On Provably Robust Meta-Bayesian Optimization (Supplementary material)

Zhongxiang Dai[1]   **Yizhou Chen**[1]   **Haibin Yu**[2]   **Bryan Kian Hsiang Low**[1]   **Patrick Jaillet**[3]

[1]Department of Computer Science, National University of Singapore, Republic of Singapore
[2]Department of Data Platform, Tencent
[3]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, USA

A   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 A 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 A.2. To simplify exposition, we omit the superscript in our notation to represent the acquisition function (2), i.e., we use $\overline{\zeta}_t$ to denote the acquisition function of RM-GP-UCB instead of $\overline{\zeta}_t^{\text{UCB}}$. 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 A.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

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

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 $\geq 1 - \delta/4$.

A.1 PROOF OF LEMMA 3

Let $\mathbf{K}_i = [k(\mathbf{x}_{i,j}, \mathbf{x}_{i,j'})]_{j,j'=1,...,N_i}$ represent the Gram matrix corresponding to the inputs of the meta-observations from meta-task i, and $\mathbf{k}_i = [k(\mathbf{x}_{i,j}, \mathbf{x})]_{j=1,...,N_i}^\top$. 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.

$$\left\| \left(\mathbf{K}_i + \sigma^2 I \right)^{-1} \right\|_2 = \sqrt{\max_{j=1,...,N_i} \lambda_j\left[\left((\mathbf{K}_i + \sigma^2 I)^{-1} \right)^\top (\mathbf{K}_i + \sigma^2 I)^{-1} \right]}$$

$$= \sqrt{\max_{j=1,...,N_i} \lambda_j\left[(\mathbf{K}_i + \sigma^2 I)^{-1} \right]^2}$$

$$\leq \frac{1}{\sigma^2}$$

$\square$

Next, define $\overline{\mathbf{f}}_i = [f_i(\mathbf{x}_{i,j})]_{j=1,...,N_i}$ (in which $f_i(\mathbf{x}_{i,j})$ represents the value of meta-function i at input $\mathbf{x}_{i,j}$), and $\widetilde{\mathbf{f}}_i = [f(\mathbf{x}_{i,j})]_{j=1,...,N_i}$ (in which $f(\mathbf{x}_{i,j})$ represents the value of target function at input $\mathbf{x}_{i,j}$). Similarly, define $\overline{\mathbf{y}}_i = [y_{i,j}]_{j=1,...,N_i}$ (in which $y_{i,j}$ represents the noisy output observation of meta-task i at input $\mathbf{x}_{i,j}$), and $\widetilde{\mathbf{y}}_i = [y(\mathbf{x}_{i,j})]_{j=1,...,N_i}$ (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* $\geq 1 - \delta/4$,

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

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

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,

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

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

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

$$\left|y_{i,j} - f_i(\mathbf{x}_{i,j})\right| \le \sqrt{2\sigma^2 \log \frac{8N_i}{\delta}},$$

$$\left|y(\mathbf{x}_{i,j}) - f(\mathbf{x}_{i,j})\right| \le \sqrt{2\sigma^2 \log \frac{8N_i}{\delta}},$$

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

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

Repeating the procedure above leads to

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

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

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

$$\left|\overline{\zeta}_t(\mathbf{x}) - \widetilde{\zeta}_t(\mathbf{x})\right| = \left|\nu_t\left[\sum_{i=1}^{M} \omega_i[\overline{\mu}_i(\mathbf{x}) + \sqrt{\tau}\overline{\sigma}_i(\mathbf{x})]\right] - \nu_t\left[\sum_{i=1}^{M} \omega_i[\widetilde{\mu}_i(\mathbf{x}) + \sqrt{\tau}\widetilde{\sigma}_i(\mathbf{x})]\right]\right|$$

$$\overset{(a)}{=} \left|\nu_t \sum_{i=1}^{M} \omega_i[\overline{\mu}_i(\mathbf{x}) - \widetilde{\mu}_i(\mathbf{x})]\right|$$

