Explorative Navigation of Mobile Sensor Networks Using Sparse Gaussian Processes

Songhwai Oh, Yunfei Xu, and Jongeun Choi

Abstract— **This paper presents an explorative navigation method using sparse Gaussian processes for mobile sensor networks. We first show that a near-optimal approximation is possible with a subset of measurements if we select the subset carefully, i.e., if the correlation between the selected measurements and the remaining measurements is small and the correlation between the prediction locations and the remaining measurements is small. An estimation method based on a subset of measurements is desirable for mobile sensor networks since we can always bound computational and memory requirements and unprocessed raw measurements can be easily shared with other agents for further processing (e.g., consensus-based distributed algorithms or distributed learning). We then present an explorative navigation method using sparse Gaussian processes with a subset of measurements. Using the explorative navigation method, mobile sensor networks can actively seek for new measurements to reduce the prediction error and maintain high-quality estimation about the field of interest indefinitely with limited memory.**

I. Introduction

Mobility in sensor networks can increase sensing coverage both in space and time and robustness against dynamic changes in the environment, showing superior performance in terms of their adaptability and high-resolution sampling capability [1]. While each agent has limited capabilities, a mobile sensor network as a group can perform various tasks, such as exploration, surveillance, and environmental monitoring. The problem of coordinated sensing in mobile sensor networks received significant attention recently [2]–[6].

The Gaussian process regression method has been widely used as a nonparametric regression technique for modeling complex physical phenomena, including nonstationary geostatistical data analysis [7], nonlinear regression [8], wireless signal strength estimation [9], indoor temperature field modeling [10], and terrain mapping [11]. Compared to parametric regression methods, nonparametric regression methods are more expressive and yield better generalizability. Hence, methods such as Gaussian processes are well suited to model complex environments. In particular, unlike other nonparametric methods, such as support vector machines (SVMs), Gaussian processes can provide the uncertainty about the predicted values and it is a highly desirable feature for mobile sensor networks. For instance, one can derive a multi-agent exploration strategy to minimize the predictive uncertainty of the field [12].

But Gaussian processes cannot be directly applied to resource-constrained mobile sensor networks due to its complexity in both time and space. In order to make predictions using Gaussian process regression, one must perform the inverse of a covariance matrix whose size grows as agents collect more measurements. In order to take the advantage of the expressiveness of Gaussian processes, we need an efficient approximation method and the need is more critical for resource-constrained mobile sensor networks. There are a number of approximation methods for Gaussian processes for handling large datasets [13]–[17]. These approximation methods aim to take the advantage of the sparsity of the covariance matrix and they can be grouped into two categories: approaches using only a subset of measurements, e.g., [13], and approaches using all measurements but with a reduced computation time, e.g., [14]–[17]. While the latter approaches perform better than the former approaches [18], we argue that the former approaches are more desirable for mobile sensor networks since we can always bound computational and memory requirements and unprocessed raw measurements can be easily shared with other agents for performing, for example, consensus based distributed algorithms or distributed learning.

In this paper, we first show conditions under which we can make a near-optimal prediction using Gaussian processes with a subset of measurements. We show that a good approximation is possible when the correlation between the selected measurements and the remaining measurements is small and the correlation between the prediction locations and the remaining measurements is small. These conditions translate to finding the most informative sparse representation of the covariance matrix using a subset of measurements. Based on this concept, we present an explorative navigation method using sparse Gaussian processes with a subset of measurements. The explorative navigation method allows mobile sensor networks to actively seek for new measurements in order to reduce the prediction error. This navigation method can be used for surveillance or monitoring a slowly changing spatio-temporal processes.

We have compared three approaches for the measurement selection problem: informative vector machine (IVM) [13], principal feature analysis (PFA) [19], and a mutual information based approach [20]. The results show that IVM performs better and requires less computation time, making it a better choice for resource-constrained mobile sensor networks.

This work has been supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2010-0013354) and the Research Settlement Fund for the new faculty of SNU.

Songhwai Oh is with the School of Electrical Engineering and Computer Science and ASRI, Seoul National University, Seoul 151-744, Korea. songhwai@snu.ac.kr.

Yunfei Xu and Jongeun Choi are with the Department of Mechanical Engineering, Michigan State University, East Lansing, MI 48824-1226, USA. xuyunfei@egr.msu.edu, jchoi@egr.msu.edu.

