the proof of RM-GP-UCB (excluding the factor of $(1-\nu_t)\beta_t$) with $C_1' \triangleq 2/\log(1+\sigma^{-2})$. □

Recall that $c_t = \mathcal{O}(\sqrt{\gamma_t}\log t)$, $c_t' = \mathcal{O}(\log t)$. Therefore, Lemma 18 can be further analyzed as:

$$\begin{aligned} R_T &= \mathcal{O}\left(c_T\sqrt{T\gamma_T} + \sum_{t=1}^{T}\nu_t(c_t + c_t' + \sum_{i=1}^{M}\omega_i d_i)\right) \\ &= \mathcal{O}\Big(\Big(\sum_{i=1}^{M}\omega_i d_i'\Big)\sum_{t=1}^{T}\nu_t + \sum_{t=1}^{T}\nu_t\sqrt{\gamma_t}\log t + \gamma_T\log T\sqrt{T}\Big). \end{aligned} \tag{39}$$

Lastly, similar to our analysis of RM-GP-UCB for the case where the ω_i's change with t (i.e., at the end of Appendix B.2), when our online meta-weight optimization is used, we simply need to slightly modify the definition of

ψ_t: $\psi_t \triangleq 2\nu_t \left(c_t + c'_t + \sum_{i=1}^M \omega_{i,t} d'_i\right)$ by allowing $\omega_{i,t}$ to change with t, and the subsequent analysis still holds by simply replacing ω_i by $\omega_{i,t}$. As a result, the no-regret guarantee of RM-GP-TS (Theorem 2) still holds (since we can simply upper-bound every $\omega_{i,t}$ by 1):

$$
\begin{aligned}
R_T &= \mathcal{O}\Big(\sum_{t=1}^T \nu_t \Big(\sum_{i=1}^M \omega_{i,t} d'_i\Big) + \sum_{t=1}^T \nu_t \sqrt{\gamma_t} \log t + \gamma_T \log T \sqrt{T}\Big) \\
&= \mathcal{O}\Big(\Big(\sum_{i=1}^M d'_i\Big) \sum_{t=1}^T \nu_t + \sum_{t=1}^T \nu_t \sqrt{\gamma_t} \log t + \gamma_T \log T \sqrt{T}\Big) \\
&= \widetilde{\mathcal{O}}\Big(\Big(\sum_{i=1}^M d'_i\Big) \sum_{t=1}^T \nu_t + \sum_{t=1}^T \nu_t \sqrt{\gamma_t} + \gamma_T \sqrt{T}\Big).
\end{aligned}
$$

D ANALYSIS OF ONLINE META-WEIGHT OPTIMIZATION

D.1 PROOF OF LEMMA 1

From the definitions of $U_{t,i,j}$ and $L_{t,i,j}$ (5), and the fact that $L_{t,i,j} \leq f(\mathbf{x}_{i,j}) \leq U_{t,i,j}, \forall t, i, j$ with probability $\geq 1 - \delta/4$ (Section 5.1), we have that

$$
\begin{aligned}
d_i &= \max_{j=1,\ldots,N_i} \left|f_i(\mathbf{x}_{i,j}) - f(\mathbf{x}_{i,j})\right| \\
&\leq \max_{j=1,\ldots,N_i} \left[\max\{\left|f_i(\mathbf{x}_{i,j}) - U_{t,i,j}\right|, \left|f_i(\mathbf{x}_{i,j}) - L_{t,i,j}\right|\}\right] \qquad \forall\, i = 1, \ldots, M, \forall t \geq 1
\end{aligned} \tag{40}
$$

which holds with probability $\geq 1 - \delta/4$. Next, we derive upper bounds on $\left|f_i(\mathbf{x}_{i,j}) - U_{t,i,j}\right|$ and $\left|f_i(\mathbf{x}_{i,j}) - L_{t,i,j}\right|$ that only consist of known or computable terms, such that the upper bounds on d_i can be efficiently calculated in practice.

Lemma 19. *With probability* $\geq 1 - \delta/4$, $\forall\, t \geq 1$, $\forall i, j$,

$$
\begin{aligned}
\left|f_i(\mathbf{x}_{i,j}) - U_{t,i,j}\right| &\leq \sqrt{2\sigma^2 \log \frac{8 \sum_{i=1}^M N_i}{\delta}} + \left|y_{i,j} - U_{t,i,j}\right|, \\
\left|f_i(\mathbf{x}_{i,j}) - L_{t,i,j}\right| &\leq \sqrt{2\sigma^2 \log \frac{8 \sum_{i=1}^M N_i}{\delta}} + \left|y_{i,j} - L_{t,i,j}\right|.
\end{aligned}
$$