$$\le \nu_t \sum_{i=1}^{M} \omega_i \left|\overline{\mu}_i(\mathbf{x}) - \widetilde{\mu}_i(\mathbf{x})\right|$$

$$\le \nu_t \sum_{i=1}^{M} \omega_i \left|\mathbf{k}_i(\mathbf{x})^\top (\mathbf{K}_i + \sigma^2 I)^{-1}(\overline{\mathbf{y}}_i - \widetilde{\mathbf{y}}_i)\right|$$

$$\overset{(b)}{\le} \nu_t \sum_{i=1}^{M} \omega_i \left\|\mathbf{k}_i(\mathbf{x})\right\|_2 \left\|(\mathbf{K}_i + \sigma^2 I)^{-1}\right\|_2 \|\overline{\mathbf{y}}_i - \widetilde{\mathbf{y}}_i\|_2$$

$$\overset{(c)}{\le} \nu_t \sum_{i=1}^{M} \omega_i \left\|\mathbf{k}_i(\mathbf{x})\right\|_2 \frac{1}{\sigma^2} \|\overline{\mathbf{y}}_i - \widetilde{\mathbf{y}}_i\|_2$$

$$\overset{(d)}{\le} \nu_t \sum_{i=1}^{M} \omega_i \sqrt{N_i} \frac{1}{\sigma^2} \|\overline{\mathbf{y}}_i - \widetilde{\mathbf{y}}_i\|_2$$

$$\le \nu_t \sum_{i=1}^{M} \omega_i \frac{\sqrt{N_i}}{\sigma^2} \left\|\overline{\mathbf{y}}_i - \overline{\mathbf{f}}_i + \overline{\mathbf{f}}_i - \widetilde{\mathbf{f}}_i + \widetilde{\mathbf{f}}_i - \widetilde{\mathbf{y}}_i\right\|_2$$

$$\le \nu_t \sum_{i=1}^{M} \omega_i \frac{\sqrt{N_i}}{\sigma^2} \left[\left\|\overline{\mathbf{y}}_i - \overline{\mathbf{f}}_i\right\|_2 + \left\|\overline{\mathbf{f}}_i - \widetilde{\mathbf{f}}_i\right\|_2 + \left\|\widetilde{\mathbf{f}}_i - \widetilde{\mathbf{y}}_i\right\|_2\right]$$

$$\overset{(e)}{\le} \nu_t \sum_{i=1}^{M} \omega_i \frac{\sqrt{N_i}}{\sigma^2} \left(2\sqrt{N_i}\sqrt{2\sigma^2 \log \frac{8N_i}{\delta}} + \left\|\overline{\mathbf{f}}_i - \widetilde{\mathbf{f}}_i\right\|_2\right)$$

$$= \nu_t \sum_{i=1}^{M} \omega_i \frac{\sqrt{N_i}}{\sigma^2} \left(2\sqrt{N_i}\sqrt{2\sigma^2 \log \frac{8N_i}{\delta}} + \sqrt{\sum_{j=1}^{N_i}\left(f_i(\mathbf{x}_{i,j}) - f(\mathbf{x}_{i,j})\right)^2}\right)$$

$$\overset{(f)}{\le} \nu_t \sum_{i=1}^{M} \omega_i \frac{\sqrt{N_i}}{\sigma^2} \left(2\sqrt{N_i}\sqrt{2\sigma^2 \log \frac{8N_i}{\delta}} + d_i\sqrt{N_i}\right)$$

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

$$\triangleq \nu_t \alpha \tag{12}$$

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: $d_i \triangleq \max_{j=1,\dots,N_i} \left| f(\mathbf{x}_{i,j}) - f_i(\mathbf{x}_{i,j}) \right|$ for $i = 1, \dots, M$. This completes the proof of Lemma 3.

