Adaptive Prior Selection in Gaussian Process Bandits with Thompson Sampling

Jack Sandberg **Morteza Haghir Chehreghani**
Department of Computer Science and Engineering
Chalmers University of Technology & University of Gothenburg
Gothenburg, Sweden
{jack.sandberg, morteza.chehreghani}@chalmers.se

Abstract

Gaussian process (GP) bandits provide a powerful framework for performing blackbox optimization of unknown functions. The characteristics of the unknown function depend heavily on the assumed GP prior. Most work in the literature assume that this prior is known but in practice this seldom holds. Instead, practitioners often rely on maximum likelihood estimation to select the hyperparameters of the prior - which lacks theoretical guarantees. In this work, we study two algorithms for joint prior selection and regret minimization in GP bandits based on GP Thompson sampling (GP-TS): Prior-Elimination GP-TS (PE-GP-TS) that disqualifies priors with poor predictive performance, and HyperPrior GP-TS (HP-GP-TS) that utilizes a bi-level Thompson sampling scheme. We theoretically analyze the algorithms and establish a sublinear regret bound for HP-GP-TS. In addition, we demonstrate the effectiveness of these algorithms compared to the alternatives through extensive experiments with synthetic and real-world data.

1 Introduction

The Gaussian process bandit problem is a variant of the multi-armed bandit problem where the arms are correlated and their expected reward is sampled from a Gaussian process (GP). The flexibility of GPs have made GP bandits applicable in a wide range of areas that need to optimize blackbox functions with noisy estimates, including hyperparameter tuning (Turner et al., 2021), online advertising (Nuara et al., 2018), and portfolio optimization (Gonzalvez et al., 2019). Most of the theoretical results in the literature assume that the GP prior is known but this is seldom the case in practical applications. Even with expert domain knowledge, selecting the exact prior to use can be a difficult task. Most practitioners tend to utilize maximum likelihood estimation (MLE) to identify suitable prior parameters. However, in a sequential decision making problem MLE is not guaranteed to recover the correct parameters which can hurt the performance.

As summarized in Table 1, previous works by Wang & de Freitas (2014); Berkenkamp et al. (2019) propose algorithms that use a decreasing sequence of lengthscales according to a fixed schedule. A drawback of these schedules is that they cannot adapt to the data and may therefore explore excessively. The Lengthscale Balancing GP-UCB algorithm of Ziomek et al. (2024) selects lengthscales such that each selected lengthscale incurs a similar amount of regret. However, this scheme relies on knowing the regret bounds, which can be impractical. Ziomek et al. (2025); Lu et al. (2023) propose algorithms that support unknown priors of (finite) arbitrary type. Prior-Elimination GP-UCB (PE-GP-UCB) (Ziomek et al., 2025) selects the prior and arm that maximize a joint upper confidence bound and eliminates priors with poor predictive performance. The joint upper confidence bound induces a double optimism in PE-GP-UCB that can lead to extra exploration. EGP-TS (Lu et al., 2023) uses bi-level Thompson sampling to select both a prior and an arm according to their posterior probabilities

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Table 1: Comparison of similar work in GP bandits with an unknown prior.

Work	Algorithm	Prior selection	MIG dependence	Supports unknown
(Wang & de Freitas, 2014)	BOHO (EI)	Schedule	$\hat{\gamma}_T^{3/2}$	Lengthscale
(Berkenkamp et al., 2019)	A-GP-UCB	Schedule	γ_{T,p_T} [†]	Lengthscale and RKHS norm
(Lu et al., 2023)	EGP-TS	Posterior sampling	$\sqrt{\lvert P\rvert\hat{\gamma}_T}$ (invalid)	Arbitrary mean and kernel
(Ziomek et al., 2024)	LB-GP-UCB	Regret balancing	$\gamma_{T,\bar{p}}$ [†]	Lengthscale and RKHS norm
(Ziomek et al., 2025)	PE-GP-UCB	Optimistic	$\sqrt{\lvert P\rvert\hat{\gamma}_T}$	Arbitrary mean and kernel
This work	**PE-GP-TS**	Optimistic	$\sqrt{\lvert P\rvert\hat{\gamma}_T}$	Arbitrary mean and kernel
This work	**HP-GP-TS** [‡]	Posterior sampling	$\sqrt{\bar{\gamma}_T(P_1)}$	Arbitrary mean and kernel

[†] p_T is the final prior selected by A-GP-UCB, and $\bar{p}$ is the prior that minimizes the frequentist regret of GP-UCB.
[‡] Equivalent to EGP-TS (Lu et al., 2023), we refer to it as HP-GP-TS.

of being the true prior and optimal arm, respectively. Among these methods, posterior sampling is the only data-adaptive prior selection rule, and provides the closest analog to MLE.

EGP-TS is an instantiation of the more general MixTS algorithm (Hong et al., 2022b), whose regret was analyzed in the standard bandit and linear setting. However, the theoretical analyses for both algorithms are flawed. The technical issues in the regret analysis of EGP-TS were recently demonstrated by Sandberg & Haghir Chehreghani (2026) and, as we show in this work, the analysis of MixTS in the linear setting contains separate technical issues that invalidate the regret bound of Hong et al. (2022b).

Motivated by the excessive exploration of double optimism, alongside the flawed theoretical guarantees of existing Thompson sampling approaches, we investigate two distinct TS-based algorithms for GP-bandits with unknown priors. The first algorithm, Prior-Elimination GP-TS (PE-GP-TS), is an extension of PE-GP-UCB that replaces the doubly optimistic selection rule with posterior sampling and one less layer of optimism. We analyze the regret of PE-GP-TS and obtain a regret bound of order $\mathcal{O}(\sqrt{T|P|\hat{\gamma}_T \log T})$ (which matches that of PE-GP-UCB) plus a term (left unbounded) depending on the uncertainty of the optimal arm under the correct prior. Here, T is the horizon, $|P|$ is the number of priors and $\hat{\gamma}_T$ is the worst-case maximum information gain. The second algorithm we study is EGP-TS, which we refer to as HyperPrior GP-TS (HP-GP-TS) to emphasize its use of a hyperprior, and it removes both levels of optimism. Our analysis of HP-GP-TS addresses the issues in the previous work and yields a regret bound of order $\mathcal{O}(\sqrt{T\bar{\gamma}_T(P_1)\log T})$ where $\bar{\gamma}_T(P_1)$ is a sum of maximum information gains with cardinality equal to the horizon T times the hyperprior probability $P_1(\cdot)$ s.t. $\bar{\gamma}_T(P_1) < |P|\hat{\gamma}_T$ generally holds.

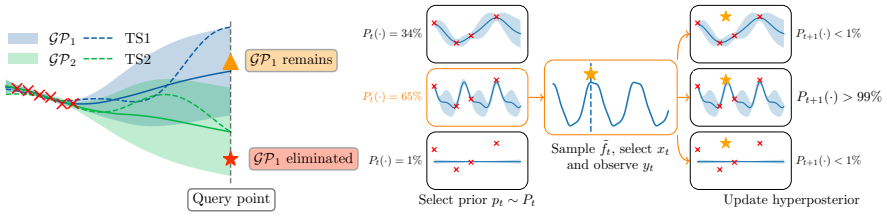
We evaluate our methods on three sets of synthetic experiments and three experiments with real-world data. Across the experiments, the Thompson sampling based methods outperform PE-GP-UCB. Additionally, we find that the regret of HP-GP-TS does not increase with $|P|$ in our scaling experiments. Finally, we analyze the priors selected by the algorithms and observe that HP-GP-TS selects the correct prior more often than the other algorithms.

The contributions of this work can be summarized as:

- We propose a Thompson sampling based algorithm for GP-bandits with unknown prior, PE-GP-TS, and theoretically analyze its regret.
- We provide a sublinear regret bound for HP-GP-TS (Lu et al., 2023, EGP-TS) that depends on $\bar{\gamma}_T(P_1)$, correcting and improving upon the bound of Lu et al. (2023).
- We identify technical issues with the proof of the regret bound for MixTS (Hong et al., 2022b) in the linear setting, preventing its direct extension to the GP-setting.
- We experimentally evaluate the TS-based algorithms on both synthetic and real-world data, demonstrating that they achieve competitive performance and that the regret of HP-GP-TS does not empirically increase with $|P|$.

2 Background and problem statement

Problem statement We consider a sequential decision making problem where an agent repeatedly selects among a set of arms and receives a random reward whose mean depends on the selected arm



(a) (b)

Figure 1: a) Elimination procedure of PE-GP-TS. The solid lines correspond to posterior means and the shaded regions are confidence intervals. The figure has been adapted from Ziomek et al. (2025). The dashed lines are samples from the posteriors. b) Overview of HP-GP-TS.

and is unknown to the agent. The goal of the agent is to maximize the cumulative sum of rewards over a finite time horizon. We assume that the distribution of the means, the *prior*, is sampled from a set of priors, the *hyperprior*. An effective agent must distinguish which prior the means are sampled from to ensure it explores efficiently.

Now, let us formally state the problem. Let $\mathcal{X} \subseteq [0, r]^d$ denote the finite set of arms and P a finite set of priors with associated prior mean and kernel functions $\mu_{1,p} : \mathcal{X} \mapsto \mathbb{R}$ and $k_{1,p} : \mathcal{X} \times \mathcal{X} \mapsto [-1, 1]$, $\forall p \in P$. Let $p^* \in P$ denote the true prior and assume the expected reward function $f : \mathcal{X} \mapsto \mathbb{R} \sim \mathcal{GP}(\mu_{1,p^*}, k_{1,p^*})$ is a sample from a Gaussian process with prior p^*. Both the function f and the true prior p^* are considered unknown. We will consider two settings: In the frequentist selection setting, the prior $p^* \in P$ is picked arbitrarily. In the Bayesian selection setting, the prior is sampled from a known hyperprior $p^* \sim P_1$. Then, for time step $t = 1, 2, \ldots, T$ where T is the horizon, the agent selects an arm $x_t \in \mathcal{X}$ and observes the reward $y_t = f(x_t) + \epsilon_t$ where $\{\epsilon_t\}_{t=1}^T$ are i.i.d. zero-mean Gaussian noise with variance σ^2. The goal of the agent is to select a sequence of arms $\{x_t\}_{t=1}^T$ that minimizes the regret $R(T) = \sum_{t \in [T]} f(x^*) - f(x_t)$ where $[T] = \{1, \ldots, T\}$ and $x^* = \arg\max_{x \in \mathcal{X}} f(x)$. In the Bayesian selection setting, we evaluate the agent based on the Bayesian regret $\mathrm{BR}(T) = \mathbb{E}[R(T)]$ where the expectation is taken over the prior p^*, the expected reward function f, the noise $\{\epsilon_t\}_{t=1}^T$ and the (potentially) stochastic selection of arms.

Gaussian processes A Gaussian process $f(x) \sim \mathcal{GP}(\mu, k)$ is a collection of random variables such that for any subset $\{x_1, \ldots, x_n\} \subset \mathcal{X}$, the vector $[f(x_1), \ldots, f(x_n)] \in \mathbb{R}^n$ has a multivariate Gaussian distribution. The probabilistic nature of GPs make them very useful for defining and solving bandit problems where the arms are correlated. Given the history $H_t = \{(x_i, y_i)\}_{i=1}^{t-1}$, the posterior mean and kernel functions of a Gaussian process $\mathcal{GP}(\mu, k)$ are given by $\mu_t(x) = \mu(x) + \mathbf{k}^\top \left(\mathbf{K} + \sigma^2 I\right)^{-1} (\mathbf{y} - \boldsymbol{\mu})$, and $k_t(x, \tilde{x}) = k(x, \tilde{x}) - \mathbf{k}^\top \left(\mathbf{K} + \sigma^2 I\right)^{-1} \tilde{\mathbf{k}}$. Above, $\mathbf{k}, \tilde{\mathbf{k}} \in \mathbb{R}^{t-1}$ are vectors such that $(\mathbf{k})_i = k(x_i, x)$ and $(\tilde{\mathbf{k}})_i = k(x_i, \tilde{x})$. Additionally, $\mathbf{y}, \boldsymbol{\mu} \in \mathbb{R}^{t-1}$ are also vectors such that $(\mathbf{y})_i = y_i$ and $(\boldsymbol{\mu})_i = \mu(x_i)$. The gram matrix is denoted by $\mathbf{K} \in \mathbb{R}^{t-1 \times t-1}$ where $(\mathbf{K})_{i,j} = k(x_i, x_j)$. Let $\mu_{t,p}$ and $k_{t,p}$ denote the posterior mean and kernel for a Gaussian process with prior $p \in P$ at time t and let $\sigma^2_{t,p}(x) = k_{t,p}(x, x)$ denote the posterior variance at time t. The kernel k determines important characteristics of the functions f, see Appendix D for more details and examples.

Information gain The maximal information gain (MIG) is a measure of the reduction in uncertainty of f after observing the most informative data points up to a specified size. The MIG commonly occurs in regret bounds for GP bandit algorithms (Srinivas et al., 2012; Vakili et al., 2021) and its growth rate is strongly determined by the prior kernel of the GP. Hence, we will define the MIG for any fixed GP prior $p \in P$. Let $\mathbf{y}_A$ denote noisy observations of f at the locations $A \subset \mathcal{X}$. Then, the MIG given prior $p \in P$ is defined as $\gamma_{T,p} := \sup_{A \subset \mathcal{X}, |A| \leq T} I_p(\mathbf{y}_A; f)$, where $I_p(\mathbf{y}_A; f) = H(\mathbf{y}_A | p) - H(\mathbf{y}_A | f, p)$ is the mutual information between $\mathbf{y}_A$ and f given p, and $H(\cdot)$ denotes the entropy. To aid our analysis later, we also define the worst-case MIG as $\hat{\gamma}_T := \max_{p \in P} \gamma_{T,p}$ and the hyperprior-weighted MIG as $\bar{\gamma}_T(P_1) := \sum_{p \in P} \Gamma_p(T P_1(p))$ for concave $\Gamma_p(\cdot)$ s.t. $\Gamma_p(t) \geq \gamma_{t,p}$ for all $t, p \in [T] \times P$. For the RBF and Matérn kernels, $\gamma_{T,p} = \mathcal{O}(\log^{d+1}(T))$ and $\gamma_{T,p} = \mathcal{O}(T^{\frac{d}{2\nu+d}} \log^{\frac{2\nu}{2\nu+d}}(T))$ (Srinivas et al., 2012; Vakili et al., 2021).

3 Algorithms

As discussed by Russo & Van Roy (2014), TS can offer advantages over UCB algorithms for problems where constructing tight confidence bounds is difficult. In addition, Thompson sampling is often observed to perform better than UCB in practice (Chapelle & Li, 2011; Wen et al., 2015; Kandasamy et al., 2018; Åkerblom et al., 2023b,a). Motivated by this, we present two algorithms for adaptive prior selection based on TS.

3.1 Prior-Elimination with Thompson sampling

Our first algorithm is an extension of PE-GP-UCB (Ziomek et al., 2025) to be employed with Thompson sampling – instead of UCB. The key difference is that instead of maximizing the upper confidence bound $U_t(x,p) = \mu_{t,p}(x) + \sqrt{\beta_t}\sigma_{t,p}(x)$ over $\mathcal{X} \times P_t$, we instead sample $\tilde{f}_{t,p}$ from the posterior $\mathcal{GP}(\mu_{t,p}, k_{t,p})$ for all priors $p \in P_t$ where P_t is the set of active priors. Then, we select the arm and prior x_t, p_t such that $x_t, p_t = \arg\max_{x,p \in \mathcal{X} \times P_t} \tilde{f}_{t,p}(x)$. Whilst PE-GP-UCB has two layers of optimism, the upper confidence bound and joint maximization of x and p, PE-GP-TS has only a single layer of optimism - which should alleviate potential overexploration issues.

Algorithm 1 Prior Elimination GP-TS (PE-GP-TS)

input Horizon T, prior functions $\{\mu_{1,p}, k_{1,p}\}_{p \in P}$, confidence parameters $\{\beta_t\}_{t=1}^T$ and $\{\xi_t\}_{t=1}^T$.

1: $P_1 = P$, $S_{0,p} = \varnothing \ \forall p \in P$
2: **for** $t = 1, 2 \dots, T$ **do**
3: Sample $\tilde{f}_{t,p} \sim \mathcal{GP}(\mu_{t,p}, k_{t,p}) \ \forall p \in P_t$
4: Set $x_t, p_t = \arg\max_{x,p \in \mathcal{X} \times P_t} \tilde{f}_{t,p}(x)$
5: $S_{t,p_t} = S_{t-1,p_t} \cup \{t\}$ and $S_{t,p} = S_{t-1,p}$ for $p \in P \setminus \{p_t\}$
6: Observe $y_t = f(x_t) + \epsilon_t$
7: Set $\eta_t = y_t - \mu_{t,p_t}(x_t)$
8: Set $V_t = \sqrt{\xi_t |S_{t,p_t}|} + \sum_{i \in S_{t,p_t}} \sqrt{\beta_i}\sigma_{i,p_t}(x_i)$
9: **if** $\left|\sum_{i \in S_{t,p_t}} \eta_i\right| > V_t$ and $|P_t| > 1$ **then**
10: $P_{t+1} = P_t \setminus \{p_t\}$
11: **else**
12: $P_{t+1} = P_t$

The elimination procedure of PE-GP-TS is illustrated in Fig. 1a. Samples $\tilde{f}_{t,p}$ are drawn from the active prior $p \in P_t$. Then, the unknown function f is queried at the selected arm x_t. If the observed value differs too much from the prediction made by the selected prior, then the selected prior is eliminated. Otherwise, it remains active.

The PE-GP-TS algorithm is presented in Algorithm 1. Similar to PE-GP-UCB, the set $S_{t,p}$ is used to store the time steps where prior p was selected up to and including time t. When prior p_t is selected, the prediction error $\eta_t = y_t - \mu_{t,p_t}(x_t)$ between the observed and predicted value made by the prior p_t is computed. If the sum of prediction errors made by the prior p_t exceeds the threshold value V_t, then p_t is eliminated from the active priors P_t, see line 9. Note that at time step t, only the selected prior p_t can be eliminated. As such, if a prior is very pessimistic it may never be selected and therefore will never be eliminated. Thus, the final set of active priors P_T should be viewed as non-eliminated priors rather than necessarily being reasonable priors.

3.2 HyperPrior Thompson sampling

In our first algorithm, we removed one layer of optimism. The second algorithm we study is a fully Bayesian algorithm that uses a hyperposterior sampling scheme where both the prior and the mean function are sampled from their respective posteriors. By shedding the optimism over the selected prior p_t, HP-GP-TS should be able to avoid costly exploration by selecting likely priors instead of optimistic ones.

Algorithm 2 HyperPrior GP-TS (HP-GP-TS)

input Horizon T, prior functions $\{\mu_{1,p}, k_{1,p}\}_{p \in P}$, hyperprior P_1.

