Spatial statistics and distributed estimation by robotic sensor networks

Rishi Graham Jorge Cortés

***Abstract*— Networks of environmental sensors are playing an increasingly important role in scientific studies of the ocean, rivers, and the atmosphere. Robotic sensors can improve the efficiency of data collection, adapt to changes in the environment, and provide a robust response to individual failures. Their operation must be driven by statistically-aware algorithms that make the most of the network capabilities for data collection and fusion. At the same time, such algorithms need to be distributed and scalable to make robotic networks capable of operating in an autonomous and robust fashion. The combination of these two objectives, complex statistical modeling and distributed coordination, presents grand technical challenges: traditional statistical modeling and inference assume full availability of all measurements and central computation. While the availability of data at a central location is certainly a desirable property, the paradigm for distributed motion coordination builds on partial, fragmented information. This work surveys recent progress at bridging the gap between sophisticated statistical modeling and distributed motion coordination.**

I. Introduction

Scientific studies of environmental phenomena often involve a data collection stage. Samples are taken of a spatially distributed process of interest, such as a temperature field or chemical concentrations. Combining these samples with a model, the scientist may make predictions about the process at unmeasured locations, or inference about the quality and accuracy of the model. This work reviews some results which lay the groundwork for cooperative control of mobile sensing devices based on statistically motivated objectives, when the underlying process is modeled as a random field.

Physical process models may be roughly divided into two categories: deterministic and stochastic. Deterministic models are often coupled with a stochastic measurement error term, e.g., [1], [2], [3], [4], but require that model parameters and initial conditions be known to a high degree of accuracy [5]. When this cannot be guaranteed, or when the parameter space of the deterministic model has high dimension, it may be desirable to treat the process itself as in some degree unknown, using a stochastic process model. A classic example is a fair coin toss. It is clear that under extremely strict monitoring of the initial conditions and model parameters, the interested physicist could exactly model the entire trajectory of the coin, culminating in its final resting position. The model which is usually used, however, is to assign a simple probability to each outcome. In this context, it is easy to allow for the possibility that the coin is not "fair". We toss the coin a few times, collect the data, and the results give us information about the model (or about future coin tosses). For this reason, stochastic modeling is sometimes called *data driven*, as opposed to the *model driven* deterministic modeling. We focus on data driven models, and particularly their explicit representations of uncertainty.

Our treatment here deals with two important tasks faced by a network of autonomous sensors: choosing the locations to take samples and incorporating those data into the global model. There is a rich literature on the use of model uncertainty to drive the placement of sensing devices, e.g. [6], [7], [8]. Most of this research has focused on choosing from discrete sets of hypothetical sampling locations, and until recently all of it has made use of centralized computational techniques. Likewise, the work on data fusion [9], [10] mostly concentrates on centralized methods where access to all of the data is allowed. A related classical problem concerns the fusion of data from multiple instruments taking noisy samples from the same deterministic process (e.g., [11]). In cooperative control, various works consider mobile sensor networks performing spatial estimation tasks. [12] introduces performance metrics for oceanographic surveys by autonomous underwater vehicles. [13] considers a robotic sensor network with centralized control estimating a static field from samples with both sensing and localization error. In [14], a deterministic model is used, where the random elements come in the form of unknown model parameters, and localization error is included. The work [15] uses a Gaussian process model where all information is globally available via all-to-all communication. [16] considers optimal sampling trajectories from a parameterized set of paths. [17] discusses the tracking of level curves in a noisy scalar field.

Here we present recent work on spatial estimation tasks that require complex statistical modeling combined with distributed computation and control. Our aim is to motivate further research at the intersection of these exciting areas. The paper is organized as follows. In Section II we describe some types of sensor networks and provide a framework for the discussion of distributed solutions. Section III outlines some of the important features of spatial statistical models as pertain to the tasks described above. Section V gives a brief introduction to some recent work in distributed optimal design and distributed sequential design (choosing locations and trajectories for sampling).

II. Network architecture

In the context of environmental sampling, the term "sensor network" may describe anything from a small number of fixed position rainfall monitors in the forest to a complex group of static flotation devices and mobile robots in the ocean. The literature on stochastic spatial modeling has traditionally dealt with sensors whose location is fixed in space. However, the ability to move about the field and

Rishi Graham is with the Department of Applied Mathematics and Statistics, University of California, Santa Cruz, rishig@ams.ucsc.edu

Jorge Cortés is with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, cortes@ucsd.edu

take samples at desired locations has obvious benefits. A network in this context is a group of agents connected by wired or wireless communication paths. For our purposes, we consider networks comprised of two types of agents: static and mobile. The term "mobile agents" describes robots with the ability to move, take samples of the spatial process, and possibly sense their immediate physical environment. Their storage and computational capabilities are assumed to be minimal. By "static agents", we refer to fixed position computational devices which may or may not take samples. Because they are static and do not require energy to move around, they may carry more equipment and thus perform more in the way of computation and storage tasks. Some limited range communication is also assumed for both types of agents. Distributed solutions to global problems are therefore defined on the *communication graph* of the system. In the examples below, we deal with networks of agents in a convex polytope $\mathcal{D} \subset \mathbb{R}^d,\ d \in \mathbb{N}$. We will call the mobile robots $\{R_1, \ldots, R_n\},\ n \in \mathbb{N}$, and denote their locations by $P = (p_1, \ldots, p_n)^T \in \mathcal{D}^n$. Where static nodes are mentioned, we will call them $\{S_1, \ldots, S_m\},\ m \in \mathbb{N}$ at locations $Q = (q_1, \ldots, q_m)^T \in \mathcal{D}^m$. We assume that robots have perfect information about their location.

