# Gaussian Process Regression Using Laplace Approximations Under Localization Uncertainty

Mahdi Jadaliha, Yunfei Xu, and Jongeun Choi

***Abstract*— In this paper, we formulate Gaussian process regression with observations under the localization uncertainty. In our formulation, effects of observations, measurement noise, localization uncertainty and prior distributions are all correctly incorporated in the posterior predictive statistics. The analytically intractable posterior predictive statistics are proposed to be approximated by Laplace approximations in different degrees of complexity. Such approximations have been carefully tailored to our problems and their approximation errors and complexity are analyzed. Simulation results demonstrate that the proposed approaches perform much better than approaches without considering the localization uncertainty correctly.**

## I. Introduction

Recently, there has been a growing interest in wireless sensor networks due to advanced embedded network technology. Exploitation of mobile sensor networks has been increased in collecting spatio-temporal data from the environment [1], [2]. Gaussian process regression (or Kriging in geostatistics) has been widely used to draw statistical inference from geostatistical and environmental data [3], [4]. Gaussian process modeling enables us to predict physical values, such as temperature or harmful algae bloom biomass, at any point and time with a predicted uncertainty level. For example, near-optimal static sensor placements with a mutual information criterion in Gaussian processes were proposed in [5], [6]. A distributed Kriged Kalman filter for spatial estimation based on mobile sensor networks was developed in [2]. Multi-agent systems that are versatile for various tasks by exploiting predictive posterior statistics of Gaussian processes were developed in [7], [8].

Localization in sensor networks is a fundamental problem in various applications to correctly fuse the spatially collected data to estimate the process of interest. However, obtaining precise localization of robotic networks under limited resources is very challenging [9], [10]. In practice, resource-constrained sensor network systems are prone to large uncertainty in localization. Most previous works on Gaussian process regression for mobile sensor networks [3], [6]–[8] have assumed that the exact sampling positions are available, which is not practical in real situations.

Therefore, motivated by the aforementioned issues, we consider correct (Bayesian) integration of uncertainties in sampling positions, observations and measurement noise for Gaussian process regression and its computation error and complexity analysis for the sensor network applications. The overall picture of our work is similar to the one in [11] in which an extended Kalman filter (EKF) was used to incorporate robot localization uncertainty and field parameter uncertainty. However, [11] relies on a parametric model, which is a radial basis function network model, and EKF while our motivation is to use more flexible non-parametric approach, viz., Gaussian process regression taking into account localization uncertainty in a Bayesian framework.

Gaussian process regression in [12], [13] integrated uncertainty in the test input position for multiple-step ahead time series forecasting. However, it did not consider uncertainty in the sampling position of the training data (or observations).

Gaussian process prediction with localization uncertainty can be obtained as a posterior predictive distribution using Bayes' rule. The main difficulty to this is that it has no analytic closed-form solution and has to be approximated either through Monte Carlo sampling [14] or other approximation techniques such as variational inference [15]. As an important analytical approximation technique, Laplace's method has been known to be useful in many such situations [16]–[19].

The contribution of this paper is as follows. First, we formulate Gaussian process regression with observations under the localization uncertainty due to the resource-constrained (possibly mobile) sensor networks. Next, approximations have been obtained for analytically intractable predictive mean and predictive variance by using Laplace approximations. Such approximation methods have been carefully tailored to our problems. In particular, a modified version of the moment generating function (MGF) approximation [17] has been developed to reduce the computational complexity. In addition, we have analyzed and compared the approximation error and the complexity. Simulation results illustrate that the proposed methods outperform the quick-and-dirty solutions often used in practice.

Standard notation will be used throughout the paper. Let $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, $\mathbb{R}_{>0}$ and $\mathbb{Z}_{>0}$ denote, respectively, the sets of real, non-negative real, positive real and positive integer numbers. $I_n$ denotes the identity matrix of size $n$ ($I$ for an appropriate dimension.) For column vectors $v_a \in \mathbb{R}^a$, $v_b \in \mathbb{R}^b$, and $v_c \in \mathbb{R}^c$, $\text{col}(v_a, v_b, v_c) := \begin{bmatrix} v_a^T & v_b^T & v_c^T \end{bmatrix}^T \in \mathbb{R}^{a+b+c}$ stacks all vectors to create one column vector, and $\|v_a\|$ denotes the Euclidean norm (or the vector 2-norm) of $v_a$. $|V|$ denotes the determinant of a matrix $V \in \mathbb{R}^{n\times n}$, and $\text{tr}(V)$ denotes trace of a square matrix $V$. A random vector $X \in \mathbb{R}^n$, which

This work has been supported, in part, by the National Science Foundation through CAREER Award CMMI-0846547.

Mahdi Jadaliha and Yunfei Xu are with the Department of Mechanical Engineering, Michigan State University, East Lansing, MI 48824, USA. E-mails: jadaliha@msu.edu and xuyunfei@egr.msu.edu.

