Active Learning for Adaptive Mobile Sensing Networks *

Aarti Singh, Robert Nowak and Parmesh Ramanathan
Department of Electrical and Computer Engineering
University of Wisconsin-Madison
singh@cae.wisc.edu, {nowak, parmesh}@engr.wisc.edu

ABSTRACT

This paper investigates data-adaptive path planning schemes for wireless networks of mobile sensor platforms. We focus on applications of environmental monitoring, in which the goal is to reconstruct a spatial map of environmental factors of interest. Traditional sampling theory deals with data collection processes that are completely independent of the target map to be estimated, aside from possible a priori specifications reflective of assumed properties of the target. We refer to such processes as passive learning methods. Alternatively, one can envision sequential, adaptive data collection procedures that use information gleaned from previous observations to guide the process. We refer to such feedback-driven processes as active learning methods. Active learning is naturally suited to mobile path planning, in which previous samples are used to guide the motion of the mobiles for further sampling. This paper presents some of the most encouraging theoretical results to date that support the effectiveness of active over passive learning, and focuses on new results regarding the capabilities of active learning methods for mobile sensing. Tradeoffs between latency, path lengths, and accuracy are carefully assessed using our theory. Adaptive path planning methods are developed to guide mobiles in order to focus attention in interesting regions of the sensing domain, thus conducting spatial surveys much more rapidly while maintaining the accuracy of the estimated map. The theory and methods are illustrated in the application of water current mapping in a freshwater lake.

Categories and Subject Descriptors

I.2.9 [**Artificial Intelligence**]: Robotics—*Sensors*; I.5.4 [**Pattern Recognition**]: Applications—*Signal Processing*

General Terms

Algorithms, Theory

Keywords

Active learning, adaptive sampling, mobile sensors, path planning, sensor networks

*This work was supported by the National Science Foundation under grants ECS-0529381, CCR-0310889 and CNS-0519824.

IPSN'06, April 19–21, 2006, Nashville, Tennessee, USA.

1. INTRODUCTION

Many environmental monitoring applications require fine resolution and high fidelity mapping of a spatial phenomena distributed over a vast physical extent, e.g. aquatic and terrestrial ecosystem studies, detection of toxic biological and chemical spreads, oil spills and weather patterns. This requires an impractically large number of sensing elements to be distributed according to a pre-computed sampling strategy over the given area. Mobile sensor networks offer an alternative by trading off the cost of large number of static sensors in exchange for increased latency, and provide flexible sampling opportunities. Efficient path planning is necessary to harness these benefits of mobile sensing systems and obtain the best possible resolution in minimum time.

Traditional data sampling methods use a precomputed, fixed strategy that involves uniform sampling of the field at regular intervals, and the sampling pattern is not updated as measurements are collected. We refer to such strategies as *passive learning* methods. When the spatial phenomena of interest is smoothly varying, passive methods are known to perform well. However, many environmental variables exhibit highly localized features in the spatial map like changepoints or edges that need to be accurately captured and tracked. For such cases, one can envision sequential, adaptive path planning where information gleaned from previous samples is used to focus mobile paths towards regions of sharp changes in the environmental field where more intensive sampling is required. Such feedback-driven approaches where future sampling locations are determined by past sampling locations and observations are referred to as *active learning* methods. Active learning has been successfully applied to standard problems in statistical inference and machine learning, however there has been very little work aimed at harnessing the power of active learning for designing efficient sampling paths for mobile sensing networks [3, 14]. In this paper we use active learning methods to design adaptive paths for mobile sensors that achieve minimax efficiency of field reconstruction and latency for environmental monitoring type applications. When the environmental phenomena

exhibits localized features, for example a sharp level transition, the proposed adaptive path algorithms significantly outperform traditional non-adaptive sampling path strategies. It is shown that for reconstruction of a $d \geq 2$ dimensional piecewise constant field to a desired error level of ϵ, active path designs require the mobile sensor to cover a pathlength $\propto \epsilon^{-\frac{d-1}{d}}$ and result in a latency that scales as $\epsilon^{-(d-1)}$. Notice that both the pathlength and latency increase as the desired error level is decreased, as one would expect. On the other hand, a non-adaptive uniform sampling path scheme would require the sensor to move over a pathlength $\propto \epsilon^{-1}$ and result in latency that scales as ϵ^{-d}. In the non-adaptive case, the pathlength and latency both increase much more rapidly as the error level is decreased. In fact, adaptive path planning provides a significant reduction in pathlength and latency (by factors of $\epsilon^{\frac{1}{d}}$ and ϵ, respectively), that translates directly to network resource savings or better capabilities for tracking the environmental phenomena of interest. Also notice that the rate exponent gain is equivalent to a dimensionality reduction of the estimation task. Since the features of interest are located in a $d-1$ dimensional subset of the original domain, one can see that active learning results in paths that adapt to the effective dimension of the data. Even more dramatic improvements are possible in one-dimensional scenarios (e.g., vertical measurements through a forest canopy or marine environment).

The path planning strategies developed in this paper can be used for applications like the NIMS (Networked Infomechanical Systems) architecture [2,3]. The NIMS system uses controlled mobility of sensors for spatiotemporal sampling of various environmental variables like solar radiation intensity, atmospheric water vapor, temperature, and chemical composition over the vast expanse of a forest. The sampling paths of these sensors need to be designed so that the scale of motion matches the environmental monitoring needs in terms of accuracy, fidelity and tracking capabilities. Many of the phenomena of interest exhibit changepoints, for example, the solar radiation intensity map exhibits edges due to shadows cast by the vegetation. Further, measurements of different environmental variables require diverse sensors with different sampling duration and velocity requirements. For example, the sampling duration required to achieve a certain SNR (Signal-to-Noise Ratio) may vary from a few seconds for solar intensity measurements to a few minutes for CO_2 sampling. We present path planning algorithms that accommodate this wide range of sensing capabilities.