III. A Bayesian approach to spatial models

Let Z denote a random spatial process taking values on $\mathcal{D}$. Let $\underline{y} = (y_1, \ldots, y_m)^T \in \mathbb{R}^m$ be $m \in \mathbb{N}$ samples taken from Z at corresponding locations $\underline{s} = (s_1, \ldots, s_m)^T \in \mathcal{D}^m$, with $s_i = (s_i, t_i),\ i \in \{1, \ldots, m\}$. Given these data, various models allow for prediction of Z at any point in $\mathcal{D}$, with associated uncertainty. In a Bayesian setting, the prediction takes the form of a distribution, called the posterior predictive [18]. If the field is modeled as a Gaussian process, we may write,

$$Z(s_0) = \mu(s_0) + \nu(s_0),$$

where $\mu : \mathcal{D} \to \mathbb{R}$ denotes the mean and $\nu : \mathcal{D} \to \mathbb{R}$ is a zero mean random field. Here, we assume second-order stationarity of the spatial process. A spatial random process δ on $\mathcal{D} \subset \mathbb{R}^d$ is second-order stationary if it has constant mean, and its covariance is of the form $\mathrm{Cov}(\delta(s_1), \delta(s_2)) = C(s_1, s_2)$, where $C : \mathcal{D} \times \mathcal{D} \to \mathbb{R}_{\geq 0}$ is a positive definite covariance function which only depends on the difference $s_1 - s_2$. This assumption is valid for spatial fields which do not exhibit abrupt changes in characteristics, such as temperature fields over a relatively small region, and is often used as an experimental first step.

Assuming the covariance of ν is known, the mean of the posterior predictive distribution corresponds to the *Best Linear Unbiased Predictor*, and its variance to the mean-squared prediction error. If the mean is known, the posterior predictive distribution of Z at location s_0 given samples $\underline{y}$ is normal with mean and variance, respectively,

$$\hat{z}_{\mathrm{S}}(s_0; \underline{s}) = \mu(s_0) + \boldsymbol{c}^T \boldsymbol{\Sigma_\nu}^{-1} (\underline{y} - \boldsymbol{\mu}), \tag{1a}$$

$$\sigma^2_{\mathrm{S}}(Z(s_0); \underline{s}) = \mathrm{Var}[Z(s_0)] - \boldsymbol{c}^T \boldsymbol{\Sigma_\nu}^{-1} \boldsymbol{c}. \tag{1b}$$

Here $\boldsymbol{\mu}$ is the m-vector whose ith element is $\mu(s_i)$, $\boldsymbol{c}$ is the vector whose ith element is $\mathrm{Cov}[\nu(s), y_i]$, $\boldsymbol{\Sigma_\nu} = \boldsymbol{\Sigma_\nu}(\underline{s}) \in \mathbb{R}^{m \times m}$ is the covariance matrix of the vector $\underline{y}$. If the mean is not known, but can be treated as an unknown expansion on a vector of $p \in \mathbb{N}$ known basis functions, we write

$$\mu(s) = \mathbf{f}(s)^T \beta, \text{ where } \mathbf{f}(s) = (f_1(s), \ldots, f_p(s))^T,$$

and $\mathbf{f}(s)$ is known for all $s \in \mathcal{D}$. The posterior predictive is again normal with mean and variance, respectively,

$$\hat{z}_{\mathrm{U}}(s_0; \underline{s}) = \left(\boldsymbol{c} + \mathbf{F}\mathbf{E}^{-1}\xi_0\right)^T \boldsymbol{\Sigma_\nu}^{-1} \boldsymbol{y}, \tag{2a}$$

$$\sigma^2_{\mathrm{U}}(Z(s_0); \underline{s}) = \mathrm{Var}[Z(s_0)] - \boldsymbol{c}^T \boldsymbol{\Sigma_\nu}^{-1} \boldsymbol{c} + \xi_0^T \mathbf{E}^{-1} \xi_0. \tag{2b}$$

Here $\mathbf{F} \in \mathbb{R}^{m \times p}$ is the matrix whose ith *row* is $\mathbf{f}(s_i)$, $\mathbf{E} = \mathbf{F}^T \boldsymbol{\Sigma_\nu}^{-1} \mathbf{F} \in \mathbb{R}^{p \times p}$, and $\xi_0 = \mathbf{f}(s_0) - \mathbf{F}\boldsymbol{\Sigma_\nu}^{-1}\boldsymbol{c}$. These first two examples of predictive distributions are well known in the literature under various names. *Kriging* [9], [5] is a standard geostatistical technique in which the distribution (1) corresponds to *simple kriging* and the distribution (2) corresponds to *universal kriging*. Here, we do not consider the case when the covariance of the field is not known. In this situation, few analytical results exist, see e.g., [19].

