Nonmyopic ϵ-Bayes-Optimal Active Learning of Gaussian Processes

Trong Nghia Hoang[†] NGHIAHT@COMP.NUS.EDU.SG
Kian Hsiang Low[†] LOWKH@COMP.NUS.EDU.SG
Patrick Jaillet[§] JAILLET@MIT.EDU
Mohan Kankanhalli[†] MOHAN@COMP.NUS.EDU.SG
[†]Department of Computer Science, National University of Singapore, Republic of Singapore
[§]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, USA

Abstract

A fundamental issue in active learning of Gaussian processes is that of the exploration-exploitation trade-off. This paper presents a novel nonmyopic *ϵ-Bayes-optimal active learning* (ϵ-BAL) approach that jointly and naturally optimizes the trade-off. In contrast, existing works have primarily developed myopic/greedy algorithms or performed exploration and exploitation separately. To perform active learning in real time, we then propose an anytime algorithm based on ϵ-BAL with performance guarantee and empirically demonstrate using synthetic and real-world datasets that, with limited budget, it outperforms the state-of-the-art algorithms.

1. Introduction

Active learning has become an increasingly important focal theme in many environmental sensing and monitoring applications (e.g., precision agriculture, mineral prospecting (Low et al., 2007), monitoring of ocean and freshwater phenomena like harmful algal blooms (Dolan et al., 2009; Podnar et al., 2010), forest ecosystems, or pollution) where a high-resolution *in situ* sampling of the spatial phenomenon of interest is impractical due to prohibitively costly sampling budget requirements (e.g., number of deployed sensors, energy consumption, mission time): For such applications, it is thus desirable to select and gather the *most informative* observations/data for modeling and predicting the spatially varying phenomenon subject to some budget constraints, which is the goal of active learning and also known as the *active sensing* problem.

To elaborate, solving the active sensing problem amounts to deriving an optimal sequential policy that plans/decides the most informative locations to be observed for minimizing the predictive uncertainty of the unobserved areas of a phenomenon given a sampling budget. To achieve this, many existing active sensing algorithms (Cao et al., 2013; Chen et al., 2012; 2013b; Krause et al., 2008; Low et al., 2008; 2009; 2011; 2012; Singh et al., 2009) have modeled the phenomenon as a *Gaussian process* (GP), which allows its spatial correlation structure to be formally characterized and its predictive uncertainty to be formally quantified (e.g., based on mean-squared error, entropy, or mutual information criterion). However, they have assumed the spatial correlation structure (specifically, the parameters defining it) to be known, which is often violated in real-world applications, or estimated crudely using sparse prior data. So, though they aim to select sampling locations that are optimal with respect to the assumed or estimated parameters, these locations tend to be sub-optimal with respect to the true parameters, thus degrading the predictive performance of the learned GP model.

In practice, the spatial correlation structure of a phenomenon is usually not known. Then, the predictive performance of the GP modeling the phenomenon depends on how informative the gathered observations/data are for both parameter estimation as well as spatial prediction given the true parameters. Interestingly, as revealed in previous geostatistical studies (Martin, 2001; Müller, 2007), policies that are efficient for parameter estimation are not necessarily efficient for spatial prediction with respect to the true model. Thus, the active sensing problem involves a potential trade-off between sampling the most informative locations for spatial prediction given the current, possibly incomplete knowledge of the model parameters (i.e., exploitation) vs. observing locations that gain more information about the parameters (i.e., exploration):

How then does an active sensing algorithm trade off between these two possibly conflicting sampling objectives?

To tackle this question, one principled approach is to frame active sensing as a sequential decision problem that jointly and naturally optimizes the above exploration-exploitation

Proceedings of the 31^{st} International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32.

trade-off while maintaining a Bayesian belief over the model parameters. This intuitively means a policy that biases towards observing informative locations for spatial prediction given the current model prior may be penalized if it entails a highly dispersed posterior over the model parameters. So, the resulting induced policy is guaranteed to be optimal in the expected active sensing performance. Unfortunately, such a nonmyopic Bayes-optimal policy cannot be derived exactly due to an uncountable set of candidate observations and unknown model parameters (Solomon & Zacks, 1970). As a result, most existing works (Diggle, 2006; Houlsby et al., 2012; Park & Pillow, 2012; Zimmerman, 2006; Ouyang et al., 2014) have circumvented the trade-off by resorting to the use of myopic/greedy (hence, sub-optimal) policies.

To the best of our knowledge, the only notable nonmyopic active sensing algorithm for GPs (Krause & Guestrin, 2007) advocates tackling exploration and exploitation separately, instead of jointly and naturally optimizing their trade-off, to sidestep the difficulty of solving the Bayesian sequential decision problem. Specifically, it performs a probably approximately correct (PAC)-style exploration until it can verify that the performance loss of greedy exploitation lies within a user-specified threshold. But, such an algorithm is sub-optimal in the presence of budget constraints due to the following limitations: (a) It is unclear how an optimal threshold for exploration can be determined given a sampling budget, and (b) even if such a threshold is available, the PAC-style exploration is typically designed to satisfy a worst-case sample complexity rather than to be optimal in the expected active sensing performance, thus resulting in an overly-aggressive exploration (Section 4.1).

This paper presents an efficient decision-theoretic planning approach to nonmyopic active sensing/learning that can still preserve and exploit the principled Bayesian sequential decision problem framework for jointly and naturally optimizing the exploration-exploitation trade-off (Section 3.1) and consequently does not incur the limitations of the algorithm of Krause & Guestrin (2007). In particular, although the exact Bayes-optimal policy to the active sensing problem cannot be derived (Solomon & Zacks, 1970), we show that it is in fact possible to solve for a nonmyopic ϵ-*Bayes-optimal active learning* (ϵ-BAL) policy (Sections 3.2 and 3.3) given a user-defined bound ϵ, which is the main contribution of our work here. In other words, our proposed ϵ-BAL policy can approximate the optimal expected active sensing performance arbitrarily closely (i.e., within an arbitrary loss bound ϵ). In contrast, the algorithm of Krause & Guestrin (2007) can only yield a sub-optimal performance bound[1]. To meet the real-time requirement in time-critical applications, we then propose an asymptotically ϵ-optimal, branch-and-bound anytime algorithm based on ϵ-BAL with performance guarantee (Section 3.4). We empirically demonstrate using both synthetic and real-world datasets that, with limited budget, our proposed approach outperforms state-of-the-art algorithms (Section 4).

[1]Its induced policy is guaranteed not to achieve worse than the optimal performance by more than a factor of $1/e$.

2. Modeling Spatial Phenomena with Gaussian Processes (GPs)

The GP can be used to model a spatial phenomenon of interest as follows: The phenomenon is defined to vary as a realization of a GP. Let $\mathcal{X}$ denote a set of sampling locations representing the domain of the phenomenon such that each location $x \in \mathcal{X}$ is associated with a realized (random) measurement z_x (Z_x) if x is observed/sampled (unobserved). Let $Z_\mathcal{X} \triangleq \{Z_x\}_{x\in\mathcal{X}}$ denote a GP, that is, every finite subset of $Z_\mathcal{X}$ has a multivariate Gaussian distribution (Chen et al., 2013a; Rasmussen & Williams, 2006). The GP is fully specified by its *prior* mean $\mu_x \triangleq \mathbb{E}[Z_x]$ and covariance $\sigma_{xx'|\lambda} \triangleq \mathrm{cov}[Z_x, Z_{x'}|\lambda]$ for all $x, x' \in \mathcal{X}$, the latter of which characterizes the spatial correlation structure of the phenomenon and can be defined using a covariance function parameterized by λ. A common choice is the squared exponential covariance function:

$$\sigma_{xx'|\lambda} \triangleq (\sigma_s^\lambda)^2 \exp\left(-\frac{1}{2}\sum_{i=1}^{P}\left(\frac{[s_x]_i - [s_{x'}]_i}{\ell_i^\lambda}\right)^2\right) + (\sigma_n^\lambda)^2\delta_{xx'}$$

where $[s_x]_i$ ($[s_{x'}]_i$) is the i-th component of the P-dimensional feature vector s_x ($s_{x'}$), the set of realized parameters $\lambda \triangleq \left\{\sigma_n^\lambda, \sigma_s^\lambda, \ell_1^\lambda, \ldots, \ell_P^\lambda\right\} \in \Lambda$ are, respectively, the square root of noise variance, square root of signal variance, and length-scales, and $\delta_{xx'}$ is a Kronecker delta that is 1 if $x = x'$ and 0 otherwise.