1: **for** $t = 1, 2 \dots, T$ **do**
2: Sample $p_t \sim P_t$
3: Sample $\tilde{f}_t \sim \mathcal{GP}(\mu_{t,p_t}, k_{t,p_t})$
4: Set $x_t = \arg\max_{x \in \mathcal{X}} \tilde{f}_t(x)$
5: Observe $y_t = f(x_t) + \epsilon_t$
6: Set $P_{t+1}(p) \propto \mathbb{P}(y_t | x_t, \{x_i, y_i\}_{i=1}^{t-1}, p) \cdot P_t(p)$ ▷ Update hyperposterior

The algorithm is visualized in Fig. 1b and presented in detail in Algorithm 2. In the first step, the current prior p_t is sampled from the hyperposterior P_t. Then, a single sample $\tilde{f}_t$ is taken from the selected posterior $\mathcal{GP}(\mu_{t,p_t}, k_{t,p_t})$ and is used to select the current arm: $x_t = \arg\max_{x \in \mathcal{X}} \tilde{f}_t(x)$. After observing y_t, the hyperposterior is updated by computing the likelihood of y_t under the different priors. Note that since the set of priors P

is finite, computing the posterior is tractable albeit computationally costly for large t with a naive complexity of $\mathcal{O}(t^3|P|)$ [1]. The algorithm can be extended to continuous priors P using MCMC sampling. In comparison to SCoreBO (Hvarfner et al., 2023) and other fully Bayesian algorithms that compute expected values over the hyperposterior through sampling, HP-GP-TS requires only one sample from the posterior and hyperposterior – potentially reducing the computational cost significantly. The likelihood $\mathbb{P}(y_t|x_t, \{x_i, y_i\}_{i=1}^{t-1}, p) = \mathcal{N}(y_t; \mu_{t,p}(x_t), \sigma^2_{t,p}(x_t) + \sigma^2)$ is simply the Gaussian likelihood of the posterior at x_t with added Gaussian noise with variance σ^2.

4 Regret analysis

In this section, we analyze the regret for the proposed algorithms. Recall from the problem statement that we consider two slightly different settings for the two algorithms. Specifically, for PE-GP-TS we assume the unknown prior p^* is selected arbitrarily from P whilst for HP-GP-TS we assume that the unknown prior p^* is selected from a known hyperprior distribution P_1.

4.1 Analysis of PE-GP-TS

Ziomek et al. (2025) structured the proof of the regret bound for PE-GP-UCB into 4 larger steps; First, showing that p^* is never eliminated with high probability. Second, establishing a bound on the instantaneous regret. Third, bounding the cumulative regret. Finally, the cumulative bound is re-expressed in terms of the worst-case MIG. For PE-GP-TS, we establish a new bound on the instantaneous regret and then adapt the steps of Ziomek et al. to accommodate the new bound. To bound the instantaneous regret in the lemma below, we require concentration inequalities to hold for the posteriors, the posterior samples and the noise (see Lemmas B.1 and B.2).

Lemma 4.1. *If the events of Lemmas B.1 and B.2 holds, then the following holds for the instantaneous regret of PE-GP-TS for all $t \in [T]$: $f(x^*) - f(x_t) \leq 2\sqrt{\beta_t}\sigma_{t,p^*}(x^*) + \sqrt{\beta_t}\sigma_{t,p_t}(x_t) - \eta_t + \epsilon_t$.*

Compared to the instantaneous regret bound for PE-GP-UCB, we obtain the additional term $2\sqrt{\beta_t}\sigma_{t,p^*}(x^*)$ which leads to the following regret bound:

Theorem 4.2. *Let $B_{p^*} = \beta_1 + \sup_{x\in\mathcal{X}} |\mu_{1,p^*}(x)|$ and $C = 2/\log(1 + \sigma^{-2})$. If $p^* \in P$ and $f \sim \mathcal{GP}(\mu_{1,p^*}, k_{1,p^*})$, then PE-GP-TS with confidence parameters $\beta_t = 2\log(2|\mathcal{X}||P|\pi^2 t^2/3\delta)$ and $\xi_t = 2\sigma^2 \log(|P|\pi^2 t^2/3\delta)$, satisfies the following regret bound with probability at least $1 - \delta$:*

$$R(T) \leq 2|P|B_{p^*} + 2\sqrt{\xi_T |P| T} + 2\sqrt{CT\beta_T\hat{\gamma}_T|P|} + 2\sqrt{T\beta_T \textstyle\sum_{t\in[T]} \sigma^2_{t,p^*}(x^*)} \tag{1}$$

The bound of the first three terms is of order $\mathcal{O}(\sqrt{T\beta_T\hat{\gamma}_T})$ w.r.t. T which matches that of PE-GP-UCB. To our knowledge, the best lower bound for standard GP bandits in the Bayesian setting, where f is sampled from a GP, is $\Omega(\sqrt{T})$ for $d = 1$ (Scarlett, 2018). This would suggest that our bound is tight up to a factor $\mathcal{O}(\sqrt{\beta_T\hat{\gamma}_T})$ when considering only the first three terms. However, note that the sublinearity of $\sum_{t\in[T]} \sigma^2_{t,p^*}(x^*)$ is not theoretically demonstrated but we observe that it behaves sublinearly in our experiments, see Fig. 13 in Appendix F.

4.2 Analysis of HP-GP-TS

We analyze the regret of HP-GP-TS by decomposing it into three terms and using the prior confidence technique. The initial regret decomposition is similar to Lu et al. (2023) and the prior confidence technique is first employed by Hong et al. (2022b) in standard and linear settings. However, as we discuss in Section 4.3, both of these works have fundamental issues making their theoretical analyses invalid.

First, note that HP-GP-TS inherits the probability matching property of GP-TS that $x_t|H_t \overset{d}{=} x^*|H_t$ where $\overset{d}{=}$ denotes equal in distribution. In addition, $p_t|H_t \overset{d}{=} p^*|H_t$ since p_t is sampled from the posterior distribution of p^*. Using this, one can derive the following decomposition of the regret:

$$\text{BR}(T) = \sum_{t\in[T]} \mathbb{E}[\underbrace{f(x^*) - U_{t,p^*}(x^*)}_{(1)} + \underbrace{(\sqrt{\beta_t} + \sqrt{\eta_T})\sigma_{t,p_t}(x_t)}_{(2)} + \underbrace{L_{t,p_t}(x_t) - f(x_t)}_{(3)}] \tag{2}$$

[1] The per-timestep complexity of computing the posterior can be reduced to $\mathcal{O}(t^2|P|)$ by using rank-one updates.

where the upper confidence bound $U_{t,p}(x) = \mu_{t,p}(x) + \sqrt{\beta_t}\sigma_{t,p}(x)$ and the lower confidence bound $L_{t,p}(x) = \mu_{t,p}(x) - \sqrt{\eta_T}\sigma_{t,p}(x)$. Term (1) can be bounded using the same steps as for standard GP-TS since the confidence bound $U_{t,p^*}(x^*)$ uses the true prior p^*. The key question for term (2) is whether a tight bound for $\sum_{t\in[T]}\sigma^2_{t,p_t}(x_t)$ can be obtained. Ziomek et al. (2025) provides the bound $\sum_{p\in P}\gamma_{N_T(p),p}$ as an intermediate step in the proof of Lemma 5.3 where $N_T(p)$ is the number of times prior p is selected in total. Due to the nature of PE-GP-UCB (and similarly for PE-GP-TS) the only guarantee on $N_T(p)$ is that it is smaller than T, thus the bound $\sum_{p\in P}\gamma_{T,p}$ is used. However, we show in Lemma 4.3 that a tighter bound can be obtained for HP-GP-TS, thereby improving the dependency upon the MIG compared to the bound of Lu et al. (2023). Under a Bayesian model, we show that $\mathbb{E}[N_T(p)] = P_1(p)T$ for HP-GP-TS and by a concavity argument we provide a bound in terms of $\bar{\gamma}_T(P_1) := \sum_{p\in P}\Gamma_p(P_1(p)T)$ where $\Gamma_p(\cdot)$ is a continuous upper bound of $\gamma_{\cdot,p}$.

Lemma 4.3. *Let $C = 2/\log(1+\sigma^{-2})$, $N_T(p) = \sum_{t\in[T]}\mathbb{1}\{p_t = p\}$ where $\mathbb{1}$ is the indicator function, and $\Gamma_p : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}_{\geq 0}$ be a concave function such that $\Gamma_p(t) \geq \gamma_{t,p}$ for all $t, p \in [T]\times P$. Furthermore, let $\bar{\gamma}_T(P_1) := \sum_{p\in P}\Gamma_p(P_1(p)T)$, then for HP-GP-TS, $\mathbb{E}[N_T(p)] = P_1(p)T$ and*

$$\mathbb{E}\Big[\sum_{t\in[T]}\sigma^2_{t,p_t}(x_t)\Big] \leq C\bar{\gamma}_T(P_1). \tag{3}$$

To bound term (3), we define the excess reward function as $G_t(p) = \sum_{s=1}^{t-1}\mathbb{1}\{p_s = p\}\left(\mu_{s,p_s}(x_s) - \sqrt{\eta}\sigma_{s,p_s} - f(x_s) - \epsilon_s\right)$ for $\eta > 0$, similar to Lu et al. (2023). Then, we define the confidence set at time t as $\mathcal{C}_t = \{p \in P : G_s(p) \leq \xi_s(p)\ \forall s \leq t\}$ where $\xi_t(p) = \sigma\sqrt{12N_{t-1}(p)\log(T)}$ where $N_t(p) = \sum_{s=1}^{t}\mathbb{1}\{p_s = p\}$ denotes how often the prior p was selected up to and including time t. Unlike Hong et al. (2022b); Lu et al. (2023), we impose a time-uniform requirement, i.e. a prior $p \in \mathcal{C}_t$ only if $p \in \mathcal{C}_s$ for all $s < t$, as we found their proofs uncompelling without this requirement, see Remark B.10. We show that $p^* \in \mathcal{C}_t$ with high probability in Lemma B.7 and split term (3) into two new terms:

$$(3) = \sum_{t\in[T]}\mathbb{E}\left[(L_{t,p_t}(x_t) - f(x_t))\mathbb{1}\{p_t \notin \mathcal{C}_t\}\right] + \sum_{t\in[T]}\mathbb{E}\left[(L_{t,p_t}(x_t) - f(x_t))\mathbb{1}\{p_t \in \mathcal{C}_t\}\right]. \tag{4}$$

Since $\mathcal{C}_t$ is defined to only consider the excess reward in the past, the right term can only be bounded up to the stopping time τ_p for each prior $p \in P$. The main hurdle is therefore to bound the expectation of the stopped excess reward for each prior $\mathbb{E}[L_{\tau_p,p}(x_{\tau_p}) - f(x_{\tau_p}) - \epsilon_{\tau_p}]$. Hong et al. (2022b); Lu et al. (2023) provide incorrect bounds for this term in the linear and GP setting respectively, see Section 4.3. We first note that the stopped value of this sequence can be bounded by the maximum over the same sequence and then provide a bound for $\mathbb{E}[\max_{t\in[T]}(L_{t,p}(x_t) - f(x_t) - \epsilon_t)]$. For the left term, we know that $\mathbb{E}[\mathbb{1}\{p_t \notin \mathcal{C}_t\}] = \mathbb{P}(p^* \notin \mathcal{C}_t) = \mathcal{O}(T^{-5})$ but the factor $L_{t,p_t}(x_t) - f(x_t)$ prevents direct application of this result. Again, the bounds provided by Hong et al. (2022b); Lu et al. (2023) do not hold. We make the observation that the two factors can be separated by the Cauchy-Schwarz inequality for expected values ($\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$) and provide bounds for $\mathbb{E}[\mu^2_{t,p_t}(x_t)]$ and $\mathbb{E}[f(x_t)^2]$ in Lemmas B.11 and B.12. Finally, we are ready to state our regret bound for HP-GP-TS.

Theorem 4.4. *Let $C = 2/\log(1+\sigma^{-2})$, $\mu_{max} = \sup_{p,x\in P\times\mathcal{X}}|\mu_{1,p}(x)|$, $M = \mathbb{E}[\sup_{x\in\mathcal{X}}|f(x)|]$, $M_p = \mathbb{E}\left[\sup_{x\in\mathcal{X}}|f(x)|\middle|p^* = p\right]$, $M_\Delta = \max_{p\in P}M_p - \min_{p\in P}M_p$, and $\bar{M} = M^2 + 1 + M^2_\Delta/4$. If $p^* \sim P_1$, $f \sim \mathcal{GP}(\mu_{1,p^*}, k_{1,p^*})$, $\beta_t = 2\log(|\mathcal{X}|t^2/\sqrt{2\pi})$, $\eta_T = 2\log|\mathcal{X}|T^6$, then the Bayesian regret of HP-GP-TS is bounded by*

$$\begin{aligned} BR(T) \leq\ & \frac{\pi^2}{6} + \sqrt{CT\bar{\gamma}_T(P_1)}(\sqrt{\beta_T} + \sqrt{\eta_T}) && \text{(Terms (1) and (2))} \\ & + \tfrac{\sqrt{3}}{T}\left(\sqrt{\sigma^{-2}(\bar{M} + 2M\mu_{max} + \mu^2_{max} + \sigma^2)} + \sqrt{\bar{M}/T}\right) \\ & + \sigma\sqrt{14T|P|\log T} + |P|\left(\sigma^{-1}\sqrt{T}(M + \mu_{max} + \sigma) + M + \sigma\sqrt{2\log T}\right) \end{aligned} \Bigg\} \text{(Term (3))} \tag{5}$$

Note that M_p is the expected supremum of $|f(x)|$ given $p^* = p$ whereas M is the expected supremum of $|f(x)|$ for the mixture $p^* \sim P_1$. Furthermore, M_Δ denotes the spread in expected supremums and $\bar{M}$ bounds the expectation of the squared supremum $\sup_x f(x)^2$, see Lemma B.12. Unlike PE-GP-TS, -UCB, and Lu et al. (2023), our regret bound of HP-GP-TS depends on the hyperprior-weighted

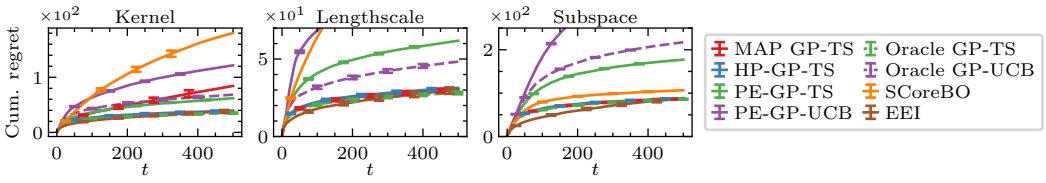


Figure 2: Cumulative regret for synthetic experiments with varying kernel, lengthscale and active subspace. The final regret for PE-GP-UCB is 114 and 389 in the lengthscale and subspace experiments, and 181 for SCoreBO in the lengthscale experiment. Errorbars correspond to ± 1 standard error.

MIG $\bar{\gamma}_T(P_1)$ rather than the worst case $|P|\hat{\gamma}_T$ which can impact the theoretical regret significantly if the complexity of the priors differ and the hyperprior is weighted towards simple priors. This is reasonable since the elimination methods assume arbitrary selection of p^* as opposed to sampling from a hyperprior. The final term in Eq. (5) is $\mathcal{O}(|P|\sqrt{T})$ whereas the term for PE-GP-TS and -UCB that is linear in $|P|$ is constant w.r.t. T. In Section 5, we empirically evaluate the dependency on $|P|$.

4.3 Comparison to MixTS and EGP-TS

Hong et al. (2022b) study MixTS, a Thompson sampling algorithm that assumes the prior is a mixture distribution, for standard and linear bandits. For the linear setting with unbounded rewards, the proof requires conditioning on the linear parameter vector θ^* to lie close to its prior mean. Under this additional event E_0, the distribution of the true parameter vector θ^*, the true prior p^* and the optimal arm x^* can shift. But, conditioned on the history H_t, MixTS is unaffected by conditioning on E_0 at time step t. Consequently, the sampled parameter vector θ_t, the selected prior p_t and the selected arm x_t maintain the same distribution. However, Hong et al. (2022b) use that $x^*, p^*|H_t, E_0 \overset{d}{=} x_t, p_t|H_t, E_0$ without proof, invalidating Theorem 1 of Hong et al. (2022b). The event E_0 bounds the maximum per-round regret by a constant, enabling $L_{t,p}(x_t) - f(x_t)$ to be conveniently bound by a constant for both terms in Eq. (4). Unfortunately, the intermediate steps contain other issues that we discuss further in Appendix C. Lu et al. (2023) study EGP-TS for sequential and parallel GP-bandit problems. Lemma 5 of Lu et al. (2023) bounds $\mathbb{E}[\mu_{t,p_t}(x_t) - f(x_t)]$ by a constant $2B$ with an incorrect proof. Even assuming a correct proof, the lemma is applied incorrectly to claim that $\mathbb{E}[(\mu_{t,p_t}(x_t) - f(x_t))\mathbb{1}\{p_t \notin \mathcal{C}_t\}] \leq \mathbb{E}[2B\mathbb{1}\{p_t \notin \mathcal{C}_t\}]$ and $\mathbb{E}[L_{\tau,p}(x_\tau) - f(x_\tau) - \epsilon_\tau] \leq 2B$. Our proof avoids the issues in previous work by separating the event $\mathbb{1}\{p_t \notin \mathcal{C}_t\}$ from the excess reward using the Cauchy-Schwarz inequality for expectations, bounding the stopped excess reward by the maximum excess reward, and bounding the expected values of $\mu_{t,p_t}(x_t)$, $\mu^2_{t,p}(x_t)$ and $f(x_t)^2$.

5 Experiments

In this section, we describe our experiments based on synthetic and real-world data. The code for all of the experiments is available at https://github.com/MachineLearning-DecisionMaking/Adaptive-Prior-Selection-GP-Bandits.

Synthetic experiments We consider three synthetic setups with different choices of priors in P. For the first setup, the priors have one of the following kernels: i) RBF kernel, ii) the rational quadratic kernel with $\alpha = 0.5$, iii) Matérn kernel with $\nu = 5/2$, iv) Matérn kernel with $\nu = 3/2$, v) periodic kernel with period $\rho = 5$, vi) linear kernel with $v = 0.05^2$. For the second setup, 8 priors use the RBF kernel with different lengthscales equidistantly spaced between $1/2$ and 4. For the third setup, the total dimensions $d = 16$ but each of the 5 priors p_i assumes $f(x)$ depends on $d_s = 4$ subdimensions. The 4 subdimensions are designed such that the priors are equally difficult to distinguish. All priors use the RBF kernel with lengthscale $\ell = 8$. For all three setups, the true prior p^* is sampled uniformly from P, the noise variance $\sigma^2 = 0.25^2$, and the horizon $T = 500$. For the first two setups, 500 arms are equidistantly spaced in $[0, 20]$ and for the third 500 arms are sampled uniformly on $[0, 20]^{16}$. All models are evaluated on 500 seeds on each setup. As baselines, we use PE-GP-UCB, SCoreBO (Hvarfner et al., 2023), fully Bayesian Expected Improvement (EEI) (Benassi et al., 2011) and Maximum A Posteriori (MAP) GP-TS. MAP GP-TS is identical to HP-GP-TS except for greedily

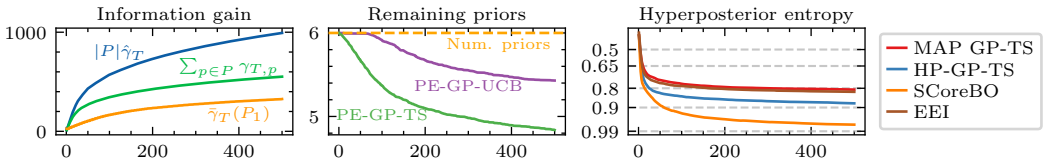


Figure 3: Analysis of the kernel experiment. Greedily maximal information gain (left). Mean number of priors remaining in P_t over time for PE-GP-UCB and -TS (middle). Entropy in the hyperposterior P_t over time for HP- and MAP GP-TS (right). The dashed reference lines correspond to entropies of discrete distributions with prob. q on one choice and prob. $\frac{1-q}{|P|-1}$ on the other $|P|-1$ choices.

selecting p_t from the posterior: $p_t = \arg\max_p P_t(p)$.[2] In addition, we compare against the oracle variants of PE-GP-TS and PE-GP-UCB that are only given the true prior: $P_1 = \{p^*\}$.