In the discussion so far, we have left out the notion that the field may evolve in time. One simple way to treat a dynamic field is to use the standard spatial methods and treat time as another dimension. This is particularly useful when the goal is to predict the value of the field at unsampled locations over a continuous time domain. If both the samples and the predictions are to be made at discrete intervals, an alternative approach may be more appropriate, as we discuss next.

IV. Distributed estimation

Here, we discuss the problem of incorporating newly collected samples into the field estimation done by a network of mobile agents following [20]. Our objective is to provide individual agents with local representations of the field that are statistically consistent with the sampled data and take into account nontrivial correlation effects among samples. At the same time, we are interested in accomplishing this in an online and distributed fashion. Once in possession of an accurate representation of the field, each agent can use this information for a variety of objectives. Here, we illustrate this idea in a scenario where the network is interested in finding the maxima of a physical process of interest.

When samples are available at a single time instant, the posterior predictive distribution is given by (2). When samples are available at several time instants, one can extrapolate these estimators using the so-called Kriged Kalman filter [21], [22]. Assume the random field is dynamic modeled as a spatio-temporal process of the form

$$Z(s, k) = \mathbf{f}(s)^T \beta(k) + \nu(s, k), \tag{3a}$$

$$\mathbf{f}(s)^T \beta(k) = b(s)^T \beta(k-1) + \eta(s, k), \tag{3b}$$

where $(s, k) \in \mathbb{R}^d \times \mathbb{Z}_{>0}$. Let us describe each one of the elements in these equations. The form of Z is the same as the universal kriging model described above, except that β and ν now evolve with time. The functions $b(s) = (b_1(s), \ldots, b_m(s))^T \in \mathbb{R}^m$ determining the evolution of β are assumed to be known. Both ν and η are stationary spatial fields that exhibit temporal variability but have no temporal dynamics associated with them. Formally, both are zero-mean Gaussian random fields with separable covariance

$$\mathrm{Cov}(\nu(s,k),\nu(s',k')) = C_\nu(s-s')\,\delta(k-k'),$$
$$\mathrm{Cov}(\eta(s,k),\eta(s',k')) = C_\eta(s-s')\,\delta(k-k'),$$

where δ denotes the Dirac delta function. Note that both ν and η are uncorrelated in time. We assume that the functions $C_\nu, C_\eta : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ have finite range $r \in \mathbb{R}_{>0}$.

After some manipulations, we can combine the equations that we obtain from (3b) with samples available at the agent positions $p_1(k), \ldots, p_n(k)$ at time k as

$$\beta(k) = \boldsymbol{H}(k)\beta(k-1) + \boldsymbol{J}(k)\boldsymbol{\eta}(k), \tag{4}$$

where, for convenience, we have introduced the notation $\boldsymbol{H}(k) = \boldsymbol{J}(k)\boldsymbol{B}(k)$, $\boldsymbol{J}(k) = (\mathbf{F}(k)^T\mathbf{F}(k))^{-1}\mathbf{F}(k)^T$, and

$$\boldsymbol{B}(k) = [b(p_1(k)), \ldots, b(p_n(k))]^T \in \mathbb{R}^{n\times m},$$
$$\mathbf{F}(k) = [\mathbf{f}(p_1(k)), \ldots, \mathbf{f}(p_n(k))]^T \in \mathbb{R}^{n\times m},$$
$$\boldsymbol{\eta}(k) = (\eta(p_1(k),k), \ldots, \eta(p_n(k),k))^T \in \mathbb{R}^n.$$

Notice that the matrices $\boldsymbol{H}$ and $\boldsymbol{J}$ driving the evolution of the parameter β change from one time instant to another only if agent positions change. Let $\boldsymbol{\Sigma_\nu}(k) \in \mathbb{R}^{n\times n}$ denote the covariance matrix of samples made at time k, and let $\boldsymbol{\Sigma_\eta}(k) \in \mathbb{R}^{n\times n}$ denote the covariance matrix of $\boldsymbol{\eta}(k)$.

A. Sequential parameter estimation via Kalman filtering

With the model (4), the parameter β can be optimally predicted via a Kalman filter. Here, instead of considering the usual Kalman filter recursion equations, we use the equivalent information filter formulation, see for instance [23].

Assume β is initially distributed according to a multivariate normal distribution $\beta(0) \sim N_p(\beta_0, \Xi)$. Given $t, s \in \mathbb{R}_{\geq 0}$, let $\hat\beta(t|s)$ denote the estimator of β at time t with data collected up to time s, and let $P(t|s)$ denote the associated mean-squared error. The usual Kalman filter equations are written in the variables $(\hat\beta(k|k-1), P(k|k-1))$ and $(\hat\beta(k|k), P(k|k))$. Instead, we define

$$\hat a(t|s) = P(t|s)^{-1}\hat\beta(t|s),$$

and write the information filter equations in the variables $(\hat a(k|k-1), P(k|k-1)^{-1})$ and $(\hat a(k|k), P(k|k)^{-1})$. Initially, $\hat a(0|0) = \Xi^{-1}\beta_0$ and $P(0|0)^{-1} = \Xi^{-1}$.