II. Gaussian Process Approximation Using a Subset of Measurements

A. Gaussian Processes

A Gaussian process[1] defines a distribution over a space of functions and it is completely specified by its mean function and covariance function. Let $X \in \mathcal{D} \subseteq \mathbb{R}^d$ denote the index vector. If $z(X) \in \mathbb{R}$ is a Gaussian process, it can be written as

$$z(X) \sim \mathcal{GP}(\mu(X), \mathcal{K}(X, X')), \tag{1}$$

where $\mu(X) = \mathbb{E}(z(X))$ is a mean function and $\mathcal{K}(X, X') = \mathbb{E}\left[(z(X) - \mu(X))(z(X') - \mu(X'))\right]$ is a covariance function of $z(X)$. An example of a covariance function is a radial basis kernel

$$\mathcal{K}(X, X') = \sigma_f^2 \exp\left(-\frac{1}{2\sigma_l^2}\|X - X'\|^2\right). \tag{2}$$

In a realistic setup, one does not observe a Gaussian process directly. Instead, we make a noisy measurement $y(X) = z(X) + w$, where w is a white Gaussian noise with variance σ_w^2. Suppose that the mean function of $z(X)$ is zero and we have made n measurements, i.e., $Y = [y_1, y_1, \ldots, y_n]^T \in \mathbb{R}^n$, at $X_1, X_2, \cdots, X_n \in \mathcal{D}$, then Y has a Gaussian distribution

$$Y \sim \mathcal{N}(0, \Sigma + \sigma_w^2 I), \tag{3}$$

where $\Sigma = [\Sigma_{ij}]$ is an $n \times n$ matrix with $\Sigma_{ij} = \mathcal{K}(X_i, X_j)$ and I is an identity matrix with an appropriate size. Given measurements, we can predict the value of the Gaussian process at an unobserved location X_* and the predicted value has a Gaussian distribution $z(X_*) \sim \mathcal{N}(\hat{z}(X_*), \mathbf{var}(z(X_*)))$, where

$$\begin{aligned} \hat{z}(X_*) &= k_*^T(\Sigma + \sigma_w^2 I)^{-1} Y \\ \mathbf{var}(z(X_*)) &= \mathcal{K}(X_*, X_*) - k_*^T(\Sigma + \sigma_w^2 I)^{-1} k_* \end{aligned}$$

with $k_* = [\mathcal{K}(X_1, X_*), \ldots, \mathcal{K}(X_n, X_*)]^T$ [18]. The predicted variance $\mathbf{var}(z(X_*))$ measures the uncertainty we have about the new location X_* and can be used in mobile sensor networks for exploration. The predicted mean can be rewritten as

$$\hat{z}(X_*) = \sum_{i=1}^{n} \alpha_i \mathcal{K}(X_i, X_*), \tag{4}$$

where $\alpha = (\Sigma + \sigma_w^2 I)^{-1} Y$.

B. Approximation Using a Subset of Measurements

The major drawback of Gaussian processes is that its computational and space complexity increase as more measurements are collected, making the method undesirable for mobile sensor networks with limited memory and computational power. Hence, it is required to develop an approximation method for Gaussian processes that is suitable for mobile sensor networks. While there are a number of approximation methods for Gaussian processes, it is unclear what the exact conditions are required for achieving a good approximation. In this section, we formalize conditions required for near-optimal approximations using a subset of measurements and motivate the sparse Gaussian process method described in this paper.

In general, if measurements are taken from nearby locations, covariance between measurements is strong and covariance decays as the distance between locations increases. If the covariance function of a Gaussian process has this property, intuitively, we can make a good prediction using only local measurements. In this section, we formalize this observation and provide a theoretical foundation for Gaussian processes with a subset of measurements.

Suppose that we want to approximate $\hat{z}(X_*)$ using only m measurements out of n measurements ($m < n$). Let this new predictor be $\hat{z}_m(\cdot)$. Without loss of generality, $\hat{z}_m(\cdot)$ is computed using the first m measurements. Let $Y_m = [y_1, \ldots, y_m]^T$. The covariance matrix Σ can be represented as

$$\Sigma = \begin{bmatrix} \Sigma_m & \Sigma_{mr} \\ \Sigma_{mr}^T & \Sigma_r \end{bmatrix},$$

where Σ_m is an $m \times m$ submatrix of Σ. Then for X_*,

$$\hat{z}_m(X_*) = K_m^T \left(\Sigma_m + \sigma_w^2 I_m\right)^{-1} Y_m, \tag{5}$$

where $K_m = [\mathcal{K}(X_1, X_*), \ldots, \mathcal{K}(X_m, X_*)]^T$.