Jongeun Choi is with the Department of Mechanical Engineering and the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA. E-mail: jchoi@egr.msu.edu.

is distributed by a multivariate Gaussian distribution of a mean $X_0$ and a variance $\Sigma$, is denoted by $X \sim \mathcal{N}(X_0, \Sigma)$. We define the first and the second derivative operators on $h(X) \in \mathbb{R}$ with respect to a vector $X \in \mathbb{R}^m$ as follow.

$$D^1 h(X) = \frac{\partial h(X)}{\partial X} \in \mathbb{R}^{m\times 1},\ D^2 h(X) = \frac{\partial^2 h(X)}{\partial X \partial X^T} \in \mathbb{R}^{m\times m}.$$

## II. Preliminaries

### A. Gaussian Process Regression

A Gaussian process defines a distribution over a space of functions and it is completely specified by its mean function and covariance function. Let $x \in \mathbb{R}^d$ denote the index vector. where $x := [\ s^T \quad t\ ]^T$ contains the sampling location $s \in \mathcal{Q} \subset \mathbb{R}^{d-1}$ and the sampling time $t \in \mathbb{R}_{\geq 0}$. A Gaussian process, $y(x) \in \mathbb{R}$, is formally defined as below.

*Definition 1:* A Gaussian process [20] is a collection of random variables, any finite number of which have a joint Gaussian distribution.

For notational simplicity, we consider a zero-mean Gaussian process[1] $y(x) \sim \mathcal{GP}(0, \mathcal{K}(x, x')) \in \mathbb{R}$. For example, one may consider a covariance function defined as $\mathcal{K}(x, x') = \sigma_f^2 \exp\left(-\frac{\|x-x'\|^2}{2\sigma_x^2}\right)$, where $x, x' \in \mathbb{R}^d$. The mean and the covariance function of a Gaussian process can be estimated a priori by maximizing the likelihood function [21].

Suppose, we have $q$ noise corrupted observations without localization error, $\mathcal{D}' = \left\{(x^{(i)}, \bar{y}^{(i)}) \,|\, i = 1, \cdots, q\right\}$. Assume that $\bar{y}^{(i)} = y^{(i)} + w^{(i)} \in \mathbb{R}$, where $w^{(i)}$ is an independent and identically distributed (i.i.d.) white Gaussian noise with variance $\sigma_w^2$. The collections of the realizations $Y = [\ y^{(1)} \quad \cdots \quad y^{(q)}\ ]^T \in \mathbb{R}^q$ and observations $\bar{Y} = [\ \bar{y}^{(1)} \quad \cdots \quad \bar{y}^{(q)}\ ]^T \in \mathbb{R}^q$ have the Gaussian distributions $Y \sim \mathcal{N}(0, K), \quad \bar{Y} \sim \mathcal{N}(0, K + \sigma_w^2 I)$, where $K \in \mathbb{R}^{q\times q}$ is the covariance matrix of $Y$ obtained by $K_{ij} = \mathcal{K}(x^{(i)}, x^{(j)})$. We can predict the value $y_\star$ of the Gaussian process at a point $x_\star$ as [20], $y_\star|\mathcal{D}' \sim \mathcal{N}(\hat{y}_\star, \hat{\sigma}_\star^2)$. The predictive mean is

$$\hat{y}_\star = \mathbb{E}[y_\star|\mathcal{D}'] = k^T (K + \sigma_w^2 I)^{-1} \bar{Y}, \tag{1}$$

and the predictive variance is given by

$$\hat{\sigma}_\star^2 = \mathrm{Var}(\hat{y}_\star|\mathcal{D}') = \mathbb{E}[(y_\star - \hat{y}_\star)^2|\mathcal{D}'] = \sigma_f^2 - k^T (K + \sigma_w^2 I)^{-1} k, \tag{2}$$

where $k \in \mathbb{R}^q$ is the covariance matrix between $Y$ and $y_\star$ obtained by $k_j = \mathcal{K}(x^{(j)}, x_\star)$, and $\sigma_f^2 = \mathcal{K}(x_\star, x_\star) \in \mathbb{R}$ is the variance at $x_\star$. (1) and (2) can be obtained from the fact that

$$\mathrm{col}(y_\star, \bar{Y})|x_\star, X \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_f^2 & k^T \\ k & (K + \sigma_w^2 I) \end{bmatrix}\right), \tag{3}$$

where $X$ is defined by $X = \mathrm{col}(x^{(1)}, x^{(2)}, \cdots, x^{(q)})$.

In this paper, we also assume that at each iteration the mobile sensor networks only needs to fuse a fixed number of truncated observations, which are near the target point of interest, to limit the computational resources [22], [23].

[1] A Gaussian process with a nonzero-mean can be treated by a change of variables. Even without a change of variables, this is not a drastic limitation, since the mean of the posterior process is not confined to zero [20].

### B. Laplace Approximations

The Laplace method is a family of asymptotic approximations that approximate an integral of a function, i.e., $\int_{\mathcal{X}} f(X)dX$, where $X \in \mathcal{X} \subset \mathbb{R}^m$. Let the function $f(X)$ be in a form $f(X) = e^{-nh(X)}$, where $1 \ll n \in \mathbb{Z}_{>0}$, and $h : \mathcal{X} \to \mathbb{R}$ is a class $C^2$. Let $\hat{X}$ denote the exact mode of $-h$, i.e., $\hat{X} = \arg\max_{X\in\mathcal{X}} -h(X)$. Then Laplace's method produces the approximation [16]:

$$\int_{\mathcal{X}} f(X)dX = \left(\frac{2\pi}{n}\right)^{\frac{m}{2}} |V|^{\frac{1}{2}} e^{-nh(\hat{X})} + O(n^{-1}), \tag{4}$$

where $V = [D^2 h(\hat{X})]^{-1}$. The Laplace approximation in (4) will produce reasonable results as long as the $-h$ is unimodal or at least dominated by a single mode.