The information filter equations have two steps.

Prediction: Using (4), the one-step-ahead prediction at time $k \in \mathbb{Z}_{>0}$ with data collected up to time $k-1$ is

$$\hat a(k|k-1) = P(k|k-1)^{-1}\boldsymbol{H}(k)P(k-1|k-1)\hat a(k-1|k-1),$$

with information matrix

$$P(k|k-1)^{-1} = (\boldsymbol{H}(k)P(k-1|k-1)\boldsymbol{H}(k)^T + \boldsymbol{J}(k)\boldsymbol{\Sigma_\eta}(k)\boldsymbol{J}(k)^T)^{-1}.$$

Correction: Under our sensor error measurement model, the optimal prediction at time $k \in \mathbb{Z}_{>0}$ with data collected up to time k can be recursively expressed as

$$\hat a(k|k) = \hat a(k|k-1) + \mathbf{F}(k)^T(\boldsymbol{\Sigma_\nu}(k))^{-1}\underline{y}(k),$$

with information matrix

$$P(k|k)^{-1} = P(k|k-1)^{-1} + \mathbf{F}(k)^T(\boldsymbol{\Sigma_\nu}(k))^{-1}\mathbf{F}(k),$$

where $\underline{y}(k) \in \mathbb{R}^n$ denotes the data collected *at time* k.

B. Sequential simple Kriging

For $k \in \mathbb{Z}_{>0}$, let $\underline{y}^{(k)}$ denote the samples taken at time k. Let $\underline{y}^{(1:k)} = (\underline{y}^{(1)}, \ldots, \underline{y}^{(k)})$ denote all of the data available up to time k. For $s \in \mathbb{R}^d$, let

$$\boldsymbol{c}(s,k)^T = (C_\nu(s-p_1(k)), \ldots, C_\nu(s-p_n(k))),$$
$$\boldsymbol{\nabla c}(s,k)^T = (\mathrm{grad}\, C_\nu(s-p_1(k)), \ldots, \mathrm{grad}\, C_\nu(s-p_n(k))).$$

The covariance structure of the spatial field has some important consequences. On the one hand, the ith components of $\boldsymbol{c}(s,k)$ and $\boldsymbol{\nabla c}(s,k)$ can only be non vanishing if $\|s - p_i(k)\| \leq r$. More importantly, the decorrelation in time of the spatial field and the sensor errors imply that only the observations collected at exactly time k play a role in the construction of the conditional predictive distribution of Z and ∇Z with observations collected up to time k. Accordingly, conditionally on the data collected up to time k and the parameter $\beta(k)$, the posterior predictive distribution is given by (1), with mean

$$\mathbf{f}(s)^T\beta(k) + \boldsymbol{c}(s,k)^T\boldsymbol{\Sigma_\nu}(k)^{-1}(\underline{y}(k) - \mathbf{F}(k)\beta(k)),$$

and variance $K(0) - \boldsymbol{c}(s,k)^T\boldsymbol{\Sigma_\nu}(k)^{-1}\boldsymbol{c}(s,k)$.

C. Distributed Kriged Kalman filter

The Bayesian universal Kriging predictor of the spatial field, which corresponds to the posterior predictive distribution conditional on the data, can be obtained in an analogous way as explained above for the simple kriging case, and hence we do not reproduce it here.

Once the statistical basics are covered, the challenge lies in developing distributed methods that allow individual agents to compute the parameter estimates and posterior predictive distributions in an online fashion. In these computations, there are several matrix-vector multiplications that involve quantities that are spatially distributed across the network. To further complicate things, some of these expressions involve the inverse of sparse correlation matrices, which are not sparse any more. Here, we do not provide a comprehensive account of the distributed methods used, but rather focus on illustrating the main idea in the computation of $\mathbf{F}(k)^T(\boldsymbol{\Sigma_\nu}(k))^{-1}\underline{y}(k)$, necessary to carry out the correction step in the parameter estimation. Note that the quantity we are interested in computing can be expressed as

$$\mathbf{F}(k)^T(\boldsymbol{\Sigma_\nu}(k))^{-1}\underline{y}(k) = \sum_{i=1}^n \mathrm{row}_i(\mathbf{F}(k))z_i(k), \tag{7}$$

where $z(k) = (\boldsymbol{\Sigma_\nu}(k))^{-1}\underline{y}(k)$ solves the linear equation

$$\boldsymbol{\Sigma_\nu}(k)z(k) = \underline{y}(k). \tag{8}$$

In this way, we have decomposed the computation of the quantity $\mathbf{F}(k)^T(\boldsymbol{\Sigma_\nu}(k))^{-1}\underline{y}(k)$ into two parts: an aggregation of n quantities, one per agent, and the solution of a linear equation determined by a sparse matrix.