Supposing λ is known and a set $z_\mathcal{D}$ of realized measurements is available for some set $\mathcal{D} \subset \mathcal{X}$ of observed locations, the GP can exploit these observations to predict the measurement for any unobserved location $x \in \mathcal{X} \setminus \mathcal{D}$ as well as provide its corresponding predictive uncertainty using the Gaussian predictive distribution $p(z_x|z_\mathcal{D}, \lambda) \sim \mathcal{N}(\mu_{x|\mathcal{D},\lambda}, \sigma_{xx|\mathcal{D},\lambda})$ with the following *posterior* mean and variance, respectively:

$$\mu_{x|\mathcal{D},\lambda} \triangleq \mu_x + \Sigma_{x\mathcal{D}|\lambda}\Sigma_{\mathcal{D}\mathcal{D}|\lambda}^{-1}(z_\mathcal{D} - \mu_\mathcal{D}) \quad (1)$$

$$\sigma_{xx|\mathcal{D},\lambda} \triangleq \sigma_{xx|\lambda} - \Sigma_{x\mathcal{D}|\lambda}\Sigma_{\mathcal{D}\mathcal{D}|\lambda}^{-1}\Sigma_{\mathcal{D}x|\lambda} \quad (2)$$

where, with a slight abuse of notation, $z_\mathcal{D}$ is to be perceived as a column vector in (1), $\mu_\mathcal{D}$ is a column vector with mean components $\mu_{x'}$ for all $x' \in \mathcal{D}$, $\Sigma_{x\mathcal{D}|\lambda}$ is a row vector with covariance components $\sigma_{xx'|\lambda}$ for all $x' \in \mathcal{D}$, $\Sigma_{\mathcal{D}x|\lambda}$ is the transpose of $\Sigma_{x\mathcal{D}|\lambda}$, and $\Sigma_{\mathcal{D}\mathcal{D}|\lambda}$ is a covariance matrix with components $\sigma_{ux'|\lambda}$ for all $u, x' \in \mathcal{D}$. When the spatial

correlation structure (i.e., λ) is not known, a probabilistic belief $b_{\mathcal{D}}(\lambda) \triangleq p(\lambda|z_{\mathcal{D}})$ can be maintained/tracked over all possible λ and updated using Bayes' rule to the posterior belief $b_{\mathcal{D}\cup\{x\}}(\lambda)$ given a newly available measurement z_x:

$$b_{\mathcal{D}\cup\{x\}}(\lambda) \;\propto\; p(z_x|z_{\mathcal{D}},\lambda)\, b_{\mathcal{D}}(\lambda)\;. \tag{3}$$

Using belief $b_{\mathcal{D}}$, the predictive distribution $p(z_x|z_{\mathcal{D}})$ can be obtained by marginalizing out λ:

$$p(z_x|z_{\mathcal{D}}) = \sum_{\lambda\in\Lambda} p(z_x|z_{\mathcal{D}},\lambda)\, b_{\mathcal{D}}(\lambda)\;. \tag{4}$$

3. Nonmyopic ϵ-Bayes-Optimal Active Learning (ϵ-BAL)

3.1. Problem Formulation

To cast active sensing as a Bayesian sequential decision problem, let us first define a sequential active sensing/learning policy π given a budget of N sampling locations: Specifically, the policy $\pi \triangleq \{\pi_n\}_{n=1}^N$ is structured to sequentially decide the next location $\pi_n(z_{\mathcal{D}}) \in \mathcal{X}\setminus\mathcal{D}$ to be observed at each stage n based on the current observations $z_{\mathcal{D}}$ over a finite planning horizon of N stages. Recall from Section 1 that the active sensing problem involves planning/deciding the most informative locations to be observed for minimizing the predictive uncertainty of the unobserved areas of a phenomenon. To achieve this, we use the entropy criterion (Cover & Thomas, 1991) to measure the informativeness and predictive uncertainty. Then, the value under a policy π is defined to be the joint entropy of its selected observations when starting with some prior observations $z_{\mathcal{D}_0}$ and following π thereafter:

$$V_1^{\pi}(z_{\mathcal{D}_0}) \triangleq \mathbb{H}[Z_\pi|z_{\mathcal{D}_0}] \triangleq -\int p(z_\pi|z_{\mathcal{D}_0}) \log p(z_\pi|z_{\mathcal{D}_0})\, \mathrm{d}z_\pi \tag{5}$$

where Z_π (z_π) is the set of random (realized) measurements taken by policy π and $p(z_\pi|z_{\mathcal{D}_0})$ is defined in a similar manner to (4).

To solve the active sensing problem, the notion of Bayes-optimality[2] is exploited for selecting observations of largest possible joint entropy with respect to all possible induced sequences of future beliefs (starting from initial prior belief $b_{\mathcal{D}_0}$) over candidate sets of model parameters λ, as detailed next. Formally, this entails choosing a sequential policy π to maximize $V_1^{\pi}(z_{\mathcal{D}_0})$ (5), which we call the *Bayes-optimal active learning* (BAL) policy π^*. That is, $V_1^*(z_{\mathcal{D}_0}) \triangleq V_1^{\pi^*}(z_{\mathcal{D}_0}) = \max_\pi V_1^{\pi}(z_{\mathcal{D}_0})$. When π^* is plugged into (5), the following N-stage Bellman equations result from the chain rule for entropy:

$$\begin{aligned} V_n^*(z_{\mathcal{D}}) &= \mathbb{H}\big[Z_{\pi_n^*(z_{\mathcal{D}})}|z_{\mathcal{D}}\big] + \mathbb{E}\big[V_{n+1}^*\big(z_{\mathcal{D}}\cup\{Z_{\pi_n^*(z_{\mathcal{D}})}\}\big)|z_{\mathcal{D}}\big] \\ &= \max_{x\in\mathcal{X}\setminus\mathcal{D}} Q_n^*(z_{\mathcal{D}},x) \end{aligned}$$

$$Q_n^*(z_{\mathcal{D}},x) \triangleq \mathbb{H}[Z_x|z_{\mathcal{D}}] + \mathbb{E}\big[V_{n+1}^*(z_{\mathcal{D}}\cup\{Z_x\})|z_{\mathcal{D}}\big]$$

$$\mathbb{H}[Z_x|z_{\mathcal{D}}] \triangleq -\int p(z_x|z_{\mathcal{D}}) \log p(z_x|z_{\mathcal{D}})\, \mathrm{d}z_x \tag{6}$$

for stage $n = 1,\ldots,N$ where $p(z_x|z_{\mathcal{D}})$ is defined in (4) and the expectation terms are omitted from the right-hand side (RHS) expressions of V_N^* and Q_N^* at stage N. At each stage, the belief $b_{\mathcal{D}}(\lambda)$ is needed to compute $Q_n^*(z_{\mathcal{D}},x)$ in (6) and can be uniquely determined from initial prior belief $b_{\mathcal{D}_0}$ and observations $z_{\mathcal{D}\setminus\mathcal{D}_0}$ using (3). To understand how the BAL policy π^* jointly and naturally optimizes the exploration-exploitation trade-off, its selected location $\pi_n^*(z_{\mathcal{D}}) = \arg\max_{x\in\mathcal{X}\setminus\mathcal{D}} Q_n^*(z_{\mathcal{D}},x)$ at each stage n affects both the immediate payoff $\mathbb{H}\big[Z_{\pi_n^*(z_{\mathcal{D}})}|z_{\mathcal{D}}\big]$ given current belief $b_{\mathcal{D}}$ (i.e., exploitation) as well as the posterior belief $b_{\mathcal{D}\cup\{\pi_n^*(z_{\mathcal{D}})\}}$, the latter of which influences expected future payoff $\mathbb{E}[V_{n+1}^*(z_{\mathcal{D}}\cup\{Z_{\pi_n^*(z_{\mathcal{D}})}\})|z_{\mathcal{D}}]$ and builds in the information gathering option (i.e., exploration).

Interestingly, the work of Low et al. (2009) has revealed that the above recursive formulation (6) can be perceived as the sequential variant of the well-known maximum entropy sampling problem (Shewry & Wynn, 1987) and established an equivalence result that the maximum-entropy observations selected by π^* achieve a dual objective of minimizing the posterior joint entropy (i.e., predictive uncertainty) remaining in the unobserved locations of the phenomenon. Unfortunately, the BAL policy π^* cannot be derived exactly because the stage-wise entropy and expectation terms in (6) cannot be evaluated in closed form due to an uncountable set of candidate observations and unknown model parameters λ (Section 1). To overcome this difficulty, we show in the next subsection how it is possible to solve for an ϵ-BAL policy π_ϵ, that is, the joint entropy of its selected observations closely approximates that of π^* within an arbitrary loss bound $\epsilon > 0$.