Another potential application arises in the North Temperate Lakes Long Term Ecological Research program in Wisconsin [1]. This project involves a sensor network system equipped on boats to take measurements of temperature, water currents, dissolved gases, algae and other biological species concentrations across the lake and at various depths. Continuous monitoring of these phenomena and change detection requires the boats to sample the 3D transect in the most efficient manner. The results presented in this paper provide guidelines for planning the path, velocity and number of boats required to achieve a certain desired accuracy of measurement in the least time. To verify the effectiveness of our methods and theoretical results, we present an illustrative example that uses adaptive and passive sensor mobility to estimate the water current velocity map for a freshwater lake in Wisconsin.

Some other applications that can benefit from fast spatial surveying adapted to the regions of interest include landscape scanning using sensor equipped aerial vehicles, estimating boundary of an oil spill using aerial or buoy sensors, prediction of weather patterns by monitoring atmospheric pressure, and geological fault-line detection [8].

While several research papers have explored the use of mobile sensors and proposed a variety of path planning algorithms (e.g., game-theoretic approaches, pursuit evasion, sensor exposure etc. [9, 11, 13]), the research has mainly being focused on target detection and tracking. In this paper we develop path planning algorithms for mobile sensors used in environmental monitoring type applications that involve mapping of a spatially distributed field. We leverage the benefits of active learning methods to design feedback-driven adaptive sensing paths that achieve minimax optimal efficiency in terms of mean square error (MSE) distortion and latency under varying sensor capabilities (velocity and sampling frequencies). A theoretical framework is provided for the analysis of tradeoffs in latency, accuracy and path lengths. We also investigate cooperative strategies for networked systems of multiple mobile sensors or combination of sensors with and without mobility. It is shown that the advantages of task distribution under cooperative strategies can also be attained while requiring minimal coordination between sensors. Such a scheme is robust to sensor failures and degrades gracefully as one or more sensors turn faulty.

The paper is organized as follows. Section II reviews the active learning methods developed for function estimation and associated theoretical results. Section III presents the application of active learning methods to adaptive path planning for a mobile sensing network and discusses the tradeoffs in accuracy, latency and path lengths. In Section IV we present the illustrative simulation study of water current velocity in Lake Wingra using both passive and adaptive sensing paths. We conclude in Section V.

2. ACTIVE LEARNING FOR SPATIAL MAPPING

Consider the task of spatial mapping of an environmental variable by mobile sensors that sample the given transect at n locations $\{X_i\}_{i=1}^n$. Denote the observed (scalar) sample values by $\{Y_i\}_{i=1}^n$, that are assumed to obey the measurement model:

$$Y_i = f(X_i) + W_i \qquad i \in \{1, \ldots, n\},$$

where the function f is the field of interest and the sensor measurement noise W_i is characterized as iid random variables that are independent of the sample locations $\{X_i\}_{i=1}^n$. The task is to reconstruct a map of the spatial field f from sample locations and noisy observations $\{X_i, Y_i\}_{i=1}^n$. Classical (passive) sampling techniques consider uniform deterministic or random sampling locations that are specified prior to gathering the observations. Active learning methods, on the other hand, select the future sampling locations online based on past locations and observations, i.e., X_i depends deterministically or randomly on the past sample locations and observations $\{X_j, Y_j\}_{j=1}^{i-1}$ [5–8, 12].

While the concept of active learning is not new, there is very little theoretical evidence to support the effectiveness of active learning, and existing theories often hinge on restrictive assumptions that are questionable in practice. However, there are a handful of key results that can be leveraged to aid

in the design of mobile wireless sensing systems. Burnashev and Zigangirov [5] investigated the problem of changepoint detection in a 1-d function using sequential sampling. They show that the error in the location of the changepoint decays exponentially in n, the number of samples, as opposed to n^{-1} for non-adaptive sampling. More recently, Castro et al. [6] investigate the fundamental limits of active learning for various nonparametric function classes including spatially homogeneous Hölder smooth functions, and piecewise constant functions that are constant except on a $d-1$ dimensional boundary set or discontinuity embedded in the d-dimensional function domain. They reveal that significantly faster rates of convergence, in the minimax sense, are achievable using active learning in cases involving a function whose complexity is highly concentrated in small regions of space. In this section we review the active learning methods presented in [5] and [6].

In [5], Burnashev and Zigangirov address the problem of estimating the confidence interval of an unknown parameter from n controlled observations. This problem is equivalent to estimating the changepoint in a 1-d function by adaptive sampling. The problem can be stated formally as follows. Consider the class of functions

$$\mathcal{F} = \{f : [0,1] \to \mathbb{R} | f(x) = 1_{[0,\theta)}(x)\}$$

where $\theta \in [0,1]$. The goal is to estimate the changepoint θ from observations $\{Y_i\}_{i=1}^n$, where

$$Y_i = \begin{cases} f(X_i) & \text{with probability } 1-p \\ 1-f(X_i) & \text{with probability } p \end{cases} = f(X_i) \oplus W_i,$$

where $\oplus$ indicates a sum *modulo* 2 and W_i represents Bernoulli noise. Clearly, if there is no noise ($p = 0$) it is easy to design an estimator $\hat{\theta}_n$ using binary bisection that attains exponential error probability i.e. $|\theta - \hat{\theta}_n| = 2^{-n}$. Burnashev and Zigangirov show that even when $p > 0$, a similar probabilistic bisection approach can be used to get $\mathbb{E}[|\hat{\theta}_n - \theta|] \leq 2^{-Cn}$, where $C > 0$ is a constant. Notice that this adaptive sampling rate is much faster than the passive rate of n^{-1}. The probabilistic bisection method is based on a bayesian posterior update. We describe here a generalization of the method [10] which does not restrict the noise W_i to be Bernoulli.

In practice, a discretized distribution for θ (e.g. a histogram form) is considered. The next sampling location is randomly chosen to be either end point of the interval in which the median lies by flipping a coin with head/tail probability that ensures the average value is the actual median. A corresponding modification is required in the distribution update. Refer to [5] for details. In the next section, we present an illustrative example using this method to design a mobile sensor's path assuming sensor readings contaminated with Gaussian noise (measurement error) and Bernoulli noise (sensor fault).