Theorem 1: Consider a zero-mean Gaussian process defined in (1). Suppose that you have collected n measurements, $y_1, \cdots, y_n$. For an unobserved input X_*, let $\hat{z}(X_*)$ be a predicted value computed using all measurements, i.e., using (4), and let $\hat{z}_m(X_*)$ be a predicted value using the first m measurements $Y_m = [y_1, \ldots, y_m]^T$. Then

$$\begin{aligned} \hat{z}(X_*) - \hat{z}_m(X_*) &= \left(K_r^T - K_m^T A^{-1} B\right) \\ &\times \left(C - B^T A^{-1} B\right)^{-1} \\ &\times \left(Y_r - B^T A^{-1} Y_m\right) \end{aligned} \tag{6}$$

where $Y_r = [y_{m+1}, \ldots, y_n]^T$, $A = \Sigma_m + \sigma_w^2 I_m$, $B = \Sigma_{mr}$, $C = \Sigma_r + \sigma_w^2 I_r$, and $K_r = [\mathcal{K}(X_{m+1}, X_*), \ldots, \mathcal{K}(X_n, X_*)]^T$.

The proof of Theorem 1 is omitted for lack of space.

Corollary 1: Under the assumptions used in Theorem 1, if $B = 0$, then $\hat{z}(X_*) - \hat{z}_m(X_*) = K_r^T C Y_r$.

Theorem 1 shows the gap between predicted values using only m measurements and all measurements. Corollary 1 is the case with an ideal partition of measurements, i.e., the first m measurements are not correlated with the remaining $(n - m)$ measurements. Hence, if the magnitude of B is small, then the error from using the first m measurements will be close to $K_r^T C Y_r$. Furthermore, if we want to reduce this error, we want K_r to be small, i.e., the correlation between X_* and the remaining $(n - m)$ measurements is small. In summary, in order to minimize the error between predictions using a subset of measurements and predictions using all measurements, following two conditions must be satisfied: (1) the correlation between the first m measurements and the remaining $(n-m)$ measurements is small and (2) the correlation between X_* and the remaining $(n - m)$

[1] A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution [18].

measurements is small. If these two conditions are met, then $\hat{z}_m(X_*)$ can be a good approximation to $\hat{z}(X_*)$, justifying a Gaussian process with a subset of measurements. We can formalize this argument as follows. Unless noted otherwise, $\|\cdot\|$ denotes the 2-norm $\|\cdot\|_2$.

Theorem 2: Consider a zero-mean Gaussian process defined in (1) with covariance function (2) and suppose that you have collected n measurements, $y_1, \ldots, y_n$. Suppose that B defined in Theorem 1 is small enough such that $\|K_m^T A^{-1} B\| \leq \delta_0 \|K_r\|$ for some $\delta_0 > 0$, $\delta_1 I_r \succeq B^T A^{-1} B$ for some $\delta_1 > 0$ with $\sigma_w^2 > \delta_1$, and $\|B^T A^{-1} Y_m\| \leq \delta_2 \|Y_r\|$ for some $\delta_2 > 0$. Given $0 < \delta_3 < 1$, choose $\bar{y}(\delta_3)$ such that $\max_{i=1}^n |y_i| < \bar{y}(\delta_3)$ with probability δ_3 and $\epsilon > 0$ such that

$$\epsilon < \frac{\sigma_f^2 (n-m)(1+\delta_0)(1+\delta_2)\bar{y}(\delta_3)}{(\sigma_w^2 - \delta_1)}.$$

For X_*, if the last $(n-m)$ data points satisfy

$$\|X_i - X_*\|^2 > 2\sigma_l^2 \log\left(\frac{\sigma_f^2 (n-m)(1+\delta_0)(1+\delta_2)\bar{y}(\delta_3)}{\epsilon(\sigma_w^2 - \delta_1)}\right),$$

then, with probability δ_3, we have $|\hat{z}(X_*) - \hat{z}_m(X_*)| < \epsilon$.