3.2. ϵ-BAL Policy

The key idea underlying the design and construction of our proposed nonmyopic ϵ-BAL policy π^ϵ is to approximate the entropy and expectation terms in (6) at every stage using a form of truncated sampling to be described next:

Definition 1 (τ-Truncated Observation) *Define random measurement $\widehat{Z}_x$ by truncating Z_x at $-\widehat{\tau}$ and $\widehat{\tau}$ as follows:*

$$\widehat{Z}_x \triangleq \begin{cases} -\widehat{\tau} & \text{if } Z_x \le -\widehat{\tau}, \\ Z_x & \text{if } -\widehat{\tau} < Z_x < \widehat{\tau}, \\ \widehat{\tau} & \text{if } Z_x \ge \widehat{\tau}. \end{cases}$$

Then, $\widehat{Z}_x$ has a distribution of mixed type with its continuous component defined as $f(\widehat{Z}_x = z_x|z_{\mathcal{D}}) \triangleq p(Z_x = z_x|z_{\mathcal{D}})$ for $-\widehat{\tau} < z_x < \widehat{\tau}$ and its discrete component defined as $f(\widehat{Z}_x = \widehat{\tau}|z_{\mathcal{D}}) \triangleq P(Z_x \ge \widehat{\tau}|z_{\mathcal{D}}) = \int_{\widehat{\tau}}^{\infty} p(Z_x =$

[2] Bayes-optimality is previously studied in reinforcement learning whose developed theories (Poupart et al., 2006; Hoang & Low, 2013) cannot be applied here because their assumptions of discrete-valued observations and Markov property do not hold.

$z_x|z_\mathcal{D})\mathrm{d}z_x$ *and* $f(\widehat{Z}_x = -\widehat{\tau}|z_\mathcal{D}) \triangleq P(Z_x \leq -\widehat{\tau}|z_\mathcal{D}) = \int_{-\infty}^{-\widehat{\tau}} p(Z_x = z_x|z_\mathcal{D})\mathrm{d}z_x$.

Let $\overline{\mu}(\mathcal{D},\Lambda) \triangleq \max_{x\in\mathcal{X}\setminus\mathcal{D},\lambda\in\Lambda}\mu_{x|\mathcal{D},\lambda}$, $\underline{\mu}(\mathcal{D},\Lambda) \triangleq \min_{x\in\mathcal{X}\setminus\mathcal{D},\lambda\in\Lambda}\mu_{x|\mathcal{D},\lambda}$, *and*

$$\widehat{\tau} \triangleq \max\left(\left|\underline{\mu}(\mathcal{D},\Lambda)-\tau\right|, \left|\overline{\mu}(\mathcal{D},\Lambda)+\tau\right|\right) \quad (7)$$

for some $0 \leq \tau \leq \widehat{\tau}$. *Then, a realized measurement of* $\widehat{Z}_x$ *is said to be a* τ*-truncated observation for location* x.

Specifically, given that a set $z_\mathcal{D}$ of realized measurements is available, a finite set of S τ-truncated observations $\{z_x^i\}_{i=1}^S$ can be generated for every candidate location $x \in \mathcal{X}\setminus\mathcal{D}$ at each stage n by identically and independently sampling from $p(z_x|z_\mathcal{D})$ (4) and then truncating each of them according to $z_x^i \leftarrow z_x^i \min\big(|z_x^i|,\widehat{\tau}\big)/|z_x^i|$. These generated τ-truncated observations can be exploited for approximating V_n^* (6) through the following Bellman equations:

$$\begin{aligned} V_n^\epsilon(z_\mathcal{D}) &\triangleq \max_{x\in\mathcal{X}\setminus\mathcal{D}} Q_n^\epsilon(z_\mathcal{D},x) \\ Q_n^\epsilon(z_\mathcal{D},x) &\triangleq \frac{1}{S}\sum_{i=1}^S -\log p\big(z_x^i|z_\mathcal{D}\big) + V_{n+1}^\epsilon\big(z_\mathcal{D}\cup\{z_x^i\}\big) \end{aligned} \quad (8)$$

for stage $n = 1,\ldots,N$ such that there is no V_{N+1}^ϵ term on the RHS expression of Q_N^ϵ at stage N. Like the BAL policy π^* (Section 3.1), the location $\pi_n^\epsilon(z_\mathcal{D}) = \arg\max_{x\in\mathcal{X}\setminus\mathcal{D}} Q_n^\epsilon(z_\mathcal{D},x)$ selected by our ϵ-BAL policy π^ϵ at each stage n also jointly and naturally optimizes the trade-off between exploitation (i.e., by maximizing immediate payoff $S^{-1}\sum_{i=1}^S -\log p(z^i_{\pi_n^\epsilon(z_\mathcal{D})}|z_\mathcal{D})$ given the current belief $b_\mathcal{D}$) vs. exploration (i.e., by improving posterior belief $b_{\mathcal{D}\cup\{\pi_n^\epsilon(z_\mathcal{D})\}}$ to maximize average future payoff $S^{-1}\sum_{i=1}^S V_{n+1}^\epsilon(z_\mathcal{D}\cup\{z^i_{\pi_n^\epsilon(z_\mathcal{D})}\})$). Unlike the deterministic BAL policy π^*, our ϵ-BAL policy π^ϵ is stochastic due to its use of the above truncated sampling procedure.

3.3. Theoretical Analysis

The main difficulty in analyzing the active sensing performance of our stochastic ϵ-BAL policy π^ϵ (i.e., relative to that of BAL policy π^*) lies in determining how its ϵ-Bayes optimality can be guaranteed by choosing appropriate values of the truncated sampling parameters S and τ (Section 3.2). To achieve this, we have to formally understand how S and τ can be specified and varied in terms of the user-defined loss bound ϵ, budget of N sampling locations, domain size $|\mathcal{X}|$ of the phenomenon, and properties/parameters characterizing the spatial correlation structure of the phenomenon (Section 2), as detailed below.

The first step is to show that Q_n^ϵ (8) is in fact a good approximation of Q_n^* (6) for some chosen values of S and τ. There are two sources of error arising in such an approximation: (a) In the truncated sampling procedure (Section 3.2), only a finite set of τ-truncated observations is generated for approximating the stage-wise entropy and expectation terms in (6), and (b) computing Q_n^ϵ does not involve utilizing the values of V_{n+1}^* but that of its approximation V_{n+1}^ϵ instead. To facilitate capturing the error due to finite truncated sampling described in (a), the following intermediate function is introduced:

$$W_n^*(z_\mathcal{D},x) \triangleq \frac{1}{S}\sum_{i=1}^S -\log p\big(z_x^i|z_\mathcal{D}\big) + V_{n+1}^*\big(z_\mathcal{D}\cup\{z_x^i\}\big) \quad (9)$$

for stage $n = 1,\ldots,N$ such that there is no V_{N+1}^* term on the RHS expression of W_N^* at stage N. The first lemma below reveals that if the error $|Q_n^*(z_\mathcal{D},x) - W_n^*(z_\mathcal{D},x)|$ due to finite truncated sampling can be bounded for all tuples $(n,z_\mathcal{D},x)$ generated at stage $n = n',\ldots,N$ by (8) to compute $V_{n'}^\epsilon$ for $1 \leq n' \leq N$, then $Q_{n'}^\epsilon$ (8) can approximate $Q_{n'}^*$ (6) arbitrarily closely:

Lemma 1 *Suppose that a set* $z_{\mathcal{D}'}$ *of observations, a budget of* $N-n'+1$ *sampling locations for* $1 \leq n' \leq N$, $S \in \mathbb{Z}^+$, *and* $\gamma > 0$ *are given. If*

$$|Q_n^*(z_\mathcal{D},x) - W_n^*(z_\mathcal{D},x)| \leq \gamma \quad (10)$$

for all tuples $(n,z_\mathcal{D},x)$ *generated at stage* $n = n',\ldots,N$ *by* (8) *to compute* $V_{n'}^\epsilon(z_{\mathcal{D}'})$, *then, for all* $x' \in \mathcal{X}\setminus\mathcal{D}'$,

$$|Q_{n'}^*(z_{\mathcal{D}'},x') - Q_{n'}^\epsilon(z_{\mathcal{D}'},x')| \leq (N-n'+1)\gamma\,. \quad (11)$$