The cumulative regret for the three synthetic experiments is shown in Fig. 2 and the final regret is shown in Table 2 in Appendix F. Across all three experiments, we observe that HP-GP-TS and EEI has lower regret than the other methods and performs close to the oracle GP-TS. For the kernel and subspace experiments, PE-GP-TS has lower regret than the oracle GP-UCB. Hence, even if PE-GP-UCB was optimized to perform as well as the oracle, it would still not achieve the regret of the TS methods. MAP GP-TS has slightly lower regret than HP-GP-TS for the lengthscale and subspace experiments but has significantly higher regret and variance for the kernel experiment. The greedy selection of MAP (MLE) leads to under-exploration for MAP GP-TS in certain instances. SCoreBO has the highest regret in the kernel and lengthscale experiment but has more comparable performance in the subspace experiment. In Fig. 10 in Appendix F, we report the regret for the two most competitive methods, HP-GP-TS and EEI, with an extended horizon $T = 1500$ where we observe that HP-GP-TS yields noticeably lower regret.

The maximum information gain, the number of priors remaining $|P_t|$ and the hyperposterior entropy for the kernel experiment is shown in Fig. 3. We note that $\bar{\gamma}_T(P_1)$ is significantly smaller than $|P|\hat{\gamma}_T$. The PE-methods eliminate at most one prior on average. In contrast, the final hyperposterior entropy across all algorithms is equivalent to 70-99% of the probability mass being assigned to one prior showing that the hyperposterior adapts more effectively. Across the experiments, SCoreBO has the lowest hyperposterior entropy followed by HP-GP-TS and EEI has the highest (except for the kernel experiment), see Figs. 11 and 12 in Appendix F. Thus, HP-GP-TS has similar regret to EEI but lower hyperposterior entropy.

In Fig. 4, we visualize how often the methods select the true prior p^* (or kernel) in the kernel experiment as confusion matrices. PE-GP-UCB selects the Matérn-3/2 kernel more than 96% of the rounds. The Matérn-3/2 kernel induces a distribution over functions that are less smooth compared to the other kernels and produces much wider confidence intervals outside the observed data leading to excessive optimistic exploration. PE-GP-TS also shows a bias towards the Matérn-3/2 kernel but does not select it as frequently as PE-GP-UCB – demonstrating that one layer of optimism has been removed. The overall "accuracy" of the selected priors, i.e. $\sum_{t\in[T]}\mathbb{1}\{p_t = p^*\}/T$, for the elimination-based methods is around 17% in the kernel experiment compared to 62.9% and 63.2% for MAP and HP-GP-TS respectively. For HP-GP-TS, we observe that it can easily identify the periodic and linear kernels. However, the RBF, Matérn and RQ kernels are often confused with each other. These kernels do not have as easily distinguishable characteristics and are likely to produce similar posteriors even with a small amount of data. See Fig. 14 in Appendix F for confusion matrices in the lengthscale and subspace experiments.

Scaling $|P|$ We perform two experiments to understand how the regret of our algorithms scale with the number of priors. In both experiments, the average difficulty of the problem is kept constant such that the regret of the oracle models is constant. In the first experiment, we increase the discretization of the lengthscale values. The lengthscales are equidistantly spaced in $[0.5, 4]$ with $|P| \in \{8, 16, 32, 64, 128\}$. As $|P|$ increases, the difference between similar priors is reduced. In the second experiment, we increase the number of priors in the subspace experiment from 5 up to 16. Each prior can share at most 3 out of 4 dimensions with other priors which ensures the priors remain meaningfully different. The total regret as the number of priors increases is shown in Fig. 5. For the

[2]Note that since the hyperprior is uniform, MAP is equivalent to discrete maximum likelihood estimation.

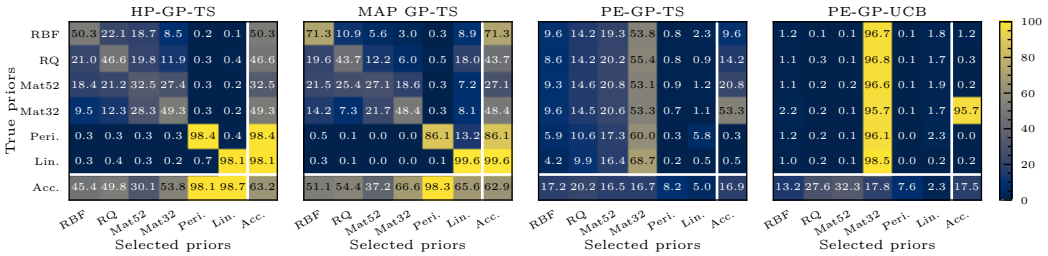


Figure 4: Confusion matrices for the true prior p^* and the selected priors p_t for the kernel experiment.

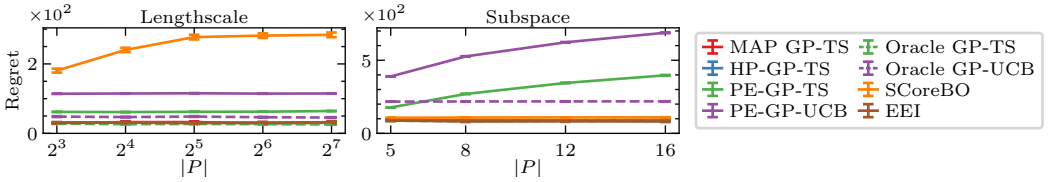


Figure 5: Total regret for the lengthscale and subspace experiments as $|P|$ increases.

lengthscale experiment, increasing the number of priors above 8 does not affect the regret for any algorithm, likely due to the increased redundancy in the priors. The one exception is SCoreBO, whose regret increases as $|P|$ is increased to 32 but levels off beyond that. In the subspace experiment, the regret of the prior elimination algorithms scales approximately as $\sqrt{|P|}$ whilst MAP- and HP-GP-TS are consistently close to the constant regret of the oracle. The regret of EEI drops initially but is otherwise constant and SCoreBO also has constant regret.

Misspecification of P_1 We evaluate the behaviour of HP-GP-TS under misspecification of the true hyperprior P_1 on the synthetic lengthscale experiment. We consider three hyperpriors: $P_{\text{short}} \propto [255, 64, 32, 16, 8, 4, 2, 1]$, $P_{\text{unif}} \propto [1, 1, 1, 1, 1, 1, 1, 1]$, and $P_{\text{long}} \propto [1, 2, 4, 8, 16, 32, 64, 128]$. The short (long) prior is heavily weighted towards short (long) lengthscales, assigning $> 75\%$ of the probability mass towards the two shortest (longest) lengthscales. We evaluate HP-GP-TS with hyperprior $P \in \{P_{\text{short}}, P_{\text{unif}}, P_{\text{long}}\}$ against the true hyperprior $Q \in \{P_{\text{short}}, P_{\text{unif}}, P_{\text{long}}\}$ and compare against Oracle TS. The cumulative regret for the misspecification experiment is shown in Fig. 6 and the total regret is reported in Table 6 in Appendix F. Aside from Oracle GP-TS, HP-GP-TS with the true hyperprior ($P = Q$) has the lowest regret in all three settings but not by a large margin. For $Q = P_{\text{unif}}$, we observe that HP-GP-TS with either of the misspecified hyperpriors is within 1 standard error of HP-GP-TS with the true hyperprior.

Real-world data We perform three experiments with real-world data from the Intel Berkeley dataset (Madden et al., 2004), California Performance Measurement System (PeMS) (Chen et al., 2001; California Department of Transportation, 2024) and Pacific Northwest (PNW) daily precipitation dataset (Widmann & Bretherton, 1999, 2000). Each dataset contains measurements from a set of sensors over time. We split each dataset into a training and test set where the test set contains the last third of the data. Hence, the distribution of the test data may have shifted from the training data allowing us to test a realistic setting where the true prior is unknown but we have a set of reasonable priors. Each training set is further split into separate buckets which we use to estimate the empirical mean and covariance of the priors. Before every run, a set of measurements f is selected randomly from the test set. During the run, these measurements are held fixed and do not vary with time t. Note

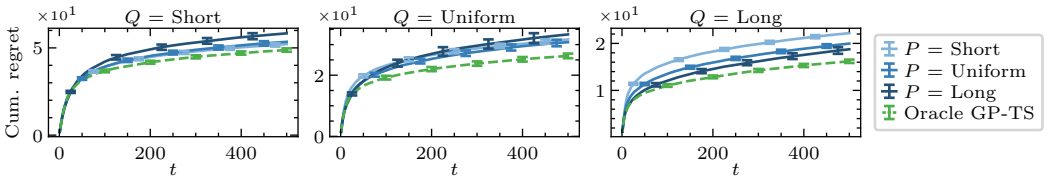


Figure 6: Regret of HP-GP-TS with misspecified hyperprior on the lengthscale. The true hyperior is denoted by Q and the assumed hyperprior is P.

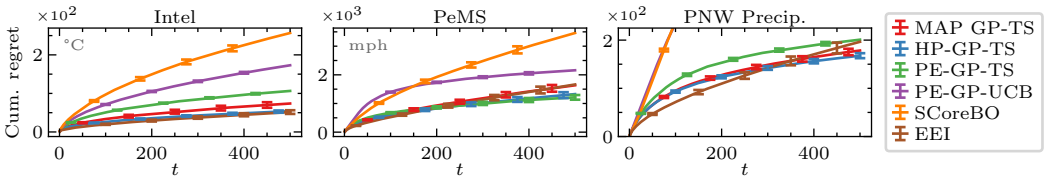


Figure 7: Cumulative regret on the real-world data experiments. Errorbars correspond to ± 1 standard error. The average final regret for SCoreBO and PE-GP-UCB is 861 and 506 on PNW.

that this differs from the time-varying experiments of Bogunovic et al. (2016); Ziomek et al. (2025) where the measurements f_t vary over time. See Appendix E for more details.

The cumulative regret for the experiments with real-world data is presented in Fig. 7. Across these experiments, HP-GP-TS has either the lowest regret or is within 1 standard error of the algorithm with the lowest regret. SCoreBO has significantly higher regret than all other methods. Notably for the PeMS data, PE-GP-TS has the lowest average regret whereas MAP GP-TS and EEI perform worse compared to the other experiments.

The number of priors remaining in $|P_t|$ and the hyperposterior entropy for the real-world data experiments is shown in Figs. 11 and 12 in Appendix F. Similar to the synthetic experiments, on average, the prior elimination methods eliminate less than 1 prior at best and no priors (across all 500 seeds) at worst. In contrast, the hyperposterior of HP-GP-TS concentrates to the equivalent of 60-80% of the probability mass to one prior. The relative standing in terms of reduced hyperposterior uncertainty between SCoreBO, HP-GP-TS and EEI remains consistent across all the experiments. SCoreBO reduces the hyperposterior uncertainty the most at the cost of significantly higher regret whereas HP-GP-TS provides a better balance between low regret and low hyperposterior uncertainty.

6 Conclusion

In this paper, we have studied two algorithms for adaptive prior selection and regret minimization in GP bandits based on GP-TS. We have analyzed the algorithms theoretically, corrected and improved upon previous work, and experimentally evaluated both algorithms on synthetic and real-world data. We find that lowering the amount of optimistic exploration leads the algorithms to obtain lower or comparable regret than previous work.

Acknowledgments and Disclosure of Funding

The work of Jack Sandberg and Morteza Haghir Chehreghani was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

References

Abbasi-Yadkori, Y., Pacchiano, A., and Phan, M. Regret Balancing for Bandit and RL Model Selection, June 2020. URL https://arxiv.org/abs/2006.05491.

Åkerblom, N., Chen, Y., and Haghir Chehreghani, M. Online Learning of Energy Consumption for Navigation of Electric Vehicles. *Artificial Intelligence*, 317:103879, April 2023a. doi: 10.1016/j.artint.2023.103879.

Åkerblom, N., Hoseini, F. S., and Haghir Chehreghani, M. Online Learning of Network Bottlenecks via Minimax Paths. *Machine Learning*, 112(1):131–150, January 2023b. doi: 10.1007/s10994-022-06270-0.

Balandat, M., Karrer, B., Jiang, D. R., Daulton, S., Letham, B., Wilson, A. G., and Bakshy, E. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. In *Advances in Neural Information Processing Systems 33*, 2020. URL http://arxiv.org/abs/1910.06403.

Basu, S., Kveton, B., Zaheer, M., and Szepesvari, C. No Regrets for Learning the Prior in Bandits. In *Advances in Neural Information Processing Systems*, volume 34, pp. 28029–28041. Curran Associates, Inc., 2021.

Benassi, R., Bect, J., and Vazquez, E. Robust Gaussian Process-Based Global Optimization Using a Fully Bayesian Expected Improvement Criterion. In Coello, C. A. C. (ed.), *Learning and Intelligent Optimization*, volume 6683, pp. 176–190. Springer Berlin, Heidelberg, 2011. ISBN 978-3-642-25565-6.

Berkenkamp, F., Schoellig, A. P., and Krause, A. No-Regret Bayesian Optimization with Unknown Hyperparameters. *Journal of Machine Learning Research*, 20(50):1–24, 2019. ISSN 1533-7928.

Bogunovic, I., Scarlett, J., and Cevher, V. Time-Varying Gaussian Process Bandit Optimization. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, pp. 314–323. PMLR, May 2016. ISSN: 1938-7228.

Boucheron, S., Lugosi, G., and Massart, P. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, February 2013. ISBN 978-0-19-953525-5. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL https://academic.oup.com/book/26549.

California Department of Transportation. Caltrans Performance Measurement System, 2024. URL https://pems.dot.ca.gov/.

California Department of Transportation. Caltrans Terms of Use, 2026. URL https://pems.dot.ca.gov/?dnode=Help&content=help_tou#ownership.

Chapelle, O. and Li, L. An Empirical Evaluation of Thompson Sampling. In *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

Chen, C., Petty, K., Skabardonis, A., Varaiya, P., and Jia, Z. Freeway Performance Measurement System: Mining Loop Detector Data. *Transportation Research Record*, 1748(1):96–102, January 2001. doi: 10.3141/1748-12.

De Ath, G., Everson, R. M., and Fieldsend, J. E. How Bayesian should Bayesian optimisation be? In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '21, pp. 1860–1869, New York, USA, July 2021. Association for Computing Machinery. ISBN 978-1-4503-8351-6.

Gardner, J. R., Pleiss, G., Weinberger, K. Q., Bindel, D., and Wilson, A. G. GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Gonzalvez, J., Lezmi, E., Roncalli, T., and Xu, J. Financial Applications of Gaussian Processes and Bayesian Optimization, 2019. URL https://arxiv.org/abs/1903.04841.

Hernández-Lobato, J. M., Hoffman, M. W., and Ghahramani, Z. Predictive Entropy Search for Efficient Global Optimization of Black-box Functions. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

Hong, J., Kveton, B., Zaheer, M., and Ghavamzadeh, M. Hierarchical Bayesian Bandits. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, pp. 7724–7741. PMLR, May 2022a.

Hong, J., Kveton, B., Zaheer, M., Ghavamzadeh, M., and Boutilier, C. Thompson Sampling with a Mixture Prior. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, pp. 7565–7586. PMLR, May 2022b.

Hvarfner, C., Hellsten, E., Hutter, F., and Nardi, L. Self-Correcting Bayesian Optimization through Bayesian Active Learning. *Advances in Neural Information Processing Systems*, 36:79173–79199, December 2023.

Kandasamy, K., Krishnamurthy, A., Schneider, J., and Poczos, B. Parallelised Bayesian Optimisation via Thompson Sampling. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, pp. 133–142. PMLR, March 2018.

Krause, A., Singh, A., and Guestrin, C. Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies. *Journal of Machine Learning Research*, 9(8): 235–284, 2008. ISSN 1533-7928.

Kveton, B., Konobeev, M., Zaheer, M., Hsu, C.-W., Mladenov, M., Boutilier, C., and Szepesvari, C. Meta-Thompson Sampling. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 5884–5893. PMLR, July 2021.

Li, H., Liang, D., and Xie, Z. Modified Meta-Thompson Sampling for Linear Bandits and Its Bayes Regret Analysis, September 2024. URL https://arxiv.org/abs/2409.06329.

Lu, Q., Polyzos, K. D., Li, B., and Giannakis, G. B. Surrogate Modeling for Bayesian Optimization Beyond a Single Gaussian Process. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):11283–11296, September 2023. ISSN 1939-3539. doi: 10.1109/TPAMI.2023.3264741. URL https://ieeexplore.ieee.org/abstract/document/10093035.

Mackay, D. J. Introduction to Gaussian processes. In *NATO ASI Series. Series F : Computer and System Sciences*, pp. 133–165, 1998. ISBN 978-3-540-64928-1.

Madden, S. et al. Intel lab data, 2004. URL https://db.csail.mit.edu/labdata/labdata.html.

Matérn, B. Spatial Variation. In Brillinger, D., Fienberg, S., Gani, J., Hartigan, J., and Krickeberg, K. (eds.), *Spatial Variation*, volume 36 of *Lecture Notes in Statistics*. Springer, New York, 1986. ISBN 978-0-387-96365-5.

Mockus, J. On Bayesian Methods for Seeking the Extremum. In Marchuk, G. I. (ed.), *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974*, pp. 400–404, Berlin, Heidelberg, 1975. Springer. ISBN 978-3-540-37497-8.

Nuara, A., Trovò, F., Gatti, N., and Restelli, M. A Combinatorial-Bandit Algorithm for the Online Joint Bid/Budget Optimization of Pay-per-Click Advertising Campaigns. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), April 2018. doi: 10.1609/aaai.v32i1.11888.

Osborne, M. A., Garnett, R., and Roberts, S. J. Gaussian Processes for Global Optimization. In *3rd International Conference on Learning and Intelligent Optimization (LION3)*, pp. 1–15, 2009.

Pacchiano, A., Dann, C., Gentile, C., and Bartlett, P. Regret Bound Balancing and Elimination for Model Selection in Bandits and RL, December 2020. URL https://arxiv.org/abs/2012.13045.

Pleiss, G., Gardner, J. R., Balandat, M., and Ament, S. Linear_Operator: Structured linear algebra in pytorch. PyTorch Conference Poster, 2022. URL https://pytorch.s3.amazonaws.com/posters/ptc2022/B05.pdf.

