Swarm Intelligence for Achieving the Global Maximum using Spatio-Temporal Gaussian Processes

Jongeun Choi, Joonho Lee, and Songhwai Oh

***Abstract*— This paper presents a novel class of self-organizing multi-agent systems that form a swarm and learn a spatio-temporal process through noisy measurements from neighbors for various global goals. The physical spatio-temporal process of interest is modeled by a spatio-temporal Gaussian process. Each agent maintains its own posterior predictive statistics of the Gaussian process based on online measurements from neighbors. A set of biologically inspired navigation strategies are identified from the posterior predictive statistics. A unified way to prescribe a global goal for the group of agents is presented. A reference trajectory state that guides agents to achieve the maximum of the objective function is proposed. A switching protocol is proposed for achieving the global maximum of a spatio-temporal Gaussian process over the surveillance region. The usefulness of the proposed multi-agent system with respect to various global goals is demonstrated by several numerical examples.**

I. Introduction

In recent years, significant enhancements have been made in the areas of sensor networks and mobile sensing agents. Emerging technologies have been reported on coordination of mobile sensing agents [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11]. Mobile sensing agents form an ad-hoc wireless communication network in which each agent operates usually under a short communication range and some computational power. Among challenging problems of distributed coordination of mobile sensing agents, tracing the maximum of an unknown field has attracted much attention of control engineers. This has numerous applications including homeland security, toxic-chemical plume tracing and environmental monitoring. For instance, the most common approach to toxic-chemical plume tracing has been biologically inspired *chemotaxis* [12], [13], in which a mobile sensing agent is driven according to a local gradient of the chemical plume concentration. However, in this case, the convergence rate can be slow and the mobile robot may easily get stuck in the local maxima of chemical plume concentration. Thus, robots require some "artificial intelligence" for them to estimate and predict *global features* of a spatio-temporal field based on samples in different time and space, which will be discussed mainly in this paper. The cooperative network of agents that performs adaptive gradient climbing in a distributed environment was presented in [14], [15]. The centralized network can adapt its configuration in response to the sensed environment in order to optimize its gradient climb.

Many of the mobility of the mobile agents are designed based on a certain field of interest. Recently distributed interpolation schemes for field estimation by mobile sensor networks are developed by [7]. Swarming sensing agents with a gradient ascent strategy for tracing the maximum of an unknown field via radial basis function learning were proposed by [10].

Our motivation is to design the mobility of sensing agents for various tasks by intelligently dealing with uncertainty in the estimation of a spatial phenomenon based on online measurements, and exploiting the posterior predictive statistics, which will lead to "learning agents". In maximizing an objective function, we are interested in achieving the global maximum rather than obtaining usual local maxima.

In our approach, the dynamical phenomenon in the surveillance region $\mathcal{R}$ will be specified by a spatio-temporal Gaussian process. A Gaussian process (or Kriging in geostatistics) has been widely used as a nonlinear regression technique to estimate and predict geostatistical data [16], [17], [18], [19], [20]. A spatio-temporal Gaussian process $z(s,t) \sim \mathcal{GP}(\mu(s,t), \mathcal{K}(\{s,t\},\{s',t'\}))$, $s(t), s'(t') \in \mathcal{R}, t, t' \in \mathbb{Z}_+$ is specified by its mean function $\mu(\cdot,t)$ and a symmetric positive definite covariance function $\mathcal{K}(\cdot,\cdot)$. Gaussian processes enable us to predict physical values, such as temperature and plume concentration, at any of spatial points with a predicted uncertainty level. A class of spatio-temporal Gaussian processes has been extensively investigated in the form of combining spatial Gaussian processes and Kalman filtering, which became so-called "space-time Kalman filter" [21], [22], [23], [24]. This model will be used in the paper. Recently near-optimal static sensor placements with a mutual information criterion in Gaussian processes were proposed by [25]. Distributed Kriged Kalman filter for spatial estimation based on mobile sensor networks are developed by [9]. Asymptotic optimality of multicenter Voronoi configurations for random field estimation is reported by [8].

The contribution of this paper is to introduce coordination algorithms that exploit posterior predictive statistics from a Gaussian process learning mechanism. Combined with a flocking algorithm [5], [6], [10] for distributed sampling, this approach provides an "artificial intelligence" to each agent to form cooperatively learning mobile agents that perform useful tasks. Depending on the tasks for the sensing agents, the collective mobility of sensing agents can be designed for various performance criterions. In particular, a set of navigation modes (tracing, avoidance, exploration with respect to the predicted field) for swarming agents is precisely prescribed based on the recursively updated Gaussian process. An interesting switching protocol is introduced to overcome the tradeoff between exploitation and exploration, which leads to the global maximum. The usefulness of the proposed coordination algorithm is demonstrated by several interesting numerical examples.

The proposed cooperatively learning control consists of motion coordination based on the recursive estimation of an unknown field of interest with measurement noise. Our

Jongeun Choi and Joonho Lee are with the Department of Mechanical Engineering, Michigan State University { jchoi, leejoon8 } @egr.msu.edu

Songhwai Oh is with the School of Engineering, University of California, Merced songhwai.oh@ucmerced.edu

strategy of the cooperative learning control can be applied to a large class of coordination algorithms for mobile agents to deal with the field of interest that requires to be recursively estimated.