Proof: Let $D = A^{-1}B$ and $E = B^T A^{-1} B$ for notational convenience. Then

$$\begin{aligned}
&|\hat{z}(X_*) - \hat{z}_m(X_*)| \\
&\quad = |(K_r^T - K_m^T D)(C-E)^{-1}(Y_r - D^T Y_m)| \\
&\quad \leq \|(K_r^T - K_m^T D)\| \|(C-E)^{-1}(Y_r - D^T Y_m)\| \\
&\quad \leq \|(K_r^T - K_m^T D)\| \\
&\qquad \times (\|(C-E)^{-1} Y_r\| + \|(C-E)^{-1} D^T Y_m\|) \\
&\quad \leq (1+\delta_0)\|K_r\|(\|(C-E)^{-1} Y_r\| \\
&\qquad + \|(C-E)^{-1} D^T Y_m\|) \\
&\quad \leq (1+\delta_0)\|K_r\|(\|(C-\delta_1 I_r)^{-1} Y_r\| \\
&\qquad + \|(C-\delta_1 I_r)^{-1} D^T Y_m\|)
\end{aligned}$$

Since Σ_r is positive definite, there is Q such that $\Sigma_r = QQ^T$. Using the matrix inversion lemma, we have

$$\begin{aligned}
(C-\delta_1 I_r)^{-1} &= (\Sigma_r + \sigma_w^2 I_r - \delta_1 I_r)^{-1} \\
&= (QQ^T + \gamma I_r)^{-1} \\
&= \gamma^{-1} I_r - \gamma^{-2} Q(I_r + \gamma^{-1} Q^T Q)^{-1} Q^T \\
&\preceq \gamma^{-1} I_r,
\end{aligned}$$

where $\gamma = \sigma_w^2 - \delta_1$. Combining this result, we get

$$\begin{aligned}
&|\hat{z}(X_*) - \hat{z}_m(X_*)| \\
&\quad \leq \frac{1}{\gamma}(1+\delta_0)\|K_r\| \left(\|Y_r\| + \|D^T Y_m\|\right) \\
&\quad \leq \frac{1}{\gamma}(1+\delta_0)(1+\delta_2)\|K_r\|\|Y_r\| \\
&\quad \leq \frac{1}{\gamma}(1+\delta_0)(1+\delta_2)\sqrt{n-m}\mathcal{K}_{\max}\|Y_r\|,
\end{aligned}$$

where $\mathcal{K}(x_i, x_*) \leq \mathcal{K}_{\max}$ for $i \in \{m+1, \ldots, n\}$. Define $\bar{y}(\delta_3)$ such that $\max_{i=m+1}^n |y_i| \leq \bar{y}(\delta_3)$ with probability δ_3. Then, with probability δ_3, we have

$$|\hat{z}(X_*) - \hat{z}_m(X_*)| \leq \frac{1}{\gamma}(1+\delta_0)(1+\delta_2)(n-m)\mathcal{K}_{\max}\bar{y}(\delta_3)$$

Hence, for $\epsilon > 0$, if

$$\mathcal{K}_{\max} < \frac{\epsilon(\sigma_w^2 - \delta_1)}{(n-m)(1+\delta_0)(1+\delta_2)\bar{y}(\delta_3)} \tag{7}$$

with probability δ_3, we have

$$|\hat{z}(X_*) - \hat{z}_m(X_*)| < \epsilon. \tag{8}$$

Now, suppose that we have a radial basis kernel (2). Let $r_i^2 = \|X_i - X_*\|^2$. Then (7) becomes, with probability δ_3,

$$\sigma_f^2 \exp\left(-\frac{r_i^2}{2\sigma_l^2}\right) \leq \mathcal{K}_{\max} < \frac{\epsilon(\sigma_w^2 - \delta_1)}{(n-m)(1+\delta_0)(1+\delta_2)\bar{y}(\delta_3)}$$

$$r_i^2 > -2\sigma_l^2 \log\left(\frac{\epsilon(\sigma_w^2 - \delta_1)}{\sigma_f^2 (n-m)(1+\delta_0)(1+\delta_2)\bar{y}(\delta_3)}\right)$$

For $\epsilon < \sigma_f^2 (n-m)(1+\delta_1)(1+\delta_2)\bar{y}(\delta_3)/(\sigma_w^2 - \delta_1)$, we have

$$r_i^2 \quad > \quad 2\sigma_l^2 \log\left(\frac{\sigma_f^2 (n-m)(1+\delta_0)(1+\delta_2)\bar{y}(\delta_3)}{\epsilon(\sigma_w^2 - \delta_1)}\right)$$

for all i and this completes the proof. ■

III. Explorative Navigation Using Sparse Gaussian Processes

In the previous section, we have found conditions for a near-optimal approximation of Gaussian processes with a subset of measurements. Since we are interested in minimizing the prediction error over the entire region of a continuous space, we cannot directly apply the conditions. However, in order to maximize the predictive power over the unvisited regions using a subset of collected measurements, we can seek for the most informative sparse representation of measurements eliminating any redundant information, i.e., a sparse representation of the covariance matrix. Hence, the conditions found in the previous section translate to finding the most informative sparse representation of the covariance matrix using a subset of measurements. To find a good subset of measurements, we consider the following three approaches: the informative vector machine [13], principal feature analysis (PFA) [19], and the mutual information based approach [20].