Pleiss, G., Gardner, J. R., Balandat, M., et al. LinearOperator, 2025. URL https://github.com/cornellius-gp/linear_operator.

Russo, D. and Van Roy, B. Learning to Optimize via Posterior Sampling. *Mathematics of Operations Research*, April 2014. doi: 10.1287/moor.2014.0650.

Sandberg, J. and Haghir Chehreghani, M. Comments on "Surrogate Modeling for Bayesian Optimization Beyond a Single Gaussian Process". Under review, 2026.

Scarlett, J. Tight Regret Bounds for Bayesian Optimization in One Dimension. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 4500–4508. PMLR, July 2018.

Snoek, J., Larochelle, H., and Adams, R. P. Practical Bayesian Optimization of Machine Learning Algorithms. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. W. Information-Theoretic Regret Bounds for Gaussian Process Optimization in the Bandit Setting. *IEEE Transactions on Information Theory*, 58(5):3250–3265, May 2012. doi: 10.1109/TIT.2011.2182033.

Turner, R., Eriksson, D., McCourt, M., Kiili, J., Laaksonen, E., Xu, Z., and Guyon, I. Bayesian Optimization is Superior to Random Search for Machine Learning Hyperparameter Tuning: Analysis of the Black-Box Optimization Challenge 2020. In *Proceedings of the NeurIPS 2020 Competition and Demonstration Track*, pp. 3–26. PMLR, August 2021.

Vakili, S., Khezeli, K., and Picheny, V. On Information Gain and Regret Bounds in Gaussian Process Bandits. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, pp. 82–90. PMLR, March 2021.

Wang, Z. and de Freitas, N. Theoretical Analysis of Bayesian Optimisation with Unknown Gaussian Process Hyper-Parameters, June 2014. URL https://arxiv.org/abs/1406.7758.

Wang, Z. and Jegelka, S. Max-value Entropy Search for Efficient Bayesian Optimization. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 3627–3635. PMLR, July 2017. ISSN: 2640-3498.

Wen, Z., Kveton, B., and Ashkan, A. Efficient Learning in Large-Scale Combinatorial Semi-Bandits. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 1113–1122. PMLR, June 2015.

Widmann, M. and Bretherton, C. S. "50" km resolution daily precipitation for the Pacific Northwest, 1949-94, May 1999. URL http://research.jisao.washington.edu/data/widmann/.

Widmann, M. and Bretherton, C. S. Validation of Mesoscale Precipitation in the NCEP Reanalysis Using a New Gridcell Dataset for the Northwestern United States. *Journal of Climate*, 13(11): 1936–1950, June 2000. ISSN 0894-8755, 1520-0442.

Williams, C. K. and Rasmussen, C. E. *Gaussian Processes for Machine Learning*, volume 2. MIT Press Cambridge, MA, 2006.

Ziomek, J., Adachi, M., and Osborne, M. A. Bayesian Optimisation with Unknown Hyperparameters: Regret Bounds Logarithmically Closer to Optimal. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, November 2024.

Ziomek, J., Adachi, M., and Osborne, M. A. Time-varying Gaussian Process Bandits with Unknown Prior. In *The 28th International Conference on Artificial Intelligence and Statistics*, February 2025.

A Extended discussion of related work

Plenty of previous work has proposed fully Bayesian approaches that integrate the acquisition function over the hyperposterior (Osborne et al., 2009; Benassi et al., 2011; Snoek et al., 2012; Hernández-Lobato et al., 2014; Wang & Jegelka, 2017; De Ath et al., 2021). A difficulty with such approaches is that to compute the expected acquisition function they must perform costly MCMC sampling over the hyperposterior. In contrast, HP-GP-TS optimizes a single hyperposterior sample instead of computing expected values over the hyperposterior. Hvarfner et al. (2023) proposed Self-Correcting Bayesian Optimization (SCoreBO) whose objective function balances reducing the uncertainty of (x^*, f^*) and reducing the uncertainty of the true prior p^*. Notably, SCoreBO explicitly tries to identify the prior rather than integrating out the uncertainty of the prior.

Wang & de Freitas (2014) first derived regret bounds for GP bandits with unknown lengthscale for the Expected Improvement algorithm (Mockus, 1975). However, the proposed algorithm requires a lower bound on the lengthscale and the regret bound depends on the worst-case MIG. Later work by Berkenkamp et al. (2019) introduced Adaptive GP-UCB (A-GP-UCB) that continually lowers the lengthscale parameter. Given a sufficiently small lengthscale, the function f lies within the reproducing kernel Hilbert space (RKHS) and the regular GP-UCB theory can be applied. However, A-GP-UCB lacks a stopping mechanism and will overexplore as the lengthscale continues to shrink. Recent work by Ziomek et al. (2025) introduced Prior-Elimination GP-UCB (PE-GP-UCB) for time-varying GP-bandits with unknown prior. Unlike the work before, the regret bound of PE-GP-UCB holds for arbitrary types of hyperparameters in the GP prior. PE-GP-UCB is doubly optimistic and selects the prior *and* arm with the highest upper confidence bound. PE-GP-UCB tracks the cumulative prediction error made by the selected priors and eliminates priors that exceed a threshold level.

Other works have introduced regret balancing algorithms that maintain a set of base learning algorithms and balance their selection frequency to achieve close to optimal regret (Abbasi-Yadkori et al., 2020; Pacchiano et al., 2020). Ziomek et al. (2024) built on this idea and introduced length-scale balancing GP-UCB which can adaptively explore smaller lengthscales but can return to longer ones, unlike A-GP-UCB.

In addition to Hong et al. (2022b); Lu et al. (2023), another line of work has studied Thompson sampling in standard and linear bandits with unknown prior distribution (Kveton et al., 2021; Basu et al., 2021; Hong et al., 2022a; Li et al., 2024). In their setting (meta or hierarchical bandits), the agent plays multiple bandit instances, either simultaneously or sequentially. The unknown means are sampled from the same (unknown) prior and by gathering knowledge across instances, the agent can solve later instances more efficiently once it has identified the prior. In contrast, in this paper we consider the setting where the agent can only access information from the instance it is facing.

B Proofs

In the following section, we state and prove the results shown in the main text.

B.1 PE-GP-TS

First, we state and prove concentration inequalities for $f(x)$ and $\tilde{f}_{t,p}(x)$. Lemma B.1 is based on Lemma 5.1 of Srinivas et al. (2012) but adapted to TS by specifying that it holds for any sequence of $x_1, \dots, x_T$, as discussed by Russo & Van Roy (2014). Additionally, we add Eq. (7) which can be shown through the same steps and an additional union bound over P.

Lemma B.1. *If $f(x) \sim \mathcal{GP}(\mu_{1,p^*}, k_{1,p^*})$ and $\beta_t = 2\log\left(\frac{|\mathcal{X}||P|\pi^2 t^2}{3\delta}\right)$. Then, with probability at least $1-\delta$, the following holds for all $t, x, p \in [T] \times \mathcal{X} \times P$:*

$$|f(x) - \mu_{t,p^*}(x)| \leq \sqrt{\beta_t}\sigma_{t,p^*}(x), \tag{6}$$

$$|\tilde{f}_{t,p}(x) - \mu_{t,p}(x)| \leq \sqrt{\beta_t}\sigma_{t,p}(x). \tag{7}$$

Proof. Follows by the same steps as Lemma 5.1 of Srinivas except we condition on the complete history H_t instead of only $\mathbf{y}_{1:t-1}$. Additionally, for Eq. (7) we must take an additional union bound over $p \in P$.

Fix $t, x, p \in [T] \times \mathcal{X} \times P$. Given the history H_t, $\tilde{f}_{t,p}(x) \sim \mathcal{N}(\mu_{t,p}(x), \sigma^2_{t,p}(x))$. Using that $\mathbb{P}(Z > c) \leq 1/2e^{-c^2/2}$ for $Z \sim \mathcal{N}(0,1)$, we get that

$$\mathbb{P}\left(\left|\frac{\tilde{f}_{t,p}(x) - \mu_{t,p}(x)}{\sigma_{t,p}(x)}\right| > \sqrt{\beta_t}\right) \leq \exp(-\beta_t/2) \tag{8}$$

$$= \frac{3\delta}{|\mathcal{X}||P|\pi^2 t^2} \tag{9}$$

Note that $\sum_{t \geq 1} \frac{1}{t^2} = \frac{\pi^2}{6}$. By taking the union bound over $\mathcal{X}$, P and $t \geq 1$, Eq. (7) holds w.p. at least $1 - \delta/2$. By the same reasoning and skipping the union bound over P, Eq. (6) holds w.p. at least $1 - \delta/2$. Thus, both events hold w.p. at least $1 - \delta$. □

Next, we state three lemmas from Ziomek et al. (2025) that are used in the proof of our regret bound.

Lemma B.2. *(Lemma 5.1 of Ziomek et al. (2025)) If $\xi_t = 2\sigma^2 \log\left(\frac{|P|\pi^2 t^2}{6\delta}\right)$, then the following holds with probability at least $1 - \delta$:*

$$\left|\sum_{i \in S_{t,p}} \epsilon_i\right| \leq \sqrt{\xi_t |S_{t,p}|} \quad \forall t, p \in [T] \times P. \tag{10}$$

Lemma B.3. *(Lemma 5.2 of Ziomek et al. (2025)) Let $B_{p^*} = \beta_1 + \sup_{x \in \mathcal{X}} |\mu_{1,p^*}(x)|$, then if μ_{1,p^*} and k_{1,p^*} satisfy $|\mu_{1,p^*}(\cdot)| < \infty$ and $k_{1,p^*}(\cdot,\cdot) \leq 1$ and Lemma B.1 holds, then*

$$\sup_{x \in \mathcal{X}} |f(x)| \leq B_{p^*}. \tag{11}$$

Lemma B.4. *(Lemma 5.3 of Ziomek et al. (2025)) For $C = 2/\log(1+\sigma^{-2})$, $\sum_{t \notin \mathcal{C}} \sqrt{\beta_t}\sigma_{t,p_t}(x_t) \leq \sqrt{CT\beta_T \hat{\gamma}_T |P|}$ where $\beta_T = \max_{p \in P} \beta_T$ and $\hat{\gamma}_T = \max_{p \in P} \gamma_{T,p}$.*

Lemma B.5. *If the events of Lemmas B.1 and B.2 hold, then PE-GP-TS never eliminates the true prior p^*.*

Proof. For any $t \in [T]$,

$$\left|\sum_{i \in S_{t,p^*}} \eta_i\right| = \left|\sum_{i \in S_{t,p^*}} (y_i - f(x_i) + f(x_i) - \mu_{i,p^*}(x_i))\right| \tag{12}$$

$$\leq \left|\sum_{i \in S_{t,p^*}} \epsilon_i\right| + \sum_{i \in S_{t,p^*}} |f(x_i) - \mu_{i,p^*}(x_i)| \qquad \text{(Triangle ineq.)} \tag{13}$$

$$\leq \sqrt{\xi_t |S_{t,p^*}|} + \sum_{i \in S_{t,p^*}} \sqrt{\beta_i}\sigma_{i,p^*}(x_i). \qquad \text{(Lemmas B.1 and B.2)} \tag{14}$$

Therefore, the elimination criteria on line 9 in Algorithm 1, $|\sum_{i \in S_{t,p_t}} \eta_i| > V_t$, always evaluates to `false` for $p_t = p^*$. □

Then, we state and prove the new instantaneous regret bound for PE-GP-TS.

Lemma 4.1. *If the events of Lemmas B.1 and B.2 holds, then the following holds for the instantaneous regret of PE-GP-TS for all $t \in [T]$: $f(x^*) - f(x_t) \leq 2\sqrt{\beta_t}\sigma_{t,p^*}(x^*) + \sqrt{\beta_t}\sigma_{t,p_t}(x_t) - \eta_t + \epsilon_t$.*

Proof. First, we upper bound $f(x^*)$ as follows

$$f(x^*) \leq \mu_{t,p^*}(x^*) + \sqrt{\beta_t}\sigma_{t,p^*}(x^*) \qquad \text{(Eq. (6))} \tag{15}$$

$$\leq \tilde{f}_{t,p^*}(x^*) + 2\sqrt{\beta_t}\sigma_{t,p^*}(x^*) \qquad \text{(Eq. (7))} \tag{16}$$

$$\leq \tilde{f}_{t,p_t}(x_t) + 2\sqrt{\beta_t}\sigma_{t,p^*}(x^*). \qquad \text{(TS selection rule and Lemma B.5)} \tag{17}$$

For the final step, note that $p^* \in P_t$ by Lemma B.5. Then, we lower bound $f(x_t)$

$$f(x_t) = \mu_{t,p_t}(x_t) + \eta_t - \epsilon_t \qquad \text{(Def. of } \eta_t\text{)} \tag{18}$$

$$\geq \tilde{f}_{t,p_t}(x_t) - \sqrt{\beta_t}\sigma_{t,p_t}(x_t) + \eta_t - \epsilon_t. \qquad \text{(Eq. (7))} \tag{19}$$

Combining, Eqs. (17) and (19) we obtain

$$f(x^*) - f(x_t) \leq 2\sqrt{\beta_t}\sigma_{t,p^*}(x^*) + \sqrt{\beta_t}\sigma_{t,p_t}(x_t) - \eta_t + \epsilon_t. \tag{20}$$

□

Finally, we state and prove the cumulative regret bound for PE-GP-TS.

Theorem 4.2. *Let $B_{p^*} = \beta_1 + \sup_{x\in\mathcal{X}} |\mu_{1,p^*}(x)|$ and $C = 2/\log(1+\sigma^{-2})$. If $p^* \in P$ and $f \sim \mathcal{GP}(\mu_{1,p^*}, k_{1,p^*})$, then PE-GP-TS with confidence parameters $\beta_t = 2\log(2|\mathcal{X}||P|\pi^2 t^2/3\delta)$ and $\xi_t = 2\sigma^2\log(|P|\pi^2t^2/3\delta)$, satisfies the following regret bound with probability at least $1-\delta$:*

$$R(T) \leq 2|P|B_{p^*} + 2\sqrt{\xi_T|P|T} + 2\sqrt{CT\beta_T\hat{\gamma}_T|P|} + 2\sqrt{T\beta_T\textstyle\sum_{t\in[T]}\sigma^2_{t,p^*}(x^*)} \tag{1}$$

Proof. To establish a bound on the cumulative regret, we separate out the rounds where priors are eliminated. Hence, define the set of critical iterations as

$$\mathcal{C} = \left\{ t \in [T] : \left| \sum_{i\in S_{t,p_t}} \eta_i \right| > \sqrt{\xi_t S_{t,p_t}} + \sum_{i\in S_{t,p_t}} \sqrt{\beta_i}\sigma_{i,p_t}(x_i) \right\}. \tag{21}$$

Note that $|\mathcal{C}| \leq |P|$. Using Lemma B.3 and Eq. (20), we can bound the cumulative regret as follows:

$$R(T) = \sum_{t\in\mathcal{C}} f(x^*) - f(x_t) + \sum_{t\notin\mathcal{C}} f(x^*) - f(x_t) \tag{22}$$

$$\leq 2|P|B_{p^*} + \sum_{t\notin\mathcal{C}} 2\sqrt{\beta_t}\sigma_{t,p^*}(x^*) + \sum_{t\notin\mathcal{C}}\sqrt{\beta_t}\sigma_{t,p_t}(x_t) + \sum_{p\in P}\sum_{t\in S_{T,p}\setminus\mathcal{C}} (\epsilon_t - \eta_t). \tag{23}$$

where $B_{p^*} := \beta_1 + \sup_{x\in\mathcal{X}} |\mu_{1,p^*}(x)|$. If $t \notin \mathcal{C}$, line 9 in Algorithm 1 evaluates to `false` and hence

$$\sum_{p\in P}\sum_{t\in S_{T,p}\setminus\mathcal{C}} -\eta_t \leq \sum_{p\in P}\sqrt{\xi_T|S_{T,p}|} + \sum_{p\in P}\sum_{t\in S_{T,p}\setminus\mathcal{C}}\sqrt{\beta_t}\sigma_{t,p}(x_t). \tag{24}$$

Additionally, using Lemma B.2, we can bound the Gaussian noise:

$$\sum_{p\in P}\sum_{t\in S_{T,p}\setminus\mathcal{C}} \epsilon_t \leq \sum_{p\in P}\left|\sum_{t\in S_{T,p}\setminus\mathcal{C}} \epsilon_t\right| \tag{25}$$

$$\leq \sum_{p\in P}\sqrt{\xi_T|S_{T,p}\setminus\mathcal{C}|} \qquad \text{(Lemma B.2)} \tag{26}$$

$$\leq \sum_{p\in P}\sqrt{\xi_T|S_{T,p}|} \tag{27}$$

$$\leq \sqrt{\xi_T|P|T}. \qquad \text{(Cauchy-Schwarz)} \tag{28}$$

Combining the above, the cumulative regret is bounded by

$$R(T) \leq 2|P|B_{p^*} + 2\sqrt{\xi_T|P|T} + 2\sum_{t\notin\mathcal{C}}\sqrt{\beta_t}\sigma_{t,p^*}(x^*) + 2\sum_{t\notin\mathcal{C}}\sqrt{\beta_t}\sigma_{t,p_t}(x_t). \tag{29}$$

Finally, applying Lemma B.4, we obtain the result

$$R(T) \leq 2|P|B_{p^*} + 2\sqrt{\xi_T|P|T} + 2\sqrt{T\beta_T\sum_{t\in[T]}\sigma^2_{t,p^*}(x^*)} + 2\sqrt{CT\beta_T\hat{\gamma}_T|P|}. \tag{30}$$

□

B.2 HP-GP-TS

In this section, we state and prove our regret bound for HP-GP-TS. We begin by proving Lemma 4.3.