In practice it might be difficult to find the exact mode of $-h$. A concept of an asymptotic mode is introduced to gauge the approximation error when the exact mode is not used [18].

*Definition 2:* $\hat{X}_a$ is called an asymptotic mode of order $O(n^{-k})$ for $-h$ if $\|\hat{X}_a - \hat{X}\| \to 0$ as $n \to \infty$, and $D^1 h(\hat{X}_a) = O(n^{-k})$.

Suppose that $\hat{X}_a$ is an asymptotic mode of order $O(n^{-1})$ for $-h$ and $\{h, \hat{X}_a\}$ satisfies the regularity conditions [18]. Then it follows that we have

$$\int_{\mathcal{X}} f(X)dX = \left(\frac{2\pi}{n}\right)^{\frac{m}{2}} |V|^{\frac{1}{2}} e^{-nh(\hat{X}_a)} C(\hat{X}_a) + O(n^{-1}), \tag{5}$$

where $C(\hat{X}_a) = e^{\left(\frac{n}{2} D^1 h(\hat{X}_a)^T [D^2 h(\hat{X}_a)]^{-1} D^1 h(\hat{X}_a)\right)}$.

More precise form with the asymptotic mode of order $O(n^{-2})$ is computed for an approximation of order $O(n^{-3})$ in [19].

In many Bayesian inference applications and as in our problem, we need to compute the ratio of two integrals. To this end, *fully exponential Laplace approximations* has been developed by [16] to compute Laplace approximations of the ratio of two integrals, i.e.,

$$\mathcal{M} = \frac{\int_{\mathcal{X}} e^{-nh^\star(X)} dX}{\int_{\mathcal{X}} e^{-nh(X)} dX}. \tag{6}$$

If each of $-h^\star$ and $-h$ has a dominant peak at its maximum, then Laplace's method may be directly applied to both the numerator and denominator of (6) separately. If the regularity conditions are satisfied, using (4) for denominator approximation and (5) for numerator approximation, Miyata obtained the following approximation and its error order (see Theorem 3 in [18]),

$$\begin{aligned} \hat{\mathcal{M}} &= \frac{|D^2 h(\hat{X})|^{1/2} e^{nh(\hat{X})}}{|D^2 h^\star(\hat{X}_a)|^{1/2} e^{nh^\star(\hat{X}_a)}} \times C^\star(\hat{X}_a), \\ \mathcal{M} &= \hat{\mathcal{M}} + O(n^{-2}), \end{aligned} \tag{7}$$

where $\hat{X}$ is the exact mode of $-h$, and $\hat{X}_a$ is the asymptotic mode of $-h^\star$, and

$$\begin{aligned} C^\star(\hat{X}_a) &= e^{\left(\frac{n}{2} D^1 h^\star(\hat{X}_a)^T [D^2 h^\star(\hat{X}_a)]^{-1} D^1 h^\star(\hat{X}_a)\right)}, \\ \hat{X}_a &= \hat{X} - [D^2 h^\star(\hat{X})]^{-1} D^1 h^\star(\hat{X}). \end{aligned} \tag{8}$$

Laplace's method typically has an error of order $O(n^{-1})$ as shown in (4) and (5). On the other hand, fully exponential Laplace approximations for the ratio form yield an error of order $O(n^{-2})$ as shown in (7) since the error terms of order $O(n^{-1})$ in the numerator and the denominator cancel each other [16].

## III. The Problem Statement

In practice, $\mathcal{D}'$ is not available due to localization uncertainty, and instead its exact sampling points will be replaced with noise corrupted sampling points.

To average out measurement and localization noises, in this paper, we propose to use a sampling scheme in which multiple measurements are taken repeatedly at a set of sampling points of a sensor network. For robotic sensors or mobile sensor networks, this sampling strategy could be efficient and inexpensive since the large energy consumption is usually due to the mobility of the sensor network. Let $q$ sensing agents be indexed by $\mathcal{J} = \{1, \cdots, q\}$. From the proposed sampling scheme, we assume that each agent takes multiple data pairs $\{(\bar{x}^{(i)}, \bar{y}^{(i)}) \,|\, i \in \mathcal{I}\}$, which are indexed by $\mathcal{I} = \{1, \cdots, n\}$ at a set of sampling points by the sensor network $\{x^{(j)} \,|\, j \in \mathcal{J}\}$. We then define the map $\phi : \mathcal{I} \to \mathcal{J}$ that takes the index of the data pair in $\mathcal{I}$ as the input and returns the index of the sensor that produced the data pair as the output. Consider the following realizations using the sampling scheme and the notation just introduced.

$$\bar{x}^{(i)} = x^{(\phi(i))} + v^{(i)} \in \mathbb{R}^d, \quad \forall i \in \mathcal{I}$$
$$\bar{y}^{(i)} = y^{(\phi(i))} + w^{(i)} \in \mathbb{R}, \quad \forall i \in \mathcal{I},$$

where $w^{(i)}$ is an i.i.d. white Gaussian noise with a zero mean and a variance of $\sigma_w^2$, i.e., $w^{(i)} \sim \mathcal{N}(0, \sigma_w^2)$ and $v^{(i)}$ is a localization error which has a multivariate normal distribution with a zero mean and a covariance matrix $\Sigma_v \in \mathbb{R}^{d\times d}$, i.e., $v^{(i)} \sim \mathcal{N}(0, \Sigma_v)$. For instance, the distribution of the localization error may be a result of the fusion of GPS and INS measurements [24], or landmark observations and robot's kinematics [25].