Each agent i has access to $\mathrm{row}_i(\boldsymbol{\Sigma_\nu}(k))$, $\underline{y}(k)_i$ and to $\mathrm{row}_i(\mathbf{F}(k))$. Knowledge of the first two quantities is all that is needed to execute a distributed Jacobi-overrelaxation (JOR) algorithm [24] to solve (8) that provides agent i with knowledge of the quantity z_i. This can then be combined with $\mathrm{row}_i(\mathbf{F}(k))$ to solve (7) via distributed averaging.

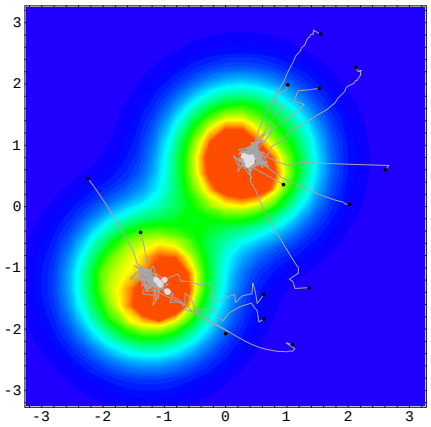


Fig. 1. Distributed gradient ascent cooperative strategy $\dot{p}_i(t) = E(\nabla Z(p_i(t),t) \mid y^{(1:t)}), i \in \{1,\dots,n\}$, implemented by a robotic sensor network of $n = 14$ agents. Individual agents converge asymptotically to the set of expected critical points of the spatial field. The field has mean $\mu(s) = .3 + 1.2\,\mathrm{e}^{-\|s-(.25,.75)\|^2} + 1.1\,\mathrm{e}^{-\|s+(1.25,1.25)\|^2}$ and covariance structure determined by $K(s) = e^{-5\|s\|^2}$ for $\|s\| \leq r = 1.75$. We depict the contour plot of the posterior mean. Initially agents know $\beta \sim N_3(0, I_3)$. The communication radius is $R = 2.75$, the control authority of each agent is bounded by $u_{\max} = .25$, and the noise sensor error variance is $\sigma = .15$. The black disks depict the (randomly generated) initial agent positions and the gray disks depict the agent positions after 38 seconds.

Similar ideas can be invoked to produce a fully distributed implementation of the Kriged Kalman filter, see [20] for details. Remarkably enough, this procedure also works for computing the posterior predictive distributions of the gradient random field associated with Z. Mobile agents, equipped with this information, can then perform a variety of motion coordination tasks with direct relevance to the random field. Fig. 1 illustrates one example of such an application, in which a network of agents seek optima of a random field.

V. Uncertainty-based sampling strategies

Whether the goal of the experiment is field prediction or model inference, a Bayesian model such as those in Section III provides a full accounting of estimation uncertainty. This uncertainty depends on the locations at which the samples are taken. *Optimal design* [7] is the process of choosing locations to sample in order to minimize the resulting uncertainty. This can have various meanings, depending on the goals of the experiment. One might be interested in minimizing the maximum or average predictive variance over the region, or in maximizing the generalized variance or predictive entropy over new sample locations. In general, the optimal sampling problem is a difficult one. In this section, we present two regimes and study the location of their optimizers.

A. Optimal static deployment under near independence

Here we follow [25] to consider the problem of where to place the agents of a mobile network in the case that a single sample is to be taken by each. In [26], an assumption of *near independence* between distinct sampling locations was suggested as a first step in gathering data in a relatively large space. The authors consider the maximum variance over the predictive region of the model (1). Out of a discrete set of sampling configurations, they show that the one which minimizes the maximum distance to the nearest agent from any point in space is asymptotically optimal in the limit of near independence. Without this assumption of near independence, even the task of choosing from discrete locations is NP-hard.

In [25], we consider two performance metrics for optimal placement of sensor networks over a continuous space based on simple kriging. We consider the maximum posterior predictive variance and a novel form of D-optimality over a bounded region,

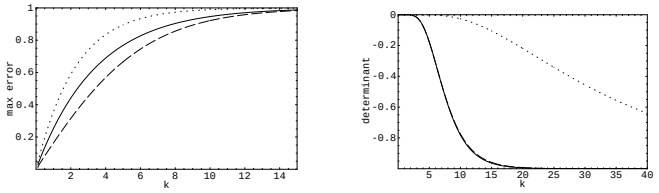
$$\mathcal{M}(P) = \max_{s \in \mathcal{D}} \sigma_{\mathrm{S}}^2(Z(s); P) \tag{9a}$$

$$\mathcal{E}(P) = -\left|\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(p_1,\dots,p_n,\gamma(p_1),\dots,\gamma(p_n))\right|, \tag{9b}$$

where $\gamma(p_i)$ denotes the reflection of location p_i over the nearest boundary of $\mathcal{D}$. We study the critical points of these criteria asymptotically, in the limit of near independence. In general, these objective functions pose nonconvex and high-dimensional optimization problems. In addition, the first criterion is nonsmooth. Our results are relevant to the extent that they guarantee that, for scenarios with small enough correlation between distinct points, circumcenter and incenter Voronoi configurations are optimal for appropriate measures of uncertainty. See [27] for details on these configurations. For the results summarized below, we consider $\mathcal{M}^{(k)}$, respectively $\mathcal{E}^{(k)}$, to denote the function $\mathcal{M}$, respectively $\mathcal{E}$, with the correlation raised to the kth power. As k increases, the correlation between distinct locations in $\mathcal{D}$ decreases in strength, but retains some aspects of the shape of the correlation function (e.g., range and smoothness). The *index* of a configuration is the cardinality of the set of minimum pairwise inter-agent distances.