Informative Vector Machine: The informative vector machine (IVM) proposed by Lawrence et al. finds the sparsity in the data set and have shown that the IVM performs as good as support vector machines (SVMs) at a fraction of computation time [13]. The IVM finds a sparse representation of a covariance matrix using an active set $\mathcal{A}$ of measurements ($|\mathcal{A}| < n$, where n is the total number of measurements). The IVM uses a heuristic called the *differential entropy score*. For measurement y_j, the differential entropy score is defined as $\Delta_j = H(z(X_j)|\mathcal{A}) - H(z(X_j)|\mathcal{A} \cup y_j)$. The algorithm starts with $\mathcal{A} = \emptyset$ and measurements are sequentially added into $\mathcal{A}$. At each step of the algorithm, Δ_j are computed for the remaining measurements in $Y \setminus \mathcal{A}$ and the measurement with the largest differential entropy score is added to $\mathcal{A}$. The algorithm continues until the desired number of measurements are selected for $\mathcal{A}$. Since

the entropy of a Gaussian random variable with variance σ^2 is $\log(2\pi e\sigma^2)/2$, the maximum differential entropy criteria reduces to choosing the measurements with largest predictive variances.

Principal Feature Analysis: Principal feature analysis (PFA) proposed in [19] is a dimensionality reduction method used for feature selection. One drawback of dimensionality reduction methods such as principal component analysis (PCA) is that the discovered mapping from the original high-dimensional feature space to a lower dimensional space uses all original features as input. This can be undesirable if each feature requires an additional sensor or if we aim to reduce preprocessing time and memory. PFA avoids this problem by selecting a subset of features with most information. PFA has been applied to select facial points for face tracking and content-based image retrieval [19] and an approach similar to PFA has been successfully applied to find a subset of features from multivariate time series data, such as human gait data and electroencephalogram (EEG) data [21].

PFA is based on PCA and exploits the structure of principal components. A principal component is a linear transformation of original features and this transformation is described by coefficients of principal components. These coefficients can be understood as contributions or weights on determining the directions of principal components [21]. Hence, when the absolute value of the i-th coefficient of a principal component is high, we can interpret this as an indication that the i-th feature is dominant in that principal component. If our goal is to choose m most representative features, we can choose features with highest coefficients of each of the first m principal components, as an approximation to PCA. If each principal component is considered separately, features with similar information or high mutual information can be chosen. PFA addresses this problem by considering all principal components together and eliminates redundant features using clustering.

Mutual information: Guestrin et al. used the mutual information (MI) for choosing sensor locations in [20]. A subset $\mathcal{A}$ of possible sensor locations is chosen to maximize the mutual information between $\mathcal{A}$ and the remaining set $Y \setminus \mathcal{A}$. Again, the algorithm starts with $\mathcal{A} = \emptyset$ and adds measurements to $\mathcal{A}$ sequentially. A measurement y_j is added to $\mathcal{A}$ if y_j maximizes $H(z(X_j)|\mathcal{A}) - H(z(X_j)|\{Y \setminus (\mathcal{A} \cup y_j)\})$. This approach can provide an approximation bound over the optimal solution using submodularity and monotonicity of the mutual information under certain conditions [20]. However, as shown below, we find that this criteria is not suited for the measurement selection problem. The method was designed to consider all potential locations of interest and this is infeasible for a continuous field considered in this paper. Although one could discretize the continuous field, the computational cost will be prohibitively high.

A. Explorative Navigation

According to the results from Section II, we can make a near-optimal approximation to a Gaussian process using only a subset of measurements if we can carefully choose a good subset of measurements. For the measurement selection problem, we can apply IVM, PFA or MI. The explorative navigation using a sparse Gaussian process is shown in Algorithm 1. The algorithm is based on the navigation strategy developed in [12], which coordinates a group of mobile agents to reduce the predictive variance of the surveillance region while maintaining a certain distance between each pair of agents using consensus and flocking. By using only a subset of measurements, we can reduce computational and memory requirements of the Gaussian process regression method and make Gaussian processes practical for mobile sensor networks. Maintaining a subset of raw measurements is also desirable for distributed mobile sensor networks since they can be easily shared with neighboring agents without further processing, e.g., consensus based distributed algorithms and distributed learning by sharing measurements of interest.