To simplify the notation, $\mathcal{D}$ is introduced to denote the data with the measurement noise and localization error as follows.

$$\mathcal{D} = \left\{ (\bar{x}^{(i)}, \bar{y}^{(i)}) \,|\, i \in \mathcal{I} \right\}. \tag{9}$$

We also define the collective sampling point vectors with and without localization uncertainty, and the cumulative localization noise vector, respectively by

$$\begin{aligned} X &= \text{col}(x^{(1)}, x^{(2)}, \cdots, x^{(q)}) \in \mathbb{R}^{dq}, \\ \bar{X} &= \text{col}(\bar{x}^{(1)}, \bar{x}^{(2)}, \cdots, \bar{x}^{(n)}) \in \mathbb{R}^{dn}, \\ V &= \text{col}(\bar{v}^{(1)}, \bar{v}^{(2)}, \cdots, \bar{v}^{(n)}) \in \mathbb{R}^{dn}. \end{aligned} \tag{10}$$

From the proposed sampling scheme, to average out the measurement and localization uncertainties, the number of measurements $n$ can be increased without increasing the number of sensors $q$, and consequently without increasing the dimension of $X \in \mathbb{R}^{dq}$. Hence, this approach may be efficient for the resource-constrained (mobile) sensor network at the cost of taking more measurements. Using collective sampling point vectors in (10), we have the following relationship.

$$\bar{X} = LX + V, \tag{11}$$

where $L = \bar{L} \otimes I_d \in \mathbb{R}^{dn\times dq}$, $\bar{L} \in \mathbb{R}^{n\times q}$ and $\bar{L}_{ij} = 1$ if $\phi(i) = j$, otherwise $\bar{L}_{ij} = 0$. The conditional probability $p(\bar{X}|X)$ can be written as follow.

$$p(\bar{X}|X) = \frac{1}{|2\pi\Sigma_V|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{X}-LX)^T\Sigma_V^{-1}(\bar{X}-LX)}.$$

From a Bayesian perspective, we can treat $X$ as a random vector to incorporate a prior distribution on $X$. For example, if we assign a prior distribution on $X$ such that $X \sim \mathcal{N}(0, \Sigma_X)$ then we have

$$p(X|\bar{X}) = \frac{1}{|2\pi\Sigma_Z|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-H\bar{X})^T\Sigma_Z^{-1}(X-H\bar{X})}. \tag{12}$$

where $H = \Sigma_Z L^T \Sigma_V^{-1}$ and $\Sigma_Z^{-1} = \Sigma_X^{-1} + L^T\Sigma_V^{-1}L$.

Evaluating posterior predictive statistics such as the density, the mean, and the variance are of critical importance for the sensor network applications.

Therefore, given the data $\mathcal{D}$ in (9), our goal is to compute the posterior predictive statistics. In particular we focused on the following two quantities given in detail. The predictive mean estimator (PME) is given by the following equation.

$$\mathbb{E}[y_\star|\mathcal{D}] = \frac{\int_{\mathcal{X}} \hat{y}_\star p(\bar{Y}|X)p(X|\bar{X})dX}{\int_{\mathcal{X}} p(\bar{Y}|X)p(X|\bar{X})dX}, \tag{13}$$

where $\hat{y}_\star$ is given by (1). The predictive variance estimator, $\text{Var}(y_\star|\mathcal{D})$ given as the following formula.

$$\begin{aligned} \text{Var}(y_\star|\mathcal{D}) = &\frac{\int_{\mathcal{X}} (\hat{\sigma}_\star^2(X) + \hat{y}_\star^2(X)) p(\bar{Y}|X)p(X|\bar{X})dX}{\int_{\mathcal{X}} p(\bar{Y}|X)p(X|\bar{X})dX} \\ &- \mathbb{E}[y_\star|\mathcal{D}]^2, \end{aligned} \tag{14}$$

where $\hat{y}_\star$ and $\hat{\sigma}_\star^2$ are given by (1) and (2), respectively.

The main challenge to our problems is the fact that there are no closed-form formulas for the posterior predictive statistics listed in (13), and (14). Therefore, in this paper, approximation techniques will be carefully applied to obtain approximate solutions. In addition, the trade-offs between the computational complexity and the precision will be investigated for the sensor networks with limited resources.

From (12), one might be tempted to use the best estimate of $X$, e.g., the conditional expectation of $X$ for given measured locations $\bar{X}$, i.e., $\mathbb{E}(X|\bar{X})$ for the Gaussian process regression. Comparison between this type of quick-and-dirty solutions and the proposed Bayesian approach as in (13) will be evaluated in Sections V.

## IV. Fully Exponential Laplace Approximations

In this section, we propose fully exponential Laplace approximations to compute the posterior predictive statistics. In the process of applying Laplace's method, we also obtain the estimation of the sampling points given $\mathcal{D}$ as a by-product. From the observations $\mathcal{D}$, we can update the estimates of the sampling points $X$.