Lemma 4.3. *Let $C = 2/\log(1+\sigma^{-2})$, $N_T(p) = \sum_{t\in[T]} \mathbb{1}\{p_t = p\}$ where $\mathbb{1}$ is the indicator function, and $\Gamma_p : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}_{\geq 0}$ be a concave function such that $\Gamma_p(t) \geq \gamma_{t,p}$ for all $t, p \in [T] \times P$. Furthermore, let $\bar{\gamma}_T(P_1) := \sum_{p\in P} \Gamma_p(P_1(p)T)$, then for HP-GP-TS, $\mathbb{E}[N_T(p)] = P_1(p)T$ and*

$$\mathbb{E}\Big[\sum_{t\in[T]} \sigma^2_{t,p_t}(x_t)\Big] \leq C\bar{\gamma}_T(P_1). \tag{3}$$

Remark B.6. *For the RBF and Matérn kernels, the known upper bounds for the maximum information gain are concave (Srinivas et al., 2012; Vakili et al., 2021), thereby satisfying the conditions of Lemma 4.3.*

Proof. We begin by showing that $\mathbb{E}[N_T(p)] = P_1(p)T$.

$$\begin{aligned}
\mathbb{E}[N_T(p)] &= \mathbb{E}\Big[\sum_{t\in[T]} \mathbb{1}\{p_t = p\}\Big] && (31)\\
&= \sum_{t\in[T]} \mathbb{E}_{H_t}\left[\mathbb{E}\left[\mathbb{1}\{p_t = p\}\middle| H_t\right]\right] && \text{(Tower rule)} \quad (32)\\
&= \sum_{t\in[T]} \mathbb{E}_{H_t}\left[\mathbb{E}\left[\mathbb{1}\{p^* = p\}\middle| H_t\right]\right] && (p_t|H_t \overset{d}{=} p^*|H_t) \quad (33)\\
&= \sum_{t\in[T]} \mathbb{E}\left[\mathbb{1}\{p^* = p\}\right] && (34)\\
&= \sum_{t\in[T]} P_1(p) = P_1(p)T. && (P_1(p) = \mathbb{P}(p^* = p)) \quad (35)
\end{aligned}$$

Then, by the intermediate steps of Lemma 5.3 of Ziomek et al. (2025) $\sum_{t\in[T]} \sigma^2_{t,p_t}(x_t) \leq C\sum_{p\in P} \gamma_{N_T(p),p}$. We include the proof here for completeness and introduce some helpful notation. Let A be a multiset over $\mathcal{X}$ s.t. $|A| < \infty$, we define

$$\sigma^2_{A,p}(x) = k_{1,p}(x,x) - \mathbf{k}_{A,p}(x)^T(\mathbf{K}_{A,p} + \sigma^2 I)^{-1}\mathbf{k}_{A,p}(x), \tag{36}$$

where $\mathbf{K}_{A,p} = [k_{1,p}(x,x')]_{x,x'\in A}$ and $\mathbf{k}_{A,p} = [k_{1,p}(x,x')]_{x'\in A}$ with elements repeated by their multiplicity in A. Then, let $A_{t,p} = \{x_i : i \in [t-1], p_i = p\}$ be the multiset of arms queried whilst selecting prior p. For any two multisets S, S' such that $S \subseteq S'$, we have that $\sigma^2_{S',p}(x) \leq \sigma^2_{S,p}(x)$ for all $x \in \mathcal{X}$. Since $A_{t,p}$ is a subset of the arms collected over the history H_t, we have that $\sigma^2_{t,p_t}(x_t) \leq \sigma^2_{A_{t,p},p}(x_t)$. By the proof of Lemma 5.4 of Srinivas et al. (2012), $\sigma^2_{A_{t,p},p}(x_t) \leq C\log(1+\sigma^{-2}\sigma^2_{A_{t,p},p}(x_t))$ for $C = 2/\log(1+\sigma^{-2})$. By reorganizing the sum over $t \in [T]$ into a sum over $p \in P$, and applying Lemma 5.3 of Srinivas et al. (2012) yields that

$$\sum_{t\in[T]} \sigma^2_{t,p_t}(x_t) \leq \sum_{p\in P}\sum_{t\in[T]:p_t=p} \sigma^2_{A_{t,p},p}(x_t) \leq C\sum_{p\in P} \gamma_{|A_{T,p}|,p} = C\sum_{p\in P} \gamma_{N_T(p),p}. \tag{37}$$

Finally, we combine the results above through a concavity argument.

$$\begin{aligned}
\mathbb{E}\left[\sum_{t\in[T]} \sigma^2_{t,p_t}(x_t)\right] &\leq C\sum_{p\in P} \mathbb{E}[\gamma_{N_T(p),p}] && \text{(Eq. (37))} \quad (38)\\
&\leq C\sum_{p\in P} \mathbb{E}[\Gamma_p(N_T(p))] && (\gamma_{t,p} \leq \Gamma_p(t), \forall t,p \in [T]\times P) \quad (39)\\
&\leq C\sum_{p\in P} \Gamma_p(\mathbb{E}[N_T(p)]) && (\Gamma_p(t) \text{ concave, Jensen's ineq.}) \quad (40)\\
&\leq C\sum_{p\in P} \Gamma_p(P_1(p)T). && \text{(Eq. (35))} \quad (41)
\end{aligned}$$

□

Next, we prove that the true prior is in the confidence set with high probability. Recall that we define the excess reward for prior p at time t as

$$G_t(p) = \sum_{s=1}^{t-1} \mathbb{1}\{p_s = p\} \left(\mu_{s,p_s}(x_s) - \sqrt{\eta_T}\sigma_{s,p_s} - f(x_s) - \epsilon_s\right) \tag{42}$$

where $\eta_T = 2\log|\mathcal{X}|T^6$. Let $\xi_t(p) = \sigma\sqrt{14N_{t-1}(p)\log(T)}$ where $N_t(p) = \sum_{s=1}^{t} \mathbb{1}\{p_s = p\}$ denotes how often the prior p was selected up to and including time t. Then, we define the confidence set at time t as

$$\mathcal{C}_t = \{p \in P : G_\tau(p) \leq \xi_\tau(p)\ \forall \tau \leq t\}. \tag{43}$$

For notational convenience, we consider the history $H_t = (p_i, x_i, y_i)_{i=1}^{t-1}$ with the selected priors $(p_i)_{i=1}^t$ augmented such that $\mathcal{C}_t$ and $(N_{t-1}(p))_{p\in P}$ are deterministic conditioned on H_t.

Lemma B.7. *For any $t \in [T]$, $\mathbb{P}(p^* \notin \mathcal{C}_t) \leq \frac{3}{T^5}$.*

Proof. Note that $\mathcal{C}_t$ is monotonically decreasing due to the time-uniform definition of $\mathcal{C}_t$, i.e. $\mathcal{C}_s \supseteq \mathcal{C}_t$ for any $s < t$. Thus, $\mathbb{P}(p^* \notin \mathcal{C}_t) \leq \mathbb{P}(p^* \notin \mathcal{C}_T)$ and we focus on bounding $\mathbb{P}(p^* \in \mathcal{C}_T)$.

Let $E = \cap_{t=1}^{T-1} E_t$ where $E_t = \{|f(x) - \mu_{t,p^*}(x)| \leq \sqrt{\eta_T}\sigma_{t,p^*}(x), \forall x \in \mathcal{X}\}$. Then, by the law of total probability

$$\mathbb{P}(p^* \notin \mathcal{C}_T) = \underbrace{\mathbb{P}(p^* \notin \mathcal{C}_T|E^c)}_{\leq 1}\underbrace{\mathbb{P}(E^c)}_{\leq 1/T^5} + \underbrace{\mathbb{P}(p^* \notin \mathcal{C}_T|E)\mathbb{P}(E)}_{\leq 2/T^5} \tag{44}$$

where the bounds for $\mathbb{P}(E^c)$ and $\mathbb{P}(p^* \notin \mathcal{C}_T|E)\mathbb{P}(E)$ are shown below.

$$\begin{aligned}
\mathbb{P}(E^c) &= \mathbb{P}\left(\exists t \in [T-1], x \in \mathcal{X} : |f(x) - \mu_{t,p^*}(x)| > \sqrt{\eta_T}\sigma_{t,p^*}(x)\right) && (45)\\
&\leq \sum_{t\in[T-1]}\sum_{x\in\mathcal{X}} \mathbb{P}\left(|f(x) - \mu_{t,p^*}(x)| > \sqrt{\eta_T}\sigma_{t,p^*}(x)\right) && (46)\\
&\leq \sum_{t\in[T-1]}\sum_{x\in\mathcal{X}} \mathbb{E}_{H_t,p^*}\left[\mathbb{P}\left(\frac{|f(x) - \mu_{t,p^*}(x)|}{\sigma_{t,p^*}(x)} > \sqrt{\eta_T}\middle| H_t, p^* = p\right)\right] && (47)\\
&\leq \sum_{t\in[T-1]}\sum_{x\in\mathcal{X}} \mathbb{E}_{H_t,p^*}\left[\exp\left(-\frac{\eta_T}{2}\right)\right] && \left(\begin{smallmatrix}\mathbb{P}(|r| > \sqrt{c}) \leq \exp(-c/2)\\ \text{for } r \sim \mathcal{N}(0,1),\, c \geq 0\end{smallmatrix}\right) \quad (48)\\
&= \sum_{t\in[T-1]}\sum_{x\in\mathcal{X}} \mathbb{E}_{H_t,p^*}\left[\frac{1}{|\mathcal{X}|T^6}\right] && \left(\eta_T = 2\log(|\mathcal{X}|T^6)\right) \quad (49)\\
&= \frac{1}{T^5}. && (50)
\end{aligned}$$

Next, we bound the right term $\mathbb{P}(p^* \notin \mathcal{C}_t|E)$. Recall that $p^* \notin \mathcal{C}_T$ is equivalent to $\exists t \in [T]$ such that $G_t(p^*) > \xi_t(p^*)$. Hence,

$$\begin{aligned}
\mathbb{P}(p^* \notin \mathcal{C}_T|E) &= \mathbb{P}(\exists t \in [T] : G_t(p^*) > \xi_t(p^*)|E) && (51)\\
&\leq \sum_{t\in[T]} \mathbb{P}(G_t(p^*) > \xi_t(p^*)|E) && \text{(Union bound)} \quad (52)\\
&= \sum_{t\in[T]} \mathbb{P}\Bigg(\sum_{s=1}^{t-1} \mathbb{1}\{p_s = p^*\}\Big(\mu_{s,p^*}(x_s) - \sqrt{\eta_T}\sigma_{t,p^*}(x_s)\\
&\qquad - f(x_s) - \epsilon_s\Big) > \xi_t(p^*)\Big|E\Bigg). && (53)
\end{aligned}$$

Given E, $\mu_{s,p^*}(x_s) - \sqrt{\eta_T}\sigma_{s,p^*}(x_s) - f(x_s) \leq 0, \forall s \in [T-1]$ and therefore

$$\mathbb{P}(p^* \notin \mathcal{C}_T|E)\mathbb{P}(E) \leq \sum_{t\in[T]} \mathbb{P}\left(\sum_{s=1}^{t-1} \mathbb{1}\{p_s = p^*\}(-\epsilon_s) > \xi_t(p^*)\Big|E\right)\mathbb{P}(E) \tag{54}$$

$$\leq \sum_{t\in[T]} \mathbb{P}\left(\left|\sum_{s=1}^{t-1} \mathbb{1}\{p_s = p^*\}(-\epsilon_s)\right| > \xi_t(p^*)\Big|E\right)\mathbb{P}(E) \tag{55}$$

$$= \sum_{t\in[T]} \mathbb{P}\left(\left|\sum_{s=1}^{t-1} \mathbb{1}\{p_s = p^*\}(-\epsilon_s)\right| > \xi_t(p^*), E\right) \tag{56}$$

$$\leq \sum_{t\in[T]} \mathbb{P}\left(\left|\sum_{s=1}^{t-1} \mathbb{1}\{p_s = p^*\}(-\epsilon_s)\right| > \xi_t(p^*)\right) \tag{57}$$

$$= \sum_{t\in[T]}\sum_{p\in P}\mathbb{P}\left(\left|\sum_{s=1}^{t-1} \mathbb{1}\{p_s = p\}\epsilon_s\right| > \sigma\sqrt{14N_{t-1}(p)\log T}\,\Big|\,p^* = p\right)$$

$$\cdot\,\mathbb{P}(p^* = p) \tag{58}$$

$$\leq \sum_{t\in[T]}\sum_{p\in P}\frac{2}{T^6}\mathbb{P}(p^* = p) \qquad \text{(Lemma B.8)} \tag{59}$$

$$\leq \frac{2}{T^5} \tag{60}$$

□

Next, we prove the self-normalizing concentration inequality for the sum of Gaussian noises as pulled by each prior that we used in Eq. (58).

Lemma B.8. *Let* $S_{t,p} = \sum_{s=1}^{t-1} \mathbb{1}\{p_s = p\}\epsilon_s$ *and* $\alpha > 0$*, then*

$$\mathbb{P}\left(|S_{t,p}| > \sigma\sqrt{2(\alpha+1)N_{t-1}(p)\log(T)}\,\Big|\,p^* = p\right) \leq \frac{2}{T^\alpha}, \quad \forall p \in P. \tag{61}$$

Remark B.9. *Note that similar results have been shown by Hong et al. (2022b, Proof of Lemma 3), Lu et al. (2023, Lemma 4), and (Ziomek et al., 2025, Lemma 5.1). We found the arguments in the proofs of Hong et al. (2022b); Lu et al. (2023) uncompelling due to their brevity. Whilst the proof of Ziomek et al. (2025) is clearer, we provide a proof using a martingale technique as a complement.*

Proof. Fix $t \in [T]$ and $p^* = p$, for the remainder of this proof all probabilities condition on $p^* = p$. We begin by defining the event

$$\mathcal{F} := \left\{S_{t,p} > \sigma\sqrt{2(\alpha+1)N_{t-1}(p)\log(T)}\right\} \tag{62}$$

$$= \bigcup_{k=1}^{t-1} \underbrace{\left\{S_{t,p} > \sigma\sqrt{2(\alpha+1)k\log(T)} \cap N_{t-1}(p) = k\right\}}_{\mathcal{F}_k :=}. \tag{63}$$

To bound the probability of the events $\mathcal{F}_k$, we introduce a martingale $M_t(\lambda)$ for $\lambda > 0$ into $\mathcal{F}_k$ as follows:

$$\mathcal{F}_k = \left\{\lambda S_{t,p} > \lambda\sigma\sqrt{2(\alpha+1)k\log(T)} \cap N_{t-1}(p) = k\right\} \quad (\lambda > 0) \tag{64}$$

$$= \left\{\lambda S_{t,p} - \frac{\lambda^2\sigma^2}{2}N_{t-1}(p) > \lambda\sigma\sqrt{2(\alpha+1)k\log(T)} - \frac{\lambda^2\sigma^2}{2}k \cap N_{t-1}(p) = k\right\} \tag{65}$$

$$= \Bigg\{\underbrace{\exp\left(\lambda S_{t,p} - \frac{\lambda^2\sigma^2}{2}N_{t-1}(p)\right)}_{M_t(\lambda):=} > \exp\left(\lambda\sigma\sqrt{2(\alpha+1)k\log(T)} - \frac{\lambda^2\sigma^2}{2}k\right) \cap N_{t-1}(p) = k\Bigg\}. \tag{66}$$

To tighten the bound of the probability of $\mathcal{F}_k$, we select $\lambda_k = \sqrt{\frac{2(\alpha+1)\log T}{\sigma^2 k}}$, yielding:

$$\mathcal{F}_k = \{M_t(\lambda_k) > \exp\left((\alpha+1)\log T\right) \cap N_{t-1}(p) = k\} \tag{67}$$

$$\subseteq \left\{M_t\left(\sqrt{\frac{2(\alpha+1)\log T}{\sigma^2 k}}\right) \geq T^{\alpha+1}\right\}. \tag{68}$$

Since $M_t(\lambda) \geq 0$, by Markov's inequality,

$$\mathbb{P}\left(\mathcal{F}_k\right) \leq \mathbb{P}\left(M_t\left(\lambda_k\right) \geq T^{\alpha+1}\right) \leq \frac{\mathbb{E}\left[M_t(\lambda_k)\right]}{T^{\alpha+1}}. \tag{69}$$

Next, it remains to show that $M_t(\lambda)$ is a martingale such that $\mathbb{E}[M_t(\lambda)] = 1$. Let $\mathcal{H}_{t-1} = \{p_s, \epsilon_s\}_{s=1}^{t-2}$ be the history of selected priors and noise up to and including time $t-2$, then

$$\mathbb{E}[M_t(\lambda)|\mathcal{H}_{t-1}] = M_{t-1}(\lambda) \cdot \mathbb{E}\left[\exp\left(\lambda\mathbb{1}\{p_{t-1} = p\}\epsilon_{t-1} - \lambda^2\sigma^2\mathbb{1}\{p_{t-1} = p\}/2\right)\big|\mathcal{H}_{t-1}\right] \tag{70}$$

$$= M_{t-1}(\lambda) \cdot \Big(\underbrace{\mathbb{E}\left[\exp(0)|p_{t-1} \neq p, \mathcal{H}_{t-1}\right]}_{=1}\mathbb{P}(p_{t-1} \neq p|\mathcal{H}_{t-1}) \tag{71}$$

$$+ \underbrace{\mathbb{E}\left[\exp(\lambda\epsilon_{t-1} - \lambda^2\sigma^2/2)|p_{t-1} = p, \mathcal{H}_{t-1}\right]}_{=1 \text{ since } \epsilon_{t-1} \perp\!\!\!\perp p_{t-1}, \mathcal{H}_{t-1}}\mathbb{P}(p_{t-1} = p, \mathcal{H}_{t-1})\Big) \tag{72}$$

$$= M_{t-1}(\lambda). \tag{73}$$

Applying the above recursively to $\mathbb{E}[M_t(\lambda)]$ and defining $M_1(\lambda) = 1$, we get that $\mathbb{E}[M_t(\lambda)] = 1$. Thus, from Eq. (69) and a union bound over $k \in [T-1]$, $\mathbb{P}(S_{t,p} > \sigma\sqrt{2(\alpha+1)N_{t-1}(p)\log(T)}) \leq 1/T^\alpha$. By symmetry of the Gaussian noise, $\mathbb{P}(|S_{t,p}| > \sigma\sqrt{2(\alpha+1)N_{t-1}(p)\log(T)}) \leq 2/T^\alpha$. □

Finally, we are ready to state and prove the regret bound for HP-GP-TS.