Its proof is given in Appendix A.1. The next two lemmas show that, with high probability, the error $|Q_n^*(z_\mathcal{D},x) - W_n^*(z_\mathcal{D},x)|$ due to finite truncated sampling can indeed be bounded from above by γ (10) for all tuples $(n,z_\mathcal{D},x)$ generated at stage $n = n',\ldots,N$ by (8) to compute $V_{n'}^\epsilon$ for $1 \leq n' \leq N$:

Lemma 2 *Suppose that a set* $z_{\mathcal{D}'}$ *of observations, a budget of* $N-n'+1$ *sampling locations for* $1 \leq n' \leq N$, $S \in \mathbb{Z}^+$, *and* $\gamma > 0$ *are given. For all tuples* $(n,z_\mathcal{D},x)$ *generated at stage* $n = n',\ldots,N$ *by* (8) *to compute* $V_{n'}^\epsilon(z_{\mathcal{D}'})$,

$$P\Big(|Q_n^*(z_\mathcal{D},x) - W_n^*(z_\mathcal{D},x)| \leq \gamma\Big) \geq 1-2\exp\left(-\frac{2S\gamma^2}{T^2}\right)$$

where $T \triangleq \mathcal{O}\left(\frac{N^2\kappa^{2N}\tau^2}{\sigma_n^2} + N\log\frac{\sigma_o}{\sigma_n} + \log|\Lambda|\right)$ *by setting*

$$\tau = \mathcal{O}\left(\sigma_o\sqrt{\log\left(\frac{\sigma_o^2}{\gamma}\left(\frac{N^2\kappa^{2N}+\sigma_o^2}{\sigma_n^2} + N\log\frac{\sigma_o}{\sigma_n} + \log|\Lambda|\right)\right)}\right)$$

with κ, σ_n^2, *and* σ_o^2 *defined as follows:*

$$\kappa \triangleq 1 + 2\max_{x',u\in\mathcal{X}\setminus\mathcal{D}:x'\neq u,\lambda\in\Lambda,\mathcal{D}} \left|\sigma_{x'u|\mathcal{D},\lambda}\right| / \sigma_{uu|\mathcal{D},\lambda}, \quad (12)$$

$$\sigma_n^2 \triangleq \min_{\lambda\in\Lambda}(\sigma_n^\lambda)^2, \;\; \textit{and} \;\; \sigma_o^2 \triangleq \max_{\lambda\in\Lambda}(\sigma_s^\lambda)^2 + (\sigma_n^\lambda)^2. \quad (13)$$

Refer to Appendix A.2 for its proof.

Remark 1. Deriving such a probabilistic bound in Lemma 2 typically involves the use of concentration inequalities for the sum of independent *bounded* random variables like the Hoeffding's, Bennett's, or Bernstein's inequalities. However, since the originally Gaussian distributed observations

are *unbounded*, sampling from $p(z_x|z_\mathcal{D})$ (4) without truncation will generate unbounded versions of $\{z_x^i\}_{i=1}^S$ and consequently make each summation term $-\log p(z_x^i|z_\mathcal{D})+V_{n+1}^*(z_\mathcal{D}\cup\{z_x^i\})$ on the RHS expression of W_n^* (9) unbounded, hence invalidating the use of these concentration inequalities. To resolve this complication, our trick is to exploit the truncated sampling procedure (Section 3.2) to generate *bounded* τ-truncated observations (Definition 1) (i.e., $|z_x^i|\leq\widehat{\tau}$ for $i=1,\ldots,S$), thus resulting in each summation term $-\log p(z_x^i|z_\mathcal{D})+V_{n+1}^*(z_\mathcal{D}\cup\{z_x^i\})$ being bounded (Appendix A.2). This enables our use of Hoeffding's inequality to derive the probabilistic bound.

Remark 2. It can be observed from Lemma 2 that the amount of truncation has to be reduced (i.e., higher chosen value of τ) when (a) a tighter bound γ on the error $|Q_n^*(z_\mathcal{D},x)-W_n^*(z_\mathcal{D},x)|$ due to finite truncated sampling is desired, (b) a greater budget of N sampling locations is available, (c) a larger space Λ of candidate model parameters is preferred, (d) the spatial phenomenon varies with more intensity and less noise (i.e., assuming all candidate signal and noise variance parameters, respectively, $(\sigma_s^\lambda)^2$ and $(\sigma_n^\lambda)^2$ are specified close to the true large signal and small noise variances), and (e) its spatial correlation structure yields a bigger κ. To elaborate on (e), note that Lemma 2 still holds for any value of κ larger than that set in (12): Since $|\sigma_{x'u|\mathcal{D},\lambda}|^2\leq\sigma_{x'x'|\mathcal{D},\lambda}\sigma_{uu|\mathcal{D},\lambda}$ for all $x'\neq u\in\mathcal{X}\setminus\mathcal{D}$ due to the symmetric positive-definiteness of $\Sigma_{(\mathcal{X}\setminus\mathcal{D})(\mathcal{X}\setminus\mathcal{D})|\mathcal{D},\lambda}$, κ can be set to $1+2\max_{x',u\in\mathcal{X}\setminus\mathcal{D},\lambda\in\Lambda,\mathcal{D}}\sqrt{\sigma_{x'x'|\mathcal{D},\lambda}/\sigma_{uu|\mathcal{D},\lambda}}$. Then, supposing all candidate length-scale parameters are specified close to the true length-scales, a phenomenon with extreme length-scales tending to 0 (i.e., with white-noise process measurements) or ∞ (i.e., with constant measurements) will produce highly similar $\sigma_{x'x'|\mathcal{D},\lambda}$ for all $x'\in\mathcal{X}\setminus\mathcal{D}$, thus resulting in smaller κ and hence smaller τ.

Remark 3. Alternatively, it can be proven that Lemma 2 and the subsequent results hold by setting $\kappa=1$ if a certain structural property of the spatial correlation structure (i.e., for all $z_\mathcal{D}$ ($\mathcal{D}\subseteq\mathcal{X}$) and $\lambda\in\Lambda$, $\Sigma_{\mathcal{D}\mathcal{D}|\lambda}$ is diagonally dominant) is satisfied, as shown in Lemma 9 (Appendix B). Consequently, the κ term can be removed from T and τ.

Lemma 3 *Suppose that a set $z_{\mathcal{D}'}$ of observations, a budget of $N-n'+1$ sampling locations for $1\leq n'\leq N$, $S\in\mathbb{Z}^+$, and $\gamma>0$ are given. The probability that $|Q_n^*(z_\mathcal{D},x)-W_n^*(z_\mathcal{D},x)|\leq\gamma$ (10) holds for all tuples $(n,z_\mathcal{D},x)$ generated at stage $n=n',\ldots,N$ by (8) to compute $V_{n'}^\epsilon(z_{\mathcal{D}'})$ is at least $1-2(S|\mathcal{X}|)^N\exp(-2S\gamma^2/T^2)$ where T is previously defined in Lemma 2.*

Its proof is found in Appendix A.3. The first step is concluded with our first main result, which follows from Lemmas 1 and 3. Specifically, it chooses the values of S and τ such that the probability of Q_n^ϵ (8) approximating Q_n^* (6) poorly (i.e., $|Q_n^*(z_\mathcal{D},x)-Q_n^\epsilon(z_\mathcal{D},x)|>N\gamma$) can be bounded from above by a given $0<\delta<1$:

Theorem 1 *Suppose that a set $z_\mathcal{D}$ of observations, a budget of $N-n+1$ sampling locations for $1\leq n\leq N$, $\gamma>0$, and $0<\delta<1$ are given. The probability that $|Q_n^*(z_\mathcal{D},x)-Q_n^\epsilon(z_\mathcal{D},x)|\leq N\gamma$ holds for all $x\in\mathcal{X}\setminus\mathcal{D}$ is at least $1-\delta$ by setting*

$$S=\mathcal{O}\left(\frac{T^2}{\gamma^2}\left(N\log\frac{N|\mathcal{X}|T^2}{\gamma^2}+\log\frac{1}{\delta}\right)\right)$$

where T is previously defined in Lemma 2. By assuming N, $|\Lambda|$, σ_o, σ_n, κ, and $|\mathcal{X}|$ as constants, $\tau=\mathcal{O}(\sqrt{\log(1/\gamma)})$ and hence $S=\mathcal{O}\left(\frac{(\log(1/\gamma))^2}{\gamma^2}\log\left(\frac{\log(1/\gamma)}{\gamma\delta}\right)\right)$.