For multidimensional settings, Castro et al. [6] developed an active learning method based on dyadic partitioning that provides near-minimax optimal error convergence for certain classes of nonparametric multivariate functions. One of the main results of that work shows that any piecewise constant d-dimensional function separated by a d-1 dimensional boundary can be accurately estimated through adaptive sampling methods using far fewer samples than required by conventional non-adaptive sampling. For $d = 1$ they obtain the same exponential rates as shown by Burnashev.

PROBABILISTIC BISECTION SAMPLING

1. *Initialization*: Assume uniform prior for θ.

$$p_\theta^0(x) = \text{uniform}([0,1])$$

2. *Repeat* i = 1,. . . ,n

Sample selection: The sample location x_i is selected to be the median of the distribution of θ.

$$x_i : \int_0^{x_i} p_\theta^{i-1}(x)dx = 1/2$$

Record noisy observation $Y_i = f(x_i) \oplus W_i$.

Posterior update: Update the distribution of θ based on the observed sample Y_i at location X_i according to Bayes rule.

$$p_\theta^i(x) = \frac{p_{Y_i}(y|\theta = x)p_\theta^{i-1}(x)}{p_{Y_i}(y)},$$

where $y \in \{0,1\}$.

3. *Estimator*:

$$\hat{\theta}_n = \arg\max_{x \in [0,1]} p_\theta^n(x).$$

Their main results (for $d \geq 2$) can be summarized in the following three theorems:

Set up: An estimation strategy consists of a pair $(\hat{f}_n, S_n)$, where $\hat{f}_n$ denotes an estimator of the d-dimensional function f using n noisy observations $\{Y_i\}_{i=1}^n$, and S_n denotes a sampling strategy for choosing the sampling locations $\{X_i\}_{i=1}^n$. Let Θ_{active} denote the set of all active estimation strategies.

Theorem 1 Let $H(\alpha)$ denote the class of spatially homogenous Hölder-α smooth functions[1], then

$$\inf_{(\hat{f}_n, S_n) \in \Theta_{active}} \sup_{f \in H(\alpha)} \mathbb{E}_{f,S_n}[||\hat{f}_n - f||^2] \asymp n^{-\frac{2\alpha}{2\alpha+d}}.$$

The notation $a_n \asymp b_n$ means that $a_n = O(b_n)$ and $b_n = O(a_n)$. This rate is the same as the minimax learning rate with passive methods, hence for this class of smooth functions that do not contain highly localized features, both passive and active methods perform equally well.

Theorem 2 Let PC denote the class of piecewise constant functions with d-1 dimensional boundaries, then

$$\inf_{(\hat{f}_n, S_n) \in \Theta_{active}} \sup_{f \in PC} \mathbb{E}_{f,S_n}[||\hat{f}_n - f||^2] \asymp n^{-\frac{1}{d-1}}.$$

Thus the error rate is determined by the effective dimension of the function. On the other hand, in this situation the non-adaptive sampling error decays like $n^{-1/d}$, and thus we see that active methods can lead to significant improvements in such cases.

Theorem 3 Let $PS(\alpha)$ denote the class of more general piecewise Hölder-α smooth functions with d-1 dimensional

[1]The function has $\lfloor\alpha\rfloor$ continuous derivatives, where $\lfloor\alpha\rfloor$ is the maximal integer $< \alpha$, and the function can be well approximated by degree-$\lfloor\alpha\rfloor$ Taylor polynomial approximation.

boundaries, then

$$\inf_{(\hat{f}_n, S_n) \in \Theta_{active}} \sup_{f \in PS(\alpha)} \mathbb{E}_{f,S_n}[||\hat{f}_n - f||^2] \asymp \max\{n^{-\frac{2\alpha}{2\alpha+d}}, n^{-\frac{1}{d-1}}\}$$

This theorem follows directly from the previous two and essentially states that active learning methods adapt to the complexity of the problem at hand, the rate of convergence being determined by the dominating cause of function complexity - dimensionality of the edge or complexity of function derivatives away from the boundary.

Castro et al. also present an active learning algorithm that nearly achieves these minimax rates. We now review the active learning strategy proposed in [6] for piecewise constant functions.

MULTISCALE ADAPTIVE SAMPLING

1. *Preview Step*: $n/2$ samples are collected at uniformly spaced points and a coarse estimate of the field f is generated from these data using a complexity-regularized tree pruning procedure. This provides a rough indication of where boundaries may exist.

2. *Refinement Step*: $n/2$ additional samples are collected at points in regions near the rough location of boundaries detected in the Preview Step. A similar tree pruning procedure is used in these regions to obtained refined estimates in the vicinity of boundaries.

3. *Fusion Step*: The estimates from Preview and Refinement Steps are combined to produce an overall estimate.

This procedure produces a non-uniform sampling pattern in which more samples are concentrated near boundaries in the field; half of the samples are focused in a small region about the boundary. Since accurately estimating f near the boundary set is key to obtaining faster rates, we expect such a strategy to outperform the passive learning technique described earlier.

The estimators used in the first two steps are built over recursive dyadic partitions (RDPs) of the function domain. An RDP can be identified with a tree structure, where each leaf of the tree corresponds to a cell of the dyadic partition (see Fig. 1). An estimate is generated on each leaf of the RDP by averaging the observations collected in each leaf. If the field is Hölder-α smooth on either side of the boundary, the estimator consists of a least square degree-$\lfloor\alpha\rfloor$ polynomial fit to the observations [6] rather than simple averaging,

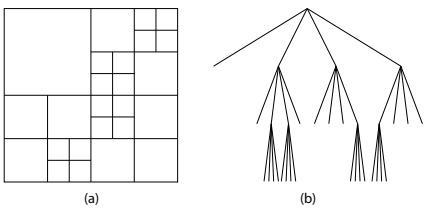


Figure 1: (a) An example Recursive Dyadic Partition (RDP) of a 2-d function domain and (b) associated tree structure.