- Let $P_{mcc} \in \mathcal{D}^n$ be a multi-circumcenter Voronoi configuration. Then, as $k \to \infty$, P_{mcc} asymptotically globally optimizes $\mathcal{M}^{(k)}$, that is, $\mathcal{M}^{(k)}(P_{\mathrm{mcc}})$ approaches a global minimum.
- Let $P_{\mathrm{mic}} \in \mathcal{D}^n$ be a multi-incenter Voronoi configuration with lowest index. Then, as $k \to \infty$, P_{mic} asymptotically globally optimizes $\mathcal{E}^{(k)}$, that is, $\mathcal{E}^{(k)}(P_{\mathrm{mic}})$ approaches a global minimum.

The work [27] describes simple, distributed algorithms which may be used to steer a mobile network towards these multi-center Voronoi configurations. Fig. 2 shows the results of some illustrative simulations. In each case, we compare



(a) (b)

Fig. 2. Value of (a) $\mathcal{M}^{(k)}$ for multi-circumcenter configuration, and (b) $\mathcal{E}^{(k)}$ for multi-incenter configuration. The multicenter results are depicted with the (solid) line, and compared against an approximated global minimum (dashed) arrived at by gradient descent for each value of k, and random (dotted) configurations of 5 agents for increasing k. The covariance function is exponential.

the multicenter configurations against a randomly chosen

configuration, as well as against a dynamic approximate local minimum. This approximate local minimum is arrived at by running a gradient descent algorithm for each value of k.

B. Adaptive design by projected gradient descent

When the goal of the experiment is to find the best *trajectories* for the mobile robots to follow in order to optimize sampling of a spatio-temporal random field, the problem becomes even more challenging. In the existing literature, a standard technique for choosing sampling locations is adaptive design. This amounts to a one step ahead, greedy optimization method where sample locations at the next step are chosen based on information known so far. In the works [28], [29], [30], we present a framework whereby a hybrid network of static nodes and mobile agents can sequentially optimize sampling for an approximation of the random field using a distributed version of the projected gradient descent technique. Assume that the mobile robots take samples synchronously at discrete timesteps $k \in \{1, \ldots, \mathcal{T}\}$, $\mathcal{T} \in \mathbb{N}$. Let $\underline{x}^{(1:k)}$ denote the vector of space-time locations at which samples have been taken up to timestep k. Between timestep k and $k+1$, R_i moves according to the discrete dynamics

$$p_i(k+1) = p_i(k) + u_i(k),$$

where $\|u_i\| \leq u_{\max}$ for some $u_{\max} \in \mathbb{R}_{>0}$. The robots collaborate with the static nodes to determine the control vector $u_i(k)$. We assume a limited communication radius, $R \in \mathbb{R}_{>0}$, for the robotic agents, with the restriction,

$$R \geq \max_{i \in \{1,\ldots,m\}} \{\mathrm{CR}(V_i(Q))\} + r_{\mathrm{s}} + u_{\max}, \qquad (10)$$

where $\mathrm{CR}(\cdot)$ denotes the *circumradius* of a polytope, and r_{s} is a maximum radius beyond which the covariance is zero. This restriction ensures communication between each robot and nodes whose Voronoi regions are correlated.

We examine two different uncertainty-based optimality criteria based on the models described in Section III. Here we use the universal kriging model. The works [29], [30] extend these results to a related Bayesian model which allows for uncertainty in the covariance, which we have omitted due to space constraints. The treatment in all cases follows a similar pattern. The optimality criteria involve optimizing at timestep k a function of the positions of the *next* set of measurements. Specifically, we try to

- maximize the *entropy* of the joint posterior predictive distribution at the new sample locations, or
- minimize the *average* over the predictive region of the posterior predictive variance.

In each case, the finite covariance radius allows an approximation of the *centralized* objective function which may be *distributed* over the hybrid network. Here we describe the process for posterior predictive entropy maximization. The posterior predictive entropy is a measure of the information which will be provided about the model by a set of new locations if samples are taken at those locations. We would like to choose sample locations at step $k+1$ such that the entropy is at a maximum. This requires calculation of the log determinant of the correlation matrix, $\mathbf{K}$, which can not be done in a distributed fashion. The second order Taylor series expansion yields an approximation of the entropy,

$$\mathcal{H}^{(k)}(P) = \log\det \Upsilon - \frac{1}{2}\mathrm{tr}\big((\mathbf{K} - \boldsymbol{I})^2 \big).$$

The value of $\mathcal{H}^{(k)}$ and of its gradient at P may be calculated using a combination of the distributed average consensus and distributed JOR algorithms.