Proof. To begin with, note that $f_i(\mathbf{x}_{i,j}) - y_{i,j} \sim \mathcal{N}(0, \sigma^2)$. Therefore, (9) suggests that

$$
\mathbb{P}\left(\left|f_i(\mathbf{x}_{i,j}) - y_{i,j}\right| > \sigma \sqrt{2 \log \frac{8 \sum_{i=1}^M N_i}{\delta}}\right) \leq \frac{\delta}{8 \sum_{i=1}^M N_i} \tag{41}
$$

which naturally leads to a high-probability upper bound on $\left|f_i(\mathbf{x}_{i,j}) - U_{t,i,j}\right|$:

$$
\begin{aligned}
\left|f_i(\mathbf{x}_{i,j}) - U_{t,i,j}\right| &= \left|f_i(\mathbf{x}_{i,j}) - y_{i,j} + y_{i,j} - U_{t,i,j}\right| \\
&\leq \left|f_i(\mathbf{x}_{i,j}) - y_{i,j}\right| + \left|y_{i,j} - U_{t,i,j}\right| \\
&\leq \sqrt{2\sigma^2 \log \frac{8 \sum_{i=1}^M N_i}{\delta}} + \left|y_{i,j} - U_{t,i,j}\right|
\end{aligned} \tag{42}
$$

which holds with probability $\geq 1 - \frac{\delta}{8 \sum_{i=1}^M N_i}$. Applying the same reasoning to $\left|f_i(\mathbf{x}_{i,j}) - L_{t,i,j}\right|$ results in a similar high-probability upper bound:

$$
\left|f_i(\mathbf{x}_{i,j}) - L_{t,i,j}\right| \leq \sqrt{2\sigma^2 \log \frac{8 \sum_{i=1}^M N_i}{\delta}} + \left|y_{i,j} - L_{t,i,j}\right|. \tag{43}
$$

Next, the proof is completed by taking a union bound over both $U_{t,i,j}$ and $L_{t,i,j}$, as well as all $\sum_{i=1}^M N_i$ observations of the meta-tasks. □

Finally, Lemma 1 follows by combining (40) and Lemma 19.

D.2 PROOF OF PROPOSITION 1

In iteration t, define $\overline{\alpha}_t$ by replacing d_i in α with $\overline{d}_{i,t}$:

$$\overline{\alpha}_t = \sum_{i=1}^{M} \omega_i \frac{N_i}{\sigma^2}(2\sqrt{2\sigma^2 \log \frac{8N_i}{\delta}} + \overline{d}_{i,t}). \tag{44}$$

Since according to Lemma 1, $d_i \leq \overline{d}_{i,t}\, \forall i = 1, \ldots, M, t \geq 1$ with probability $\geq 1 - \delta/2$, we have that $\alpha \leq \overline{\alpha}_t\, \forall t \geq 1$, which also holds with probability $\geq 1 - \delta/2$.

Therefore, Theorem 1 implies that, with probability $\geq 1 - \delta$,

$$R_T \leq 2\underline{\sum_{t=1}^{T} \overline{\alpha}_t \nu_t} + 2\tau \sum_{t=1}^{T} \nu_t + \beta_T \sqrt{C_1 T \gamma_T}. \tag{45}$$

In (45), only the underlined term depends on the ω_i's. Define two column vectors $\overline{\boldsymbol{\alpha}} = [\overline{\alpha}_t]_{t=1,\ldots,T}^\top$ and $\boldsymbol{\nu} = [\nu_t]_{t=1,\ldots,T}^\top$. Then, the underlined term in (45) can be further decomposed as

$$2\sum_{t=1}^{T} \overline{\alpha}_t \nu_t \triangleq 2\overline{\boldsymbol{\alpha}}^\top \boldsymbol{\nu} \overset{\text{(a)}}{\leq} 2\|\overline{\boldsymbol{\alpha}}\|_2 \|\boldsymbol{\nu}\|_2 \overset{\text{(b)}}{\leq} 2\|\overline{\boldsymbol{\alpha}}\|_1 \|\boldsymbol{\nu}\|_1 \overset{\text{(c)}}{=} 2\underline{\sum_{t=1}^{T} \overline{\alpha}_t} \sum_{t=1}^{T} \nu_t \tag{46}$$

in which (a) results from Cauchy-Schwarz inequality, (b) follows because the L2 norm is upper-bounded by the L1 norm, and (c) is obtained because $\overline{\alpha}_t > 0, \nu_t \geq 0, \forall t \geq 1$.