Theorem 4.4. *Let $C = 2/\log(1+\sigma^{-2})$, $\mu_{max} = \sup_{p,x\in P\times\mathcal{X}}|\mu_{1,p}(x)|$, $M = \mathbb{E}[\sup_{x\in\mathcal{X}}|f(x)|]$, $M_p = \mathbb{E}\left[\sup_{x\in\mathcal{X}}|f(x)||p^* = p\right]$, $M_\Delta = \max_{p\in P} M_p - \min_{p\in P} M_p$, and $\bar{M} = M^2 + 1 + M_\Delta^2/4$. If $p^* \sim P_1$, $f \sim \mathcal{GP}(\mu_{1,p^*}, k_{1,p^*})$, $\beta_t = 2\log(|\mathcal{X}|t^2/\sqrt{2\pi})$, $\eta_T = 2\log|\mathcal{X}|T^6$, then the Bayesian regret of HP-GP-TS is bounded by*

$$BR(T) \leq \frac{\pi^2}{6} + \sqrt{CT\bar{\gamma}_T(P_1)}(\sqrt{\beta_T} + \sqrt{\eta_T}) \quad \textit{(Terms (1) and (2))}$$
$$+\tfrac{\sqrt{3}}{T}\left(\sqrt{\sigma^{-2}(\bar{M} + 2M\mu_{max} + \mu_{max}^2 + \sigma^2)} + \sqrt{\bar{M}/T}\right)$$
$$+\sigma\sqrt{14T|P|\log T} + |P|\left(\sigma^{-1}\sqrt{T}(M + \mu_{max} + \sigma) + M + \sigma\sqrt{2\log T}\right) \Bigg\} \textit{(Term (3))} \tag{5}$$

Proof. Recall that $p^*, x^*|H_t \overset{d}{=} p_t, x_t|H_t$ and that $U_{t,p}(x)$ is a deterministic function w.r.t. p and x conditioned on the history H_t, therefore $\mathbb{E}[U_{t,p^*}(x^*)] = \mathbb{E}[U_{t,p_t}(x_t)]$ follows by the tower rule. We

begin by decomposing the Bayesian regret into three terms and show the bounds that we will later obtain for each of them.

$$
\begin{aligned}
\mathrm{BR}(T) &= \sum_{t\in[T]} \mathbb{E}[f(x^*) - f(x_t)] && (74)\\
&= \sum_{t\in[T]} \mathbb{E}[f(x^*) - U_{t,p^*}(x^*) + U_{t,p_t}(x_t) - f(x_t)] && \left(p^*, x^* | H_t \overset{d}{=} p_t, x_t | H_t\right) \quad (75)\\
&= \sum_{t\in[T]} \mathbb{E}\Big[f(x^*) - U_{t,p^*}(x^*) + (\sqrt{\beta_t} + \sqrt{\eta_T})\sigma_{t,p_t}(x_t) \\
&\qquad + \mu_{t,p_t}(x_t) - \sqrt{\eta_T}\sigma_{t,p_t}(x_t) - f(x_t)\Big] && (\pm\sqrt{\eta_T}\sigma_{t,p_t}(x_t)) \quad (76)\\
&= \underbrace{\sum_{t\in[T]} \mathbb{E}\left[f(x^*) - U_{t,p^*}(x^*)\right]}_{A_1} + \underbrace{\sum_{t\in[T]} \mathbb{E}\left[(\sqrt{\beta_t} + \sqrt{\eta_T})\sigma_{t,p_t}(x_t)\right]}_{A_2} && (77)\\
&\quad + \underbrace{\sum_{t\in[T]} \mathbb{E}\left[\mu_{t,p_t}(x_t) - \sqrt{\eta_T}\sigma_{t,p_t}(x_t) - f(x_t)\right]}_{A_3}. && (78)\\
&\le \underbrace{\frac{\pi^2}{6}}_{A_1} + \underbrace{\sqrt{CT\bar{\gamma}_T(P_1)}(\sqrt{\beta_T} + \sqrt{\eta_T})}_{A_2} \\
&\qquad + \underbrace{\frac{\sqrt{3}}{T}\left(\sqrt{\sigma^{-2}(\bar{M} + 2M\mu_{\max} + \mu_{\max}^2 + \sigma^2)} + \sqrt{\bar{M}/T}\right)}_{A_{3,1}} \\
&\quad + \underbrace{\sigma\sqrt{14T|P|\log T} + |P|\left(\sigma^{-1}\sqrt{T}(M + \mu_{\max} + \sigma) + M + \sigma\sqrt{2\log T}\right)}_{A_{3,2}}. && (79)
\end{aligned}
$$

Next, we will prove the bounds for the terms A_1, A_2, and A_3 where the bound for A_3 is given by the sum of $A_{3,1}$ and $A_{3,2}$.

Bounding A_1 Since the upper confidence term in A_1, $U_{t,p^*}(x^*)$, corresponds to the confidence bound of the true prior p^*, the bound for A_1 follows by standard techniques (Russo & Van Roy, 2014):

$$
\begin{aligned}
A_1 &= \sum_{t\in[T]} \mathbb{E}\left[f(x^*) - U_{t,p^*}(x^*)\right] && (80)\\
&\le \sum_{t\in[T]} \mathbb{E}\left[\left[f(x^*) - U_{t,p^*}(x^*)\right]_+\right] && ([\cdot]_+ := \max(\cdot, 0)) \quad (81)\\
&\le \sum_{t\in[T]}\sum_{x\in\mathcal{X}} \mathbb{E}\left[\left[f(x) - U_{t,p^*}(x)\right]_+\right] && (x^* \in \mathcal{X}, [\cdot]_+ \ge 0) \quad (82)\\
&= \sum_{t\in[T]}\sum_{x\in\mathcal{X}} \mathbb{E}_{p^*,H_t}\left[\mathbb{E}_t\left[\left[f(x) - \mu_{t,p^*}(x) - \sqrt{\beta_t}\sigma_{t,p^*}(x)\right]_+ \,|\, p^*, H_t\right]\right]. && \text{(Tower rule)} \quad (83)
\end{aligned}
$$

Recall that for $Z \sim \mathcal{N}(\mu, \sigma)$ with $\mu \le 0$, $\mathbb{E}[[Z]_+] \le \frac{\sigma}{\sqrt{2\pi}} \exp\left(\frac{-\mu^2}{2\sigma^2}\right)$. In our case, note that $f(x)|p^*, H_t \sim \mathcal{N}(\mu_{t,p^*}(x), \sigma^2_{t,p^*}(x))$ and $-\mu_{t,p^*}(x) - \sqrt{\beta_t}\sigma_{t,p^*}(x)$ is deterministic given p^*, H_t.

Hence,

$$A_1 \le \sum_{t\in[T]}\sum_{x\in\mathcal{X}} \mathbb{E}_{p^*,H_t}\left[\frac{\sigma_{t,p^*}(x)}{\sqrt{2\pi}}\exp\left(\frac{-\beta_t}{2}\right)\right] \tag{84}$$

$$\le \sum_{t\in[T]}\sum_{x\in\mathcal{X}} \mathbb{E}_{p^*,H_t}\left[\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-\beta_t}{2}\right)\right] \qquad (\sigma_{t,p^*}(x)\le\sigma_{0,p^*}(x)\le 1) \tag{85}$$

$$= \sum_{t\in[T]}\sum_{x\in\mathcal{X}} \frac{1}{\sqrt{2\pi}}\exp(-\beta_t/2) \tag{86}$$

$$= \sum_{t\in[T]}\frac{1}{t^2} \le \frac{\pi^2}{6}. \qquad (\beta_t = 2\log(|\mathcal{X}|t^2/\sqrt{2\pi})) \tag{87}$$

Bounding A_2 To bound A_2, we separate it into two terms and apply Cauchy-Schwarz to each term:

$$A_2 = \mathbb{E}\left[\sum_{t\in[T]}\sqrt{\beta_t}\sigma_{t,p_t}(x_t)\right] + \mathbb{E}\left[\sum_{t\in[T]}\sqrt{\eta_T}\sigma_{t,p_t}(x_t)\right] \tag{88}$$

$$\le \mathbb{E}\left[\sqrt{\sum_{t\in[T]}\beta_t\sum_{t\in[T]}\sigma^2_{t,p_t}(x_t)}\right] + \mathbb{E}\left[\sqrt{\sum_{t\in[T]}\eta_T\sum_{t\in[T]}\sigma^2_{t,p_t}(x_t)}\right] \qquad \text{(Cauchy-Schwarz)} \tag{89}$$

$$\le \mathbb{E}\left[\sqrt{T\beta_T\sum_{t\in[T]}\sigma^2_{t,p_t}(x_t)}\right] + \mathbb{E}\left[\sqrt{T\eta_T\sum_{t\in[T]}\sigma^2_{t,p_t}(x_t)}\right] \qquad (\beta_t\le\beta_T, \forall t\in[T]) \tag{90}$$

$$= \sqrt{T}(\sqrt{\beta_T}+\sqrt{\eta_T})\mathbb{E}\left[\sqrt{\sum_{t\in[T]}\sigma^2_{t,p_t}(x_t)}\right] \tag{91}$$

$$\le \sqrt{T}(\sqrt{\beta_T}+\sqrt{\eta_T})\sqrt{\mathbb{E}\Big[\sum_{t\in[T]}\sigma^2_{t,p_t}(x_t)\Big]}. \qquad \text{(Jensen's inequality)} \tag{92}$$

By Lemma 4.3, we have that $\mathbb{E}\big[\sum_{t\in[T]}\sigma^2_{t,p_t}(x_t)\big] \le C\bar{\gamma}_T(P_1)$ and therefore

$$A_2 \le \sqrt{CT\bar{\gamma}_T(P_1)}\left(\sqrt{\beta_T}+\sqrt{\eta_T}\right). \tag{93}$$

Bounding A_3 We further split A_3 based on whether $p_t \in \mathcal{C}_t$ holds:

$$A_3 = \underbrace{\sum_{t\in[T]}\mathbb{E}\left[(\mu_{t,p_t}(x_t) - \sqrt{\eta_T}\sigma_{t,p_t}(x_t) - f(x_t))\mathbb{1}\{p_t\notin\mathcal{C}_t\}\right]}_{A_{3,1}} \tag{94}$$

$$+ \underbrace{\sum_{t\in[T]}\mathbb{E}\left[(\mu_{t,p_t}(x_t) - \sqrt{\eta_T}\sigma_{t,p_t}(x_t) - f(x_t))\mathbb{1}\{p_t\in\mathcal{C}_t\}\right]}_{A_{3,2}}. \tag{95}$$

Bounding $A_{3,1}$: To bound $A_{3,1}$, we apply the Cauchy-Schwarz inequality for expectations to separate the factors $\mu_{t,p_t}(x_t) - f(x_t)$ and $\mathbb{1}\{p_t \notin \mathcal{C}_t\}$ into different expectations as follows:

$$A_{3,1} \leq \sum_{t\in[T]} \mathbb{E}\left[(\mu_{t,p_t}(x_t) - f(x_t))\mathbb{1}\{p_t \notin \mathcal{C}_t\}\right] \qquad (\sqrt{\eta_T}\sigma_{t,p_t}(x_t) \geq 0) \quad (96)$$

$$= \sum_{t\in[T]} \left(\mathbb{E}\left[\mu_{t,p_t}(x_t)\mathbb{1}\{p_t \notin \mathcal{C}_t\}\right] + \mathbb{E}\left[-f(x_t)\mathbb{1}\{p_t \notin \mathcal{C}_t\}\right]\right) \quad (97)$$

$$\leq \sum_{t\in[T]} \left(\sqrt{\mathbb{E}\left[(\mu_{t,p_t}(x_t))^2\right]\mathbb{E}\left[(\mathbb{1}\{p_t \notin \mathcal{C}_t\})^2\right]}\right.$$
$$\left. + \sqrt{\mathbb{E}\left[(f(x_t))^2\right]\mathbb{E}\left[(\mathbb{1}\{p_t \notin \mathcal{C}_t\})^2\right]}\right) \qquad \left(\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}\right) \quad (98)$$

$$\leq \sum_{t\in[T]} \sqrt{\mathbb{E}\left[\mathbb{1}\{p^* \notin \mathcal{C}_t\}\right]}\left(\sqrt{\mathbb{E}\left[(\mu_{t,p_t}(x_t))^2\right]} + \sqrt{\mathbb{E}\left[\left(\sup_{x\in\mathcal{X}}|f(x)|\right)^2\right]}\right). \quad (p^*|H_t \overset{d}{=} p_t|H_t)$$
(99)

To bound the three expectations above, we have from Lemma B.7 that $\mathbb{E}\left[\mathbb{1}\{p^* \notin \mathcal{C}_t\}\right] \leq 3T^{-5}$ and from Lemma B.11 that $\mathbb{E}\left[(\mu_{t,p_t}(x_t))^2\right] \leq \sigma^{-2}T(\bar{M} + 2M\mu_{\max} + \mu_{\max}^2 + \sigma^2)$ for all $t \in [T]$. Similarly, by Lemma B.12 we have that $\mathbb{E}\left[\left(\sup_{x\in\mathcal{X}}|f(x)|\right)^2\right] \leq \bar{M}$. Put together, we arrive at the following bound for $A_{3,1}$:

$$A_{3,1} \leq \frac{\sqrt{3}}{T}\left(\sqrt{\sigma^{-2}(\bar{M} + 2M\mu_{\max} + \mu_{\max}^2 + \sigma^2)} + \sqrt{\bar{M}/T}\right). \quad (100)$$

Bounding $A_{3,2}$: Then, $A_{3,2}$ can be bound as follows:

$$A_{3,2} = \sum_{t\in[T]} \mathbb{E}\left[(\mu_{t,p_t}(x_t) - \sqrt{\eta_T}\sigma_{t,p_t}(x_t) - f(x_t) - \epsilon_t)\,\mathbb{1}\{p_t \in \mathcal{C}_t\}\right] \qquad (\epsilon_t \perp\!\!\!\perp \mathbb{1}\{p_t \in \mathcal{C}_t\}) \quad (101)$$

$$\leq \sum_{p\in P} \mathbb{E}\left[\sum_{t\in[T]} (\mu_{t,p}(x_t) - \sqrt{\eta_T}\sigma_{t,p}(x_t) - f(x_t) - \epsilon_t)\,\mathbb{1}\{p_t = p\}\mathbb{1}\{p \in \mathcal{C}_t\}\right] \quad (102)$$

We define the final time step where prior p is selected and is in the confidence set as $\tau_p := \max\{t \in [T] : p_t = p, p \in \mathcal{C}_t\}$. Then, $\sum_{t\in[\tau_p-1]}(\mu_{t,p_t}(x_t) - \sqrt{\eta_T}\sigma_{t,p_t}(x_t) - f(x_t) - \epsilon_t)\mathbb{1}\{p_t = p\}\mathbb{1}\{p \in \mathcal{C}_t\} = G_{\tau_p}(p)$ since $\mathcal{C}_t$ is a shrinking sequence of sets. By definition of $p \in \mathcal{C}_{\tau_p}$ (Eq. (43)), $G_{\tau_p}(p) \leq \sigma\sqrt{14N_{\tau_p-1}(p)\log T}$ and

$$A_{3,2} \leq \sum_{p\in P} \mathbb{E}\left[\sigma\sqrt{14N_{\tau_p-1}(p)\log T} + \left(\mu_{\tau_p,p}(x_t) - \sqrt{\eta_T}\sigma_{\tau_p,p}(x_{\tau_p}) - f(x_{\tau_p}) - \epsilon_{\tau_p}\right)\right] \quad (103)$$

$$\leq \sum_{p\in P} \mathbb{E}\left[\sigma\sqrt{14N_T(p)\log T}\right] + \sum_{p\in P} \mathbb{E}\left[\left(\mu_{\tau_p,p}(x_{\tau_p}) - f(x_{\tau_p}) - \epsilon_{\tau_p}\right)\right] \quad (104)$$

since $\sqrt{\eta_T}\sigma_{t,p}(x) \geq 0$, and $N_t(p) \leq N_T(p)$, $\forall t, p, x \in [T] \times P \times \mathcal{X}$. To bound the left term in Eq. (104), we apply the Cauchy-Schwarz inequality such that $\sum_{p\in P}\sqrt{N_T(p)} \leq \sqrt{T|P|}$. For the right term in Eq. (104), we note that $\tau_p \in [T]$ and consider the maximum:

$$\mathbb{E}\left[\left(\mu_{\tau_p,p}(x_{\tau_p}) - f(x_{\tau_p}) - \epsilon_{\tau_p}\right)\right] \quad (105)$$

$$\leq \sum_{p\in P} \mathbb{E}\left[\max_{t\in[T]}\left(\mu_{t,p}(x_t) - f(x_t) - \epsilon_t\right)\right] \quad (106)$$

$$\leq \sum_{p\in P}\left(\mathbb{E}\left[\max_{t\in[T]}\mu_{t,p}(x_t)\right] + \mathbb{E}\left[\sup_{x\in\mathcal{X}}|f(x)|\right] + \mathbb{E}\left[\max_{t\in[T]} -\epsilon_t\right]\right) \quad (107)$$

$$\leq |P|\left(\sigma^{-1}\sqrt{T}(M + \mu_{\max} + \sigma) + M + \sigma\sqrt{2\log T}\right). \qquad \text{(Lemma B.11)} \quad (108)$$

The bound $\mathbb{E}[\max_{t\in[T]} -\epsilon_t] \leq \sigma\sqrt{2\log T}$ follows by standard results for independent zero-mean Gaussians (Boucheron et al., 2013, Section 2.5). Combined, we get that

$$A_{3,2} \leq \sigma\sqrt{14T|P|\log T} + |P|\left(\sigma^{-1}\sqrt{T}(M + \mu_{\max} + \sigma) + M + \sigma\sqrt{2\log T}\right). \tag{109}$$

□

Remark B.10. *Unlike Hong et al. (2022b); Lu et al. (2023), our definition of the confidence set $\mathcal{C}_t$ includes a condition that the excess reward $G_s(p)$ is below the threshold $\xi_s(p)$ for all $s \leq t$, not just $s = t$. This guarantees that the sets are non-increasing in size, and therefore if $p \in \mathcal{C}_t$ then $p \in C_s$ for all $s < t$. Furthermore, $\sum_{s=1}^{t-1}(\mu_{s,p_s}(x_s) - \sqrt{\eta_T}\sigma_{s,p_s}(x_s) - y_s)\mathbb{1}\{p = p_s\}\mathbb{1}\{p \in \mathcal{C}_s\} = G_t(p)$ if $p \in \mathcal{C}_{t-1}$ which is critical to go from Eq. (102) to Eq. (103). Without the time-uniform requirement, $p \notin C_s$ could hold for some $s < t$ s.t. $\mu_{s,p_s}(x_s) - \sqrt{\eta_T}\sigma_{s,p_s}(x_s) - y_s < 0$. Then, $\sum_{s=1}^{t-1}(\mu_{s,p_s}(x_s) - \sqrt{\eta_T}\sigma_{s,p_s}(x_s) - y_s)\mathbb{1}\{p = p_s\}\mathbb{1}\{p \in \mathcal{C}_s\} > G_t(p)$ which prevents bounding $G_t(p)$ by $\xi_t(p)$.*

B.3 Auxiliary lemmas

In this section, we state and prove auxiliary lemmas that bound the expectations of $\mu_{t,p}(x_t)$ and $\sup_{x\in\mathcal{X}}|f(x)|^2$.