Its proof is provided in Appendix A.4.

Remark. It can be observed from Theorem 1 that the number of generated τ-truncated observations has to be increased (i.e., higher chosen value of S) when (a) a lower probability δ of Q_n^ϵ (8) approximating Q_n^* (6) poorly (i.e., $|Q_n^*(z_\mathcal{D},x)-Q_n^\epsilon(z_\mathcal{D},x)|>N\gamma$) is desired, and (b) a larger domain $\mathcal{X}$ of the phenomenon is given. The influence of γ, N, $|\Lambda|$, σ_o, σ_n, and κ on S is similar to that on τ, as previously reported in Remark 2 after Lemma 2.

Thus far, we have shown in the first step that, with high probability, Q_n^ϵ (8) approximates Q_n^* (6) arbitrarily closely for some chosen values of S and τ (Theorem 1). The next step uses this result to probabilistically bound the performance loss in terms of Q_n^* by observing location $\pi_n^\epsilon(z_\mathcal{D})$ selected by our ϵ-BAL policy π^ϵ at stage n and following the BAL policy π^* thereafter:

Lemma 4 *Suppose that a set $z_\mathcal{D}$ of observations, a budget of $N-n+1$ sampling locations for $1\leq n\leq N$, $\gamma>0$, and $0<\delta<1$ are given. $Q_n^*(z_\mathcal{D},\pi_n^*(z_\mathcal{D}))-Q_n^*(z_\mathcal{D},\pi_n^\epsilon(z_\mathcal{D}))\leq 2N\gamma$ holds with probability at least $1-\delta$ by setting S and τ according to that in Theorem 1.*

See Appendix A.5 for its proof. The final step extends Lemma 4 to obtain our second main result. In particular, it bounds the *expected* active sensing performance loss of our stochastic ϵ-BAL policy π^ϵ relative to that of BAL policy π^*, that is, policy π^ϵ is ϵ-Bayes-optimal:

Theorem 2 *Given a set $z_{\mathcal{D}_0}$ of prior observations, a budget of N sampling locations, and $\epsilon>0$, $V_1^*(z_{\mathcal{D}_0})-\mathbb{E}_{\pi^\epsilon}[V_1^{\pi^\epsilon}(z_{\mathcal{D}_0})]\leq\epsilon$ by setting and substituting $\gamma=\epsilon/(4N^2)$ and $\delta=\epsilon/(2N(N\log(\sigma_o/\sigma_n)+\log|\Lambda|))$ into S and τ in Theorem 1 to give $\tau=\mathcal{O}(\sqrt{\log(1/\epsilon)})$ and $S=\mathcal{O}\left(\frac{(\log(1/\epsilon))^2}{\epsilon^2}\log\left(\frac{\log(1/\epsilon)}{\epsilon^2}\right)\right)$.*

Its proof is given in Appendix A.6.

Remark 1. The number of generated τ-truncated observations and the amount of truncation have to be, respectively,

increased and reduced (i.e., higher chosen values of S and τ) when a tighter user-defined loss bound ϵ is desired.

Remark 2. The deterministic BAL policy π^* is Bayes-optimal among all candidate stochastic policies π since $\mathbb{E}_\pi[V_1^\pi(z_{\mathcal{D}_0})] \leq V_1^*(z_{\mathcal{D}_0})$, as proven in Appendix A.7.

3.4. Anytime ϵ-BAL ($\langle\alpha,\epsilon\rangle$-BAL) Algorithm

Unlike the BAL policy π^*, our ϵ-BAL policy π^ϵ can be derived exactly because its time complexity is independent of the size of the set of all possible originally Gaussian distributed observations, which is uncountable. But, the cost of deriving π^ϵ is exponential in the length N of planning horizon since it has to compute the values $V_n^\epsilon(z_\mathcal{D})$ (8) for all $(S|\mathcal{X}|)^N$ possible states $(n, z_\mathcal{D})$. To ease this computational burden, we propose an anytime algorithm based on ϵ-BAL that can produce a good policy fast and improve its approximation quality over time, as discussed next.

The key intuition behind our *anytime* ϵ-BAL *algorithm* ($\langle\alpha,\epsilon\rangle$-BAL of Algo. 1) is to focus the simulation of greedy exploration paths through the most uncertain regions of the state space (i.e., in terms of the values $V_n^\epsilon(z_\mathcal{D})$) instead of evaluating the entire state space like π^ϵ. To achieve this, our $\langle\alpha,\epsilon\rangle$-BAL algorithm maintains both lower and upper heuristic bounds (respectively, $\underline{V}_n^\epsilon(z_\mathcal{D})$ and $\overline{V}_n^\epsilon(z_\mathcal{D})$) for each encountered state $(n, z_\mathcal{D})$, which are exploited for representing the uncertainty of its corresponding value $V_n^\epsilon(z_\mathcal{D})$ to be used in turn for guiding the greedy exploration (or, put differently, pruning unnecessary, bad exploration of the state space while still guaranteeing policy optimality).

To elaborate, each simulated exploration path (EXPLORE of Algo. 1) repeatedly selects a sampling location x and its corresponding τ-truncated observation z_x^i at every stage until the last stage N is reached. Specifically, at each stage n of the simulated path, the next states $(n+1, z_\mathcal{D} \cup \{z_x^i\})$ with uncertainty $|\overline{V}_{n+1}^\epsilon(z_\mathcal{D}\cup\{z_x^i\}) - \underline{V}_{n+1}^\epsilon(z_\mathcal{D}\cup\{z_x^i\})|$ exceeding α (line 6) are identified (lines 7-8), among which the one with largest lower bound $\underline{V}_{n+1}^\epsilon(z_\mathcal{D} \cup \{z_x^i\})$ (line 10) is prioritized/selected for exploration (if more than one exists, ties are broken by choosing the one with most uncertainty, that is, largest upper bound $\overline{V}_{n+1}^\epsilon(z_\mathcal{D} \cup \{z_x^i\})$ (line 11)) while the remaining unexplored ones are placed in the set $\mathcal{U}$ (line 12) to be considered for future exploration (lines 3-6 in $\langle\alpha,\epsilon\rangle$-BAL). So, the simulated path terminates if the uncertainty of every next state is at most α; the uncertainty of a state at the last stage N is guaranteed to be zero (14).

Then, the algorithm backtracks up the path to update/tighten the bounds of previously visited states (line 7 in $\langle\alpha,\epsilon\rangle$-BAL and line 14 in EXPLORE) as follows:

$$\begin{aligned}\overline{V}_n^\epsilon(z_\mathcal{D}) &\leftarrow \min\left(\overline{V}_n^\epsilon(z_\mathcal{D}),\ \max_{x\in\mathcal{X}\setminus\mathcal{D}} \overline{Q}_n^\epsilon(z_\mathcal{D}, x)\right)\\ \underline{V}_n^\epsilon(z_\mathcal{D}) &\leftarrow \max\left(\underline{V}_n^\epsilon(z_\mathcal{D}),\ \max_{x\in\mathcal{X}\setminus\mathcal{D}} \underline{Q}_n^\epsilon(z_\mathcal{D}, x)\right)\end{aligned} \tag{14}$$

Algorithm 1 $\langle\alpha,\epsilon\rangle$-BAL$(z_{\mathcal{D}_0})$

$\langle\alpha,\epsilon\rangle$-BAL$(z_{\mathcal{D}_0})$
1: $\mathcal{U} \leftarrow \{(1, z_{\mathcal{D}_0})\}$
2: **while** $|\overline{V}_1^\epsilon(z_{\mathcal{D}_0}) - \underline{V}_1^\epsilon(z_{\mathcal{D}_0})| > \alpha$ **do**
3: $\quad\mathcal{V} \leftarrow \arg\max_{(n,z_\mathcal{D})\in\mathcal{U}} \underline{V}_n^\epsilon(z_\mathcal{D})$
4: $\quad(n', z_{\mathcal{D}'}) \leftarrow \arg\max_{(n,z_\mathcal{D})\in\mathcal{V}} \overline{V}_n^\epsilon(z_\mathcal{D})$
5: $\quad\mathcal{U} \leftarrow \mathcal{U} \setminus \{(n', z_{\mathcal{D}'})\}$
6: $\quad$EXPLORE$(n', z_{\mathcal{D}'}, \mathcal{U})$ $\quad/*\ \mathcal{U}$ is passed by reference $*/$
7: $\quad$UPDATE$(n', z_{\mathcal{D}'})$
8: **return** $\pi_1^{\langle\alpha,\epsilon\rangle}(z_{\mathcal{D}_0}) \leftarrow \arg\max_{x\in\mathcal{X}\setminus\mathcal{D}_0} \underline{Q}_1^\epsilon(z_{\mathcal{D}_0}, x)$