This paper is organized as follows. In Section II, we briefly review the mobile sensing network model, notations related to a graph, and artificial potentials to form a swarming behavior. A recursive learning algorithm for estimating parameters and predicting a spatio-temporal Gaussian process is presented in Section III. Section IV explains the biologically inspired navigation and a unified way to prescribe the global goal for agents. A reference trajectory state that guides agents to achieve the global maximum of the objective function is presented as well. In Section V, the resulting cooperatively learning control is described. In Section VI, we numerically test agents under different navigation strategies and a switching protocol with respect to several configurations and different Gaussian processes.

II. Preliminaries

In this section, we explain notations and concepts that will arise throughout the paper.

A. Mobile Sensing Agent Network

First, we explain the mobile sensing network and sensor models used in this paper. Let N_s be the number of sensing agents distributed over the surveillance region $\mathcal{R} \subset \mathbb{R}^2$. Assume that $\mathcal{R}$ is a compact set. The identity of each agent is indexed by $\mathcal{I} := \{1, 2, \cdots, N_s\}$. Let $q_i(t) \in \mathcal{R}$ be the location of the i-th sensing agent at time $t \in \mathbb{R}_+$ and let $q := \text{col}(q_1, q_2, \cdots, q_{N_s}) \in \mathbb{R}^{2N_s}$ be the configuration of the swarm system. The discrete time, high-level dynamics of agent i is modeled by

$$\left\{ \begin{array}{l} q_i(t+1) = q_i(t) + \epsilon p_i(t), \\ p_i(t+1) = p_i(t) + \epsilon u_i(t) \end{array} \right. , \tag{1}$$

where ϵ is the iteration step size (or sampling time). $q_i, p_i, u_i \in \mathbb{R}^2$ are, respectively, the position, the velocity, and the input of the mobile agent. We assume that the measurement $y(q_i(t), t)$ of sensor i includes the scalar value of the field $z(q_i(t), t)$ and sensor noise $w(t)$, at its position $q_i(t)$ and the measurement time t,

$$y(q_i(t), t) := z(q_i(t), t) + w(t). \tag{2}$$

B. A Graph

The group behavior of mobile sensing agents and their complicated interactions with neighbors are best treated by a graph with edges. Let $G(q) := (\mathcal{I}, \mathcal{E}(q))$ be a communication graph such that an edge $(i, j) \in \mathcal{E}(q)$ if and only if agent i can communicate with agent $j \neq i$. We assume that each agent can communicate with its neighboring agents within a limited transmission range given by a radius of r. Therefore, $(i, j) \in \mathcal{E}(q)$ if and only if $\|q_i(t) - q_j(t)\| \leq r$. We define the neighborhood of agent i with a configuration of q by $N(i, q) := \{j : (i, j) \in \mathcal{E}(q), i \in \mathcal{I}\}$. The adjacency matrix $A := [a_{ij}]$ of an undirected graph G is a symmetric matrix such that $a_{ij} = k_3 > 0$ if vertex i and vertex j are neighbors and $a_{ij} = 0$ otherwise, where k_3 is a positive scalar. The scalar graph Laplacian $L = [l_{ij}] \in \mathbb{R}^{N_s \times N_s}$ is a matrix defined as $L := D(A) - A$, where $D(A)$ is a diagonal matrix whose diagonal entries are row sums of A, i.e., $D(A) := \text{diag}(\sum_{j=1}^{N_s} a_{ij})$. The 2-dimensional graph Laplacian is defined as $\hat{L} := L \otimes I_2$, where $\otimes$ is the Kronecker product. A quadratic disagreement function [6] can be obtained via the Laplacian $\hat{L}$:

$$\Psi_G(p) := p^T \hat{L} p = \frac{1}{2} \sum_{(i,j) \in \mathcal{E}(q)} a_{ij} ||p_j - p_i||^2, \tag{3}$$

where $p := \text{col}(p_1, p_2, \cdots, p_{N_s}) \in \mathbb{R}^{2N_s}$.

C. A Swarming Behavior

In order for agents to sample measurements of a scalar field at spatially distributed locations simultaneously, a group of mobile agents will be coordinated by a flocking algorithm ([6], [5], [10]). We use attractive and repulsive smooth potentials similar to those used in [5], [6], [10] to generate a swarming behavior. To enforce a group of agents to satisfy a set of algebraic constraints $\|q_i - q_j\| = d$ for all $j \in \mathcal{N}(i, q)$, we introduce a collective potential

$$U_1(q) := \sum_i \sum_{j \neq i} U_{ij}(\|q_i - q_j\|^2) = \sum_i \sum_{j \neq i} U_{ij}(r_{ij}), \tag{4}$$

where $r_{ij} := \|q_i - q_j\|^2$. U_{ij} in (4) is defined by

$$U_{ij}(r_{ij}) := \frac{1}{2} \left(\log(\alpha + r_{ij}) + \frac{\alpha + d^2}{\alpha + r_{ij}} \right), \text{ if } r_{ij} < d_0^2, \tag{5}$$

otherwise (i.e., $r_{ij} \geq d_0^2$), it is defined according to the gradient of the potential, which will be described shortly. Here $\alpha, d \in \mathbb{R}_+$ and $d < d_0$. The gradient of the potential w.r.t. q_i for agent i is given by