A.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 $f(\mathbf{x})$:

$$\widetilde{\zeta}_t(\mathbf{x}) - f(\mathbf{x}) = \widetilde{\zeta}_t(\mathbf{x}) - \left[\nu_t \sum_{i=1}^{M} \omega_i f(\mathbf{x}) + (1 - \eta_t) f(\mathbf{x}) \right]$$

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

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

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

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. $\square$

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

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

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. $\square$

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 $\beta_t \ \forall t \geq 1$ 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

$$
4(1-\nu_t)^2\beta_t^2\sigma_{t-1}^2(\mathbf{x}_t) \overset{(a)}{\le} 4(1-\nu_T)^2\beta_T^2\sigma^2\left(\sigma^{-2}\sigma_{t-1}^2(\mathbf{x}_t)\right)
$$

$$
\overset{(b)}{\le} 4(1-\nu_T)^2\beta_T^2\sigma^2\left(\frac{\sigma^{-2}}{\log(1+\sigma^{-2})}\log\left(1+\sigma^{-2}\sigma_{t-1}^2(\mathbf{x}_t)\right)\right) \tag{16}
$$

$$
= (1-\nu_T)^2\beta_T^2\frac{8}{\log(1+\sigma^{-2})}\left[\frac{1}{2}\log\left(1+\sigma^{-2}\sigma_{t-1}^2(\mathbf{x}_t)\right)\right]
$$

in which (a) follows since β_t is non-decreasing in t and ν_t is non-increasing in t, (b) follows since $\sigma^{-2}x \le \frac{\sigma^{-2}}{\log(1+\sigma^{-2})}\log(1+\sigma^{-2}x)$ 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

$$
\sum_{t=1}^{T}\left[2(1-\nu_t)\beta_t\sigma_{t-1}(\mathbf{x}_t)\right]^2 \overset{(a)}{\le} (1-\nu_T)^2\beta_T^2\frac{8}{\log(1+\sigma^{-2})}\sum_{t=1}^{T}\left[\frac{1}{2}\log\left(1+\sigma^{-2}\sigma_{t-1}^2(\mathbf{x}_t)\right)\right]
$$

$$
\overset{(b)}{=} (1-\nu_T)^2\beta_T^2\frac{8}{\log(1+\sigma^{-2})}I(\mathbf{y}_T;\mathbf{f}_T)
$$

$$
\overset{(c)}{\le} (1-\nu_T)^2 C_1\beta_T^2\gamma_T
$$

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. $\qquad\square$

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

$$
R_T = \sum_{t=1}^{T} r_t \overset{(a)}{\le} \sum_{t=1}^{T}\left[2\nu_t(\alpha+\tau) + 2\left(1-\nu_t\right)\beta_t\sigma_{t-1}(\mathbf{x}_t)\right]
$$

$$
= 2(\alpha+\tau)\sum_{t=1}^{T}\nu_t + \sum_{t=1}^{T} 2(1-\nu_t)\beta_t\sigma_{t-1}(\mathbf{x}_t)
$$

$$
\overset{(b)}{\le} 2(\alpha+\tau)\sum_{t=1}^{T}\nu_t + \sqrt{T}\sqrt{\sum_{t=1}^{T}\left[2(1-\nu_t)\beta_t\sigma_{t-1}(\mathbf{x}_t)\right]^2} \tag{17}
$$

$$
\overset{(c)}{\le} 2(\alpha+\tau)\sum_{t=1}^{T}\nu_t + \sqrt{C_1 T(1-\nu_T)^2\beta_T^2\gamma_T}
$$

$$
\overset{(d)}{\le} 2(\alpha+\tau)\sum_{t=1}^{T}\nu_t + \beta_T\sqrt{C_1 T\gamma_T}
$$

which holds with probability $\ge 1 - 3\delta/4$. (a) is a result of Lemma 8, (b) follows from Cauchy-Schwarz inequality, (c) is obtained using Lemma 10, and (d) follows since $1 - \nu_T \le 1$. 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: $R_T \le 2\tau\sum_{t=1}^{T}\nu_t + 2\sum_{t=1}^{T}\nu_t\alpha_t + \beta_T\sqrt{C_1 T\gamma_T}$. 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 $R_T \le 2(\alpha'+\tau)\sum_{t=1}^{T}\nu_t + \beta_T\sqrt{C_1 T\gamma_T}$, with $\alpha' \triangleq \sum_{i=1}^{M}\frac{N_i}{\sigma^2}(2\sqrt{2\sigma^2\log\frac{8N_i}{\delta}} + d_i)$.