EXPLORE$(n, z_\mathcal{D}, \mathcal{U})$
1: $\mathcal{T} \leftarrow \emptyset$
2: **for all** $x \in \mathcal{X} \setminus \mathcal{D}$ **do**
3: $\quad\{z_x^i\}_{i=1}^S \leftarrow$ sample from $p(z_x|z_\mathcal{D})$ (4)
4: $\quad$**for** $i = 1, \ldots, S$ **do**
5: $\quad\quad z_x^i \leftarrow z_x^i \min(|z_x^i|, \widehat{\tau})/|z_x^i|$
6: $\quad\quad$**if** $|\overline{V}_{n+1}^\epsilon(z_\mathcal{D} \cup \{z_x^i\}) - \underline{V}_{n+1}^\epsilon(z_\mathcal{D} \cup \{z_x^i\})| > \alpha$ **then**
7: $\quad\quad\quad\mathcal{T} \leftarrow \mathcal{T} \cup \{(n+1, z_\mathcal{D} \cup \{z_x^i\})\}$
8: $\quad\quad\quad$parent$(n+1, z_\mathcal{D} \cup \{z_x^i\}) \leftarrow (n, z_\mathcal{D})$
9: **if** $|\mathcal{T}| > 0$ **then**
10: $\quad\mathcal{V} \leftarrow \arg\max_{(n+1,z_\mathcal{D}\cup\{z_x^i\})\in\mathcal{T}} \underline{V}_{n+1}^\epsilon(z_\mathcal{D} \cup \{z_x^i\})$
11: $\quad(n+1, z_{\mathcal{D}'}) \leftarrow \arg\max_{(n+1,z_\mathcal{D}\cup\{z_x^i\})\in\mathcal{V}} \overline{V}_{n+1}^\epsilon(z_\mathcal{D} \cup \{z_x^i\})$
12: $\quad\mathcal{U} \leftarrow \mathcal{U} \cup (\mathcal{T} \setminus \{(n+1, z_{\mathcal{D}'})\})$
13: $\quad$EXPLORE$(n+1, z_{\mathcal{D}'}, \mathcal{U})$
14: Update $\overline{V}_n^\epsilon(z_\mathcal{D})$ and $\underline{V}_n^\epsilon(z_\mathcal{D})$ using (14)

UPDATE$(n, z_\mathcal{D})$
1: Update $\overline{V}_n^\epsilon(z_\mathcal{D})$ and $\underline{V}_n^\epsilon(z_\mathcal{D})$ using (14)
2: **if** $n > 1$ **then**
3: $\quad(n-1, z_{\mathcal{D}'}) \leftarrow$ parent$(n, z_\mathcal{D})$
4: $\quad$UPDATE$(n-1, z_{\mathcal{D}'})$

$$\overline{Q}_n^\epsilon(z_\mathcal{D}, x) \triangleq \frac{1}{S}\sum_{i=1}^S -\log p(z_x^i|z_\mathcal{D}) + \overline{V}_{n+1}^\epsilon(z_\mathcal{D} \cup \{z_x^i\})$$

$$\underline{Q}_n^\epsilon(z_\mathcal{D}, x) \triangleq \frac{1}{S}\sum_{i=1}^S -\log p(z_x^i|z_\mathcal{D}) + \underline{V}_{n+1}^\epsilon(z_\mathcal{D} \cup \{z_x^i\})$$

for stage $n = 1, \ldots, N$ such that there is no $\overline{V}_{N+1}^\epsilon$ ($\underline{V}_{N+1}^\epsilon$) term on the RHS expression of $\overline{Q}_N^\epsilon$ ($\underline{Q}_N^\epsilon$) at stage N. When the planning time runs out, we provide the greedy policy induced by the lower bound: $\pi_1^{\langle\alpha,\epsilon\rangle}(z_{\mathcal{D}_0}) \triangleq \arg\max_{x\in\mathcal{X}\setminus\mathcal{D}_0} \underline{Q}_1^\epsilon(z_{\mathcal{D}_0}, x)$ (line 8 in $\langle\alpha,\epsilon\rangle$-BAL).

Central to the anytime performance of our $\langle\alpha,\epsilon\rangle$-BAL algorithm is the computational efficiency of deriving informed initial heuristic bounds $\underline{V}_n^\epsilon(z_\mathcal{D})$ and $\overline{V}_n^\epsilon(z_\mathcal{D})$ where $\underline{V}_n^\epsilon(z_\mathcal{D}) \leq V_n^\epsilon(z_\mathcal{D}) \leq \overline{V}_n^\epsilon(z_\mathcal{D})$. Due to lack of space, we have shown in Appendix A.8 how they can be derived efficiently. We have also derived a theoretical guarantee similar to that of Theorem 2 on the expected active sensing performance of our $\langle\alpha,\epsilon\rangle$-BAL policy $\pi^{\langle\alpha,\epsilon\rangle}$, as shown in Appendix A.9. We have analyzed the time complexity of simulating k exploration paths in our $\langle\alpha,\epsilon\rangle$-BAL algorithm to be $\mathcal{O}(kNS|\mathcal{X}|(|\Lambda|(N^3 + |\mathcal{X}|N^2 + S|\mathcal{X}|) + \Delta + \log(kNS|\mathcal{X}|)))$ (Appendix A.10) where $\mathcal{O}(\Delta)$ denotes the cost of initializing the heuristic bounds at each state. In practice, $\langle\alpha,\epsilon\rangle$-BAL's planning horizon can be shortened to reduce its computational cost further by limiting the depth of each simulated path to strictly less than N. In that case,

although the resulting $\pi^{\langle\alpha,\epsilon\rangle}$'s performance has not been theoretically analyzed, Section 4.1 demonstrates empirically that it outperforms the state-of-the-art algorithms.

4. Experiments and Discussion

This section evaluates the active sensing performance and time efficiency of our $\langle\alpha,\epsilon\rangle$-BAL policy $\pi^{\langle\alpha,\epsilon\rangle}$ (Section 3) empirically under limited sampling budget using two datasets featuring a simple, simulated spatial phenomenon (Section 4.1) and a large-scale, real-world traffic phenomenon (i.e., speeds of road segments) over an urban road network (Section 4.2). All experiments are run on a Mac OS X machine with Intel Core i7 at 2.66 GHz.

4.1. Simulated Spatial Phenomenon

The domain of the phenomenon is discretized into a finite set of sampling locations $\mathcal{X} = \{0, 1, \ldots, 99\}$. The phenomenon is a realization of a GP (Section 2) parameterized by $\lambda^* = \{\sigma_n^{\lambda^*} = 0.25, \sigma_s^{\lambda^*} = 10.0, \ell^{\lambda^*} = 1.0\}$. For simplicity, we assume that $\sigma_n^{\lambda^*}$ and $\sigma_s^{\lambda^*}$ are known, but the true length-scale $\ell^{\lambda^*} = 1$ is not. So, a uniform prior belief $b_{\mathcal{D}_0=\emptyset}$ is maintained over a set $\mathcal{L} = \{1, 6, 9, 12, 15, 18, 21\}$ of 7 candidate length-scales ℓ^{λ}. Using *root mean squared prediction error* (RMSPE) as the performance metric, the performance of our $\langle\alpha,\epsilon\rangle$-BAL policies $\pi^{\langle\alpha,\epsilon\rangle}$ with planning horizon length $N' = 2, 3$ and $\alpha = 1.0$ are compared to that of the state-of-the-art GP-based active learning algorithms: (a) The *a priori greedy design* (APGD) policy (Shewry & Wynn, 1987) iteratively selects and adds $\arg\max_{x\in\mathcal{X}\setminus\mathcal{S}_n} \sum_{\lambda\in\Lambda} b_{\mathcal{D}_0}(\lambda)\mathbb{H}[Z_{\mathcal{S}_n\cup\{x\}}|z_{\mathcal{D}_0},\lambda]$ to the current set $\mathcal{S}_n$ of sampling locations (where $\mathcal{S}_0 = \emptyset$) until $\mathcal{S}_N$ is obtained, (b) the *implicit exploration* (IE) policy greedily selects and observes sampling location $x^{\text{IE}} = \arg\max_{x\in\mathcal{X}\setminus\mathcal{D}} \sum_{\lambda\in\Lambda} b_{\mathcal{D}}(\lambda)\mathbb{H}[Z_x|z_{\mathcal{D}},\lambda]$ and updates the belief from $b_{\mathcal{D}}$ to $b_{\mathcal{D}\cup\{x^{\text{IE}}\}}$ over $\mathcal{L}$; if the upper bound on the performance advantage of using π^* over APGD policy is less than a pre-defined threshold, it will use APGD with the remaining sampling budget, and (c) the *explicit exploration via independent tests* (ITE) policy performs a PAC-based binary search, which is guaranteed to find ℓ^{λ^*} with high probability, and then uses APGD to select the remaining locations to be observed.

