

Boundary Estimation in Sensor Networks: Theory and Methods

Robert Nowak* and Urbashi Mitra†

* Department of Electrical & Computer Engineering, Rice University
e-mail: nowak@rice.edu

† Electrical Engineering Department, University of Southern California
e-mail: ubli@usc.edu

Abstract

Sensor networks have emerged as a fundamentally new tool for monitoring spatially distributed phenomena. This paper investigates a strategy by which sensor nodes detect and estimate non-localized phenomena such as “boundaries” and “edges” (e.g., temperature gradients, variations in illumination or contamination levels). A general class of boundaries, with mild regularity assumptions, is considered, and theoretical bounds on the achievable performance of sensor network based boundary estimation are established. A hierarchical boundary estimation algorithm is proposed that achieves a near-optimal balance between mean-squared error and energy consumption.

1 Introduction

Sensor networks have emerged as a fundamentally new tool for monitoring inaccessible environments such as non-destructive evaluation of buildings and structures; contaminant tracking in the environment; habitat monitoring in the jungle; and surveillance in military zones. These *ad hoc* networks are envisioned to be a collection of embedded sensors, actuators and processors. We shall assume that communication between sensors is done in a wireless fashion. Sensor networks are distinguished from more classical networks due to strict limitations on energy consumption, the density of nodes, the simplicity of the processing power of nodes and possibly high environmental dynamics. An important problem in sensor networking applications is *boundary estimation*. Consider a network sensing a field composed of two or more regions of distinct behavior (e.g., differing

*Supported by the National Science Foundation, grant nos. MIP-9701692 and ANI-0099148, the Office of Naval Research, grant no. N00014-00-1-0390, and the Army Research Office, grant no. DAAD19-99-1-0290.

†Supported by the Texas Instruments Visiting Professorship.

mean values for the sensor measurements). An example of such a field is depicted in Figure 1(a). Boundary estimation is the process of determining the delineation between homogeneous regions.

There are two fundamental limitations in the boundary estimation problem. First, the accuracy of a boundary estimate is limited by the spatial density of sensors in the network and by the amount of noise associated with the measurement process. Second, energy constraints may limit the complexity of the boundary estimate that is ultimately transmitted to a desired destination. The trade-off between accuracy and energy consumption can be characterized as follows. Assume that n sensor nodes are arranged on an $\sqrt{n} \times \sqrt{n}$ square lattice (assuming a planar, square sensor field). Suppose that the field being sensed consists of two homogeneous regions separated by a one-dimensional boundary (like the case depicted in Figure 1(a)). A broad class of boundaries is considered in this paper. Specifically, we only assume that the boundary is a Lipschitz function [6, 3] or, more generally, has a box-counting dimension of one [9]. This class includes linear boundaries and other parametric curves, but also includes boundaries that cannot be described parametrically.

Each sensor node makes a (noisy) measurement of the field. Under these assumptions, there will be $O(\sqrt{n})$ nodes lying on the boundary. The boundary nodes provide a description of the boundary to within a resolution of $1/\sqrt{n}$. Noise present in the measurements limits the achievable accuracy of a boundary estimate. It is known that, under the assumptions on the class of boundaries above, the mean-square error (MSE) cannot, in general, decay faster than $O(1/\sqrt{n})$ [6, 3]. That is, no estimator (based on centralized or distributed processing) can exceed this convergence speed-limit. It is important to point out that if one restricts the class of boundaries, then faster decay rates are certainly possible. For example, if one assumes that the boundary is a line, then the problem is a parametric estimation problem and the rate of decay is $O(1/n)$. Assuming a line or parametric curve is, of course, very restrictive (and probably unreasonable for natural phenomena), and therefore this paper focuses on a much more general class of boundaries.

To quantify the total energy required to transmit a boundary estimate of this accuracy, note that each boundary node must send one message to the desired destination (indicating that it is on the boundary). Thus, the total energy required to transmit the boundary description is $O(\sqrt{n})$. Combining these results yields a fundamental trade-off between accuracy and energy of the form

$$\text{MSE} \sim \frac{1}{\text{Energy}}.$$

This tradeoff does not take into consideration the additional energy required to determine whether a sensor is in fact a boundary. It is important to note that this relation should not be interpreted to mean that a fixed number of sensor nodes using more energy can provide more accuracy. Rather, both the MSE and the energy consumption are functions of the number of sensor nodes, and the above relation indicates how the accuracy and energy consumption behave as the density of nodes increases. Also, note that if a boundary can be described

parametrically, then the energy required to transmit the description is proportional to the number of parameters, and does not depend on n . However, as discussed above, the aim here is to avoid such restrictive parametric assumptions. The boundaries of interest may not admit exact parametric descriptions, and therefore the accuracy of the boundary description and transmission cost both grow as density of nodes increases.