as for piecewise constant field. If we consider Π to be the set of all RDPs that can be generated on the input domain, the best RDP is chosen according to the following complexity-regularized estimation rule:

$$\hat{\pi} = \arg\min_{\pi \in \Pi} \left\{ \sum_{i=1}^{n} (Y_i - \hat{f}_n^{(\pi)}(X_i))^2 + \lambda(|\pi|) \right\}, \qquad (1)$$

where $\lambda(\cdot)$ is a penalty term that depends on $|\pi|$, the number of leaves in RDP π and $\hat{f}_n^{(\pi)}$ is the estimator containing average of all observations in each leaf of the RDP. The final estimate for each step, $\hat{f}_n$ is given by

$$\hat{f}_n = \hat{f}_n^{(\hat{\pi})}. \qquad (2)$$

The computation of $\hat{f}_n$ can be done efficiently using bottom-up tree pruning algorithms, in the spirit of CART [4]. In the Preview Step, leaves that are not pruned back indicate regions of the field where further averaging (through pruning and aggregation) would have led to large data fitting errors. Thus, these leaves indicate regions that probably contain boundaries or other sharply varying characteristics of the field, and these are the focus regions for sampling in the Refinement Step. There are a few additional subtleties involved in the procedure [6], but this gives one the main idea of the method.

The MSE of $\hat{f}_n$ can be shown to decay like $n^{-1/(d-1+1/d)}$, much faster than the best passive rate of $n^{-1/d}$. Furthermore, performing repetitive refinement steps leads to an improved rate of $n^{-1/(d-1+\delta)}$, where $0 < \delta \leq 1/d$ decreases with the number of refinements. Thus, this method can be arbitrarily close to the minimax rate of Theorem 2.

We use this algorithm to plan the path of a current velocity sensor equipped boat and generate a spatial map of the water current velocity profile for a freshwater lake in Wisconsin. Simulation results reveal that huge savings in time can be achieved for reconstructing the spatial map to same accuracy as a passive path. In the next section we show how the active learning methods discussed here can be used to design fast adaptive spatial survey paths for mobile sensing networks and explore the tradeoffs in path length, accuracy and latency under varying sensor capabilities.

3. ACTIVE LEARNING FOR MOBILE SENSING

Research efforts in designing mobile sensor paths have primarily been focused on target detection, tracking and evasion [9,11,13]. Spatial mapping of a vast area is a fundamentally different task that confronts the challenges of unpredictable variability of environmental variables and demands for high spatial resolution. For such applications, active learning can be used to guide the mobile sensor along regions of interest resulting in shorter and faster survey paths, for a desired resolution accuracy. We now apply the active learning methods discussed in the last section to design efficient adaptive paths for a mobile sensing network that is required to generate a high fidelity, fine resolution estimate of a spatially varying environmental phenomena.

3.1 Single mobile sensor

We start with the simplest case of a single mobile sensor that needs to efficiently sample a 1-d field containing a change point. This situation is well-suited, for example, to a

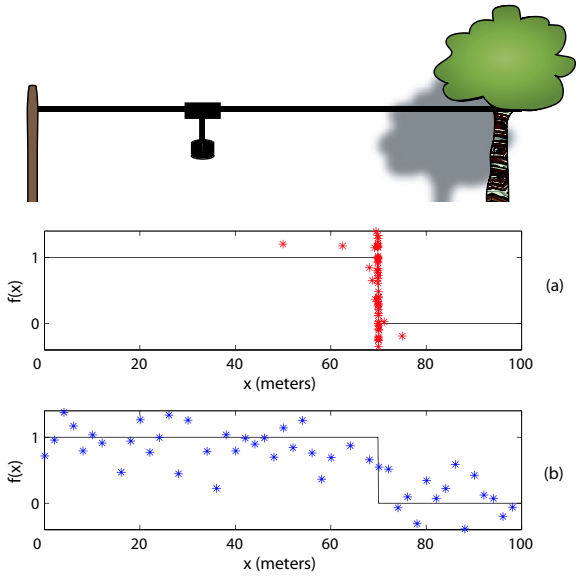


Figure 2: A toy example of estimating a 1-d solar intensity map containing a single changepoint at 70 m using a NIMS type mobile sensor. The sample locations and observations are shown in (a) for adaptive survey and (b) for passive survey.

NIMS [2] like sensor suspended on cableways that measures solar radiation intensity within the forest transect with the shadow cast by an object constituting a changepoint. We model a toy example (Figure 2) for estimating this 1-d function under two scenarios:

I. **Measurement noise** - Typically the sensor measurements are noisy due to environmental fluctuations and slight uncertainties in sensor readings. Such noise can be modelled as a Gaussian random variable. Figure 2 shows the sampling locations and the corresponding noisy observations made by the mobile sensor with (a) adaptive sampling using probabilistic bisection and (b) passive sampling. In the adaptive case, though the sensor takes longer excursions for the initial samples, it quickly "homes-in" on the more interesting feature of the field concentrating most of its measurements around the changepoint with only a few samples where the field is constant. For a 100 m transect containing a changepoint at 70 m that needs to be mapped to a resolution of 0.1 m and assuming noise variance of 0.1, the number of samples required in the adaptive scheme are exponentially less - only about 10 as opposed to 700 samples for passive case. A NIMS sensor typically moves at 1 m/s and takes 1 sec to record a sample of the solar intensity, which implies that an adaptive mobile sensor would take 2 mins instead of 13 mins to accomplish the task.

II. **Sensor fault** - We model this case by assuming the noise W_i is a Bernoulli(p) random variable with p denoting the probability that either the sensor fails to record a reading or the reading is corrupted and has to be discarded. We simulate the same example as before for Bernoulli noise with $p = 0.2$. It was observed that 20 samples were required with the adaptive path to achieve the desired resolution of 0.1 m, as opposed to nearly 700 for non-adaptive path.