In (46), the dependence on the ω_i's appears in the underlined term, which can be further decomposed as

$$\begin{aligned}
\sum_{t=1}^{T} \overline{\alpha}_t &= \sum_{t=1}^{T} \left[\sum_{i=1}^{M} \omega_i \frac{N_i}{\sigma^2} \left(2\sqrt{2\sigma^2 \log \frac{8N_i}{\delta}} + \overline{d}_{i,t} \right) \right] \\
&\triangleq \frac{1}{\sigma^2} \sum_{t=1}^{T} \left[\sum_{i=1}^{M} \omega_i l_{i,t} \right] \\
&\triangleq \frac{1}{\sigma^2} \sum_{t=1}^{T} \boldsymbol{\omega}^\top \boldsymbol{l}_t
\end{aligned} \tag{47}$$

in which we have defined $\boldsymbol{\omega} \triangleq [\omega_i]_{i=1,\ldots,M}$, $\boldsymbol{l}_t \triangleq [l_{i,t}]_{i=1,\ldots,M}$, with

$$l_{i,t} \triangleq N_i \left(2\sqrt{2\sigma^2 \log \frac{8N_i}{\delta}} + \overline{d}_{i,t} \right). \tag{48}$$

Plugging (46) and (47) in to (45) completes the proof.

D.3 DERIVATION OF EQUATION 7

Recall that our objective is to minimize

$$\sum_{s=1}^{t-1} \boldsymbol{\omega}'^\top \boldsymbol{l}_s + \frac{1}{\eta} \sum_{i=1}^{M} \omega'_i \log \omega'_i$$

subject to the constraint that $\boldsymbol{\omega}'$ forms a probability simplex: $\sum_{i=1}^{M} \omega_i' = 1.0$ and $\omega_i' \geq 0$ for all $i = 1, \ldots, M$. Define the Lagrangian as

$$L(\boldsymbol{\omega}, \lambda) = \sum_{s=1}^{t-1} \boldsymbol{\omega}'^{\top} \boldsymbol{l}_s + \frac{1}{\eta} \sum_{i=1}^{M} \omega_i' \log \omega_i' + \lambda \left(1 - \sum_{i=1}^{M} \omega_i' \right). \tag{49}$$

Taking the derivative of $L(\boldsymbol{\omega}, \lambda)$ with respect to ω_i', we get

$$\frac{\partial L(\boldsymbol{\omega}, \lambda)}{\partial \omega_i'} = \sum_{s=1}^{t-1} l_{i,s} + \frac{1}{\eta} \left(\log \omega_i' + 1 \right) - \lambda. \tag{50}$$

Setting (50) to 0 gives us

$$\omega_i' = e^{\eta\lambda - 1} e^{-\eta \sum_{s=1}^{t-1} l_{i,s}} \propto e^{-\eta \sum_{s=1}^{t-1} l_{i,s}}. \tag{51}$$

Normalizing the ω_i''s for all $i = 1 \ldots, M$ to form a probability simplex leads to (7).

D.4 ANALYSIS FOR RM-GP-TS

Here we use the function gap d_i to approximate d_i' (defined in Theorem 2), i.e., $d_i' \approx d_i, \forall i = 1, \ldots, M$. Combining Lemma 1 and Theorem 2, we have for RM-GP-TS that with probability of $\geq 1 - \delta$,

$$\begin{aligned} R_T &= \mathcal{O}\Big(\sum_{t=1}^{T} \nu_t \left(\sum_{i=1}^{M} \omega_i \overline{d}_{i,t} \right) + \sum_{t=1}^{T} \nu_t \sqrt{\gamma_t} \log t + \gamma_T \log T \sqrt{T} \Big) \\ &\leq \mathcal{O}\Big(\underline{\left(\sum_{t=1}^{T} \sum_{i=1}^{M} \omega_i \overline{d}_{i,t} \right)} \left(\sum_{t=1}^{T} \nu_t \right) + \sum_{t=1}^{T} \nu_t \sqrt{\gamma_t} \log t + \gamma_T \log T \sqrt{T} \Big), \end{aligned} \tag{52}$$

in which the inequality can be proved in a similar way as equation (46). Next, define $\boldsymbol{\omega} \triangleq [\omega_i]_{i=1,\ldots,M}$, $\boldsymbol{d}_t \triangleq [\overline{d}_{i,t}]_{i=1,\ldots,M}$, then the underlined term above can be denoted as:

$$\sum_{t=1}^{T} \sum_{i=1}^{M} \omega_i \overline{d}_{i,t} = \sum_{t=1}^{T} \boldsymbol{\omega}^{\top} \boldsymbol{d}_t. \tag{53}$$

Therefore, equation (52) can be further upper-bounded as:

$$R_T = \mathcal{O}\Big(\underline{\left(\sum_{t=1}^{T} \boldsymbol{\omega}^{\top} \boldsymbol{d}_t \right)} \left(\sum_{t=1}^{T} \nu_t \right) + \sum_{t=1}^{T} \nu_t \sqrt{\gamma_t} \log t + \gamma_T \log T \sqrt{T} \Big). \tag{54}$$

Next, applying similar derivations as Appendix D.3 (treating the underlined term above as the loss to be minimized) leads to the same update rule for the meta-weights as equation (7). Approximating d_i' using d_i also allows us to derive the same update rule for ν_t (Sec. 5.2).