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

Here, we utilize the analysis in Appendix A.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: $d_i = \max_{j=1,\ldots,N_i} |f(\mathbf{x}_{i,j}) - f_i(\mathbf{x}_{i,j})| = 0, \forall i = 1,\ldots,M$. 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.

B   PROOF OF THEOREM 2

Our theoretical analysis of RM-GP-TS shares similarity with the works of [Dai et al., 2020b, 2021] but has important differences, e.g., unlike the works of [Dai et al., 2020b, 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 communication efficiency and retaining (hence not transmitting) the raw data.

Based on the acquisition function $\overline{\zeta}_t$ for RM-GP-TS (3) (we have again removed the superscript for simplicity), 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:

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

$\square$

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 $\geq 1 - \delta/4$. $\square$

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 ($\delta/4$). $\qquad\square$

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 instantaneous 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:

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

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. $\square$

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. $\square$

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

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

$$\leq 2B + c_t \left(1 + \frac{2}{(1-\nu_1)p} \right) + \psi_t,$$

(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):

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

$$\sqrt{2 \log \frac{10}{\delta} \sum_{t=1}^{T} \left(2B + c_t \left(1 + \frac{2}{(1-\nu_1)p} \right) + \psi_t \right)^2}$$

$$\leq c_T \left(1 + \frac{2}{(1-\nu_1)p} \right) \sum_{t=1}^{T} \sigma_{t-1}(\mathbf{x}_t) + \sum_{t=1}^{T} \psi_t +$$

$$\left(2B + c_T \left(1 + \frac{2}{(1-\nu_1)p} \right) + \psi_1 \right) \sqrt{2T \log \frac{12}{\delta}}$$

(38)

$$\leq c_T \left(1 + \frac{2}{(1-\nu_1)p} \right) \sqrt{C'_1 \gamma_T T} + \sum_{t=1}^{T} \psi_t +$$

$$\left(2B + c_T \left(1 + \frac{2}{(1-\nu_1)p} \right) + \psi_1 \right) \sqrt{2T \log \frac{12}{\delta}}$$

$$\leq \left(2B + c_T \left(1 + \frac{2}{(1-\nu_1)p} \right) + \psi_1 \right) \sqrt{T(C'_1 \gamma_T + 2\log(12/\delta))} +$$

$$2 \sum_{t=1}^{T} \nu_t (c_t + c'_t + \sum_{i=1}^{M} \omega_i d_i).$$

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 $C'_1 \triangleq 2/\log(1 + \sigma^{-2})$. $\qquad \square$

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:

$$R_T = \mathcal{O} \left(c_T \sqrt{T \gamma_T} + \sum_{t=1}^{T} \nu_t (c_t + c'_t + \sum_{i=1}^{M} \omega_i d_i) \right)$$

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

(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 A.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):

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

$$= \mathcal{O}\Big(\Big(\sum_{i=1}^M d'_i \Big) \sum_{t=1}^T \nu_t + \sum_{t=1}^T \nu_t \sqrt{\gamma_t} \log t + \gamma_T \log T \sqrt{T} \Big)$$

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

C   ANALYSIS OF ONLINE META-WEIGHT OPTIMIZATION

C.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

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

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* $\geq 1 - \delta/4$, $\forall\, t \geq 1$, $\forall i, j$,

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

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

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 $\left| f_i(\mathbf{x}_{i,j}) - U_{t,i,j} \right|$:

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

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. $\qquad \square$

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

C.2 PROOF OF PROPOSITION 1

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

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

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

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

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.

C.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).