Algorithm 1 Explorative Navigation Using Sparse GPs

Input: m (size of selected measurements), t_s (update interval), r (resolution)

$\mathcal{A} = \emptyset$
while true **do**
  $t = t + 1$ (update the current time)
  Collect measurements y_t from all neighboring agents.
  if $t > m$ and $\text{mod}(t, t_s) = 0$ **then**
    (Measurement Update) Let $\mathcal{A}$ be the m measurements selected from $\mathcal{A} \cup \{y_{t-t_s+1}, \ldots, y_t\}$ using a measurement selection algorithm.
    Discard unselected measurements.
  else if $t \leq m$ **then**
    $\mathcal{A} = \mathcal{A} \cup y_t$
  end if
  Compute predicted variances at evenly divided locations at resolution r using $\mathcal{A}$ and additional measurements collected since the last measurement update.
  Let p be the location with the highest predicted variance.
  Compute the control law to move towards p while maintaining distances and velocities among agents based on the gradient based control described in [12].
  Apply the control and move to the new location.
end while

IV. Simulation Results

A. Sparse Gaussian Processes

We first study performance of sparse Gaussian processes using informative vector machine (IVM), principal feature analysis (PFA), and mutual information based measurement selection algorithm (MI). We consider a static field as shown in Figure 1. The field is constructed using the radial basis kernel (2), where $\sigma_f^2 = 1$ and $\sigma_l^2 = 2$. From this field, we made 500 measurements at random locations with measurement noises ($\sigma_w = 0.2$). We first ran three measurement selection algorithms for a different number of selected measurements

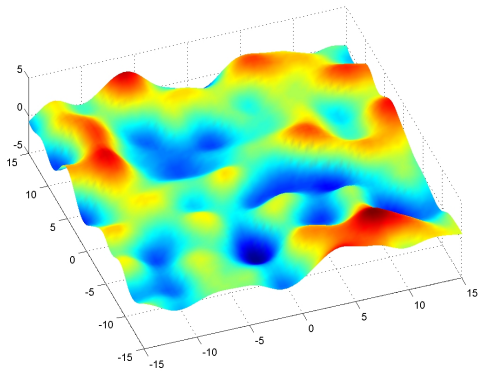


Fig. 1. A field modeled by a Gaussian process with the radial basis kernel (2), where $\sigma_f^2 = 1$ and $\sigma_l^2 = 2$.

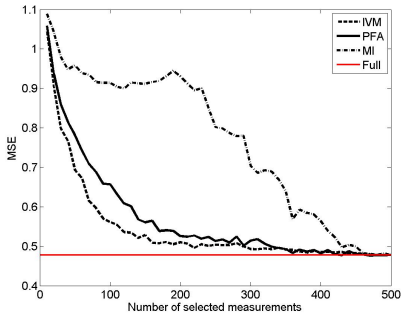


Fig. 2. Mean square errors of three measurement selection algorithms as a function of the number of selected measurements. All values are the averages of 10 independent runs. (IVM: informative vector machine, PFA: principal feature analysis, MI: mutual information based measurement selection algorithm, and Full: All 500 measurements are used.)

(from 5 to 500 with an increment of 5). To improve the performance of the mutual information based measurement selection algorithm, we used the local kernels as suggested in [20]. We then measured the mean square error of predictions made by each approach at 3,721 evenly discretized locations, i.e., a grid with unit length 0.5. The resulting mean square errors are shown in Figure 2.

As shown in the figure, the informative vector machine approach gives the best performance in terms of the mean square error while the principal feature analysis approach gives a comparable performance. But the mutual information approach showed the poorest performance. It indicates that the mutual information criteria is not suitable for the measurement selection problem. This is an expected result since the mutual information criteria was designed to consider all potential sensing locations [20]. One could consider including all the 3,721 distinct sites. But it is not a computationally feasible approach for a continuous field considered here since it would require more sites if a finer resolution of the field is desired. We have found that IVM and PFA can provide a good approximation with a smaller number of measurements and sparse Gaussian processes are well suited for resource-constrained mobile sensor networks.