E MORE EXPERIMENTAL DETAILS AND RESULTS

In every experiment, the same set of random initializations are used for all methods to ensure fair comparisons. The kernel bandwidth parameter ρ in TAF is set to $\rho = 0.5$ in all experiments, but we have observed that other values of ρ (such as 0.1 and 0.9) lead to similar performances. $S = 500$ posterior samples are used to compute the ensemble weights in RGPE. All experiments are run on a server with 16 cores of Intel Xeon processor, 256G of RAM and 5 NVIDIA GTX1080 Ti GPUs.

E.1 OPTIMIZATION OF SYNTHETIC FUNCTIONS

E.1.1 Synthetic Functions Sampled from GPs

The objective functions are drawn from GP's with the Squared Exponential kernel (with the length scale of 0.05) from the domain $\mathcal{D} = [0, 1]$. Fig. 3 shows an example of such synthetic functions. The meta-functions and meta-tasks are generated

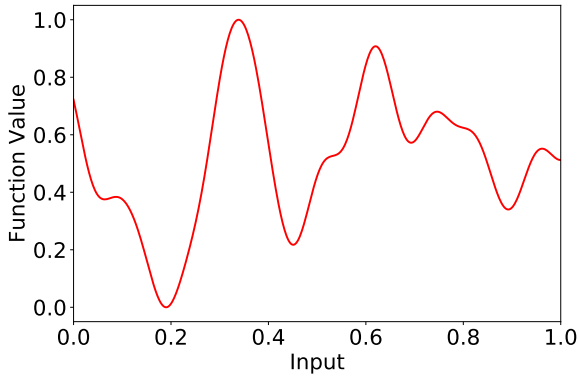
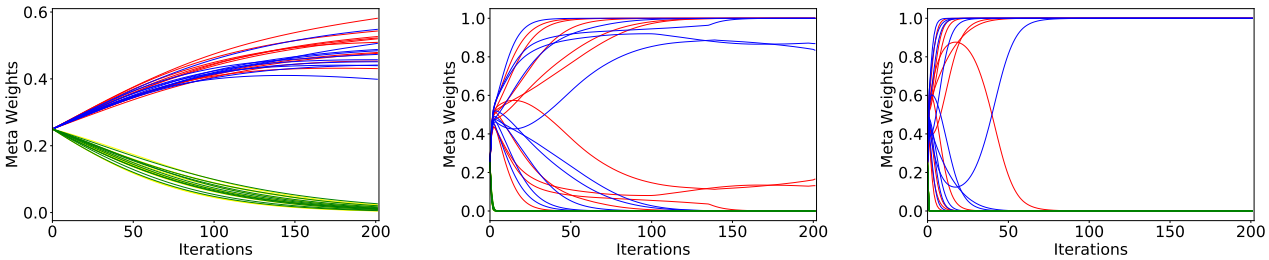


Figure 3: An example synthetic function sampled from a GP.



(a) $\eta N^{3/2} = 0.01$. (b) $\eta N^{3/2} = 1.0$. (c) $\eta N^{3/2} = 5.0$.

Figure 4: Evolution of the meta-weights with different learning rate, η, for online meta-weight optimization in the synthetic experiments. In each figure, the red and blue curves represent the meta-weights of the two meta-tasks that are more similar to the target task (i.e., the first two meta-tasks), whereas the green and yellow curves correspond to the meta-weights of the other two dissimilar meta-tasks. Every color has 10 curves in each figure, which correspond to 10 independent runs of the algorithm with different random initializations.

in the following way. To begin with, we fix the number of meta-tasks $M = 4$, the number of observations (input-output pairs) for each meta-task $N = N_i = 20$ for $i = 1 \dots M$, and the function gaps: $d_1 = d_2 = 0.05$, $d_3 = d_4 = 4.0$. For the i-th meta-task, firstly, N_i inputs are randomly drawn from the entire domain $\mathcal{D} = [0, 1]$. Then for each of the N_i inputs $\mathbf{x}_{i,j}$, a number is randomly drawn from $[-d_i, d_i]$, which is added to the value of the target function $f(\mathbf{x}_{i,j})$ to produce the corresponding function value of the meta-function $f_i(\mathbf{x}_{i,j})$. Subsequently, a zero-mean Gaussian noise (with a noise variance of 0.01) is added to $f_i(\mathbf{x}_{i,j})$, resulting in the corresponding output of the meta-observation $y_i(\mathbf{x}_{i,j})$. The above-mentioned procedure is repeated for each of the $M = 4$ meta-tasks. Note that according to the specified function gaps, meta-tasks 1 and 2 are relatively more similar to the target task, whereas meta-tasks 3 and 4 are dissimilar to the target task due to the larger function gaps.