This paper explores the basic trade-off between MSE and energy consumption, as functions of node density. We propose and develop a boundary estimation algorithm based on multiscale partitioning methods. The algorithm is quite practical and maps nicely onto a sensor network architecture. Moreover, we demonstrate theoretically that our method nearly achieves the optimal MSE/Energy trade-off discussed above. The theory hinges on an application of our extension [5] of the Li-Barron bound for complexity regularized model selection [8] to bound the MSE and on a recent concentration inequality for chi-squared distributions to bound the expected energy consumption [7]. Since our method (nearly) achieves the optimal trade-off above, no other scheme can be devised that will (asymptotically) perform significantly better. Simulation experiments verify the predicted theoretical performance of our method.

1.1 Related Work

Due to the nascence of sensor network research, there is a limited literature concerning boundary estimation for such networks. At first glance, boundary estimation (or boundary detection) has goals that are similar to that of edge detection in image processing. However, a major distinction exists. Due to energy constraints, processing the entire “image” simultaneously is impractical, and hence a single node does not have access to all of the sensor measurements. In [2], several techniques based on averaging and thresholds are developed and compared for boundary detection. All of the techniques rely on the collection of measurements from sensor neighbors within a *probing radius*, R . The authors note that the performance of their methods will improve as the probing radius increases at the expense of communication cost. To contrast with our approach, we systematically increase the probing radius, however our communication cost does not increase as $O(R^2)$ due to the fact that lower dimensional statistics (versus all measurements) are passed to nodes within the sensor network hierarchy; and, furthermore, messages are only passed to clusterheads rather than all nodes.

The data collection algorithm in [4] shares many features with our proposed boundary estimation method. A hierarchical compression scheme is considered where clusterheads aggregate measurements from children nodes and then pass signal estimates to the next layer in the hierarchy. Our objective, herein, is to analytically determine the estimation capability of a tree-based boundary estimation scheme which is penalized by communication costs. We note that the scheme of [4] does not explicitly optimize the description of the phenomena being encoded (in our case, a boundary) and thus suffers in terms of the error between the estimated boundary and the true boundary; however, the communication

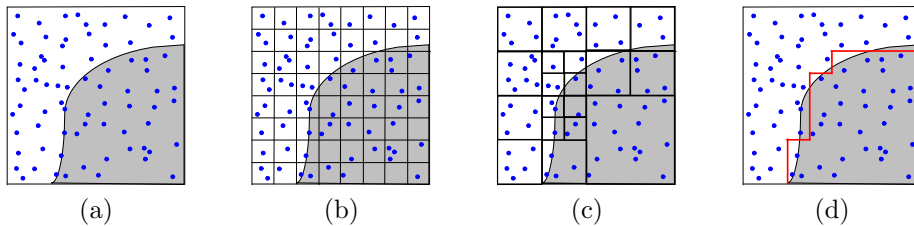


Figure 1: *Sensing an inhomogeneous field. (a) Points are sensor locations. The environment has two conditions indicated by the gray and white regions of the square. (b) the sensor network domain is partitioned into square cells. (c) Sensors within the network operate collaboratively to determine a pruned partition that matches the boundary. (d) Final approximation to the boundary between the two regions which is transmitted to a remote point.*

cost is lessened. With our scheme, we can systematically tradeoff between communication cost and reconstruction error by increasing the penalty associated with communication.

2 Problem Formulation and Approach

The basic problem is illustrated in Figure 1. Our objective is to consider measurements from a collection of sensors and determine the boundary between two fields of relatively homogeneous measurements.

We presume a hierarchical structure of “clusterheads” (see *e.g.* [4]) which manage measurements from nodes below them in the hierarchy. Thus, the nodes in each square of the partition communicate their measurements to a clusterhead in the square. Index the squares at the finest scale by row and column (i, j) . The clusterhead in square (i, j) computes the average of these measurements to obtain a value $x_{i,j} \sim \mathcal{N}\left(\mu_{i,j}, \frac{\sigma^2}{m_{i,j}}\right)$, where $\mu_{i,j}$ is the mean value, σ^2 is the noise variance for each sensor measurement, and $m_{i,j}$ is the number of nodes in square (i, j) . Thus we assume sensor measurements that have a Gaussian distribution. For simplicity, we assume $m_{i,j} = 1$. The random distribution is to account for noise in the system as well as for the small probability of node failure (outlier measurements).