For a 2-d transect, the multiscale adaptive sampling technique developed by Castro et al. [6] can be used for designing adaptive paths for a mobile sensor. We start with the case of a single mobile sensor and later extend the result to multiple mobile sensing network and a network system comprised of heterogeneous sensors with and without mobility.

The design of the sensing path for a single mobile sensor can be described as follows. For simplicity, we assume that the coordinates are scaled so that the area to be monitored can be described as a unit square or hypercube $[0,1]^d$. Also, we only describe the case where the field is piecewise constant, i.e. the variable of interest exhibits a level transition taking on constant values on either side of the boundary. If the field is not constant, but say Hölder-α smooth on either side of the boundary, the path planning is same as described for the piecewise constant case, except that the estimator involves a least square degree-$\lfloor\alpha\rfloor$ polynomial fit to the observations rather than simple averaging [6].

1. *Coarse survey* - With no prior knowledge of the field, the mobile sensor starts by doing a coarse passive survey of the field in a raster scan fashion. The sensor collects $n/2$ samples at regular intervals along the coarse survey path of length ℓ. A complexity penalized estimate $\hat{f}^c$ is constructed using the $n/2$ samples over recursive dyadic partitions of the field according to equations (1) and (2), with penalty $\lambda(|\pi|) = C\sigma^2\log(n/2)|\pi|$, that is proportional to the number of leaves $|\pi|$ and sensor measurement noise σ^2. $C > 0$ is a constant that depends on the dimension d and smoothness of the field. Notice that this penalizes RDPs with fine partitions and leaves at maximum depth are retained only if pruning by averaging the observations would lead to large data fitting errors. The estimator averages out the noise where the field is smooth (reduces the variance where bias is low), while performing no averaging on the samples around the boundary (where bias is high). This yields an RDP estimate of the field with leaves at the greatest depth J providing a rough location of the boundary.

$$\mathbb{B} = \{\text{Regions of } [0,1]^d \text{ where the coarse estimate contains leaves at maximum depth}\}$$

Since at each depth in the tree, the sidelength of cells is halved, at maximum depth J the leaves have sidelength 2^{-J}, which is equal to the finest resolution of $1/\ell$ provided by a pathlength of ℓ. The volume at this maximum depth is ℓ^{-d}, which implies that the maximum possible number of leaves at this depth is ℓ^d. Of these, only $O(\ell^{d-1})$ intersect the boundary since the boundary occupies a d-1 dimensional subspace in the monitored region. Thus, $|\mathbb{B}| = O(\ell^{d-1})$.

2. *Refinement pass* - The mobile sensor is now guided along the regions identified as containing the boundary in the coarse survey (set $\mathbb{B}$) to collect an additional $n/2$ samples in these regions, again traversing a pathlength ℓ. A complexity penalized estimator $\hat{f}^r$ is generated on each region in the set $\mathbb{B}$, similar to the estimator built on the entire area in the coarse survey. This confines

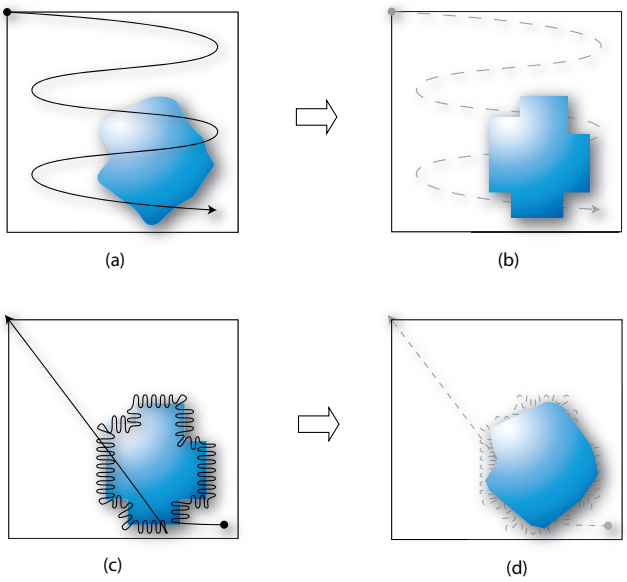


Figure 3: Multiscale adaptive path of a mobile sensor for mapping a field containing a boundary. In the first step, the mobile sensor follows a coarse survey path (a) and produces a rough estimate of the field (b). In the refinement pass (c), the mobile follows a path adapted to the interesting regions of the field and produces a fine resolution estimate (d).

the boundary location (and hence bias) to even smaller regions, while averaging out the noise and lowering the variance in the smooth regions of set $\mathbb{B}$.

3. *Field reconstruction* - A final estimate of the field is now generated by fusing the estimates obtained in the coarse survey and the refinement pass as follows:

$$\hat{f}_{active}(x) = \begin{cases} \hat{f}^r(x) & \text{if } x \in \mathbb{B} \\ \hat{f}^c(x) & \text{otherwise} \end{cases}$$

Figure 3 shows the two stages of the adaptive mobile sensing path. The advantage of a two-step method is that the computation involved can be done either on the mobile platform or if enough processing power is not available on board the mobile, the data can be dumped to a fusion center after each pass where processing and design of future paths may be carried out.

The mean square error (MSE) of the estimator can be upper bounded using the Craig-Bernstein inequality as in [6]:

$$\mathbb{E}[||\hat{f}_{active} - f||^2] \leq C_1 \left(\frac{\log n'}{n'}\right)^{1/d} 2^{-J} + C_2 2^{J(d-1)} \frac{\log n}{n}$$

where J denotes the maximum depth of the RDP tree estimator in the coarse step and $n' = n/(2|\mathbb{B}|)$ denotes the number of samples in each preview leaf collected in the refinement pass and $C_1, C_2 > 0$ are constants. Since $1/\ell = 2^{-J}$ as discussed before, in terms of pathlength the above bound can be expressed as:

$$MSE \leq C_1 \left(\frac{\log n'}{n'}\right)^{1/d} \ell^{-1} + C_2 \ell^{d-1} \frac{\log n}{n} \qquad (3)$$