Both nonmyopic IE and ITE policies are proposed by Krause & Guestrin (2007): IE is reported to incur the lowest prediction error empirically while ITE is guaranteed not to achieve worse than the optimal performance by more than a factor of $1/e$. Fig. 1a shows results of the active sensing performance of the tested policies averaged over 20 realizations of the phenomenon drawn independently from the underlying GP model described earlier. It can be observed that the RMSPE of every tested policy decreases with a larger budget of N sampling locations. Notably, our $\langle\alpha,\epsilon\rangle$-BAL policies perform better than the APGD, IE, and ITE policies, especially when N is small. The performance gap between our $\langle\alpha,\epsilon\rangle$-BAL policies and the other policies decreases as N increases, which intuitively means that, with a tighter sampling budget (i.e., smaller N), it is more critical to strike a right balance between exploration and exploitation.

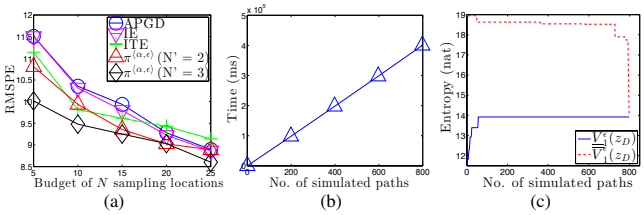
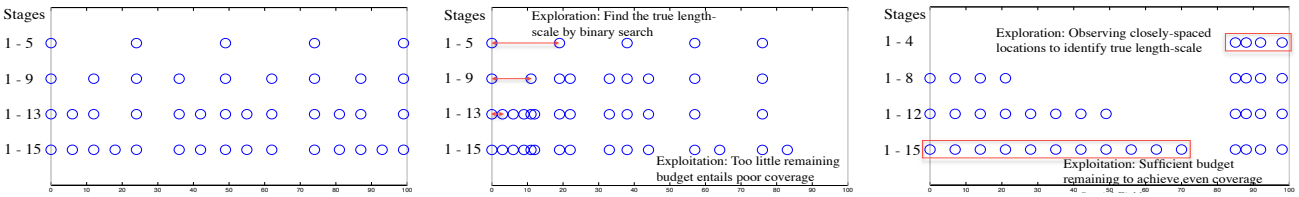


Figure 1. Graphs of (a) RMSPE of APGD, IE, ITE, and $\langle\alpha,\epsilon\rangle$-BAL policies with planning horizon length $N' = 2, 3$ vs. budget of N sampling locations, (b) stage-wise online processing cost of $\langle\alpha,\epsilon\rangle$-BAL policy with $N' = 3$ and (c) gap between $\overline{V}_1^{\epsilon}(z_{\mathcal{D}_0})$ and $\underline{V}_1^{\epsilon}(z_{\mathcal{D}_0})$ vs. number of simulated paths.

Fig. 2 shows the stage-wise sampling designs produced by the tested policies with a budget of $N = 15$ sampling locations. It can be observed that our $\langle\alpha,\epsilon\rangle$-BAL policy achieves a better balance between exploration and exploitation and can therefore discern ℓ^{λ^*} much faster than the IE and ITE policies while maintaining a fine spatial coverage of the phenomenon. This is expected due to the following issues faced by IE and ITE policies: (a) The myopic exploration of IE tends not to observe closely-spaced locations (Fig. 2a), which are in fact informative towards estimating the true length-scale, and (b) despite ITE's theoretical guarantee in finding ℓ^{λ^*}, its PAC-style exploration is too aggressive, hence completely ignoring how informative the posterior belief $b_{\mathcal{D}}$ over $\mathcal{L}$ is during exploration. This leads to a sub-optimal exploration behavior that reserves too little budget for exploitation and consequently entails a poor spatial coverage, as shown in Fig. 2b.

Our $\langle\alpha,\epsilon\rangle$-BAL policy can resolve these issues by jointly and naturally optimizing the trade-off between observing the most informative locations for minimizing the predictive uncertainty of the phenomenon (i.e., exploitation) vs. the uncertainty surrounding its length-scale (i.e., exploration), hence enjoying the best of both worlds (Fig. 2c). In fact, we notice that, after observing 5 locations, our $\langle\alpha,\epsilon\rangle$-BAL policy can focus 88.10% of its posterior belief on ℓ^{λ^*} while IE only assigns, on average, about 18.65% of its posterior belief on ℓ^{λ^*}, which is hardly more informative than the prior belief $b_{\mathcal{D}_0}(\ell^{\lambda^*}) = 1/7 \approx 14.28\%$. Finally, Fig. 1b shows that the online processing cost of $\langle\alpha,\epsilon\rangle$-BAL per sampling stage grows linearly in the number of simulated paths while Fig. 1c reveals that its approximation quality improves (i.e., gap between $\overline{V}_1^{\epsilon}(z_{\mathcal{D}_0})$ and $\underline{V}_1^{\epsilon}(z_{\mathcal{D}_0})$ decreases) with increasing number of simulated paths. Interestingly, it can be observed from Fig. 1c that although $\langle\alpha,\epsilon\rangle$-BAL needs about 800 simulated paths (i.e., 400 s) to close the gap between $\overline{V}_1^{\epsilon}(z_{\mathcal{D}_0})$ and $\underline{V}_1^{\epsilon}(z_{\mathcal{D}_0})$, $\underline{V}_1^{\epsilon}(z_{\mathcal{D}_0})$



(a) IE policy (b) ITE policy (c) $\langle\alpha,\epsilon\rangle$-BAL policy

Figure 2. Stage-wise sampling designs produced by (a) IE, (b) ITE, and (c) $\langle\alpha,\epsilon\rangle$-BAL policy with a planning horizon length $N'=3$ using a budget of $N=15$ sampling locations. The final sampling designs are depicted in the bottommost rows of the figures.

only takes about 100 simulated paths (i.e., 50 s). This implies the actual computation time needed for $\langle\alpha,\epsilon\rangle$-BAL to reach $V_1^\epsilon(z_{\mathcal{D}_0})$ (via its lower bound $\underline{V}_1^\epsilon(z_{\mathcal{D}_0})$) is much less than that required to verify the convergence of $\underline{V}_1^\epsilon(z_{\mathcal{D}_0})$ to $V_1^\epsilon(z_{\mathcal{D}_0})$ (i.e., by checking the gap). This is expected since $\langle\alpha,\epsilon\rangle$-BAL explores states with largest lower bound first (Section 3.4).

4.2. Real-World Traffic Phenomenon

Fig. 3a shows the traffic phenomenon (i.e., speeds (km/h) of road segments) over an urban road network $\mathcal{X}$ comprising 775 road segments (e.g., highways, arterials, slip roads, etc.) in Tampines area, Singapore during lunch hours on June 20, 2011. The mean speed is 52.8 km/h and the standard deviation is 21.0 km/h. Each road segment $x\in\mathcal{X}$ is specified by a 4-dimensional vector of features: length, number of lanes, speed limit, and direction. The phenomenon is modeled as a relational GP (Chen et al., 2012) whose correlation structure can exploit both the road segment features and road network topology information. The true parameters $\lambda^*=\{\sigma_n^{\lambda^*},\sigma_s^{\lambda^*},\ell^{\lambda^*}\}$ are set as the maximum likelihood estimates learned using the entire dataset. We assume that $\sigma_n^{\lambda^*}$ and $\sigma_s^{\lambda^*}$ are known, but ℓ^{λ^*} is not. So, a uniform prior belief $b_{\mathcal{D}_0=\emptyset}$ is maintained over a set $\mathcal{L}=\{\ell^{\lambda_i}\}_{i=0}^6$ of 7 candidate length-scales $\ell^{\lambda_0}=\ell^{\lambda^*}$ and $\ell^{\lambda_i}=2(i+1)\ell^{\lambda^*}$ for $i=1,\ldots,6$.