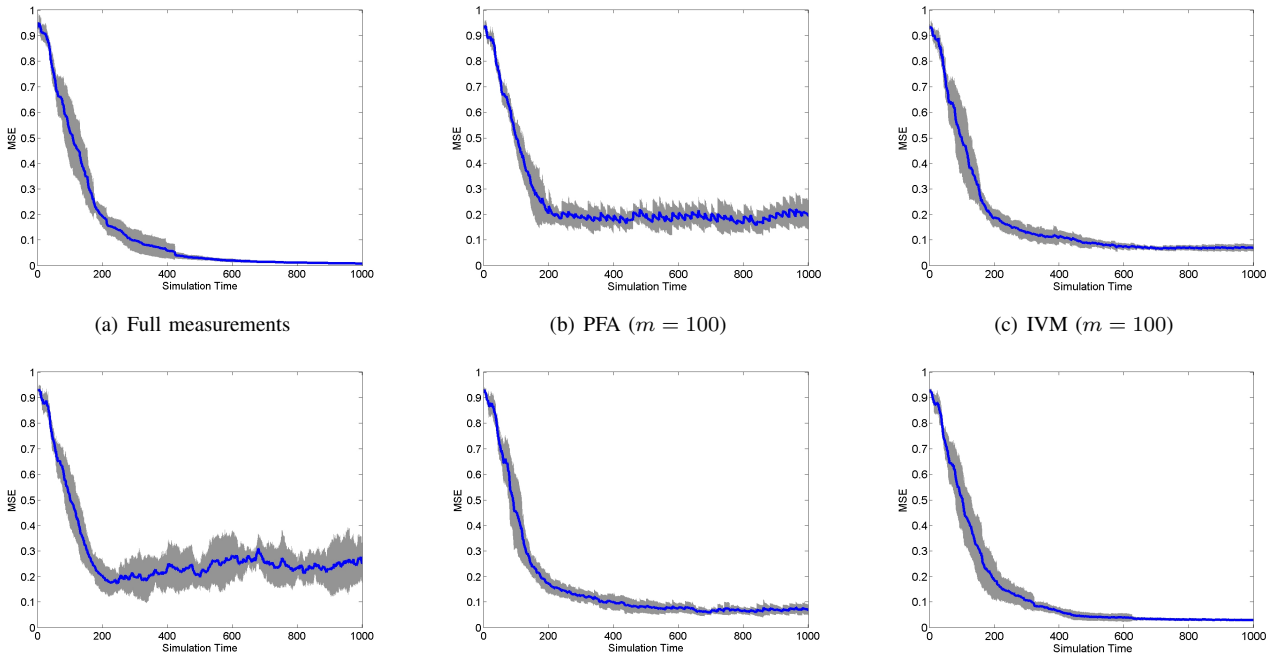
B. Explorative Navigation Using Sparse Gaussian Processes

We consider a field similar to Figure 1 with the same set of parameters except the surveillance region is now $[-10, 10]^2$. The explorative navigation algorithm (Algorithm 1) is used to coordinate a group of five agents with $r = 0.5$. The mean square error of predictions at 1,681 evenly discretized locations, i.e., a grid with unit length 0.5, is measured at each simulation time. Since the order in which measurements are considered in the IVM may influence its performance, we randomly shuffled the order of measurements at each simulation time.

Results are shown in Figure 3. Figure 3(a) shows the case when all past measurements are used to make predictions. We can see that the explorative navigation strategy rapidly reduces the mean square error over the entire region. Figure 3(b) and Figure 3(c) show mean square errors from navigating with PFA and IVM with $m = 100$ and $t_s = 20$, respectively (at most 120 measurements are used at each time). It should be noted that the mean square error keeps decreasing after the simulation time of 100, which means the algorithm continuously selects the most informative measurements. Figure 3(d) shows the case when a measurement selection algorithm is not used but recent 220 measurements are used for making predictions. The mean square error does not stabilize and there is a large variability in the quality of prediction. This case shows the poor performance of uninformative selection of measurements and clearly indicates the benefits of sparse Gaussian processes. Figure 3(e) and Figure 3(f) show mean square errors from navigating with PFA and IVM with $m = 200$ and $t_s = 20$, respectively (at most 220 measurements are used at each time). We find that the explorative navigation approach using IVM is more stable than the approach with PFA. Even with 100 measurements, the mobile sensor network using IVM can maintain a high quality estimation of the entire region, showing its stability.