### A. Predictive mean

The fully exponential Laplace approximations which are presented in [16] are limited for the posterior expectation of positive functions. Then, Tierney et al. [17] proposed a second-order approximation to the expectation of a general function $g(X)$ (not necessarily positive) by applying the fully exponential method to approximate $\mathcal{M}(\tau) = \mathbb{E}[e^{\tau g(X)}|\mathcal{D}]$ and then differentiating the approximated $\mathcal{M}(\tau)$.

Let $h(X) = -\frac{1}{n}\ln\left(p(\bar{Y}|X)p(X|\bar{X})\right)$, $h^\star(X) = -\frac{\tau}{n}g(X)+h(X)$, where $g(X) = y_\star(X)$. $\frac{d}{d\tau}\hat{\mathcal{M}}(\tau)$ evaluated at $\tau = 0$ yields a second-order approximation to $\mathbb{E}[y_\star|\mathcal{D}]$ and its order of the error is as follow.

$$\mathbb{E}[y_\star|\mathcal{D}] = \frac{d}{d\tau}\mathcal{M}(\tau)\bigg|_{\tau=0} = \frac{d}{d\tau}\hat{\mathcal{M}}(\tau)\bigg|_{\tau=0} + O(n^{-2}). \tag{15}$$

This method, which is called moment generating function (MGF) Laplace approximation, gives a standard-form approximation using the exact mode of $-h(X)$ [17].

Miyata [18], [19] extended the MGF method for one without computing the exact mode of $-h(X)$. Let $\hat{X}$ be an asymptotic mode of order $O(n^{-1})$ for $-h(X)$. Suppose that $\{h, \hat{X}\}$ satisfies the regularity conditions for the asymptotic-mode Laplace method, which are given in [18]. By using Theorem 5 in [18], the MGF-PME of our problem $\hat{\mathbb{E}}[y_\star|\mathcal{D}] = \hat{\mathbb{E}}[g(X)|\mathcal{D}]$ and its error order are given as

$$\begin{aligned}\hat{\mathbb{E}}[y_\star|\mathcal{D}] &= g(\hat{X}) + \frac{1}{2n}\text{tr}\left(D^2 g(\hat{X})V\right) \\ &\quad - \frac{1}{2n}\sum_{ijkq=1}^{n} h^{[ijk]}(\hat{X})V_{iq}V_{jk}\frac{\partial g(\hat{X})}{\partial X_q} \\ &\quad - D^1 g(\hat{X})^T V D^1 h(\hat{X}), \\ \mathbb{E}[y_\star|\mathcal{D}] &= \hat{\mathbb{E}}[y_\star|\mathcal{D}] + O(n^{-2}),\end{aligned} \tag{16}$$

where $V = [D^2 h(\hat{X})]^{-1}$ and $V_{ij}$ is the $i$-th row, $j$-th column element of the matrix $V$, and $h^{[ijk]} = \frac{\partial^3 h(\hat{X})}{\partial X_i \partial X_j \partial X_k}$ is the third partial derivative of $h(X)$ respect to $X_i, X_j$ and $X_k$ at the point $X = \hat{X}$.

Furthermore, if $\hat{X}$ is the exact mode of $-h$ (see [17]), then approximation has a simpler form because the terms that include $D^1 h(\hat{X})$ vanish

$$\begin{aligned}\mathbb{E}[y_\star|\mathcal{D}] \approx& g(\hat{X}) + \frac{1}{2n}\text{tr}\left(D^2 g(\hat{X})V\right) \\ &- \frac{1}{2n}\sum_{ijkq=1}^{n} h^{[ijk]}(\hat{X})V_{iq}V_{jk}\frac{\partial}{\partial X_q}g(\hat{X}).\end{aligned} \tag{17}$$

However, the MGF-PME given by (16) and (17) needs the computation of the third derivative of $h$, which increases the complexity of the algorithm.

In this paper, another MGF method has been developed in order not to use the third derivative of $h$. To approximate the derivative of $\mathcal{M}(\cdot)$ at a point $\tau$, we utilize a three-point estimation, which is the slope of a nearby secant line through the points $(\tau-\delta, \mathcal{M}(\tau-\delta))$ and $(\tau+\delta, \mathcal{M}(\tau+\delta))$. Approximating the derivative in (15) with the three-point estimation, we can avoid the third derivative in (16) or (17). We summarize our results in the following theorem.

*Theorem 3:* Let $\hat{X}$ be the exact mode of $-h(X)$. The three-point predictive mean estimator (TP-PME) and its order of the error are given by

$$\begin{aligned}\hat{\mathbb{E}}[y_\star|\mathcal{D}] &= \frac{1}{2}n^{3/4}\left|D^2 h(\hat{X})\right|^{1/2} e^{nh(\hat{X})} \\ &\quad \times \left\{\left|D^2 h_+(\hat{X}_+)\right|^{-1/2} C_+(\hat{X}_+)e^{-nh_+(\hat{X}_+)}\right. \\ &\quad \left. - \left|D^2 h_-(\hat{X}_-)\right|^{-1/2} C_-(\hat{X}_-)e^{-nh_-(\hat{X}_-)}\right\} \\ &= \mathbb{E}[y_\star|\mathcal{D}] + O(n^{-3/2}),\end{aligned} \tag{18}$$

where we have used the following definitions

$$\begin{aligned}h_+(X) &= h(X) - n^{-7/4}g(X), \\ h_-(X) &= h(X) + n^{-7/4}g(X), \\ \hat{X}_+ &= \hat{X} - [D^2 h_+(\hat{X})]^{-1}D^1 h_+(\hat{X}), \\ \hat{X}_- &= \hat{X} - [D^2 h_-(\hat{X})]^{-1}D^1 h_-(\hat{X}), \\ C_+(\hat{X}_+) &= e^{\frac{n}{2}D^1 h_+(\hat{X}_+)^T[D^2 h_+(\hat{X}_+)]^{-1}D^1 h_+(\hat{X}_+)}, \\ C_-(\hat{X}_-) &= e^{\frac{n}{2}D^1 h_-(\hat{X}_-)^T[D^2 h_-(\hat{X}_-)]^{-1}D^1 h_-(\hat{X}_-)}.\end{aligned}$$