$$\nabla U_1(q_i) := \frac{\partial U_1(q)}{\partial \tilde{q}_i}\Big|_{\tilde{q}_i = q_i} = \sum_{j \neq i} \frac{\partial U_{ij}(r)}{\partial r}\Big|_{r = r_{ij}} (q_i - q_j)$$
$$= \left\{ \begin{array}{ll} \sum_{j \neq i} \frac{(r_{ij} - d^2)(q_i - q_j)}{(\alpha + r_{ij})^2} & \text{if } r_{ij} < d_0^2 \\ \sum_{j \neq i} \rho\left(\frac{\sqrt{r_{ij}} - d_0}{|d_1 - d_0|}\right) \frac{\|d_0^2 - d^2\|}{(\alpha + d_0^2)^2} (q_i - q_j) & \text{otherwise,} \end{array} \right. \tag{6}$$

where $\rho : \mathbb{R}_+ \to [0, 1]$ is the bump function

$$\rho(z) := \left\{ \begin{array}{ll} 1, & z \in [0, h); \\ \frac{1}{2}\left[1 + \cos\left(\pi \frac{(z-h)}{(1-h)}\right)\right], & z \in [h, 1]; \\ 0, & \text{otherwise,} \end{array} \right.$$

that smoothly varies from 1 to 0 as the scalar input increases. In equations (4), (5), and (6), α was introduced to prevent the reaction force from diverging at $r_{ij} = \|q_i - q_j\|^2 = 0$. This potential yields a reaction force that is attracting when the agents are too far and repelling when a pair of two agents are too close. It has an equilibrium point at a distance of d. We also introduce a potential U_2 to model the environment. U_2 enforces each agent to stay inside the closed and connected surveillance region $\mathcal{R}$ and prevents collisions with obstacles in $\mathcal{R}$. We construct U_2 such that it is radially unbounded in q, i.e.,

$$U_2(q) \to \infty \text{ as } \|q\| \to \infty. \tag{7}$$

Define the total artificial potential by

$$U(q) := k_1 U_1(q) + k_2 U_2(q), \tag{8}$$

where $k_1 > 0$ and $k_2 > 0$ are weighting factors.

III. Learning Agents for Gaussian Processes

To model the noisy measurement of the dynamical random field $z(s,t), s \in \mathcal{R}$, consider the space-time Kalman filter model or the spatio-temporal Gaussian process [22], [23], [24], [9]:

$$\begin{aligned} z(s,t) &= \mu(s,t) + \varsigma(s,t) \\ &\sim \mathcal{GP}(\mu(s,t), \mathcal{K}(\{s,t\},\{s^*,t^*\})), \end{aligned} \tag{9}$$

where $s, s^* \in \mathcal{R}, t, t^* \in \mathbb{Z}_+$. The mean field $\mu(\cdot,t)$ is a linear function of the temporal state $\theta(t)$:

$$\mu(s,t) := \sum_{j=1}^{m} \phi_j(s) x_j(t) = \phi^T(s)\theta(t), \tag{10}$$

where $\phi^T(s)$ and $\theta(t)$ are respectively by

$$\begin{aligned} \phi^T(\nu) &:= \begin{pmatrix} \phi_1(s) & \phi_2(s) & \cdots & \phi_m(s) \end{pmatrix} \in \mathbb{R}^{1\times m} \\ \theta(t) &:= \begin{pmatrix} \theta_1(t) & \theta_2(t) & \cdots & \theta_m(t) \end{pmatrix}^T \in \mathbb{R}^{m\times 1}. \end{aligned}$$

Gaussian kernels $\phi_j(s)$ are given by

$$\phi_j(s) := \frac{1}{Z_\mu} \exp\left(\frac{-\|s-\nu_j^c\|^2}{\sigma_\mu^2} \right), \tag{11}$$

where σ_μ is the width of the Gaussian basis and Z_μ is a normalizing constant. ν_j^c for $j \in \{1,\cdots,m\}$ are uniformly distributed in the surveillance region $\mathcal{R}$. We also specify *a prior* over $\theta(0)$ by $\theta(0) \sim \mathcal{N}(\theta_0, \Sigma_\theta(0))$. The spatial correlation can be further prescribed by the zero-mean Gaussian process $\varsigma(s,t) \sim \mathcal{GP}(0, \mathcal{K}(\{s_i,t_i\},\{s_j,t_j\}))$ with a covariance matrix $\mathcal{K}(\{s_i,t_i\},\{s_j,t_j\}) := \kappa(s_i,s_j)\delta_{(t_i,t_j)}$, where $\delta_{(\cdot,\cdot)}$ is the Kronecker delta and

$$\kappa(s_i,s_j) := \frac{1}{Z_\kappa} \exp\left(\frac{-\|s_i-s_j\|^2}{\sigma_\kappa^2} \right). \tag{12}$$

The dynamical part of the spatio-temporal Gaussian process and the noisy observation are given by

$$\begin{aligned} \theta(t+1) &= F(t)\theta(t) + G(t)u(t) \in \mathbb{R}^m, \\ y(s,t) &= \phi^T(s)\theta(t) + \varsigma(s,t) + w(t) \in \mathbb{R}, \end{aligned} \tag{13}$$

where $u(t) \sim \mathcal{N}(0, Q(t))$ and $w(t) \sim \mathcal{N}(0, \sigma_w^2)$ are respectively the usual zero-mean system noise and the uncorrelated measurement noise.