Lemma B.11. *Let $\mu_{max} = \sup_{p,x\in P\times\mathcal{X}}|\mu_{1,p}(x)|$, $M = \mathbb{E}[\sup_{x\in\mathcal{X}}|f(x)|]$, $M_p = \mathbb{E}[\sup_{x\in\mathcal{X}}|f(x)||p^* = p]$, $M_\Delta = \max_{p\in P} M_p - \min_{p\in P} M_p$, and $\bar{M} = M^2 + 1 + \frac{M_\Delta^2}{4}$. If $k_p(x,x) : \mathcal{X}\times\mathcal{X} \mapsto [-1,1]$, $\forall p \in P$, then*

$$\mathbb{E}[\mu_{t,p_t}(x_t)^2] \leq \frac{T}{\sigma^2}(\bar{M} + 2M\mu_{max} + \mu_{max}^2 + \sigma^2), \tag{110}$$

$$\mathbb{E}[\max_{t\in[T]}\mu_{t,p}(x_t)] \leq \frac{\sqrt{T}}{\sigma}(M + \mu_{max} + \sigma). \tag{111}$$

Proof. To begin, recall that $\mu_{t,p}(x) = \mathbf{k}_{t,p}(x)^T\left(\mathbf{K}_{t,p} + \sigma^2 I\right)^{-1}(\mathbf{f}_{1:t-1} + \boldsymbol{\epsilon}_{1:t-1} - \boldsymbol{\mu}_{1:t-1,p})$ where $\mathbf{f}_{1:t} = [f(x_1),\ldots,f(x_{t-1})]^T$, $\boldsymbol{\epsilon}_{1:t} = [\epsilon_1,\ldots,\epsilon_{t-1}]$, and $\boldsymbol{\mu}_{1:t,p} = [\mu_{1,p}(x_1),\ldots,\mu_{1,p}(x_{t-1})]$. Additionally, we note that

$$\left\|\mathbf{k}_{t,p}(x)^T\left(\mathbf{K}_{t,p} + \sigma^2 I\right)^{-\frac{1}{2}}\right\|_2^2 = \mathbf{k}_{t,p}(x)^T\left(\mathbf{K}_{t,p} + \sigma^2 I\right)^{-1}\mathbf{k}_{t,p}(x) \leq k_p(x,x) \leq 1, \tag{112}$$

$\forall t,p,x \in [T]\times P\times\mathcal{X}$ by the definition of the posterior variance $\sigma_{t,p}^2(x)$ and since the posterior variance is non-negative $\sigma_{t,p}^2(x) \geq 0$. Similarly, note that $\left\|\left(\mathbf{K}_{t,p} + \sigma^2 I\right)^{-\frac{1}{2}}\right\|_2 \leq \sigma^{-1}$ since $\mathbf{K}_{t,p}$ is positive semi-definite for any t and p. Therefore,

$$\mu_{t,p}(x) \leq \left\|\mathbf{k}_{t,p}(x)^T\left(\mathbf{K}_{t,p} + \sigma^2 I\right)^{-\frac{1}{2}}\right\|_2 \cdot \left\|\left(\mathbf{K}_{t,p} + \sigma^2 I\right)^{-\frac{1}{2}}\right\|_2 \cdot \|\mathbf{f}_{1:t} + \boldsymbol{\epsilon}_{1:t} - \boldsymbol{\mu}_{1:t,p}\|_2 \tag{113}$$

$$\leq \frac{1}{\sigma}\|\mathbf{f}_{1:t} + \boldsymbol{\epsilon}_{1:t} - \boldsymbol{\mu}_{1:t,p}\|_2. \tag{114}$$

Then, we bound $\mathbb{E}[\max_{t\in[T]} \mu_{t,p}(x_t)]$.

$$\mathbb{E}[\max_{t\in[T]} \mu_{t,p}(x_t)] \le \frac{1}{\sigma}\left(\mathbb{E}\left[\max_{t\in[T]}\|\mathbf{f}_{1:t}\|_2 + \max_{t\in[T]}\|\boldsymbol{\epsilon}_{1:t}\|_2 + \max_{t\in[T]}\|\boldsymbol{\mu}_{1:t,p}\|_2\right]\right) \tag{115}$$

$$\le \frac{1}{\sigma}\mathbb{E}\left[\max_{t\in[T]}\sqrt{\sum_{s=1}^{t} f(x_s)^2} + \max_{t\in[T]}\sqrt{\sum_{s=1}^{t}\epsilon_s^2} + \max_{t\in[T]}\sqrt{\sum_{s=1}^{t}\mu_{1,p}^2(x_s)}\right] \tag{116}$$

$$\le \frac{1}{\sigma}\left(\mathbb{E}\left[\sqrt{T\sup_{x\in\mathcal{X}} f(x)^2}\right] + \mathbb{E}\left[\sqrt{\sum_{t=1}^{T}\epsilon_t^2}\right] + \sqrt{T\sup_{x\in\mathcal{X}}\mu_{1,p}^2(x)}\right) \tag{117}$$

$$\le \frac{1}{\sigma}\left(\sqrt{T}\mathbb{E}\left[\sup_{x\in\mathcal{X}}|f(x)|\right] + \sqrt{\mathbb{E}\left[\sum_{t=1}^{T}\epsilon_t^2\right]} + \sqrt{T}\mu_{\max}\right) \text{ (Jensen's ineq.)} \tag{118}$$

$$\le \frac{\sqrt{T}}{\sigma}\left(M + \sigma + \mu_{\max}\right). \qquad \left(\textstyle\sum_{t\in[T]}\epsilon_t^2 \sim \sigma^2\chi_T^2\right) \tag{119}$$

Similarly, we bound $\mathbb{E}[\mu_{t,p}(x_t)^2]$:

$$\mathbb{E}[\mu_{t,p_t}(x_t)^2] \le \frac{1}{\sigma^2}\mathbb{E}\left[\|\mathbf{f}_{1:t} + \boldsymbol{\epsilon}_{1:t} - \boldsymbol{\mu}_{1:t,p_t}\|_2^2\right] \tag{120}$$

$$\le \frac{1}{\sigma^2}\mathbb{E}\left[\sum_{t\in[T]}\left(f(x_t) + \epsilon_t - \mu_{1,p_t}(x_t)\right)^2\right] \tag{121}$$

$$= \frac{1}{\sigma^2}\mathbb{E}\Bigg[\sum_{t\in[T]} f(x_t)^2 + \epsilon_t^2 + \mu_{1,p_t}^2(x_t) + 2(f(x_t)\epsilon_t - f(x_t)\mu_{1,p_t}(x_t) - \epsilon_t\mu_{1,p_t}(x_t))\Bigg] \tag{122}$$

$$= \frac{1}{\sigma^2}\mathbb{E}\left[\sum_{t\in[T]} f(x_t)^2 + \epsilon_t^2 + \mu_{1,p_t}^2(x_t) - 2f(x_t)\mu_{1,p_t}(x_t)\right] \quad \begin{pmatrix} f(x_t), \mu_{1,p_t}(x_t) \perp\!\!\!\perp \epsilon_t, \\ \text{and } \mathbb{E}[\epsilon_t] = 0 \end{pmatrix} \tag{123}$$

$$\le \frac{1}{\sigma^2}\mathbb{E}\left[\sum_{t\in[T]}\sup_{x\in\mathcal{X}} f(x)^2 + \epsilon_t^2 + \mu_{1,p_t}^2(x_t) + 2\sup_{x\in\mathcal{X}}|f(x)||\mu_{1,p_t}(x)|\right] \tag{124}$$

$$\le \frac{1}{\sigma^2}T\left(\mathbb{E}\left[\left(\sup_{x\in\mathcal{X}}|f(x)|\right)^2\right] + \sigma^2 + \mu_{\max}^2 + 2M\mu_{\max}\right) \qquad \left(\textstyle\sum_{t\in[T]}\epsilon_t^2 \sim \sigma^2\chi_T^2\right) \tag{125}$$

$$\le \frac{1}{\sigma^2}T\left(\bar{M} + \sigma^2 + \mu_{\max}^2 + 2M\mu_{\max}\right). \qquad \text{(Lemma B.12)} \tag{126}$$

□

Lemma B.12. *Let* $M = \mathbb{E}[\sup_{x\in\mathcal{X}}|f(x)|]$, $M_p = \mathbb{E}[\sup_{x\in\mathcal{X}}|f(x)||p^* = p]$, *and* $M_\Delta = \max_{p\in P} M_p - \min_{p\in P} M_p$. *If* $k_p(x,x) : \mathcal{X}\times\mathcal{X} \mapsto [-1,1]$, $\forall p \in P$, *then*

$$\mathbb{E}\left[\left(\sup_{x\in\mathcal{X}}|f(x)|\right)^2\right] \le M^2 + 1 + \frac{M_\Delta^2}{4} =: \bar{M}. \tag{127}$$

Proof. First, by the variance formula $\mathbb{V}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$,

$$\mathbb{E}[(\sup_{x\in\mathcal{X}}|f(x)|)^2] = M^2 + \mathbb{V}\left(\sup_{x\in\mathcal{X}}|f(x)|\right). \tag{128}$$

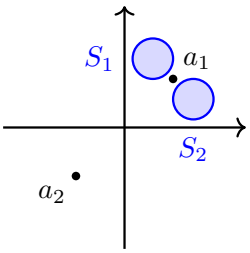


Figure 8: Potential counterexample to Eq (9) in Hong et al. (2022b). The blue regions represent when the event E_0 holds and the black dots represent the two arms a_1 and a_2.

The variance $\mathbb{V}[\sup_{x\in\mathcal{X}}|f(x)|]$ can be bounded by the law of total variance as follows:

$$\mathbb{V}\left[\sup_{x\in\mathcal{X}}|f(x)|\right] = \mathbb{E}_{p^*}\left[\mathbb{V}\left(\sup_{x\in\mathcal{X}}|f(x)||p^*\right)\right] + \mathbb{V}_{p^*}\left(\underbrace{\mathbb{E}\left[\sup_{x\in\mathcal{X}}|f(x)||p^*\right]}_{M_{p^*}:=}\right) \tag{129}$$

$$\overset{(a)}{\leq} \mathbb{E}_{p^*}\left[\sup_{x\in\mathcal{X}}\sigma^2_{1,p^*}(x)\right] + \mathbb{V}_{p^*}(M_{p^*}) \tag{130}$$

$$\leq 1 + \frac{(\max_p M_p - \min_p M_p)^2}{4}. \qquad \left(\begin{matrix}\sigma^2_{1,p}(x)\leq 1,\\ \text{and Popoviciu's ineq.}\end{matrix}\right) \tag{131}$$

Note that $\mathbb{V}\left(\sup_{x\in\mathcal{X}}|f(x)|\,\middle|\,p^*\right) \leq \sup_{x\in\mathcal{X}}\sigma^2_{1,p^*}(x)$, used in (a), follows from the Gaussian Poincaré inequality applied to $\sup_{x\in\mathcal{X}}|f(x)|$, see Boucheron et al. (2013, Theorem 3.20 and Exercise 3.24). □

C Technical issues with MixTS regret bound in the linear setting

Theorem 1 in Hong et al. (2022b) provides a regret bound for MixTS in the linear setting. The linear setting assumes that the true parameter $\theta^*|S_* \sim \mathcal{N}(\theta_{0,S_*}, \Sigma_{0,S_*})$ where the latent state S_* is sampled from a discrete prior P_1. The proof of Theorem 1 in Hong et al. (2022b) contains non-obvious steps that seem difficult to motivate. We use the notation of Hong et al. (2022b).

First, Eq. (9) in Hong et al. (2022b) uses the TS property that the true prior and optimal arm is equal in distribution to the selected prior and selected arm given the history: $A_{t,*}^\top\bar{\theta}_{t,S_*}|H_t \overset{d}{=} A_t^\top\bar{\theta}_{t,S_t}|H_t$ (equivalent to $\mu_{t,p^*}(x^*)|H_t \overset{d}{=} \mu_{t,p_t}(x_t)|H_t$ in our notation). However, Eq. (9) additionally conditions on the event $E_0 = \{\|\theta_* - \theta_{0,S_*}\|_{\Sigma^{-1}_{0,S_*}} \leq \sqrt{2d\log(dn)}\}$ where θ_* lies close to its prior mean. The TS property does not hold under this event since it modifies the distribution of the linear parameter θ^* but not the sampled parameters θ_t, thus changing the distribution of the optimal arm $A_{t,*}$ but not the selected arm A_t. Consider the example in Fig. 8, if E_0 holds then θ_* lies in the blue regions and thus a_2 is optimal w.p. 0. If E_0^c holds, then a_2 is optimal with a non-zero probability. However, MixTS is oblivious to E_0 given the history and thus $A_{t,*}|H_t, E_0 \overset{d}{\neq} A_t|H_t$. This counterexample illustrates the overall idea but we have not validated that the scale of the arms and the blue regions are feasible.

Second, five lines above Eq. (9) in Hong et al. (2022b), it is stated that the *regret* is upper-bounded by a constant M whenever E_0 occurs. However, from Eq (9) to the first term in step 3 of their analysis (page 15), the bound of M is applied implicitly to $A_t^\top\bar{\theta}_{t,S_t} - A_t^\top\theta_*$ without motivation. For the setting with bounded rewards, then $\bar{\theta}_{t,S_t}$ is also bounded but for Gaussian rewards $\bar{\theta}_{t,S_t}$ can be unbounded.

Third, the second term in Eq. (9) contains the indicator function $\mathbf{1}\{E_0\}$: $\mathbb{E}[(A_t^\top\bar{\theta}_{t,S_t} - A_t^\top\theta_*)\mathbf{1}\{E_0\}]$. In step 3 (page 15), this indicator function is dropped without motivation: $\mathbb{E}[\langle A_t^\top\bar{\theta}_{t,S_t} - A_t^\top\theta_*\rangle_M]$ where $\langle\cdot\rangle_M = \min(\cdot, M)$ for the bound M. If the expression inside is non-negative w.p. 1, then this step would be valid but this is not the case.

Fourth, from our understanding, the final equation on page 15 adds and subtracts the confidence bound and adds a zero-mean Gaussian inside a minimum. However, adding a zero-mean Gaussian inside a minimum reduces the expectation but the analysis seems to assume that it would increase the expectation. I.e. it is seemingly assumed that $\mathbb{E}[\min(M, X)] \leq \mathbb{E}[\min(M, X + \epsilon_t)$ for a constant M and random variable X. However, the reverse inequality is true.

D Description of kernels

The RBF kernel, $k(x, \tilde{x}) = \exp(-||x - \tilde{x}||^2/\ell^2)$ guarantees that f is smooth. The lengthscale parameter $\ell > 0$ determines how quickly f changes, smaller values lead to more fluctuations. The rational quadratic (RQ) kernel $k(x, \tilde{x}) = \left(1 + \frac{||x-\tilde{x}||^2}{2\alpha\ell^2}\right)^{-\alpha}$ where $\alpha > 0$ is a mixture of RBF kernels with varying lengthscales. The Matérn kernel (Matérn, 1986) $k(x, \tilde{x}) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}||x-\tilde{x}||}{\ell}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}||x-\tilde{x}||}{\ell}\right)$ where $\nu > 0$ is the smoothness parameter that imposes that f is k-times differentiable if $\nu > k$ for integer k. The functions $\Gamma(\nu)$ and K_v correspond to the gamma function and a modified Bessel function (Williams & Rasmussen, 2006). The periodic kernel $k(x, \tilde{x}) = \exp\left(-\frac{1}{2}\sum_{i=1}^{d} \sin^2(\frac{\pi}{\rho}(x_i - \tilde{x}_i))/\ell\right)$ generates smooth and periodic functions with period $\rho > 0$ (Mackay, 1998). The linear kernel $k(x, \tilde{x}) = vx^\top \tilde{x}$ generates linear functions where v is the variance parameter.

E Additional experimental details

In this section, we provide some additional details about the experiments. All experiments were run on a compute cluster with a mix of GPUs (Nvidia A100, A40, T4 and V100). The GPU used was decided based on availability at the time and no implementation depends on a specific GPU. The algorithms were run in parallel in a single job for each seed. Each job in the synthetic and real-world data experiments ran for approximately 5 minutes. With the 500 seeds, this leads to a combined 250 GPU-hours. Running all algorithms for one seed in the lengthscale scaling experiment with $|P| = 128$ took approximately 40 minutes, and is in total equivalent to around 330 GPU-hours.

E.1 Synthetic experiments

For the kernel experiment, all kernels use a lengthscale of 1.0 and are scaled s.t. $k(x, \tilde{x}) \leq 1$. In addition, the mean function for all priors is zero everywhere. For the subspace experiment, the total dimensions $d = 16$ but each prior p_i assumes $f(x)$ depends on $d_s = 4$ subdimensions: $[i, i+1, i+2, i+3]$ for $i \in [5]$. Dimensions larger than 5 are wrapped around 1, i.e. $((j-1) \mod 5)+1$, such that the priors are equally difficult to distinguish and optimize. The prior elimination methods use $\delta = 0.05$ across all experiments, including the oracle methods. During every iteration t, SCoreBO samples M priors from the hyperposterior P_t and samples N optimizers x^*, f^* for each prior sampled through posterior sampling. In all experiments, we use $M = 16$ and $N = 12$ for SCoreBO. While our M value matches that of Hvarfner et al. (2023, Table 3), we increase the N value from 8 to 12. We use the implementation of the SCoreBO acquisition function in BoTorch (Community) (Balandat et al., 2020). To make the implementation fast with GPUs, we set `linear_operator.settings.stable_qr_cpu_threshold` to 8 in order to avoid QR-factorization being performed on CPU (Gardner et al., 2018; Pleiss et al., 2022, 2025). To avoid out of memory issues, we replace the default `torch.matmul` in `DefaultPredictionStrategy._exact_predictive_covar_inv_quad_form_root` (from `gpytorch.models.exact_prediction_strategies`) with an equivalent `torch.einsum` (Gardner et al., 2018). Since the priors in our experiments are discrete, we compute the hyperposterior exactly and sample from it directly. Similarly, the expectation with respect to the hyperposterior is computed exactly for EEI.

E.2 Real-world data experiments

As discussed in Section 5, each dataset is split into a training and test set. The training sets are split into separate buckets to define our priors. For each bucket p, we compute the empirical mean $\hat{\mu}_p$ and covariance $\hat{\Sigma}_p$ which defines the prior $\mathcal{GP}(\hat{\mu}_p, \hat{\Sigma}_p)$. The buckets in the Intel data corresponds to the

12 days in the training dataset. For the PeMS data, each hour between 06:00 and 13:00 defines one bucket, giving 7 priors. For the daily precipitation data, each month in the year constitutes a bucket, yielding 12 priors. When running the experiments, we select a measurement of all sensors from the test data uniformly at random. The selected measurements correspond to the unknown function $f(x)$ where x is the sensor index and the goal is then to identify sensors measuring large temperatures, small speeds or high precipitation respectively for the three datasets. When the algorithms select an arm to evaluate, we add Gaussian noise with variance σ^2 around 5% of the signal variance, similar to Srinivas et al. (2012); Bogunovic et al. (2016).

For all the real-world datasets, sensors containing any null measurements are filtered out.

The Intel Berkeley dataset consists of measurements from 46 temperature sensors across 19 days. The training set consists of the first 12 days of measurements and the remaining 7 days constitute the test set. The noise variance is set to $\sigma^2 = 0.7^2$.

The PeMS data is considered in the public domain (California Department of Transportation, 2026) and consists of measurements from 211 sensors along the I-880 highway from all of 2023. The goal is to find the sensors with low speeds to identify congestions. We negate the speed values to obtain a maximization problem. We use the 5-min averages provided by PeMS. Data between 2023-01-01 and 2023-09-01 is put into the training set whilst the data until 2023-12-31 is put into the test set. The noise variance is set to $\sigma^2 = 2.25^2$.

The PNW precipitation data consists of daily precipitation data from 1949 to 1994 across 167 50×50 km regions in the Pacific Northwest. The goal is to find the region with the highest precipitation for any given day. The training data consists of the measurements made prior to 1980 and the test data consists of the measurements between 1980 and 1994. The original data is stated to be given in mm/day however the data seems to be off by a factor of 10. We rescale the data to a log-scale using $\log(\cdot/10 + 0.1)$, similar to Krause et al. (2008). The noise variance is set to $\sigma^2 = 0.41^2$.

In the Intel experiment, we removed one outlier seed. All methods had a final cumulative regret around 6000°C on this instance, note that the average for the worst performing model across the other seeds was ≈ 250°C. The outlier is shown in Fig. 9. We can see that one of the sensors display very high temperatures compared to all other sensors, which is why all methods performed poorly on this seed. It should be noted that many of the sensors in the Intel data logged degrees above 100°C after a certain time - likely due to sensor failure rather than boiling temperatures in an office environment. Also note that these days were excluded from both our training and test data. The outlier could be an indication that this particular sensor was starting to fail earlier than others.