*Proof:* The proof is omitted due to the page limit. ■

### B. Predictive variance

We now apply Laplace's method to approximate prediction error variance in a similar way. The prediction error variance is given by (14). In this case, we define

$$\begin{aligned}h(X) &= -\frac{1}{n}\ln\left(p(\bar{Y}|X)p(X|\bar{X})\right) \\ h^\star(X) &= -\frac{1}{n}\ln(\hat{\sigma}_\star^2(X) + \hat{y}_\star^2(X)) + h(X).\end{aligned} \tag{19}$$

Applying (7) to this case, the approximate of $\text{Var}(y_\star|\mathcal{D})$ and its order of the error are given by

$$\begin{aligned}\widehat{\text{Var}}(y_\star|\mathcal{D}) =& \frac{|D^2 h(\hat{X})|^{1/2}e^{nh(\hat{X})}}{|D^2 h^\star(\hat{X}_a)|^{1/2}e^{nh^\star(\hat{X}_a)}}C^\star(\hat{X}_a) - \hat{\mathbb{E}}[y_\star|\mathcal{D}]^2 \\ =& \text{Var}(y_\star|\mathcal{D}) + O(n^{-2}),\end{aligned} \tag{20}$$

where $\hat{X}$ is the exact mode of $-h$, and $\hat{X}_a$ is the asymptotic mode of $-h^\star$. $C^\star(\hat{X}_a)$ and $\hat{X}_a$ are given by (8).

### C. Simple Laplace approximations

To minimize the computational complexity, one may prefer a simpler approximation. In this paper, we propose such a simple approximation at the cost of precision, which is summarized in the following theorem.

*Theorem 4:* Let $\hat{X}$ be an asymptotic mode of order $O(n^{-1})$ for $-h$ given by (19). Assume that $\{h, \hat{X}\}$ satisfies the regularity conditions. Consider the following approximations for $\mathbb{E}[y_\star|\mathcal{D}]$ and $\text{Var}(y_\star|\mathcal{D})$

$$\hat{\mathbb{E}}[y_\star|\mathcal{D}] = k^T(\hat{X})(K(\hat{X}) + \sigma_w^2 I)^{-1}\bar{Y}, \tag{21}$$

$$\widehat{\text{Var}}(y_\star|\mathcal{D}) = \sigma_f^2 - k^T(\hat{X})(K(\hat{X}) + \sigma_w^2 I)^{-1}k(\hat{X}), \tag{22}$$

TABLE I

ERROR AND COMPLEXITY ANALYSIS

| Estimator | Method | Error | Complexity |
|---|---|---|---|
| MGF-PME in (16) | Laplace MGF | $O(n^{-2})$ | $O(n^6)$ |
| TP-PME in (18) | Laplace MGF | $O(n^{-3/2})$ | $O(n^5)$ |
| S-PME in (21) | Laplace MGF | $O(n^{-1})$ | $O(n^3)$ |

TABLE II

SIMULATION PARAMETERS

| Description | Parameter | Value |
|---|---|---|
| Number of agents with different positions | $q$ | 20 |
| Number of measurements | $n$ | 40 |
| The variability at a fixed point | $\sigma_f$ | $\sqrt{2}$ |
| Bandwidth | $\sigma_x$ | $\sqrt{2}$ |
| Noise covariance matrix of localization | $\Sigma_v$ | $\sqrt{0.1} \times I$ |
| Measurement noise level | $\sigma_w$ | 0.01 |

where $K(\hat{X})$ and $k(\hat{X})$ are covariance matrices as in (1) but obtained with $\hat{X}$. We have then the following order of errors.

$$\hat{\mathbb{E}}[y_\star|\mathcal{D}] = \mathbb{E}[y_\star|\mathcal{D}] + O(n^{-1}),$$
$$\widehat{\text{Var}}(y_\star|\mathcal{D}) = \text{Var}(y_\star|\mathcal{D}) + O(n^{-1}).$$

*Proof:* The proof is omitted due to the page limit. ■

*Remark 5:* As we previously mentioned, $\hat{X}$ is the mode of $-h$ given by (19) and is the MAP estimator of $X$, i.e., $\hat{X} = \hat{X}_{MAP}$, which is given by

$$\hat{X}_{MAP} = \arg\max_{X \in \mathcal{X}} p(\bar{Y}|X)p(X|\bar{X}).$$

Therefore, the difference in the simple Laplace approximations with respect to a quick-and-dirty solution in which the measured location vector $\bar{X}$ is used is that simple Laplace approximations use $\hat{X}$ instead of $\bar{X}$.