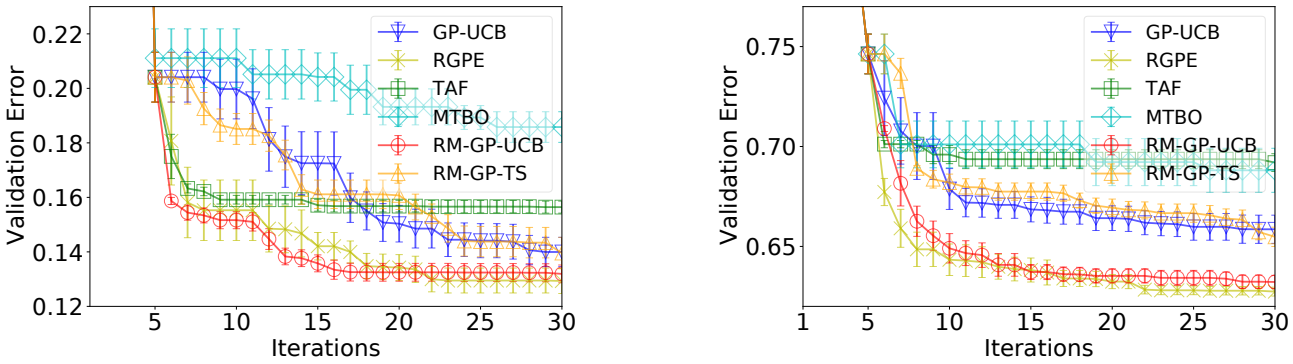
Fig. 4 plots the evolution of the meta-weights for each of the 4 meta-tasks in the experiments exploring the impact of η, i.e., corresponding to Fig. 1c in Section 6.1. These figures are used to demonstrate the observations that overly large and excessively small values of η can both degrade the performance of RM-GP-UCB.

E.2 REAL-WORLD EXPERIMENTS

All datasets and software used in our real-world experiments are publicly available, and none of them contains personally identifiable information or offensive content. More details on these datasets and software are included below.

Hyperparameter Tuning for Convolutional Neural Networks (CNNs). The MNIST, CIFAR-10 and CIFAR-100 datasets can all be directly downloaded using the Keras Python package[6], and the SVHN dataset can be downloaded from `http://ufldl.stanford.edu/housenumbers/`. The MNIST dataset is under the GNU General Public License, CIFAR-

[6] `https://keras.io/`



(a) SVHN.

(b) CIFAR-100.

Figure 5: Best validation error of CNN (both averaged over 10 random initializations).

10 adn CIFAR-100 are under the MIT License, and SVHN is under the Custom (non-commercial) License. The image pixel values are all normalized into the range $[0, 1]$. The CNN hyperparameters being optimized in this set of experiments are the learning rate, learning rate decay, and the L2 regularization parameter, all of which have the search space from 10^{-7} to 10^{-2}. Other than these hyperparameters, a common CNN architecture is used for all datasets, i.e., a CNN containing two convolutional layers (both with 32 filters and each filter has a size of 3×3) each of which is followed by a Max pooling layer (with a pooling size of 3×3), followed by two fully connected layers (both with 64 hidden units); all non-linear activations are ReLU. The size of the training set and validation set for the four datasets are: 60,000/10,000 for MNIST, 73,257/26,032 for SVHN, 50,000/10,000 for both CIFAR-10 and CIFAR-100. For the evaluation of a set of selected hyperparameters, the CNN model is trained using the RMSprop algorithm for 20 epochs, and the final validation error is used as the corresponding output observation. Fig. 5 presents the results when the SVHN and CIFAR-100 datasets are used to produce the target functions.

Comparing Figs. 1e, 1f and Fig. 5 shows that our RM-GP-UCB performs similarly to RGPE for the CIFAR-10, CIFAR-100 and SVHN datasets, and outperforms RGPE for MNIST. After inspection, we found that this is because for the first three datasets (Fig. 1f and Fig. 5), both RM-GP-UCB and RGPE assign most meta-weights to the same meta-task. On the other hand, for MNIST (Fig. 1e), RM-GP-UCB (and RM-GP-TS) is able to assign most weights to SVHN which is indeed more similar to MNIST since they both contain images of digits. In contrast, RGPE mistakenly assigns more meta-weights to CIFAR-10. The reason is that RGPE chooses the weights based on how accurately each meta-task's GP surrogate predicts the pairwise ranking of the target observations (more details in Appendix A, second paragraph). However, for MNIST, most target observations have very similar values since the overall accuracy is very high due to the simplicity of the MNIST dataset. Therefore, the predicted pairwise rankings become unreliable, thus rendering the weights learned by RGPE inaccurate and deteriorating the performance.