C.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 $\geq 1 - \delta$,

$$R_T = \mathcal{O}\Big(\sum_{t=1}^{T} \nu_t \left(\sum_{i=1}^{M} \omega_i \bar{d}_{i,t} \right) + \sum_{t=1}^{T} \nu_t \sqrt{\gamma_t} \log t + \gamma_T \log T \sqrt{T} \Big)$$

$$\leq \mathcal{O}\Big(\Big(\underline{\sum_{t=1}^{T} \sum_{i=1}^{M} \omega_i \bar{d}_{i,t}} \Big) \Big(\sum_{t=1}^{T} \nu_t \Big) + \sum_{t=1}^{T} \nu_t \sqrt{\gamma_t} \log t + \gamma_T \log T \sqrt{T} \Big), \tag{52}$$

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

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

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

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

Next, applying similar derivations as Appendix C.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).

D   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.

D.1   OPTIMIZATION OF SYNTHETIC FUNCTIONS

D.1.1   Synthetic Functions Sampled from GPs

The objective functions are drawn from GP's with the Squared Exponential kernel (with a 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
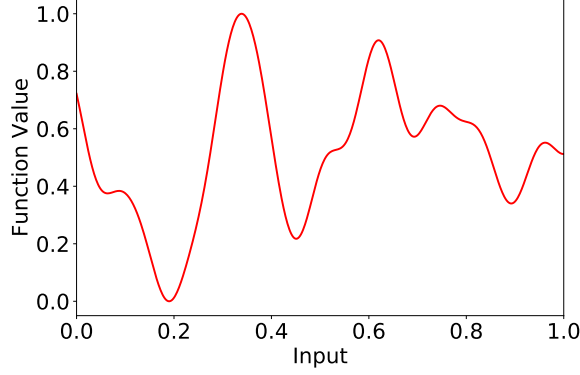
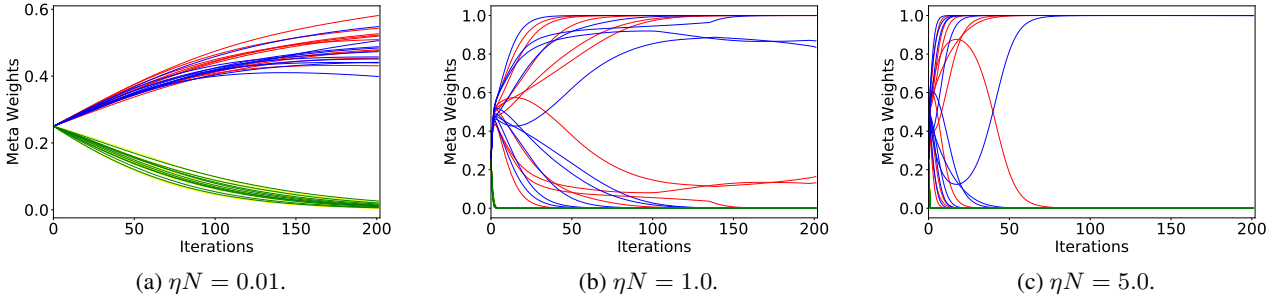Figure 3: An example synthetic function sampled from a GP.

(a) $\eta N = 0.01$.

(b) $\eta N = 1.0$.

(c) $\eta N = 5.0$.

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 $i = 1 \ldots M$, 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.

Moreover, we have added another experiment where the N_i's (i.e., the number of observations from the meta-tasks) are different. Specifically, we use the same experimental setting involving $M = 4$ meta-tasks as described above, and let $N_1 = 15, N_2 = 25, N_3 = 10, N_4 = 30$, where $d_1 = d_2 = 0.05$, $d_3 = d_4 = 4.0$. The results (Fig. 5) show that when the N_i's are different, our RM-GP-UCB algorithm, despite performing worse than the setting where all N_i's are equal, is still able to significantly outperform standard GP-UCB.
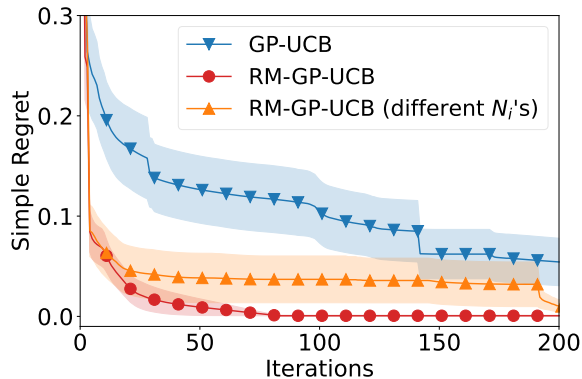
Figure 5: The performance of RM-GP-UCB when the N_i's are different.