In applying Laplace's method, using the one step Newton gradient method to compute asymptotic modes, e.g., $\hat{X}_a$ required in (20) or $\hat{X}_+$ and $\hat{X}_-$ required in (18) may not satisfy the regularity conditions. In this case, one needs to continue the Newton gradient optimization until the regularity conditions are satisfied.

The order of the error and the computational complexity for the proposed approximation methods are summarized in Table I. A tradeoff between approximation error and complexity can be chosen taking into account the performance requirements and constrained resources for a specific sensor network application. For Laplace's method, the order of the error ranges from $O(n^{-1})$ to $O(n^{-2})$ at the cost of complexity from $O(n^3)$ to $O(n^6)$ as shown in Table I.

## V. SIMULATION RESULTS

In this section, we provide simulation results to evaluate the performances of different estimation methods. To this end, a snapshot of a realization of a Gaussian process that will serve as ground truth is shown in Fig. 1. The parameter values used to generate this Gaussian process are shown in Table II. $q = 20$ and $n = 40$ imply that each robot takes measurements twice at each sampling position.

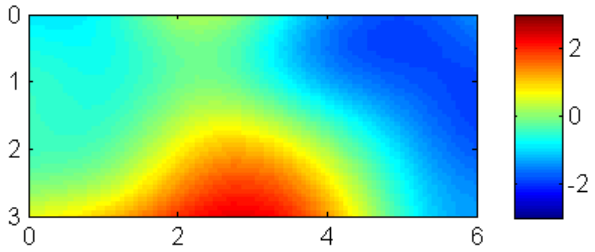


Fig. 1. A snapshot of a realization of a Gaussian process (ground truth).

The prediction results are summarized for the four methods of prediction described as follows.

- *Case 1:* The predicted field and the prediction error variance from applying the Gaussian process regression using the noiseless positions $X$ and the noisy measurement $\bar{Y}$ are shown in Fig. 2-(a) and (e), respectively.
- *Case 2:* When the measurements are taken repeatedly as suggested in Section III, To compare proposed approximation with typical quick-and-dirty solutions (QDS) to deal with noisy locations $\bar{X}$ in practice we use the conditional expectation of sampling positions $X$ given $\bar{X}$ as in (12) and the least squares solution of $Y$ for given $\bar{Y}$, which shall be plugged into Gaussian process regression, i.e.,

  $$\hat{y}_\star = k^T(H\bar{X})(K(H\bar{X}) + \sigma_w^2 I)^{-1}[(\bar{L}^T\bar{L})^{-1}\bar{L}^T\bar{Y}],$$

  where $H$ is from (12) and $\bar{L}$ is from (11). Figs. 2-(b) and (f) show the predicted field and the prediction error variance by applying QDS on $\bar{X}$ and $\bar{Y}$.
- *Case 3:* The predicted field and the prediction error variance, using Laplace approximations, are shown in Fig. 2-(c) and (g), respectively.
- *Case 4:* Figs. 2-(d) and (h) show the result of applying the simple Laplace approximations on noisy localizations and measurements data.

To numerically quantify the performance of each approach, we have computed the root mean square (RMS) error between the predicted and true fields over the two dimensional space for all methods, which are summarized as follows. The RMS errors for case 1, 2, 3 and 4 are 0.1281, 0.7374, 0.3320, and 0.3503, respectively. This RMS error analysis could be done since we know the true realization of the Gaussian process exactly in this simulation study. As expected, Gaussian process regression with the true locations perform best. The proposed approaches, i.e., Laplace approximations and simple Laplace approximations outperform QDS in terms of RMS errors as well. In addition, Fig. 2 clearly shows that the Laplace's method outperforms QDS.

## VI. CONCLUSION

We have formulated Gaussian process regression with observations under the localization uncertainty due to (possibly mobile) sensor networks with limited resources. Effects of observations, measurement noise, localization uncertainty and prior distributions have been all correctly incorporated

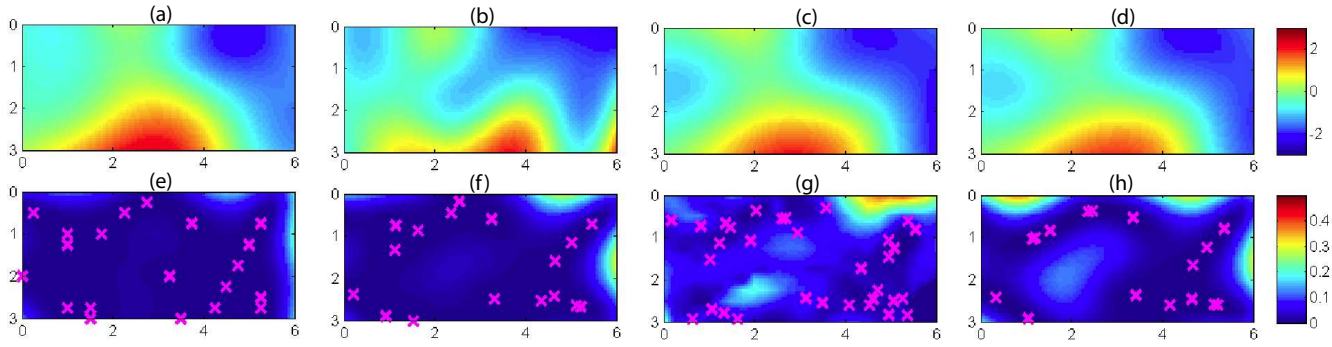


Fig. 2. The predicted fields and the prediction error variances for all 4 cases are shown in the first and second rows, respectively. The results of case 1, 2, 3, and 4 are shown in the first, second, third, and fourth columns. Pink crosses in (e), (f), (g), and (h) represent $X$, $\mathbb{E}(X|\bar{X})$, $\bar{X}$, and $\hat{X}$, respectively.