Suppose that at time t, agent i can collect observations $Y(t) := [y(q_1(t),t), \cdots, y(q_n(t),t)]^T$ taken at the n sites $\{q_1(t),\cdots,q_n(t)\}$ by itself and its $n-1$ number of neighbors, then we have:

$$Y(t) = \Phi(t)\theta(t) + v(t) \in \mathbb{R}^n, \tag{14}$$

where $\Phi(t) := [\phi(q_1(t)),\cdots,\phi(q_n(t))]^T$ and $v(t) \sim \mathcal{N}(0, \Sigma_v(t))$. Here we assume that agent i can compute the covariance function by

$$\Sigma_v(t) := \left([\kappa(q_i(t) - q_j(t))] + \text{diag}(\sigma_{w1}^2, \cdots, \sigma_{wn}^2) \right) \in \mathbb{R}^{n\times n}, \tag{15}$$

where σ_{wj}^2 are due to sensor noise and communication noise between agent i and neighboring agents.

Let $\hat{\theta}(t|t-1)$ and $\hat{\theta}(t|t)$ be the estimates of $\theta(t)$ based on the observations obtained to times $t-1$ and t. The estimation error is $\tilde{\theta}(t|t-1) := \theta(t) - \hat{\theta}(t|t-1)$. Let $P(t|t-1)$ and $P(t|t)$ be the covariance matrices of $\hat{\theta}(t|t-1)$ and $\hat{\theta}(t|t)$ respectively. For (13) and (14), the standard Kalman filter [26] provides the measurement updates:

$$\begin{aligned} K_f(t) &= P(t|t-1)\Phi^T(t)\left[\Sigma_v(t) + \Phi(t)P(t|t-1)\Phi^T(t)\right]^{-1}, \\ \hat{\theta}(t|t) &= \hat{\theta}(t|t-1) + K_f(t)\left[Y(t) - \Phi(t)\hat{\theta}(t|t-1)\right], \\ P(t|t) &= [I - K_f(t)\Phi(t)]P(t|t-1), \end{aligned} \tag{16}$$

and the predictions:

$$\begin{aligned} \hat{\theta}(t+1|t) &= F(t)\hat{\theta}(t|t-1) \\ &\quad + F(t)K_f(t)[Y(t) - \Phi(t)\hat{\theta}(t|t-1)], \\ P(t+1|t) &= F(t)P(t|t-1)F^T(t) + G(t)Q(t)G^T(t) \\ &\quad - F(t)K_f(t)[\Sigma_v(t) + \Phi(t)P(t|t-1)\Phi^T(t)]^{-1} \\ &\quad K_f^T(t)F^T(t). \end{aligned} \tag{17}$$

For a fixed $\theta(t)$, we have the following :

$$\begin{aligned} \Sigma_z(t) &:= \text{Cov}(z(s,t), z(s,t)|\theta(t)) = \kappa(s,s), \\ \Sigma_Y(t) &:= \text{Cov}(Y(t), Y(t)|\theta(t)) = \Sigma_v(t), \\ \Sigma_{Yz}(t) &= \Sigma_{zY}^T(t) := \text{Cov}(Y(t), z(s,t)|\theta(t)) = \psi(s), \end{aligned} \tag{18}$$

where $\text{Cov}(x,y) := \mathbb{E}(x-\mathbb{E}x)(y-\mathbb{E}y)^T$ and $\psi(s) := [\kappa(s_i,s)] \in \mathbb{R}^n$. From the Kalman filter, we have

$$\theta(t)|Y_{\leq t} := \{Y(t),\cdots,Y(1)\} \sim \mathcal{N}(\hat{\theta}(t|t), P(t|t)).$$

The posterior predictive distribution of $z(s,t)$ conditioned on $Y_{\leq t}$ can be obtained by marginalizing $p(z(s,t)|\theta(t), Y_{\leq t})$ over $p(\theta(t)|Y_{\leq t})$:

$$z(s,t|t) := z(s,t)\,|\,Y_{\leq t} \sim \mathcal{N}\left(\hat{z}(s,t|t), \sigma^2(s,t|t)\right), \tag{19}$$

where $\hat{z}(s,t|t) := \mathbb{E}\{z(s,t|t)\}$ is:

$$\begin{aligned} \hat{z}(s,t|t) &:= \phi^T(s)\hat{\theta}(t|t) + \Sigma_{zY}(t)\Sigma_Y^{-1}(t)\left(Y(t) - \Phi(t)\hat{\theta}(t|t)\right), \\ &= \phi^T(s)\hat{\theta}(t|t) + \psi^T(s)\Sigma_v^{-1}(t)\left(Y(t) - \Phi(t)\hat{\theta}(t|t)\right), \end{aligned}$$

and $\sigma^2(s,t|t)$ is given by

$$\begin{aligned} \sigma^2(s,t|t) &:= \Sigma_z(t) - \Sigma_{zY}(t)\Sigma_Y^{-1}(t)\Sigma_{zY}^T(t) \\ &\quad + (\phi^T(s) - \Sigma_{zY}\Sigma_Y^{-1}\Phi(t))P(t|t)(\phi^T(s) - \Sigma_{zY}\Sigma_Y^{-1}\Phi(t))^T \\ &= \kappa(s,s) - \psi^T(s)\Sigma_v^{-1}(t)\psi(s) \\ &\quad + [\phi^T(s) - \psi^T(s)\Sigma_v^{-1}\Phi(t)]P(t|t)[\phi^T(s) - \psi^T(s)\Sigma_v^{-1}\Phi(t)]^T. \end{aligned}$$

The last term is due to using the MMSE estimate $\hat{\theta}(t)$ as compared to applying a simple kriging or a prediction of the Gaussian process for a known $\theta(t)$. This formulation is a popular way to embed a finite number of deterministic kernels to represent a mean trend [27], [21], [20], [9]. This algorithm combines parametric and nonparametric estimations, which is robust w.r.t possible mismatches in the selected radial basis functions that parameterize the mean trend. In the following section, navigation strategies based on the spatial prediction and the estimated uncertainty in (19) are presented.