Between timestep k and $k+1$, we restrict the robots to a convex region, $\Omega^{(k)}$, defined by the maximum movement rate and a minimum inter-agent distance. The network may use the following steps to perform a distributed version of projected gradient method at timestep k to optimize $\mathcal{H}^{(k)}$ over $\Omega^{(k)}$,

- calculate $\mathcal{H}^{(k)}(P)$ and $\nabla\mathcal{H}^{(k)}(P)$ using distributed average consensus and distributed JOR
- run a distributed version of an Armijo-type line search to find a stepsize, α
- find the projection, P', on $\Omega^{(k)}$ of $P + \alpha\nabla\mathcal{H}^{(k)}$
- repeat above for $P = P'$.

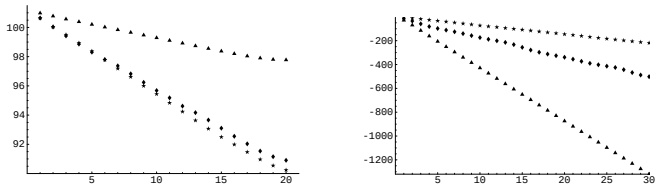
Note that these steps are meant in a distributed way across the network of static nodes. Thus all information is not known to all nodes at once. For example, S_j will only know the partial derivatives of $\mathcal{H}^{(k)}$ corresponding to the robots within communication range of $\mathcal{V}_j(Q)$. Using this projected gradient ascent algorithm, the network can be guaranteed that the location chosen for the next set of measurements is at a local maximum of $\mathcal{H}^{(k)}$ over $\Omega^{(k)}$. The overall adaptive design algorithm for the network then follows these steps:

- at timestep $k \in \{1, \ldots, T\}$, R_i executes the following
  - take sample
  - send sample and location to nearby nodes
  - receive next location
  - move to next location before next timestep
- at timestep $k \in \{1, \ldots, T\}$, S_j executes the following
  - collect samples and locations from nearby robots
  - using the method described above, run the distributed gradient ascent algorithm to find the next sample locations
  - send resulting next location to each robot in $V_j(Q)$

In simulation, we compared our gradient method against two a priori methods. The first was a static configuration where the robots spread out and remain in position. The second was a naive lawnmower approach, in which the robots began evenly spaced in the vertical direction and marched back and forth horizontally across the region. In all cases, some of the agents dropped communication during the course of the simulation to illustrate the robustness to failure of the gradient approach. Fig. 3 shows the resulting objective function values as a function of timestep for both the entropy condition and the average error variance condition.

VI. Distributed estimation and control

In mobile robotics, it is well known that performance and robustness can be improved through cooperative control. To extend these benefits to the paradigm of sensor networks using random field models, guidelines must be established for data collection and representation in such distributed



(a) (b)

Fig. 3. Objective values as a function of timestep with the static (triangle), lawnmower (diamond), and gradient descent (star) approaches. (a) $\tilde{\mathcal{A}}^{(k)}$, run with 5 static nodes and 10 robotic agents. (b) $\mathcal{H}^{(k)}$, run with 10 static nodes and 20 robotic agents. The covariance function is exponential, and the model used is a more general version of the kriging models in Section III (see [29], [30]).

systems. Here, we have reviewed work that provides a basis for this study, but there is much more yet to be explored. We briefly outline some exciting research topics below.

Within the context of the Gaussian process model discussed here, asymptotic results similar to those in Section V may be useful in finding optimal *trajectories* for sampling dynamic random fields. Other distributed optimization methods and criteria should be examined, as well as the asynchronous sampling regime. In addition, the statistical assumptions within the Gaussian process model should be challenged. Other established spatial models such as Gaussian Markov random fields and graphical models should also be examined with an eye towards distributed implementation,

A comprehensive solution to the distributed approach to data collection in random fields should include development of new statistical models which take the distributed nature of the problem into account directly. Recent work in Bayesian statistics has focused on the use of hierarchical models to represent non-stationary or even discontinuous random fields. We believe that similar techniques may be used to combine accurate process models with the distributed operational context required for cooperative control.

VII. Conclusions

We have motivated the need for statistically aware distributed algorithms for the estimation and control of robotic sensor networks. This exciting area of research draws from cutting edge spatial statistics and the fields of distributed computation and control. We have outlined some of the main problems and given examples of recent work on distributed data fusion and sequential optimal design. We believe the coming years will see a fertile activity on this area.

Acknowledgements

Research supported by NSF CAREER Award ECS-0546871.

References

[1] K. M. Lynch, I. B. Schwartz, P. Yang, and R. A. Freeman, "Decentralized environmental modeling by mobile sensor networks," *IEEE Transactions on Robotics*, vol. 24, no. 3, pp. 710–724, 2008.