where $n' = n/(2\ell^{(d-1)})$, $n/2$ samples distributed over $|\mathbb{B}| = O(\ell^{d-1})$ leaves that contain rough location of the boundary. The first term denotes the error incurred in the refinement pass in regions close to the boundary and the second term denotes the error incurred in the coarse survey in regions away from the boundary. The two terms can be balanced by appropriately choosing the number of samples and path length:

$$\ell \sim O\left(n^{\frac{d-1}{(d-1)^2+d}}\right), \qquad (4)$$

where $\sim$ denotes polynomial order dependence, dropping any log terms. This represents the classical bias-variance tradeoff since the length ℓ primarily determines the bias in estimating the boundary location, while the number of samples n determines the variance. If a spatial map with a MSE accuracy ϵ is desired, then by setting the MSE bound in (3) to ϵ and combining with (4), we find that the minimum number of samples required and the optimal sensing path length would scale with the desired accuracy as:

$$n_{opt} \sim O(\epsilon^{-(d-1+\frac{1}{d})})$$

and,

$$\ell_{opt} \sim O(\epsilon^{-\frac{d-1}{d}}),$$

where the latter expression is obtained by substituting n_{opt} into (4). This implies that for a 2-d field, as in the water current mapping in a lake, increasing the accuracy by a factor of 10 would require the mobile sensor to traverse roughly 3 times longer pathlength and collect about 30 times more samples. This is a significant improvement over the passive approach, where the mobile would have to cover 10 times longer path and collect 100 times more samples.

If the mobile sensor moves at a velocity v and requires time T to record one sample of the environmental variable to a desired Signal to Noise Ratio (SNR), the total time required to generate the estimate is given by

$$t = nT + \frac{\ell}{v}$$

Thus for a fixed sampling time T and fixed velocity v,

$$\begin{aligned} t &\sim O(\max\{n_{opt}, l_{opt}\}) \\ &\sim O(\epsilon^{-(d-1+\frac{1}{d})}) \end{aligned}$$

for all $0 \leq \epsilon < 1$. As discussed in section II, multiple refinement steps can be carried out to reduce the time required to $O(\epsilon^{-(d-1+\delta)})$, where $0 < \delta \leq \frac{1}{d}$ decreases with successive refinement passes.

If the mobile sensor was to move along a passive path, the time required to achieve the same accuracy would be $O(\epsilon^{-d})$, and the mobile would need to traverse a length of $O(\epsilon^{-1})$. Thus, the improvement provided by adaptive path planning is essentially equivalent to a dimensionality reduction of the field estimation problem. This shows that active learning adapts to the complexity of the problem, since the effective dimension of the field in the piecewise constant case is the dimension of the boundary (d-1). This represents a huge reduction in latency and pathlengths, that directly translates to network resource savings as well as allows for more accurate tracking of the temporal changes in the spatial map of the environmental variable.

In practice, sensors for measuring different environmental variables may have sampling time anywhere from a few seconds for measuring water current velocity to few minutes for measuring CO_2 level in a forest canopy. Also sensors might be restricted in mobility to a certain maximum velocity. For example, NIMS type sensors that are suspended on cableways are limited to a maximum velocity of about 1 m/s, on the other hand sensors mounted on aerial vehicles can move at hundreds of m/s. Thus for a desired small accuracy of ϵ, one can identify two distinct regimes of operation based on sensor capabilities.

I. *Sampling time constrained* - If the sensor mobility is constrained by the sampling time, the latency incurred is primarily due to the time spent in recording samples. This represents the **variance limited** regime since the noise in sensor measurements limits achievable accuracy, while the sensor can move at high enough velocity to cover the path length required to reduce bias. In this situation the latency scales according to

$$t \sim O(n_{opt}) \sim O(\epsilon^{-(d-1+\frac{1}{d})})$$

Balancing the two terms contributing to latency ($nT = \ell/v$), we find that the mininum velocity required for a sampling time constrained sensor scales like

$$v_{opt} \sim O(\epsilon^{d-2+\frac{2}{d}}).$$

The sensor can move at any velocity greater than or equal to the optimal velocity. However if the sensor's velocity is limited to less than v_{opt}, we are in the *velocity constrained* regime.

II. *Velocity constrained* - If the sensor mobility is velocity limited (i.e., if the platform is not capable of moving at the minimum required velocity above), then the time to move from one sampling location to the other is the primary cause of latency. This corresponds to the **bias limited** regime since enough samples can be collected to reduce the variance, while the path length that the sensor can cover and hence reduction in bias, is limited. In this regime the latency scales like

$$t \sim O(\ell_{opt}) \sim O(\epsilon^{-(\frac{d-1}{d})})$$

Again this rate is faster than that for passive paths $O(\epsilon^{-1})$, leading to time savings of the same order as in the variance limited regime where the sensor is sampling time constrained.

Notice that even for the **noise-free case** (i.e. when the sensor reading has high enough SNR), latency scales with pathlength since the variance is negligible (ideally zero) and the error is completely due to bias in locating the boundary. In this case, no error is incurred in cells away from the boundary, and the pathlength determines the resolution or bias. Thus the adaptive rates for bias-limited regime also characterize the noise-free case and one sees that multiscale adaptive scheme performs better than passive whether noise is present or not.

It should be noted that if the environmental variable does not exhibit much variability and the spatial map is smooth, the path planning strategies described above will yield performance similar to a uniform passive scan of the field. Thus active learning paths are data adaptive, resulting in shorter localized paths if the interesting phenomena are confined to certain regions of space and longer uniform paths if the phenomena is spatially well distributed.

3.2 Mobile Sensing Network

So far we have developed strategies to design adaptive paths for a single mobile sensor. Now we consider a network of mobile sensors. Intuitively, it is clear that if there are k mobile sensors, the field estimation task can be distributed amongst them leading to reduced latency (by a factor of k). However, the way the task distribution is done can have significant impact on the robustness of the system. For example, an obvious approach is to divide the field into k equal regions and require each mobile to perform field estimation on one region. Assuming, for example, that the mobile sensors have a fixed sampling time and are moving at the optimal velocity, the MSE distortion is reduced by a factor of $k^{-1/(d-1+\frac{1}{d})}$ for a desired latency, or equivalently, for a given distortion it yields a factor k improvement in latency.