TABLE I

A LIST OF COMMON GOALS AND THEIR RELATED SMOOTH PERFORMANCE CRITERIONS TO BE MAXIMIZED.

Goals	Smooth objective function
Avoidance (Negative prediction)	$\beta_0 = -\hat{z}(s,t\|t)$
Tracing (Prediction)	$\beta_1 = \hat{z}(s,t\|t)$
Exploration (Variance)	$\beta_2 = \sigma^2(s,t\|t)$
Exploration (Entropy)	$\beta_3 = H(z(s,t\|t))$

IV. NAVIGATION STRATEGIES

Depending on the tasks for the sensing agents, the collective mobility of sensing agents is designed to maximize a specified performance criterion. In this section, we introduce a set of different objective functions for navigation, their parameterization, and a way to maximize such an objective function via a reference trajectory state.

A. Biologically Inspired Navigation

We propose a set of useful, biologically inspired navigation modes (tracing, avoidance, and exploration) for agents: (i) for tracing, agents move to the maximum location of the estimated field:

$$q_i(t) = \arg\max_{s\in\mathcal{R}} \hat{z}(s,t|t), \tag{20}$$

(ii) for avoidance, we also have:

$$q_i(t) = \arg\max_{s\in\mathcal{R}} -\hat{z}(s,t|t), \tag{21}$$

(iii) for exploration, the agents can move to a location associated to the maximum variance

$$q_i(t) = \arg\max_{s\in\mathcal{R}} \sigma^2(s,t|t), \tag{22}$$

or differential entropy of the Gaussian process:

$$q_i(t) = \arg\max_{s\in\mathcal{R}} H(z(s,t|t)) = \frac{1}{2}\ln(2\pi e\sigma^2(s,t|t)). \tag{23}$$

By using (22) and (23), we expect the variance and the entropy of the Gaussian process in the surveillance region $\mathcal{R}$ to decrease. Notice that prediction in (19) and gradients in (20), (21), (22) and (23) are smooth functions of a location s, which ensures the existence of extreme values over a compact set.

In Table I, a list of common goals and their related performance criterion functions for coordinating sensing agents are summarized.

B. The Parameterization of a Global Goal

The optimal balance between exploitation and exploration is commonly observed in biological searchers [28], [29] as well as in learning theory. The balance can be achieved by switching amongst or a convex combination of such goals with different nature. For instance, a global goal can be parameterized in the following way:

$$J_i(\Lambda(t);s,t) := \frac{\sum_{k=1}^{3}\lambda_{ik}(t)\beta_{ik}(s,t)}{\sum_{k=1}^{3}\lambda_{ik}(t)}, \text{ for all } i\in\mathcal{I}, \tag{24}$$

where $\beta_{i1}(s,t) := \hat{z}_i(s,t|t)$, $\beta_{i2}(s,t) := \sigma_i^2(s,t|t)$ and $\beta_{i3}(s,t) := H(z_i(s,t|t))$ are specifically chosen for the later simulation study. A global goal is a function of a navigation strategy $\Lambda(t) := [\lambda_{ik}(t)] \in \mathbb{R}_+^{|\mathcal{I}|\times 3}$ that sets the individual weights on the all possible performance criterions (typical ones are shown in Table I).

C. Reference Trajectory State

It is important to notice that agents have access to the maximum location of posterior predictive entities (such as prediction, variance, and entropy) over $s\in\mathcal{R}$, using the Gaussian process. Hence, instead of using local gradient ascent strategy [10], [11] that converges to a local maximum, we introduce a reference trajectory state, which will guide agents to maximize the objective function in (24) directly for the global maximum. For each of agents at each time instant, the maximum location of the corresponding posterior predictive entity will be evaluated and this value will be fed into a lowpass filter to generate a reference trajectory state. Therefore, agents can directly locate the global maximum of the posterior predictive entity.

$$\eta_i(t) = \arg\max_{s\in\mathcal{R}} J_i(\Lambda(t);s,t). \tag{25}$$

The reference trajectory $r_i(t)$ is generated by a lowpass filter with a time constant T, i.e., $\frac{1}{Ts+1}$ fed by $\eta_i(t)$. The discretized version is given by:

$$r_i(t) = r_i(t-1) + \epsilon\left[-\frac{1}{T}r_i(t-1) + \frac{1}{T}\eta_i(t)\right]. \tag{26}$$

Agent i then can perform a gradient ascent strategy with respect to the following objective function:

$$\begin{aligned} \hat{J}_i(\Lambda(t);q_i,t) &:= -\frac{1}{2}\|q_i(t)-r_i(t)\|^2, \\ \nabla\hat{J}_i(\Lambda(t);q_i,t) &= -(q_i(t)-r_i(t)). \end{aligned} \tag{27}$$