[2] M. A. Demetriou and I. I. Hussein, "Estimation of spatially distributed processes using mobile spatially distributed sensor network," *SIAM Journal on Control and Optimization*, vol. 48, no. 1, pp. 266–291, 2009.

[3] G. M. Hoffmann and C. J. Tomlin, "Mobile sensor network control using mutual information methods and particle filters," *IEEE Transactions on Automatic Control*, vol. 55, no. 1, pp. 32–47, 2010.

[4] S. Martínez, "Distributed interpolation schemes for field estimation by mobile sensor networks," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, 2010. To appear.

[5] N. D. Lee and J. V. Zidek, *Statistical Analysis of Environmental Space-Time Processes*. Springer Series in Statistics, New York: Springer, 2006.

[6] K. Chaloner and I. Verdinelli, "Bayesian experimental design, a review," *Statistical Science*, vol. 10, no. 3, pp. 273–304, 1995.

[7] F. Pukelsheim, *Optimal Design of Experiments*, vol. 50 of *Classics in Applied Mathematics*. Philadelphia, PA: SIAM, 2006.

[8] E. P. Liski, N. K. Mandal, K. R. Shah, and B. K. Sinha, *Topics in Optimal Design*, vol. 163 of *Lecture Notes in Statistics*. New York: Springer, 2002.

[9] M. L. Stein, *Interpolation of Spatial Data. Some Theory for Kriging*. Springer Series in Statistics, New York: Springer, 1999.

[10] J. S. Clark and A. E. Gelfand, *Hierarchical modelling for the environmental sciences*. Oxford University Press, 2006.

[11] D. A. Castanon and D. Teneketzis, "Distributed estimation algorithms for nonlinear systems," *IEEE Transactions on Automatic Control*, vol. AC-30, pp. 418–425, May 1985.

[12] J. S. Willcox, J. G. Bellingham, Y. Zhang, and A. B. Baggeroer, "Performance metrics for oceanographic surveys with autonomous underwater vehicles," *IEEE Journal of Oceanic Engineering*, vol. 26, no. 4, pp. 711–725, 2001.

[13] D. O. Popa, K. Sreenath, and F. L. Lewis, "Robotic deployment for environmental sampling applications," in *International Conference on Control and Automation*, (Budapest, Hungary), pp. 197–202, June 2005.

[14] M. F. Mysorewala, *Simultaneous robot localization and mapping of parameterized spatio-temporal fields using multi-scale adaptive sampling*. PhD thesis, University of Texas at Arlington, 2008.

[15] J. Choi, J. Lee, and S. Oh, "Biologically-inspired navigation strategies for swarm intelligence using spatial Gaussian processes," in *IFAC World Congress*, (Seoul, Korea), July 2008.

[16] N. E. Leonard, D. Paley, F. Lekien, R. Sepulchre, D. M. Fratantoni, and R. Davis, "Collective motion, sensor networks and ocean sampling," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 48–74, 2007.

[17] F. Zhang and N. E. Leonard, "Cooperative filters and control for cooperative exploration," *IEEE Transactions on Automatic Control*, vol. 55, pp. 650–663, March 2010.

[18] T. Leonard and J. S. J. Hsu, *Bayesian Methods*, vol. 1 of *Cambridge series in statistical and probabilistic mathematics*. Cambridge, UK: Cambridge University Press, 1999.

[19] M. Gaudard, M. Karvson, E. Linder, and D. Sinha, "Bayesian spatial prediction," *Environmental and Ecological Statistics*, vol. 6, pp. 147–171, 1999.

[20] J. Cortés, "Distributed Kriged Kalman filter for spatial estimation," *IEEE Transactions on Automatic Control*, vol. 54, no. 12, pp. 2816–2827, 2009.

[21] 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. With discussion.

[22] N. A. C. Cressie and C. K. Wikle, "Space-time Kalman filter," in *Encyclopedia of Environmetrics* (A. H. El-Shaarawi and W. W. Piegorsch, eds.), vol. 4, pp. 2045–2049, New York: Wiley, 2002.

[23] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Prentice Hall, 1979.

[24] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.

[25] R. Graham and J. Cortés, "Asymptotic optimality of multicenter Voronoi configurations for random field estimation," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 153–158, 2009.

[26] M. E. Johnson, L. M. Moore, and D. Ylvisaker, "Minimax and maximin distance designs," *Journal of Statistical Planning and Inference*, vol. 26, pp. 131–148, 1990.

[27] J. Cortés 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.

[28] R. Graham and J. Cortés, "A cooperative deployment strategy for optimal sampling in spatiotemporal estimation," in *IEEE Conf. on Decision and Control*, (Cancun, Mexico), pp. 2432–2437, Dec. 2008.

[29] R. Graham and J. Cortés, "Distributed sampling of random fields with unknown covariance," in *American Control Conference*, (St. Louis, MI), pp. 4543–4548, 2009.

[30] R. Graham and J. Cortés, "Cooperative adaptive sampling via approximate entropy maximization," in *IEEE Conf. on Decision and Control*, (Shanghai, China), pp. 7055–7060, Dec. 2009.