in the posterior predictive statistics in a Bayesian approach. We have reviewed Laplace's method, which has been applied to compute the analytically intractable posterior predictive statistics of the Gaussian processes with localization uncertainty. The approximation error and complexity of all proposed approaches have been analyzed. In particular, we have provided tradeoffs between approximation error and complexity of Laplace's method and its different degrees such that one can choose a tradeoff taking into account the performance requirements and computation complexity due to the resource-constrained sensor network. The simulation study demonstrated that the proposed approaches perform much better than approaches without considering the localization uncertainty properly.

## REFERENCES


[1] J. Choi, S. Oh, and R. Horowitz, "Distributed learning and cooperative control for multi-agent systems," *Automatica*, vol. 45, pp. 2802–2814, 2009.

[2] J. Cortés, "Distributed Kriged Kalman filter for spatial estimation," *IEEE Transactions on Automatic Control*, vol. 54, no. 12, pp. 2816–2827, 2010.

[3] N. Cressie, "Kriging nonstationary data," *Journal of the American Statistical Association*, vol. 81, no. 395, pp. 625–634, 1986.

[4] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. The MIT Press, Cambridge, Massachusetts, London, England, 2006.

[5] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg, "Near-optimal sensor placements: maximizing information while minimizing communication cost," in *Proceedings of the 5th international conference on Information processing in sensor networks*. ACM, 2006, pp. 2–10.

[6] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: theory, efficient algorithms and empirical studies," *The Journal of Machine Learning Research*, vol. 9, pp. 235–284, 2008.

[7] J. Choi, J. Lee, and S. Oh, "Swarm intelligence for achieving the global maximum using spatio-temporal Gaussian processes," in *Proceedings of the 27th American Control Conference (ACC)*, 2008.

[8] ——, "Biologically-inspired navigation strategies for swarm intelligence using spatial Gaussian processes," in *Proceedings of the 17th International Federation of Automatic Control (IFAC) World Congress*, 2008.

[9] L. Hu and D. Evans, "Localization for mobile sensor networks," in *Proceedings of the 10th annual international conference on Mobile computing and networking*. ACM, 2004, pp. 45–57.

[10] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte carlo localization for mobile robots," in *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 2, 1999, pp. 1322–1328.

[11] M. Mysorewala, D. Popa, and F. Lewis, "Multi-scale adaptive sampling with mobile agents for mapping of forest fires," *Journal of Intelligent and Robotic Systems*, vol. 54, no. 4, pp. 535–565, 2009.

[12] A. Girard, C. Rasmussen, J. Candela, and R. Murray-Smith, "Gaussian process priors with uncertain inputs-application to multiple-step ahead time series forecasting," *Advances in Neural Information Processing Systems*, pp. 545–552, 2003.

[13] M. Deisenroth, C. Rasmussen, and J. Peters, "Gaussian process dynamic programming," *Neurocomputing*, vol. 72, no. 7-9, pp. 1508–1524, 2009.

[14] O. Malve, M. Laine, H. Haario, T. Kirkkala, and J. Sarvala, "Bayesian modelling of algal mass occurrences–using adaptive MCMC methods with a lake water quality model," *Environmental Modelling & Software*, vol. 22, no. 7, pp. 966–977, 2007.

[15] M. Wainwright and M. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends® in Machine Learning*, vol. 1, no. 1-2, pp. 1–305, 2008.

[16] L. Tierney and J. Kadane, "Accurate approximations for posterior moments and marginal densities," *Journal of the American Statistical Association*, vol. 81, no. 393, pp. 82–86, 1986.

[17] L. Tierney, R. Kass, and J. Kadane, "Fully exponential Laplace approximations to expectations and variances of nonpositive functions," *Journal of the American Statistical Association*, vol. 84, no. 407, pp. 710–716, 1989.

[18] Y. Miyata, "Fully exponential Laplace approximations using asymptotic modes," *Journal of the American Statistical Association*, vol. 99, no. 468, pp. 1037–1049, 2004.

[19] ——, "Laplace approximations to means and variances with asymptotic modes," *Journal of Statistical Planning and Inference*, vol. 140, no. 2, pp. 382–392, 2010.

[20] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. Springer, 2006.

[21] Y. Xu and J. Choi, "Mobile sensor networks for learning anisotropic gaussian processes," in *Proceedings of the American Control Conference*. IEEE, 2009, pp. 5049–5054.

[22] Y. Xu, J. Choi, and S. Oh, "Mobile sensor network coordination using Gaussian processes with truncated observations," *IEEE Transactions on Robotics*, 2011.

[23] A. Brix and P. Diggle, "Spatiotemporal prediction for log-gaussian cox processes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 4, pp. 823–841, 2001.

[24] D. Titterton and J. Weston, *Strapdown inertial navigation technology*. Peter Peregrinus Ltd, 2004, vol. 17.

[25] H. Durrant-Whyte and T. Bailey, "Simultaneous localisation and mapping (SLAM): Part I the essential algorithms," *Robotics and Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.