D.2 REAL-WORLD EXPERIMENTS

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[1], and the SVHN dataset can be downloaded from `http://ufldl.stanford.edu/housenumbers/`. The MNIST dataset is under the GNU General Public License, CIFAR-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. 6 presents the results when the SVHN and CIFAR-100 datasets are used to produce the target functions.

Comparing Figs. 1e, 1f and Fig. 6 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. 6), 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 Sec. 7, 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, 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

---

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

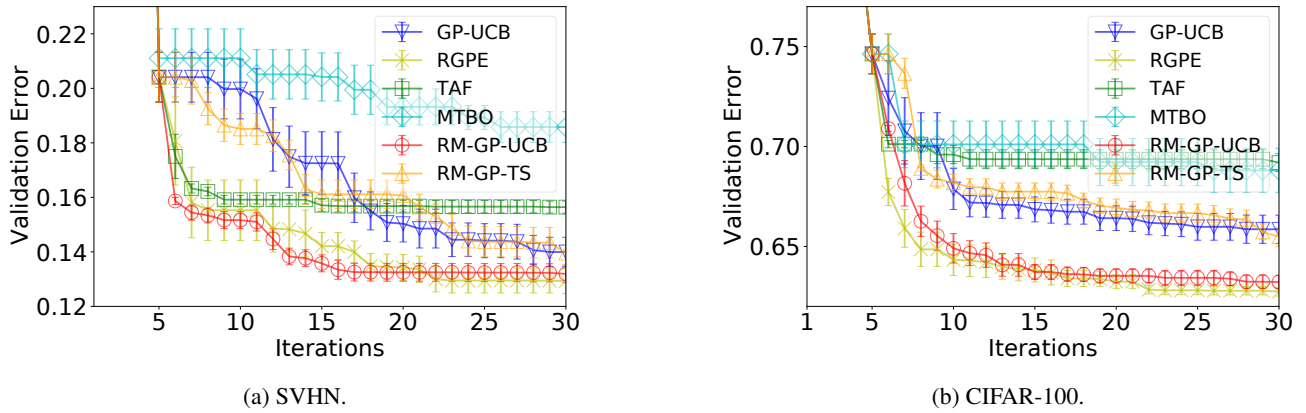|	
(a) SVHN.	(b) CIFAR-100.

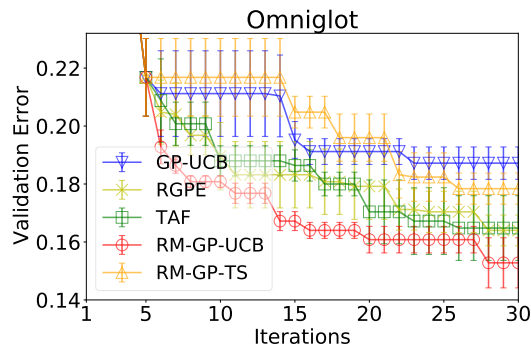Figure 6: Best validation error of CNN (both averaged over 10 random initializations).

Figure 7: 2-way validation error on the Omniglot dataset.