Hyperparameter Tuning for CNNs Using the Omniglot Dataset. The Omniglot dataset can be downloaded from `https://github.com/brendenlake/omniglot`, and it is under the MIT License. The dataset consists of 50 alphabets, 30 from the background set and 20 from the evaluation set. Each alphabet includes a number of characters, and all alphabets combine to have 1623 characters. Every character only consists of 20 example images, each drawn by a different person. To perform one-shot classification, we use a Siamese neural network [Koch et al., 2015], which takes two images as inputs and outputs a score indicating whether the pair of input images are predicted to be the same character. The evaluation metric we use in the experiment is 2-way validation error. That is, we compare a test image in the validation set with two other images, only one of which is the same character as the test image, and evaluate whether the Siamese network is able to output a higher predictive score for the correct image which is the same character; we do this using every test image, and use the percentage of errors as the 2-way validation error. In our setting, each task represents tuning 3 hyperparameters of the Siamese network (the same hyperparameters and ranges as the CNN experiments above) using one alphabet. For each task, we use 75% of the characters in the alphabet to produce the training set, and the remaining 25% to generate the validation set. We use 10 alphabets from the background set as 10 meta-tasks. For each meta-task, we generate 30 meta-observations by running BO (using GP-UCB) for 30 iterations. This in total produces $10 \times 30 = 300$ meta-observations. We use one of the alphabets from the evaluation set as the target task.

Hyperparameter Tuning for Support Vector Machines (SVMs). This benchmark dataset, which was originally intro-

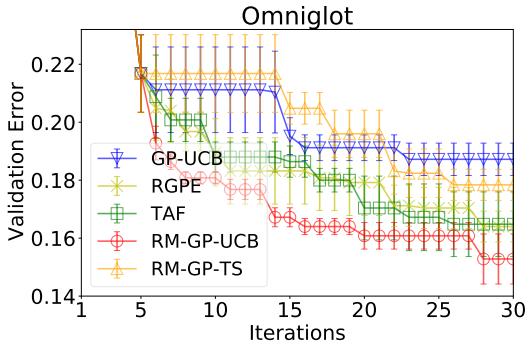


Figure 6: 2-way validation error on the Omniglot dataset.

duced by [Wistuba et al., 2015a] and can be downloaded from `https://github.com/wistuba/TST`, is created by performing hyperparameter tuning of SVM using 50 diverse datasets. 6 hyperparameters are tuned: 3 binary parameters indicating whether a linear, polynomial or radial basis function (RBF) kernel is used, the penalty parameter, the degree of the polynomial kernel, and the bandwidth parameter for the RBF kernel. A fixed grid of hyperparameters of size 288 is created. For each dataset, every hyperparameter configuration on the grid is evaluated and the corresponding validation accuracy is recorded as the observed output of the objective function. In our experiments, each dataset corresponds to a task. We treat one of the 50 tasks as the target task, and the remaining tasks as 49 meta-tasks. For each meta-task, the meta-observations are produced by randomly sampling 50 points (hyperparameter configurations) from the grid. The results reported in the main paper (Fig. 2c) are averaged over 25 trials, each trial treating a different task as the target task; for each trial/target task, we again average the results over 5 random initializations.

Human Activity Recognition (HAR). The dataset used in this experiment can be downloaded from `https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones`.

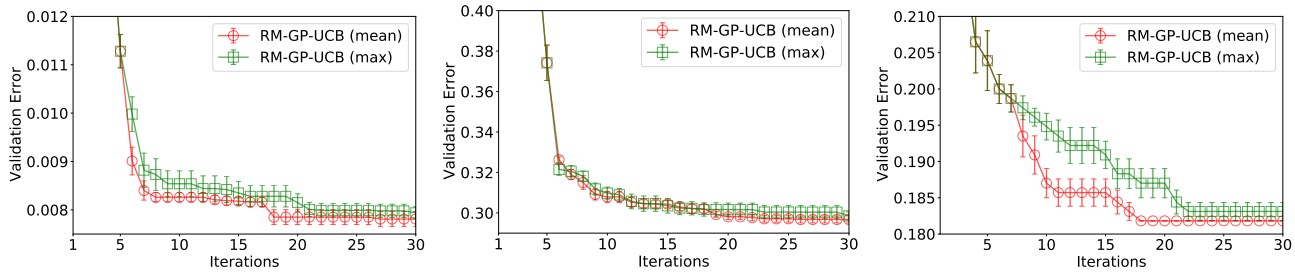
In this experiment of human activity prediction, each data instance (input-output pair) is characterized by a feature vector of length 561 and a label corresponding to one of the 6 activities. The SVM hyperparameters being optimized are the penalty parameter C (from 0.01 to 10) and the radial basis function (RBF) kernel coefficient γ (from 0.01 to 1). There are in total 7,352 data instances for the 21 subjects that are used to generate the meta-tasks, and 2,947 instances for the 9 subjects used for performance validation. For each subject, half of the instances are used as the training set, with the other half being used for validation.