However, this approach is clearly not robust to sensor failure. We wish to devise a path planning strategy that would degrade gracefully with node failures. Certainly, as long as there is at least one working sensor, the error should be no larger than the error in the single mobile sensor case. A cooperative strategy needs to be followed where sensors coordinate their movements so that even with sensor failures there is adequate sampling of the entire field, though at a reduced resolution. We outline such a collaborative strategy below.

Cooperative strategy for k mobile sensors

1. *Initialization* - The mobile sensors start at uniformly spaced locations in the field. This provides robustness if sensors are likely to fail with progression in time.

2. *Coarse survey* - The mobile sensors survey the whole field in a raster scan fashion moving along paths that are interleaved such that spacing between adjacent mobiles' paths is $1/\ell$, where ℓ is the pathlength a sensor would traverse if acting alone (see Fig. 3). Notice that rows of each raster scan are now k/ℓ apart since there are k sensors. Thus, each mobile covers a k times shorter (and hence less dense) path and collects n/k measurements. The measurements are transmitted to a fusion center where a coarse RDP estimate of the field is constructed using measurements from all k mobiles. The rough location of the boundary is then conveyed back to all the mobiles.

3. *Refinement pass* - The mobile sensors move along coordinated paths, similar to the coarse survey (each path is again widely spaced by a factor of k), along the regions identified as possibly containing the boundary to collect additional n/k measurements each. The final refined estimate is then generated using the collective measurements.

If the failures are random, with p the probability of a sensor failure, it can be shown that under this cooperative strategy the distortion will be reduced by a factor of

$(pk)^{-1/(d-1+\frac{1}{d})}$, with pk reflecting the effective number of active sensors. (The proof is omitted due to space constraints.) Thus this cooperative strategy is robust and degrades gracefully with sensor failures. The coordination requirements of this scheme during the data collection process can actually be reduced. It can be shown that if the mobile sensors start at uniformly random locations and collect uniformly random samples independent of other sensors in each pass, we can still guarantee distortion improvement by the same factor, provided the estimates in each step are formed using collective measurements and the fusion center transmits the rough location of the boundary to all sensors after the first step.

It is envisioned that in many cases mobile sensors will be used to complement a static network of sensors. In this case, the static sensor network can be deployed uniformly across the region of interest to form a low resolution estimate of the spatial map. This would provide continuous monitoring of the environmental variable, and if an event of interest occurs the mobile sensor(s) can be dispatched to collect guided measurements and form a refined estimate of the changepoints.

4. CASE STUDY: LAKE MAPPING

In this section, we use the path planning strategies developed in the paper for a real-life problem of water current mapping in a freshwater lake. The Lake Ecology research group in Wisconsin [1] is interested in studying hydrodynamics in lakes and how the biophysical setting of the lake influences these dynamics. One of the primary requirements for such studies is the spatial mapping of the water current velocity in the lake. Currently, such measurements are taken by a sensor called Acoustic Doppler Current Profiler mounted on a boat that moves around the lake taking samples at regular intervals. However, the time requirements for producing a spatial map of the entire lake have restricted the researchers to base their inference on field measurements collected from a small part of the lake, and at coarse resolutions. The techniques developed in this paper for designing adaptive paths that yield low latency, high resolution spatial map can greatly benefit such research efforts.

The water current velocity in a lake is influenced by many factors including the lake bathymetry, littoral-zone vegetation, stratification induced by temperature and the wind profile generated by the lake surroundings. The circulation velocity map is shown in Figure 4 for a freshwater lake. The map exhibits two distinct regions with significant velocity gradient between the two. Hence, the field can be characterized as a level transition with an irregular boundary. This situation is ideally suited for the multiscale adaptive sampling based path planning approach developed in the last section. We investigate the performance of the multiscale adaptive approach against a passive approach by designing the path for the sensor-carrying boat using both strategies and comparing the accuracies obtained for the reconstructed maps.

To apply our algorithms to the lake model, we consider a square region around the lake with the values at the edges of the lake interpolated to fill the square area. This is necessary to prevent the algorithm from focusing samples along the boundary of the lake itself, rather than the boundary of the velocity gradient that we really seek. We use the data from a three-dimensional non-hydrostatic and stratified flow model (3DNHYS) [16] for our experiments. For simplicity we model the lake with a piecewise constant function. In general, platelet or polynomial fits can be used to approximate the velocity on either side of the boundary. The platelet fit has been used in [15] for a network of static sensors and is shown to provide good approximation to a smooth field using the active learning method of [6].

Figure 4: Simulated low resolution water current velocity profile in Lake Wingra. Notice there are two distinct regions characterized by low or high velocity, with significant gradient between them.

To study the circulation pattern, spatial map of the water current velocity at a resolution of $\sim 10\,\text{m}$ is required. The original discrete model data provided by the limnology group is for a $2\,\text{km} \times 1\,\text{km}$ lake with a resolution of 25 m between samples. We interpolate this data using nearest neighbor interpolation to get a sampling resolution of $25/3 = 8.3\,\text{m}$. These samples can be thought of as the data collected using a sensor mounted on a boat that moves along a dense uniform passive path around the entire lake. Of the 256×256 samples in our square region, 37323 samples actually lie in the lake region and are meaningful to evaluate the accuracy, latency and pathlength of a boat recording these measurements. In the field experiments done by the lake research group, the sampling rate of data collection using the acoustic doppler current profile sensor is 2 Hz (i.e. 0.5 secs/sample), and the boat speed is about 2 m/s. This implies that the boat needs to cover an extensive pathlength of 310 km and takes nearly 48 hrs to collect all the samples. Analysis of the latency shows that only 5 hrs are spent in the measurement process, while most of the time is consumed in moving around the lake. Thus for this problem, we are in the "velocity-constrained" regime. The reconstructed estimate for the passive case is shown in Figure 5(a) with MSE $= 3 \times 10^{-4}$.