The performance of our $\langle\alpha,\epsilon\rangle$-BAL policies with planning horizon length $N'=3,4,5$ are compared to that of APGD and IE policies (Section 4.1) by running each of them on a mobile probe to direct its active sensing along a path of adjacent road segments according to the road network topology; ITE cannot be used here as it requires observing road segments separated by a pre-computed distance during exploration (Krause & Guestrin, 2007), which violates the topological constraints of the road network since the mobile probe cannot "teleport". Fig. 3 shows results of the tested policies averaged over 5 independent runs: It can be observed from Fig. 3b that our $\langle\alpha,\epsilon\rangle$-BAL policies outperform APGD and IE policies due to their nonmyopic exploration behavior. In terms of the total online processing cost, Fig. 3c shows that $\langle\alpha,\epsilon\rangle$-BAL incurs <4.5 hours given a budget of $N=240$ road segments, which can be afforded by modern computing power. To illustrate the behavior of each policy, Figs. 3d-f show, respectively, the road segments observed (shaded in black) by the mobile

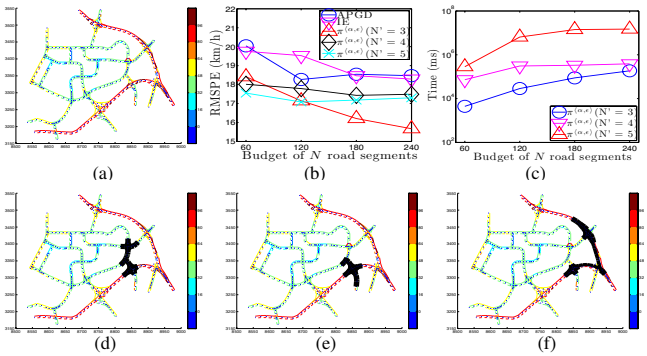


Figure 3. (a) Traffic phenomenon (i.e., speeds (km/h) of road segments) over an urban road network in Tampines area, Singapore, graphs of (b) RMSPE of APGD, IE, and $\langle\alpha,\epsilon\rangle$-BAL policies with horizon length $N'=3,4,5$ and (c) total online processing cost of $\langle\alpha,\epsilon\rangle$-BAL policies with $N'=3,4,5$ vs. budget of N segments, and (d-f) road segments observed (shaded in black) by respective APGD, IE, and $\langle\alpha,\epsilon\rangle$-BAL policies ($N'=5$) with $N=60$.

probe running APGD, IE, and $\langle\alpha,\epsilon\rangle$-BAL policies with $N'=5$ given a budget of $N=60$. It can be observed from Figs. 3d-e that both APGD and IE cause the probe to move away from the slip roads and highways to low-speed segments whose measurements vary much more smoothly; this is expected due to their myopic exploration behavior. In contrast, $\langle\alpha,\epsilon\rangle$-BAL nonmyopically plans the probe's path and can thus direct it to observe the more informative slip roads and highways with highly varying measurements (Fig. 3f) to achieve better performance.

5. Conclusion

This paper describes and theoretically analyzes an ϵ-BAL approach to nonmyopic active learning of GPs that can jointly and naturally optimize the exploration-exploitation trade-off. We then provide an anytime $\langle\alpha,\epsilon\rangle$-BAL algorithm based on ϵ-BAL with real-time performance guarantee and empirically demonstrate using synthetic and real-world datasets that, with limited budget, it outperforms the state-of-the-art GP-based active learning algorithms.

Acknowledgments. This work was supported by Singapore National Research Foundation in part under its International Research Center @ Singapore Funding Initiative and administered by the Interactive Digital Media Programme Office and in part through the Singapore-MIT Alliance for Research and Technology Subaward Agreement No. 52.

References

Cao, N., Low, K. H., and Dolan, J. M. Multi-robot informative path planning for active sensing of environmental phenomena: A tale of two algorithms. In *Proc. AAMAS*, pp. 7–14, 2013.

Chen, J., Low, K. H., Tan, C. K.-Y., Oran, A., Jaillet, P., Dolan, J. M., and Sukhatme, G. S. Decentralized data fusion and active sensing with mobile sensors for modeling and predicting spatiotemporal traffic phenomena. In *Proc. UAI*, pp. 163–173, 2012.

Chen, J., Cao, N., Low, K. H., Ouyang, R., Tan, C. K.-Y., and Jaillet, P. Parallel Gaussian process regression with low-rank covariance matrix approximations. In *Proc. UAI*, pp. 152–161, 2013a.

Chen, J., Low, K. H., and Tan, C. K.-Y. Gaussian process-based decentralized data fusion and active sensing for mobility-on-demand system. In *Proc. RSS*, 2013b.

Cover, T. and Thomas, J. *Elements of Information Theory*. John Wiley & Sons, NY, 1991.

Diggle, P. J. Bayesian geostatistical design. *Scand. J. Statistics*, 33(1):53–64, 2006.

Dolan, J. M., Podnar, G., Stancliff, S., Low, K. H., Elfes, A., Higinbotham, J., Hosler, J. C., Moisan, T. A., and Moisan, J. Cooperative aquatic sensing using the telesupervised adaptive ocean sensor fleet. In *Proc. SPIE Conference on Remote Sensing of the Ocean, Sea Ice, and Large Water Regions*, volume 7473, 2009.

Hoang, T. N. and Low, K. H. A general framework for interacting Bayes-optimally with self-interested agents using arbitrary parametric model and model prior. In *Proc. IJCAI*, pp. 1394–1400, 2013.

Houlsby, N., Hernandez-Lobato, J. M., Huszar, F., and Ghahramani, Z. Collaborative Gaussian processes for preference learning. In *Proc. NIPS*, pp. 2105–2113, 2012.

Krause, A. and Guestrin, C. Nonmyopic active learning of Gaussian processes: An exploration-exploitation approach. In *Proc. ICML*, pp. 449–456, 2007.

Krause, A., Singh, A., and Guestrin, C. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *JMLR*, 9:235–284, 2008.

Low, K. H., Gordon, G. J., Dolan, J. M., and Khosla, P. Adaptive sampling for multi-robot wide-area exploration. In *Proc. IEEE ICRA*, pp. 755–760, 2007.

Low, K. H., Dolan, J. M., and Khosla, P. Adaptive multi-robot wide-area exploration and mapping. In *Proc. AAMAS*, pp. 23–30, 2008.

Low, K. H., Dolan, J. M., and Khosla, P. Information-theoretic approach to efficient adaptive path planning for mobile robotic environmental sensing. In *Proc. ICAPS*, pp. 233–240, 2009.

Low, K. H., Dolan, J. M., and Khosla, P. Active Markov information-theoretic path planning for robotic environmental sensing. In *Proc. AAMAS*, pp. 753–760, 2011.

Low, K. H., Chen, J., Dolan, J. M., Chien, S., and Thompson, D. R. Decentralized active robotic exploration and mapping for probabilistic field classification in environmental sensing. In *Proc. AAMAS*, pp. 105–112, 2012.

Martin, R. J. Comparing and contrasting some environmental and experimental design problems. *Environmetrics*, 12(3):303–317, 2001.

Müller, W. G. *Collecting Spatial Data: Optimum Design of Experiments for Random Fields*. Springer, 3rd edition, 2007.

Ouyang, R., Low, K. H., Chen, J., and Jaillet, P. Multi-robot active sensing of non-stationary Gaussian process-based environmental phenomena. In *Proc. AAMAS*, 2014.

Park, M. and Pillow, J. W. Bayesian active learning with localized priors for fast receptive field characterization. In *Proc. NIPS*, pp. 2357–2365, 2012.

Podnar, G., Dolan, J. M., Low, K. H., and Elfes, A. Telesupervised remote surface water quality sensing. In *Proc. IEEE Aerospace Conference*, 2010.

Poupart, P., Vlassis, N., Hoey, J., and Regan, K. An analytic solution to discrete Bayesian reinforcement learning. In *Proc. ICML*, pp. 697–704, 2006.

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

Shewry, M. C. and Wynn, H. P. Maximum entropy sampling. *J. Applied Statistics*, 14(2):165–170, 1987.

Singh, A., Krause, A., Guestrin, C., and Kaiser, W. J. Efficient informative sensing using multiple robots. *J. Artificial Intelligence Research*, 34:707–755, 2009.

Solomon, H. and Zacks, S. Optimal design of sampling from finite populations: A critical review and indication of new research areas. *J. American Statistical Association*, 65(330):653–677, 1970.

Zimmerman, D. L. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics*, 17(6):635–652, 2006.