V. COOPERATIVELY LEARNING CONTROL

Each of mobile agents receives measurements from neighbors, then updates its estimation of the Gaussian process in $\mathcal{R}$ via the recursive algorithm presented in (16), (17) and an update in (19). Subsequently, based on the update of a gradient of a performance criterion in Table I, the control for its coordination will be decided. For agent i, the update becomes:

$$\begin{aligned} \Sigma_{Yi}^*(t) &= \Sigma_i(t) + \Phi_i(t)P_i(t|t-1)\Phi_i(t)^T \\ K_{fi}(t) &= P_i(t|t-1)\Phi_i(t)^T{\Sigma_{Yi}^*}^{-1}(t), \\ \hat{\theta}_i(t|t) &= \hat{\theta}_i(t|t-1) + K_{fi}(t)\left[Y_i(t) - \Phi_i(t)\hat{\theta}_i(t|t-1)\right], \\ P_i(t|t) &= [I - K_{fi}(t)\Phi_i(t)]P_i(t|t-1), \\ \hat{\theta}_i(t+1|t) &= F_i(t)\hat{\theta}_i(t|t-1) \\ &\quad + F_i(t)K_{fi}(t)[Y_i(t) - \Phi_i(t)\hat{\theta}_i(t|t-1)], \\ P_i(t+1|t) &= F_i(t)P_i(t|t-1)F_i^T(t) + G_i(t)Q_i(t)G_i^T(t) \\ &\quad - F_i(t)K_{fi}(t){\Sigma_{Yi}^*}^{-1}(t)K_{fi}^T(t)F_i^T(t), \end{aligned} \tag{28}$$

where $Y_i(t)$ is the collection of collaboratively measured data at iteration time t. Based on the gradient of the performance function $\nabla\hat{J}_i(\cdot;\cdot,\cdot)$ in (27) updated by (28), a distributed control for agent i is decided by

$$\begin{aligned} u_i(t) &:= -\nabla U(q_i(t)) - k_{di}p_i(t) + k_4\nabla\hat{J}_i(\Lambda(t);q_i(t),t) \\ &+ \sum_{j\in N(i,q(t))} a_{ij}(q(t))(p_j(t)-p_i(t)), \end{aligned} \tag{29}$$

here $k_4 \in \mathbb{R}_+$ is a gain for the global objective function and $k_{di} \in \mathbb{R}_+$ is a gain for the velocity feedback. The first term in (29) is the gradient of (8) which attracts agents while avoiding collisions among them. Also it restricts the movements of agents inside $\mathcal{R}$. The last term in (29) is an effort for agent i to match its velocity with those of neighbors. This term is also called a "velocity consensus" and serves as a damping force among agents. Incorporating the closed-loop discrete time model in (1) along with the proposed control in (29) and the reference trajectory state (27) gives

$$\begin{aligned} r_i(t+1) &= r_i(t) + \epsilon \left[-\frac{1}{T} r_i(t) + \frac{1}{T} \eta_i(t+1) \right], \\ q_i(t+1) &= q_i(t) + \epsilon [p_i(t)] \\ p_i(t+1) &= p_i(t) + \epsilon \Big[- \nabla U(q_i(t)) - k_{di} p_i(t) \\ &\quad - \nabla \Psi_G(p_i(t)) + k_4 (r_i(t) - q_i(t)) \Big], \end{aligned} \tag{30}$$

where the iteration step or the sampling time ϵ is sufficiently small so that the trajectories of states can be approximated by the associated ODE (see more details in [30]). A velocity saturation can be imposed in (30), which can be thought of a projection onto a space with a bounded velocity [30].

VI. Simulation Results

To demonstrate the proposed learning agents, we applied the control (29) to spatio-temporal Gaussian processes introduced in Section III under various global goals generated by (24).

Hereafter plots contain updated posterior predictive values of agent 1 along with trajectories of all agents. Mode 1, Mode 2, and Mode 3 denote the tracing in (20), exploration via variance in (22), exploration via entropy in (23) respectively.

A. Tracing

Nine swarming learning agents performed a tracing strategy as shown in Fig. 1. Here $\eta(t)$ in (25) is generated by the location associated to the maximum of the predicted value $\hat{z}(s, t|t)$ over $\mathcal{R}$. The maximum is obtained by evaluating $\hat{z}(s, t|t)$ at a fine grid over $\mathcal{R}$. The reference trajectory generated by (26) is plotted by the smaller stars and a dotted black line. The white star represents the latest reference point for agent 1 to track. Learning agents with the tracing mode alone converge to a configuration for a local maximum. Agents with the tracing strategy alone can not find the maximum point if they are far from it.

B. Exploration

Learning agents performed group exploration by tracking the maximal variance location as shown in Fig. 2. The plots show $\sigma^2(s, t|t)$ of agent 1. Stars represents the reference state trajectory for agent 1 to follow. As depicted in Fig. 2, the reference state can travel between alternating maximum points, since the maximum point decreases as agents collect samples near the maximum variance location. The worst-case error variance $\max_{s \in \mathcal{R}} \sigma^2(s, t|t)$ and the RMS value of a error field between $\mu(s, t|t)$ and $\hat{\mu}(s, t|t)$ quickly converges to a small number. In the same way, the entropy driven exploration strategy is shown with $H(z(s, t|t))$ in Fig. 3.