Non-stationary Bayesian Optimization. The clinical diagnosis dataset used in this experiment can be found at `https://www.kaggle.com/uciml/pima-indians-diabetes-database`, and it is associated with the CC0 License. The hyperparameters of the logistic regression (LR) model being optimized are the batch size (20 to 60), the L2 regularization parameter (10^{-6} to 0.01) and the learning rate (0.01 to 0.1). The dataset represents a binary classification problem (whether a patient has diabetes or not), with each input instance consisting of 8 diagnostic features: number of pregnancies, plasma glucose concentration, blood pressure, skin thickness, insulin, BMI, diabetes pedigree function, and age. The entire dataset consists of 768 data instances, among which 77 instances are set aside to measure the validation accuracy. The sizes of the 5 progressively growing training datasets (i.e., corresponding to the 4 meta-tasks and the target task respectively) are 138, 276, 414, 552, and 691.

Policy Search for Reinforcement Learning. In this experiment, we use the Cart-Pole environment from OpenAI Gym (`https://github.com/openai/gym`), which is under the MIT License. We adopt the linear softmax policy which linearly maps a state vector of length 4 to an action vector of length 2, followed by a softmax operator. As a result, for a particular state, the action with the largest softmax value is taken. With this setting, $4 \times 2 = 8$ parameters are tuned in this experiment. The performance metric used in the experiment is the cumulative rewards (normalized to the range $[0, 1]$) in an episode (averaged over 10 independent episodes), and the maximum length of each episode is set to 200.

E.3 IMPACTS OF MAX VS MEAN IN FUNCTION GAP ESTIMATION

Here we explore the impact of the choice between using $\max$ (the outer $\max$ operator over $j = 1, ..., N_i$) or the empirical mean in the estimated upper bound on the function gap (Lemma 1), as mentioned in the first paragraph of Section 6. Fig. 7 plots the different performances using these two choices in the MNIST, CIFAR-10 and clinical diagnosis (non-stationary



(a) (b) (c)

Figure 7: Impacts of using max vs empirical mean in estimating the upper bound on the function gaps, using the (a) MNIST, (b) CIFAR-10 and (c) clinical diagnosis experiments.

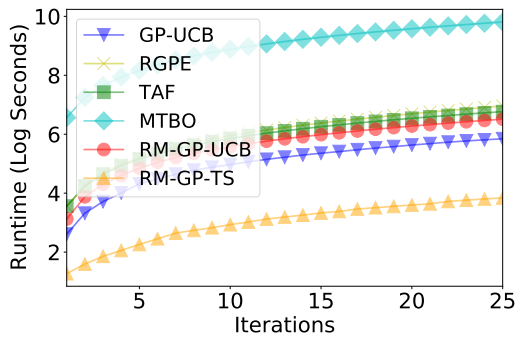


Figure 8: Runtime of different algorithms in the non-stationary BO (clinical diagnosis) experiment.

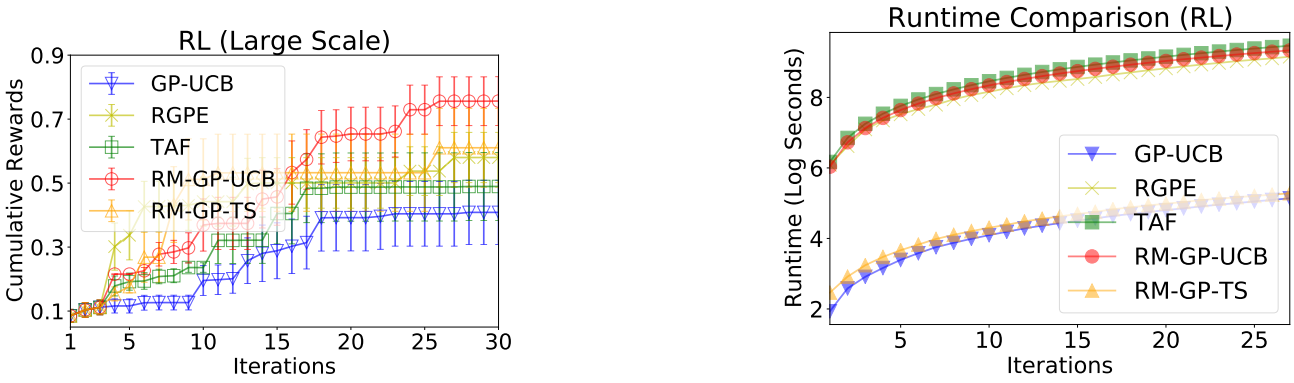
BO) experiments. The results show that the performance deficit resulting from the use of the $\max$ operator is marginal in some experiments (Fig. 7a and b), whereas the difference can be larger in some other experiments (Fig. 7c). Therefore, it is recommended to use the empirical mean when estimating the upper bound on the function gap in practice.