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 introduced 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.
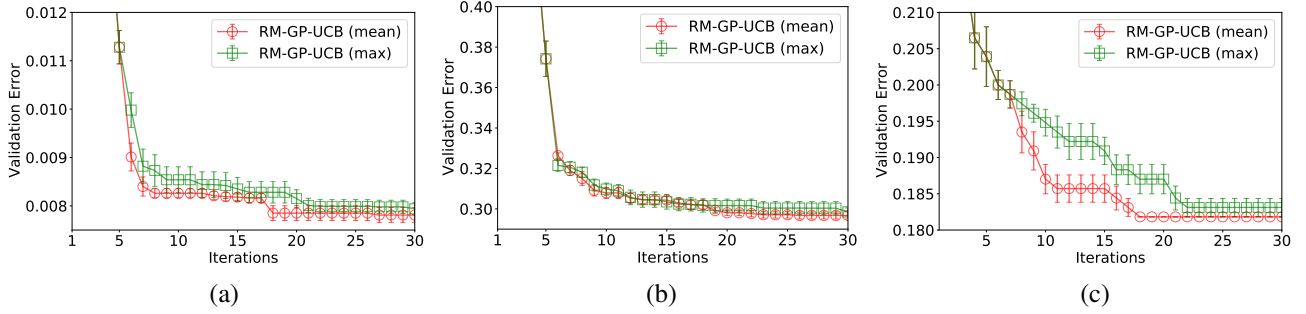
Figure 8: Impacts of using max vs empirical mean in estimating the upper bound on the function gaps, using the (a) MNIST, (b) CIFAR-10 and (c) non-stationary BO (clinical diagnosis) experiments.

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.

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.

### D.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. 8 plots the different performances using these two choices in the MNIST, CIFAR-10 and clinical diagnosis (non-stationary BO) experiments. The results show that the performance deficit resulting from the use of the $\max$ operator is marginal in some experiments (Fig. 8a and b), whereas the difference can be larger in some other experiments (Fig. 8c). Therefore, it is recommended to use the empirical mean when estimating the upper bound on the function gap in practice.

D.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).

D.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., 2020b, 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 $\{\mathbf{s}_i\}_{i=1,...,m}$ from the D-dimensional Gaussian distribution: $\mathcal{N}(0, \frac{1}{l^2}I)$, and sample m scalar values $\{b_i\}_{i=1,...,m}$ from the uniform
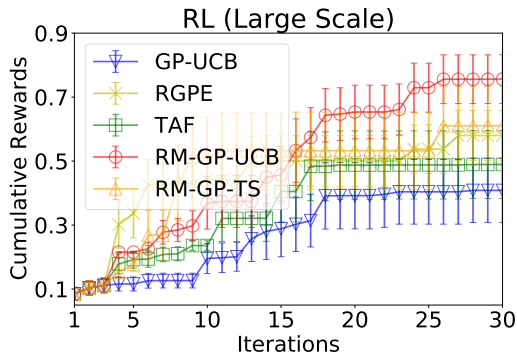
(a) Cumulative rewards.
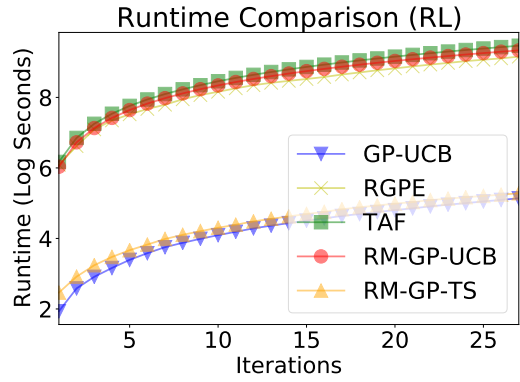
(b) Runtime.

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

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 $\phi(\mathbf{x}) = [\sqrt{2/m} \cos(\mathbf{s}_i^\top \mathbf{x} + b_i)]_{i=1,\ldots,m}^\top$. Every $\phi(\mathbf{x})$ is then normalized such that $\|\phi(\mathbf{x})\|_2^2 = \sigma_k^2, \forall \mathbf{x} \in \mathcal{D}$. 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 $\boldsymbol{\Phi}_t = [\phi(\mathbf{x}_1, \ldots, \mathbf{x}_t)]^\top$. Finally, we can use the sampled $\boldsymbol{\omega}$ to construct the sampled function such that $f^t(\mathbf{x}) = \phi(\mathbf{x})^\top \boldsymbol{\omega}, \forall \mathbf{x} \in \mathcal{D}$. As a result, as 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.