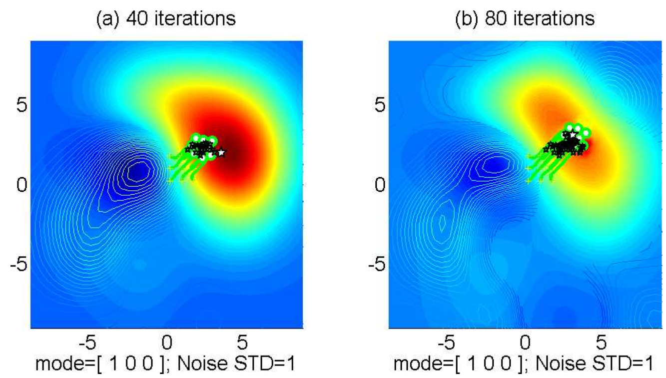

Fig. 1. Tracing: The plot of $\hat{z}(s, t|t)$. The error field between $\hat{z}(s, t|t)$ and $z(s, t|t)$ is plotted by colored contours.

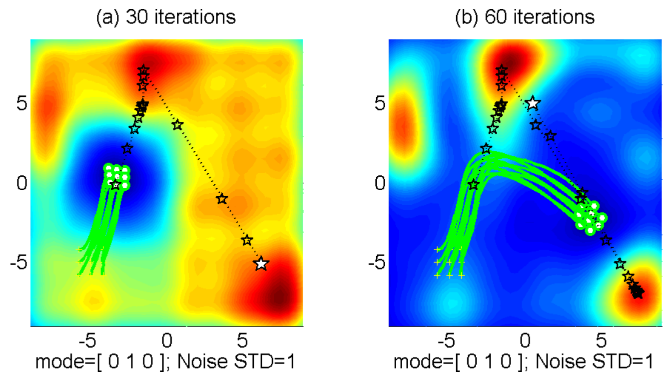

Fig. 2. Variance driven exploration: The plot of $\sigma^2(s, t|t)$ updated by agent 1 (blue-lowest, red-highest).

C. A Switching Protocol

We propose a switching protocol between exploration (Mode 2 or Mode 3) and exploitation (Mode 1). Each agent starts with an exploration strategy, when an agent obtains a maximum prediction error variance within a specified tolerance, it switches to the tracing strategy in order to locate the maximum of the predicted field within the variance error tolerance. Fig. 4 shows that the convergence rates of the trace of $P(t|t)$, the RMS value of the error field, and the maximum value $\eta(t)$ under the switching protocol. Agent 1 switches from Mode 1 to Mode 2 when the maximum variance error becomes less than 10 as in Fig. 4. As shown in Fig. 5,

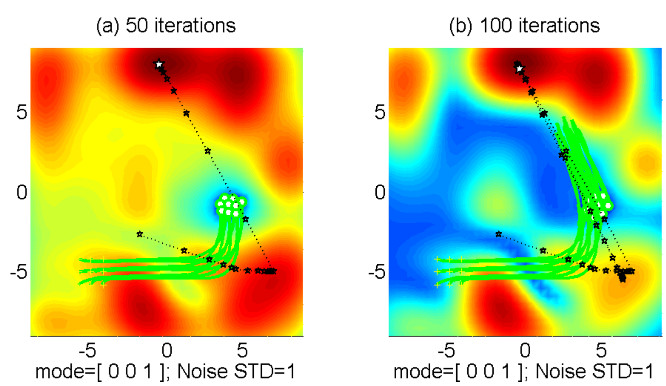

Fig. 3. Entropy driven exploration: The plot of entropy $H(z(s, t|t))$.

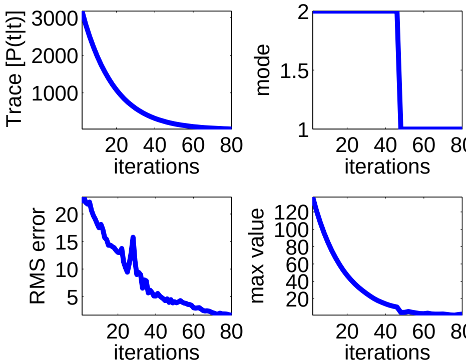

Fig. 4. A switching protocol: The convergence rates of the trace of $P(t|t)$, the RMS value of the error field, and maximum value $\eta(t)$ under a switching protocol. Agent 1 switched from Mode 1 to Mode 2 around iteration time 47.

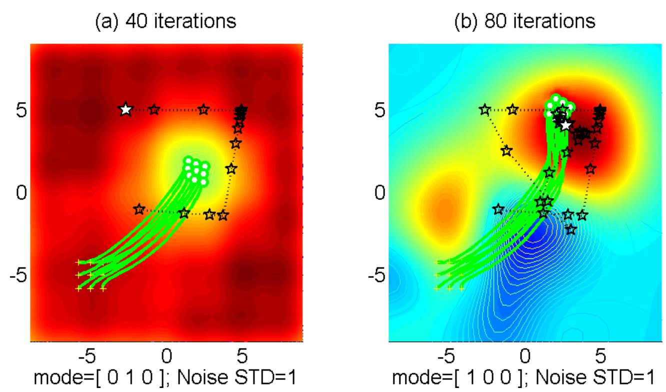

Fig. 5. Plots for a switching protocol. Left: the plot of $\sigma^2(s,t|t)$ under variance driven exploration. Right: the plot of $\hat{z}(s,t|t)$ under the tracing mode.

learning agents under this switching protocol successfully locate the global maximum point in $\mathcal{R}$.