In the two-step adaptive approach, the boat makes an initial coarse survey collecting samples at 33.2 m resolution. Figure 5(b) shows the estimate obtained after the coarse survey and 5(c) shows the cells corresponding to the rough boundary locations (as indicated by the coarse estimator), superimposed on the noisy field. In the refinement pass, the boat is guided along the rough boundary location identified by the coarse survey to collect additional samples at finer resolution of 8.3 m. The final reconstructed map is shown in 5(d). The adaptive approach achieves nearly the same accuracy as the passive with MSE $= 4 \times 10^{-4}$ and requires only 8077 samples in the lake. The pathlength is thus reduced by a factor of 3 to 92 km and the time reduced to nearly 14 hrs. This represents huge savings in time, and makes the desired task much more feasible. Multiple sensors, if available, can

be used to reduce the latency further, e.g. 5 such sensors would reduce the time to less than 3 hrs.

5. CONCLUSION

This paper proposes adaptive path planning approaches to design paths of mobile sensors used for mapping a spatially distributed environmental phenomena. Using ideas from the active learning literature, fast spatial surveying methods are developed to guide the mobile sensors along interesting regions of the sensing domain. A theoretical framework is developed for evaluating the accuracy, pathlength and latency tradeoffs under varying sensor capabilities. It is shown that the proposed data-adaptive path planning procedures achieve significant improvements in latency and pathlength, for a given desired accuracy, over a non-adaptive raster-scan approach. Application of the developed methods to water current mapping in a lake results in latency improvement from 48 hrs to 14 hrs.

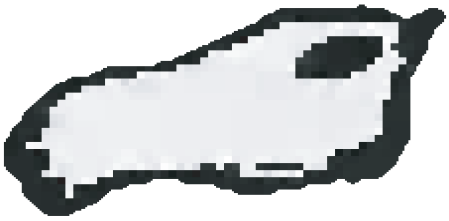
Acknowledgments

The authors would like to thank Henry Yuan and Prof. Chin Wu for providing the lake current velocity data, Rui Castro for his helpful remarks and Rebecca Willett for providing the pruned RDP estimator code.

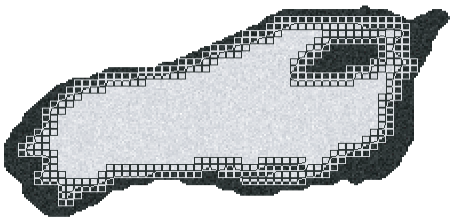
6. REFERENCES

[1] *http://limnosun.limnology.wisc.edu/*.
[2] *http://www.cens.ucla.edu/portal/nims/*.
[3] M. Batalin, M. Rahimi, Y. Yu, D. Liu, A. Kansal, G. Sukhatme, W. Kaiser, M. Hansen, G. Pottie, M. Srivastava, and D. Estrin. Call and response: Experiments in sampling the environment. In *Proceedings of ACM SenSys*, 2004.
[4] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth, Belmont, CA, 1983.
[5] M. Burnashev and K. S. Zigangirov. An interval estimation problem for controlled observations. *Problems in Information Transmission*, 10:223–231, 1974.
[6] R. Castro, R. Willett, and R. Nowak. Faster rates in regression via active learning. In *Proceedings of NIPS*, 2005. (also technical report available at http://homepages.cae.wisc.edu/∼rcastro/ECE-05-3.pdf).
[7] G. Golubev and B. Levit. Sequential recovery of analytic periodic edges in the binary image models. *Mathematical Methods of Statistics*, 12:95–115, 2003.
[8] P. Hall and I. Molchanov. Sequential methods for design-adaptive estimation of discontinuities in regression curves and surfaces. *The Annals of Statistics*, 31(3):921-941, 2003.
[9] J. P. Hespanha, H. J. Kim, and S. Sastry. Multiple-agent probabilistic pursuit-evasion games. In *Proceedings of the Conference on Decision and Control*, Dec. 1999.
[10] M. Horstein. Sequential decoding using noiseless feedback. *IEEE Trans. Information Theory*, 9(3):136–143, 1963.
[11] G. Kesidis, T. Konstantopoulos, and S. Phoha. Surveillance coverage of sensor networks under a random mobility strategy. In *Proceedings of IEEE Sensors*, pages 961–965, 2003.
[12] A. Korostelev. On minimax rates of convergence in image models under sequential design. *Statistics and Probability Letters*, 43:369–375, 1999.
[13] T. D. Parsons. Pursuit-evasion in a graph. In Y. Alani and D. R. Lick, editors, *Theory and Application of Graphs*, pages 426–441. Springer-Verlag, 1976.
[14] M. Rahimi, R. Pon, W. J. Kaiser, G. S. Sukhatme, D. Estrin, and M. Srivastava. Adaptive sampling for environmental robotics. In *Proceedings of IEEE Int. Conf. on Robotics and Automation, ICRA*, 2004.
[15] R. Willett, A. Martin, and R. Nowak. Backcasting: Adaptive sampling for sensor networks. In *Proceedings of IPSN*, 2004.
[16] H. Yuan and C. Wu. An implicit 3d fully non-hydrostatic model for free-surface flows. *International J. for Numerical Methods in Fluids*, 46:709–733, 2004.

(a) Estimate generated by a boat moving along a uniform passive path. MSE $= 3 \times 10^{-4}$, latency $= 48$ hrs.



(b) Estimate generated after the coarse survey in a 2-step adaptive path approach.



(c) Cells denoting rough location of the boundary as indicated by the coarse estimate, superimposed on the noisy field.

(d) Final estimate generated after the refinement pass over the cells in 5(c). MSE $= 4 \times 10^{-4}$, latency $= 14$ hrs.

Figure 5: Comparison of passive and adaptive path planning approaches for water current velocity mapping in a freshwater lake. The adaptive strategy requires only 14 hrs for mapping the nearly 2 km× 1 km lake to a resolution of < 10 m, as opposed to 48 hrs using the passive method.