V. Conclusions

In this paper, we have shown formal conditions under which a near-optimal approximation is possible for Gaussian processes using a subset of measurements. A near-optimal approximation is possible when (1) the correlation between the selected measurements and the remaining measurements is small and (2) the correlation between the prediction locations and the remaining measurements is small. Based on these principles, we have presented a sparse Gaussian process based explorative navigation algorithm using a subset of measurements. There are two major benefits of using a subset of measurements to approximate Gaussian processes. First, we can greatly reduce the computational and memory requirements. Second, unprocessed raw measurements can be easily shared with other agents for performing consensus based distributed algorithms or distributed learning. In our future work, we plan to develop a path planning algorithm based on sparse Gaussian processes.



(a) Full measurements (b) PFA ($m = 100$) (c) IVM ($m = 100$)

(d) Recent 220 measurements (e) PFA ($m = 200$) (f) IVM ($m = 200$)

Fig. 3. Results from explorative navigation. Mean square errors are shown as a function of simulation time. The explorative navigation algorithm (Algorithm 1) is used to coordinate a group of five agents with $r = 0.5$ and $t_s = 20$. The mean square error of predictions at 1,681 evenly discretized locations is measured at each time. The mean square errors are the averages of 10 independent runs and the gray area indicates the standard deviation of those runs.

REFERENCES

[1] A. Singh, M. Batalin, M. Stealey, V. Chen, Y. Lam, M. Hansen, T. Harmon, G. S. Sukhatme, and W. Kaiser, "Mobile robot sensing for environmental applications," in *Proceedings of the International Conference on Field and Service Robotics*, 2007.

[2] J. Cortes, S. Martinez, T. Karatas, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 2, pp. 243–255, 2004.

[3] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[4] H. G. Tanner, A. Jadbabaie, and G. J. Pappas, "Stability of flocking motion," University of Pennsylvania," Technical Report, 2003.

[5] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.

[6] W. Ren and R. W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Transactions on Automatic Control*, vol. 50, no. 5, pp. 655–661, May 2005.

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

[8] M. Gibbs and D. MacKay, "Efficient implementation of Gaussian processes for interpolation," Department of Physics, Cavendish Laboratory, Cambridge University, Tech. Rep., 1996. [Online]. Available: http://wol.ra.phy.cam.ac.uk/mackay/GP

[9] B. Ferris, D. Fox, and N. Lawrence, "WiFi-SLAM using Gaussian process latent variable models," in *Proc. of the 20th International Joint Conference on Artificial Intelligence*, 2007.

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

[11] H. Durrant-Whyte, F. Ramos, E. Nettleton, A. Blair, and S. Vasudevan, "Gaussian process modeling of large scale terrain," in *Proc. of the 2009 IEEE International Conference on Robotics and Automation*, 2009.

[12] J. Choi, J. Lee, and S. Oh, "Swarm intelligence for achieving the global maximum using spatio-temporal Gaussian processes," in *Proc. of the American Control Conference*, Seattle, WA, June 2008.

[13] N. Lawrence, M. Seeger, and R. Herbrich, "Fast sparse Gaussian process methods: The informative vector machine," *Advances in Neural Information Processing Systems*, pp. 625–632, 2003.

[14] A. J. Smola and P. L. Bartlett, "Sparse greedy Gaussian process regression," in *Advances in Neural Information Processing Systems 13*, 2001, pp. 619–625.

[15] C. Williams and M. Seeger, "Using the Nystrom method to speed up kernel machines," in *Advances in Neural Information Processing Systems 13*. MIT Press, 2001, pp. 682–688.

[16] M. Seeger, "Bayesian Gaussian process models: PAC-Bayesian generalisation error bounds and sparse approximations," Ph.D. dissertation, School of Informatics, University of Edinburgh, 2003.

[17] V. Tresp, "A Bayesian committee machine," *Neural Computation*, vol. 12, pp. 2719–2741, 2000.

[18] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, Massachusetts, London, England, 2006.

[19] Y. Lu, I. Cohen, X. Zhou, and Q. Tian, "Feature selection using principal feature analysis," in *Proc. of the 15th international conference on Multimedia*, 2007.

[20] C. Guestrin, A. Krause, and A. Singh, "Near-optimal sensor placements in gaussian processes," in *Proc. of the International Conference on Machine Learning*, 2005.

[21] H. Yoon, K. Yang, and C. Shahabi, "Feature subset selection and feature ranking for multivariate time series," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 9, pp. 1186–1198, 2005.