VII. Conclusions

In this paper, we presented a novel class of self-organizing multi-agent systems that form a swarm and learn a spatio-temporal process through noisy measurements from neighbors for various global goals. The physical spatio-temporal process of interest is modeled by a spatio-temporal Gaussian process. Each agent maintains its own posterior predictive statistics of the Gaussian process based on online measurements from neighbors. The proposed learning agents perform a prescribed task by tracking the maximum location of an objective function. A switching protocol was proposed for achieving the global maximum of a spatio-temporal Gaussian process. The usefulness of the proposed multi-agent system with respect to various global goals was demonstrated by several numerical examples.

References

[1] J. Cortes and F. Bullo, "Coordination and geometric optimization via distributed dynamical systems," *SIAM Journal on Control and Optimization*, vol. 44, no. 5, pp. 1543–1574, 2005.

[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] J. Cortes, S. Martinez, and F. Bullo, "Spatially-distributed coverage optimization and control with limited-range interactions," *ESAIM. Control, Optimisation and Calculus of Variations*, vol. 11, pp. 691–719, 2005.

[4] A. Jadbabie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomuous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, June 2003.

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

[6] Olfati-Saber, "Flocking for multi-agent dynamic systems: Algorithm and theory," *IEEE Transactions on Automatic Control*, vol. 51, no. 3, pp. 401–420, March 2006.

[7] S. Martinez, "Distributed interpolation schemes for field estimation by mobile sensor networks," *submitted to IEEE Transactions on Control Systems Technology*, July 2007.

[8] R. Graham and J. Cortes, "Asymptotic optimality of multicenter voronoi configurations for random field estimation," *submitted to IEEE Transactions on Automatic Control*, August 2007.

[9] J. Cortes, "Distributed kriged kalman filter for epatial estimation," *submitted to IEEE Transactions on Automatic Control*, May 2007.

[10] J. Choi, S. Oh, and R. Horowitz, "Cooperatively learning mobile agents for gradient climbing," in *Proceedings of the 46th IEEE Conference on Decision and Control*, 2007.

[11] J. Choi, J. Lee, and S. Oh, "Biologically-inspired navigation strategies for swarm intelligence using spatial gaussian processes," *to appear in the International Federation of Automatic Control (IFAC) World Congress*, 2008.

[12] J. Adler, "Chemotaxis in bacteria," *Journal of Supramolecular Structure*, vol. 4, no. 3, pp. 305–317, 1966.

[13] A. Dhariwal, G. S. Sukhatme, and A. A. G. Requicha, "Bacterium-inspired robots for environmental monitoring," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2004.

[14] P. Őgren, E. Fiorelli, and N. E. Leonard, "Cooperative control of mobile sensor networks: Adaptive gradient climing in a distributed envrironment," *IEEE Transaction on Automatic Control*, vol. 49, no. 8, p. 1292, August 2004.

[15] N. Leonardo and A. Robinson, "Adaptive sampling and forecasting plan AOSN-II MB'03 project," Available: http://www.princeton.edu/ dcsl/aosn/documents/ASFP.pdf, Tech. Rep., 2003.

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

[17] N. A. C. Cressie, *Statistics for Spatial Data*. A Wiley-Interscience Publication, John Wiley and Sons, Inc., 1991.

[18] M. N. Gibbs and D. J. C. MacKay, "Efficient implementation of Gaussian processes for interpolation."

[19] D. J. C. MacKay, "Introduction to Gaussian processes," in *Neural Networks and Machine Learning*, ser. NATO ASI Series, C. M. Bishop, Ed. Kluwer, 1998, pp. 133–166.

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

[21] N. Cressie and C. K. Wikle, *Encyclopedia of Environments*. John Wiley and Sons, Ltd, Chichester, 2002, vol. 4, ch. Space-time Kalman filter, pp. 2045–2049.

[22] K. V. Mardia, C. Goodall, E. J. Redfern, and F. J. Alonso, "The kriged kalman filter," *Test*, vol. 7, no. 2, pp. 217–285, 1998.

[23] N. Cressie and C. C. K. Wikle, "Strategies for dynamic space-time statistical modeling: Discussion of paper by k. v. mardia et al." *Test*, vol. 7, no. 2, pp. 257–285, 1998.

[24] C. K. Wikle and N. Cressie, "A dimensino-reduced approach to space-time kalman filtering," *Biometrika*, vol. 86, no. 4, pp. 815–829, 1999.

[25] C. Guestrin, A. Krause, and A. P. Singh, "Near-optimal sensor placement in gaussian processes," in *Proceeding of the 22nd International Conference on Machine Learning*, 2005.

[26] T. Kailath, A. H. sayed, and B. Hassibi, *Linear Estimation*, A. Dworkin, Ed. Upper Saddle River, New Jersey: Prentice-Hall, Inc., 2000.

[27] B. J. N. Blight and L. Ott, "A bayesian approach to model inadequacy for polynomial regression," *Biometrika*, vol. 19, no. 3, pp. 79–88, 1975.

[28] D. Grünbaum, "Schooling as a strategy for taxis in a noisy environment," *Evolutionary Ecology*, vol. 12, pp. 503–522, 1998.

[29] M. Vergassola, E. Villermaux, and B. I. Shraiman, "'infotaxis' as a strategy for searching without gradients," *Naure*, vol. 445, no. 25, pp. 406–409, 2007.

[30] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*. Springer, 1997.