E.4 SCALABILITY OF OUR ALGORITHMS

Here we further demonstrate the scalability of our RM-GP-UCB and RM-GP-TS algorithms. by showing that our algorithms can be applied to experiments with a very large scale, and still performs competitively. Specifically, we construct a much larger version of the experiment on policy search for RL, with 60 meta-tasks each containing 130 meta-observations. Fig. 9a and b plot the performance and runtime in this large-scale experiment. Consistent with Fig. 2e in the main text, our RM-GP-UCB algorithm still performs the best among all algorithms (Fig. 9a). RM-GP-TS has a better performance here than in Fig. 2e, performing comparably with RGPE (Fig. 9a). Moreover, RM-GP-TS is again significantly more scalable than RM-GP-UCB, RGPE and TAF, and its computational cost is comparable with standard GP-UCB (Fig. 9b).

E.5 MORE DETAILS ON RM-GP-TS

In this section, we present more details on the practical implementation of our RM-GP-TS algorithm. In all experiments, when sampling a function from the GP posterior, we use random Fourier features (RFF) [Dai et al., 2020, Rahimi and Recht, 2008] with $m = 120$ random Fourier features. Firstly, we need to construct a set of random features. For an SE kernel with hyperparameters l and σ_k (i.e., $k(\mathbf{z}) = \sigma_k^2 e^{-\frac{\|\mathbf{z}\|_2^2}{2l^2}}$, with $\mathbf{z} = \mathbf{x}_1 - \mathbf{x}_2, \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$), we firstly sample m vectors $\{\mathbf{s}_i\}_{i=1,\dots,m}$ from the D-dimensional Gaussian distribution: $\mathcal{N}(0, \frac{1}{l^2} I)$, and sample m scalar values $\{b_i\}_{i=1,\dots,m}$ from the uniform distribution within the domain $[0, 2\pi]$. Next, for any input $\mathbf{x} \in \mathcal{D}$, its corresponding m-dimensional random features can be constructed as $\boldsymbol{\phi}(\mathbf{x}) = [\sqrt{2/m}\cos(\mathbf{s}_i^\top \mathbf{x} + b_i)]_{i=1,\dots,m}^\top$. Every $\boldsymbol{\phi}(\mathbf{x})$ is then normalized such that $\left\|\boldsymbol{\phi}(\mathbf{x})\right\|_2^2 = \sigma_k^2, \forall \mathbf{x} \in \mathcal{D}$. Based on these, in order to (approximately) sample a function from the GP posterior, we firstly sample a vector $\boldsymbol{\omega}$ from the Gaussian distribution $\boldsymbol{\omega} \sim \mathcal{N}(\boldsymbol{\nu}_t, \sigma^2 \boldsymbol{\Sigma}_t)$, with $\boldsymbol{\Sigma}_t = (\boldsymbol{\Phi}_t^\top \boldsymbol{\Phi}_t + \sigma^2 \boldsymbol{I})^{-1}$, $\boldsymbol{\nu}_t = \boldsymbol{\Sigma}_t \boldsymbol{\Phi}_t^\top \mathbf{y}_t$, and $\boldsymbol{\Phi}_t = [\boldsymbol{\phi}(\mathbf{x}_1, \dots, \mathbf{x}_t)]^\top$. Finally, we can use the sampled $\boldsymbol{\omega}$ to construct the sampled function such that $f^t(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^\top \boldsymbol{\omega}, \forall \mathbf{x} \in \mathcal{D}$. As a result, as



(a) Cumulative rewards.

(b) Runtime.

Figure 9: Results demonstrating that our algorithms can be applied to experiments of a very large scale, using a larger version of the RL experiment (with $60 \times 130 = 7800$ meta-observations).

mentioned in Sec. 3, for a meta-task i, in order to sample multiple functions from the meta-function f_i before the algorithm starts, we simply need to draw multiple samples of the vector $\boldsymbol{\omega}$ from the corresponding multivariate Gaussian distribution using the observations from meta-task i. For both the target function and every meta-function, the kernel hyperparameters (l and σ_k) used in the posterior sampling steps above are learned by maximizing the marginal likelihood (using full GP without RFF approximation), which is a common practice in BO.