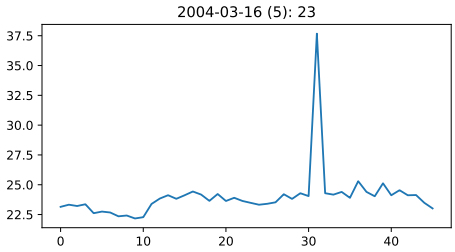


Figure 9: Removed sample from the test data in the Intel experiment. One of the sensors displays very high temperatures.

F Additional experimental results

In this section, we provide some additional experimental results.

First, we provide the average total regret for the synthetic and real-world data experiments in Table 2. We observe that HP-GP-TS either has the lowest regret or is within 1 standard error of the algorithm with the lowest regret across all the experiments. In Fig. 2, it can be noted that EEI had low regret early in the synthetic experiment but HP-GP-TS either catches up or almost catches up later in the experiments. We compare both algorithms with an extended time horizon of $T = 1500$, the results are shown in Fig. 10 and Table 3. With the extended time horizon, HP-GP-TS achieves the lowest regret

Table 2: Average total regret and $\pm$ 1 standard error for the synthetic and real-world data experiments. The algorithms with the lowest regret (excluding oracle algorithms) are highlighted in bold.

Algorithm	Synthetic			Real-world data		
	Kernel	Lengthscale	Subspace	Intel	PeMS	PNW Precip.
MAP GP-TS	84.3 ± 8.4	$\mathbf{30.2 \pm 1.2}$	$\mathbf{87.2 \pm 1.0}$	73.8 ± 7.7	1635.0 ± 129.3	$\mathbf{178.4 \pm 6.9}$
HP-GP-TS	$\mathbf{39.2 \pm 1.4}$	$\mathbf{31.4 \pm 1.0}$	$\mathbf{88.3 \pm 0.9}$	$\mathbf{54.1 \pm 3.0}$	$\mathbf{1327.8 \pm 107.9}$	$\mathbf{167.7 \pm 5.2}$
PE-GP-TS	62.0 ± 0.6	61.8 ± 0.5	177.1 ± 1.4	106.5 ± 2.1	$\mathbf{1214.2 \pm 81.5}$	200.9 ± 4.0
PE-GP-UCB	121.6 ± 1.2	114.2 ± 0.6	389.0 ± 1.5	173.0 ± 2.7	2159.2 ± 48.4	506.2 ± 2.6
Oracle GP-TS	35.0 ± 1.1	28.1 ± 0.8	86.0 ± 1.0			
Oracle GP-UCB	68.5 ± 1.9	48.3 ± 1.2	217.3 ± 1.0			
SCoreBO	180.4 ± 7.7	180.8 ± 5.8	106.6 ± 0.9	256.8 ± 9.3	3460.2 ± 163.7	861.3 ± 21.0
EEI	$\mathbf{39.0 \pm 2.6}$	$\mathbf{30.1 \pm 2.1}$	$\mathbf{88.3 \pm 4.2}$	$\mathbf{51.6 \pm 4.8}$	1664.1 ± 137.7	196.5 ± 12.3

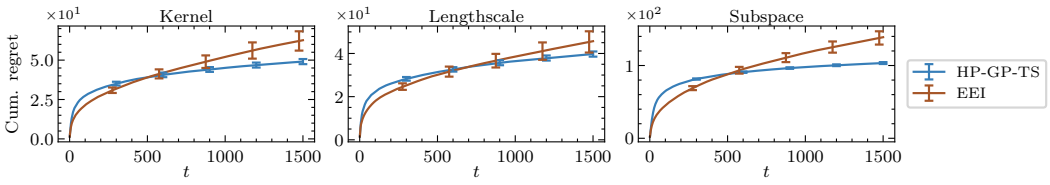


Figure 10: Cumulative regret for synthetic experiments extended time horizon $T = 1500$ with varying kernel (left), lengthscale (center) and mean function (right). Errorbars correspond to ± 1 standard error.

across all synthetic experiments. Although, EEI is still within 1 standard error on the lengthscale experiment.

Next, we include the mean number of priors in P_t for all experiments in Fig. 11. Similarly, we include the average entropy of the hyperposterior for all experiments in Fig. 12. For the lengthscale, subspace, PeMS and PNW precipitation experiments, hardly any priors are eliminated. In contrast, the hyperposterior entropy concentrates rapidly across all experiments with the subspace and PNW precipitation having the most and least concentrated hyperposteriors.

We include the full set of confusion matrices for the lengthscale and subspace experiments in Fig. 14. In the lengthscale experiments, we observe that PE-GP-UCB and -TS oversample the shortest lengthscale. This is similar to the kernel experiment where the Matérn 3/2 kernel was also oversampled. However, we see that HP-GP-TS and MAP GP-TS do not suffer from this optimistic bias. In the subspace experiment, HP- and MAP GP-TS have an accuracy of around 96% whereas PE-GP-TS and -UCB have accuracies 30% and 36% respectively. Even though PE-GP-UCB has a higher accuracy than PE-GP-TS, it still has significantly higher regret. Additionally, the priors are equivalent up to coordinate permutations and therefore generate functions that are equally difficult to optimize. Unlike the kernel and lengthscale experiments, the PE-methods do not oversample any specific prior but commit too much time to exploring along the irrelevant dimensions.

In Fig. 13, the sum of variances $\sum_{s=1}^{t} \sigma^2_{t,p^*}(x^*)$ is shown for PE-GP-TS in the synthetic experiments. This term is unbounded in Theorem 4.2 but we observe experimentally that it is sublinear across all three experiments.

In Tables 4 and 5, the total regret for the lengthscale and subspace scaling experiments are shown.

Table 3: Average total regret and $\pm$ 1 standard error for the synthetic experiments with longer horizon $T = 1500$. The algorithms with the lowest regret (excluding oracle algorithms) are highlighted in bold.

Algorithm	Kernel	Lengthscale	Subspace
HP-GP-TS	$\mathbf{49.1 \pm 1.6}$	$\mathbf{39.7 \pm 1.2}$	$\mathbf{103.4 \pm 1.3}$
EEI	62.6 ± 6.2	$\mathbf{45.7 \pm 5.0}$	138.9 ± 9.2

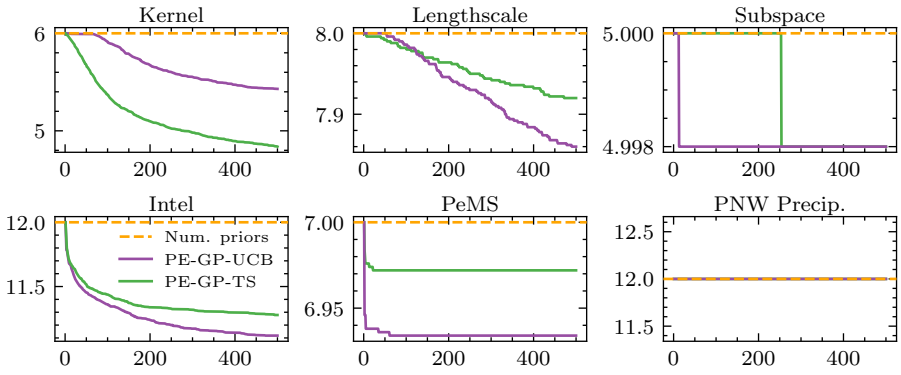


Figure 11: Mean number of priors remaining in P_t over time for PE-GP-UCB and -TS.

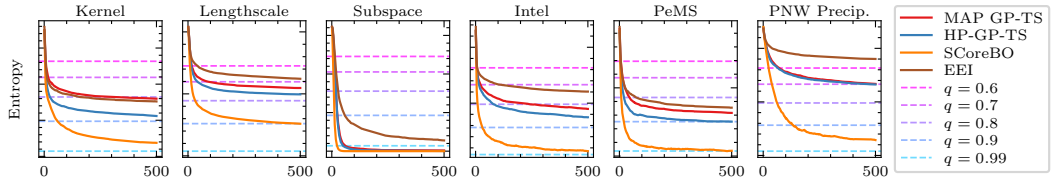


Figure 12: Average entropy in the hyperposterior P_t over time for HP- and MAP GP-TS. The dashed reference values correspond to entropies of discrete distributions with prob. q on one choice and prob. $\frac{1-q}{|P|-1}$ on the other $|P| - 1$ choices.

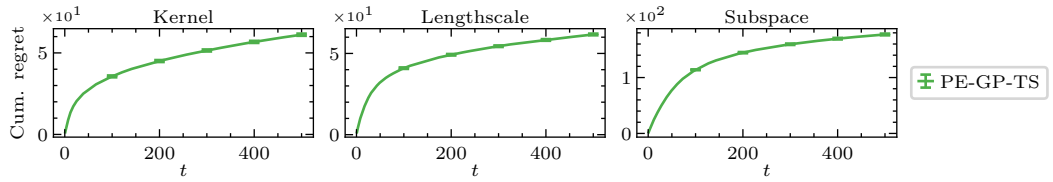
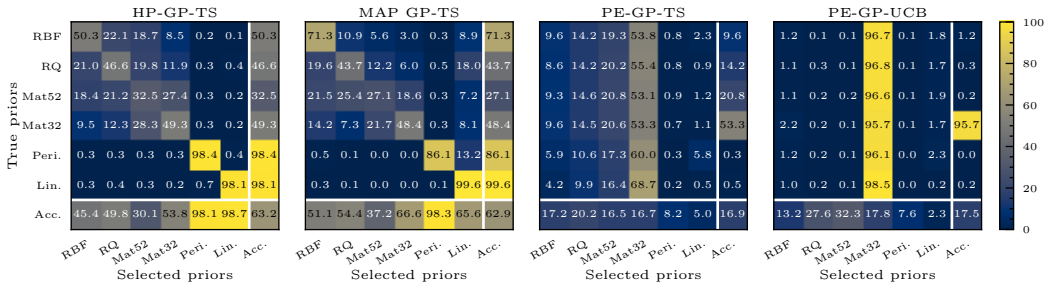


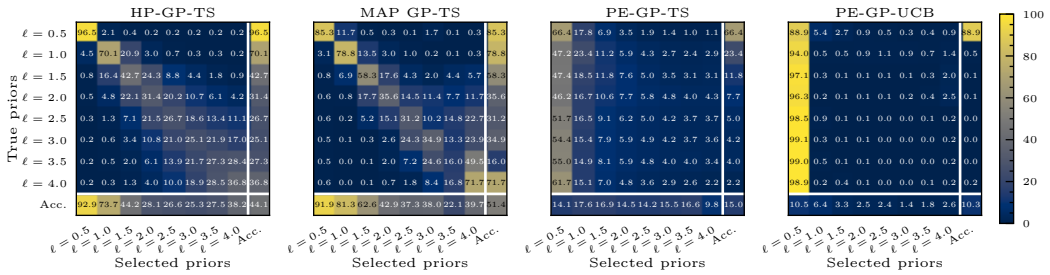
Figure 13: The sum of variances for the optimum x^* under the true prior p^* up to time t ($\sum_{s=1}^{t} \sigma^2_{t,p^*}(x^*)$) for PE-GP-TS.

Table 4: Average total regret and ± 1 standard error for the lengthscale experiment as $|P|$ increases. The algorithms with the lowest regret (excluding oracle algorithms) are highlighted in bold.

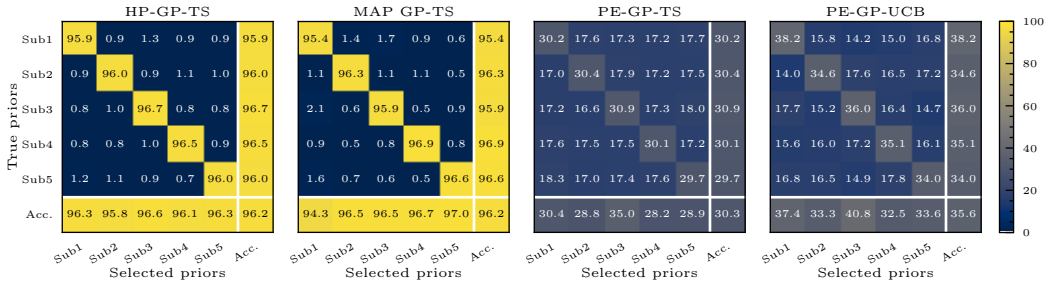
Algorithm	Lengthscales, $\lvert P\rvert$				
	8	16	32	64	128
MAP GP-TS	**30.2 ± 1.2**	**32.4 ± 2.5**	**32.5 ± 2.1**	**28.7 ± 1.1**	**30.8 ± 1.9**
HP-GP-TS	**31.4 ± 1.0**	**31.7 ± 0.9**	**30.8 ± 0.8**	**30.7 ± 1.0**	**31.0 ± 1.4**
PE-GP-TS	61.8 ± 0.5	61.3 ± 0.5	62.2 ± 0.5	62.4 ± 0.4	64.3 ± 0.4
PE-GP-UCB	114.2 ± 0.6	114.8 ± 0.6	115.5 ± 0.6	114.5 ± 0.6	114.8 ± 0.6
Oracle GP-TS	28.1 ± 0.8	26.4 ± 0.8	27.3 ± 0.8	26.5 ± 0.7	25.7 ± 0.7
Oracle GP-UCB	48.3 ± 1.2	46.9 ± 1.1	48.4 ± 1.1	46.5 ± 1.0	45.6 ± 1.0
SCoreBO	180.8 ± 5.8	240.3 ± 6.3	277.2 ± 6.8	281.6 ± 6.7	283.9 ± 6.8
EEI	**30.1 ± 2.1**	**30.9 ± 2.2**	**29.4 ± 2.2**	**30.8 ± 2.6**	**32.1 ± 2.8**



(a) Kernel experiment



(b) Lengthscale experiment



(c) Subspace experiment

Figure 14: Confusion matrices for the true prior p^* and p_t across all time steps of the synthetic experiments.

Table 5: Average total regret and ± 1 standard error for the subspace experiment as $|P|$ increases. The algorithms with the lowest regret (excluding oracle algorithms) are highlighted in bold.

Algorithm	Subspaces, $\lvert P\rvert$			
	5	8	12	16
MAP GP-TS	**87.2 ± 1.0**	89.9 ± 1.1	89.1 ± 0.9	90.9 ± 1.2
HP-GP-TS	**88.3 ± 0.9**	88.8 ± 0.9	89.5 ± 0.9	90.8 ± 0.9
PE-GP-TS	177.1 ± 1.4	269.5 ± 1.9	344.7 ± 2.3	396.9 ± 2.5
PE-GP-UCB	389.0 ± 1.5	526.0 ± 1.8	622.4 ± 2.3	688.0 ± 2.7
Oracle GP-TS	86.0 ± 1.0	84.1 ± 0.9	84.6 ± 1.0	84.8 ± 1.0
Oracle GP-UCB	217.3 ± 1.0	218.2 ± 1.0	218.6 ± 1.0	218.9 ± 0.9
SCoreBO	106.6 ± 0.9	108.2 ± 0.8	108.9 ± 0.7	109.5 ± 0.7
EEI	**88.3 ± 4.2**	**81.3 ± 3.8**	**82.5 ± 3.6**	**81.3 ± 3.8**

Table 6: Average total regret and ± 1 standard error for the misspecification experiment for HP-GP-TS. The algorithms with the lowest regret (excluding the oracle algorithm) are highlighted in bold.

	P = Short	P = Uniform	P = Long	Oracle GP-TS
Q = Short	**52.1** $\pm$ **1.0**	**53.5** $\pm$ **1.3**	58.3 $\pm$ 1.8	48.8 $\pm$ 1.0
Q = Uniform	**31.7** $\pm$ **0.7**	**30.8** $\pm$ **1.1**	**33.3** $\pm$ **1.7**	26.3 $\pm$ 0.8
Q = Long	22.2 $\pm$ 0.3	20.0 $\pm$ 0.4	**18.8** $\pm$ **0.7**	16.2 $\pm$ 0.4

In Table 6, we provide the average total regret for the misspecification experiment for HP-GP-TS with unknown lengthscale. We note that misspecification does increase the regret however not by a big margin.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: To our knowledge, PE-GP-TS has not been considered in the literature before and our regret analysis is found in Section 4 with complete proofs in Appendix B. We analyze the regret of HP-GP-TS in Section 4 and provide a sublinear regret bound in Theorem 4.4. Section 4 contrasts our proof with those of Hong et al. (2022b); Lu et al. (2023) and describes their respective issues. We provide a more detailed discussion of the issues with the proof of Hong et al. (2022b) in Appendix C. The TS-based algorithms are evaluated in Section 5, with results shown in Figs. 2 to 5 and 7. Further results are presented in Appendix F, in particular, Table 2 shows that HP-GP-TS has the lowest regret or is within 1 standard error of the algorithm with the lowest regret across the synthetic and real-world data experiments.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Some of the limitations discussed in the paper include outlining that we only consider a finite set of priors (Section 2), highlighting the computational complexity of HP-GP-TS (Section 3.2), and noting that sublinearity is not established in Theorem 4.2 for PE-GP-TS.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.

- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The setting considered is described in Section 2 with the algorithms described in Section 3. The overall idea of the theoretical proofs are presented in Section 4 with complete proofs provided in Appendix B.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The experimental setup is described in Section 5 and more specific details needed for reproducibility are provided in Appendix E. In addition the code is available at https://github.com/MachineLearning-DecisionMaking/Adaptive-Prior-Selection-GP-Bandits.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.

(b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The code is available at https://github.com/MachineLearning-DecisionMaking/Adaptive-Prior-Selection-GP-Bandits, and includes instructions on how to access and preprocess the data.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The experimental setup is described in Section 5 and more specific details are provided in Appendix E.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: The plots and tables contain error bars corresponding to $\pm$ 1 standard error.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: In Appendix E, we describe the computational setup used and provide estimates for the compute used. Note that all of our experiments were conducted with 500 repetitions and that reproducing our experiments can be done with significantly less compute. In addition, the experiments do not depend on any specific hardware, such as GPUs, and can be reproduced on CPUs.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The research in the paper is mainly theoretical and does not involve human subjects. The societal impact is discussed in question 10. The datasets used comply with the available licenses. The original website hosting the PNW precipitation dataset was taken down sometime in 2025. Based on personal communications, the dataset was not taken down by the original author and usage of it complies with the policy of deprecated datasets. To ensure reproducibility, we provide the dataset in our code repository.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: The work of this paper studies general algorithms for efficient blackbox optimization when parameters of the blackbox cannot be uniquely determined beforehand. Overall, blackbox optimization is used in many different disciplines where the typical goal is to improve allocation of resources or increase resource output. Overall, we believe such applications are beneficial to society at large and believe our work will have a positive impact but we believe that an extensive discussion is not warranted.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: For the three real-world datasets, we cite the original papers (Chen et al., 2001; Widmann & Bretherton, 2000) and provide URLs in the corresponding references (Madden et al., 2004; California Department of Transportation, 2024; Widmann & Bretherton, 1999). The Intel and PNW precipitation datasets do not specify any licenses or terms of use. The PNW precipitation dataset is no longer available from here, but a snapshot is available here. With the permission of the original authors, we provide the PNW precipitation dataset in our code repository.

    The PEMS dataset is considered in the public domain (California Department of Transportation, 2026).

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The provided code repository contains descriptions of the preprocessing procedure alongside instructions for running the experiments.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs are not a component of the core methods in this research.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.