Our approach to the boundary estimation problem is to devise a hierarchical processing strategy that enables the nodes to collaboratively determine a non-uniform rectangular partition of the sensor domain that is adapted to the boundaries. Specifically, the desired partition will have high, fine resolution along the boundary, and low, coarse resolution in homogeneous regions of the field, as depicted in Figure 1. The partition effectively provides a “staircase”-like approximation to the boundary. Similar strategies have been recently investigated to handle edges in images [3, 10] and decision boundaries in classification problems [9]. The advantage of our approach is that, under mild conditions

on the smoothness of the boundary curve, we can establish upper bounds on the MSE of the estimator using theoretical tools we have developed in previous work. These upper bounds can be used to tune the trade-off between data fitting and the complexity of the boundary estimate. The complexity of the boundary estimate relates directly to energy consumption in the network.

Our approach is as follows. Let us take the sensor domain to be the unit square $[0, 1]^2$. Partition the domain into n sub-squares of sidelength $\frac{1}{\sqrt{n}}$, as shown in Figure 1(b). The sidelength $\frac{1}{\sqrt{n}}$ is the finest resolution of our analysis. In principle, this initial partition can be generated by a *recursive dyadic partition* (RDP). First divide the domain into four sub-squares of equal size. Repeat this process again on each sub-square. Repeat this $1/2 \log_2 n = J$ times. This gives rise to a *complete* RDP of resolution $\frac{1}{\sqrt{n}}$ (the rectangular partition of the sensing domain shown above in Figure 1(b)). The RDP process can be represented with a quadtree structure. The quadtree can be pruned back to produce an RDP with non-uniform resolution as shown in Figure 1(c). The key issues are: (1) How to implement the pruning process in the sensor network; (2) How to determine the best pruned tree. Here, we discuss the first issue, and the second issue will be investigated in later sections of the paper.

Let \mathcal{P}_n denote the set of all RDPs, including the initial complete RDP and all possible prunings. For each RDP $P \in \mathcal{P}_n$, there is an associated quadtree structure (generally of non-uniform depth corresponding to the non-uniform resolution of most RDPs). The leaves of each quadtree represent dyadic (sidelength equal to a negative power of 2) square regions of the associated partition. For a given RDP and quadtree, each sensor node belongs to a certain dyadic square. We consider these squares “clusters” and assume that one of the nodes in each square serves as a “clusterhead,” which will assimilate information from the other nodes in the square. Notice that if one considers all RDPs in \mathcal{P}_n , then each sensor node actually belongs to a nested hierarchy of $1/2 \log_2 n$ dyadic squares of sidelengths $\frac{1}{\sqrt{n}}, \frac{2}{\sqrt{n}}, \frac{4}{\sqrt{n}}, \dots, 1$, respectively. Thus, we have a hierarchy of clusters and clusterheads.

Consider a certain RDP $P \in \mathcal{P}_n$. Define the estimator of the field as follows. On each square of the partition, average the measurements from the sensors in that square and set the estimate of the field to that average value. This results in a piecewise constant estimate, denoted by θ , of the field. This estimator will be compared with the data $x = \{x_{i,j}\}$. The data themselves are undesirable for two reasons. First, they are noisy and averaging over larger regions will reduce the noise. Second, the unprocessed data x will require the maximum amount of energy to transmit to the destination. Our empirical measure of performance is the sum-of-squared errors between $\theta = \theta(P)$ and the data $x = \{x_{i,j}\}$.

$$R(\theta, x) = \sum_{i,j=1}^{\sqrt{n}} (\theta(i, j) - x_{i,j})^2, \quad (1)$$

Define the complexity penalized estimator

$$\hat{\theta}_n = \arg \min_{\theta(P): P \in \mathcal{P}_n} R(\theta(P), x) + 2\sigma^2 p(n)|\theta(P)|, \quad (2)$$

where σ^2 is the noise variance, $|\theta(P)|$ denotes the total number of squares in the partition P , and $p(n)$ is a certain monotonically increasing function of n that discourages unnecessarily high resolution partitions (appropriate choices of $p(n)$ will be discussed in the sequel). It is well known that the optimization in (2) can be solved using a bottom-up tree pruning algorithm in $O(n)$ operations [1, 3, 10]. This is possible because both the sum-of-squared errors and the penalty are additive functions, and therefore the squared error plus penalty cost can be separated into terms associated with each individual square of the partition θ . The hierarchy of clusterheads facilitates this process in the sensor network. At each level of the hierarchy, the clusterhead receives the best sub-partition/sub-tree estimates from the four clusterheads below it, and compares the total cost of these estimates with the cost of the estimate equal to the average of all sensors in that cluster.

3 Upper Bounds on Achievable Accuracy

We begin by recalling a fundamental upper bound on expected error of complexity penalized estimators, like that in (2). This particular bound was originally developed for mixture density modeling [8], and we later extended it to more general settings [5]. Here we state a specialized version of the bound, tailored to the estimator proposed in (2).

Let Θ_n denote the set of all possible models of the field. This set contains piecewise constant models (constant on the dyadic squares corresponding to one of the partitions in \mathcal{P}_n). The constant values are in a prescribed range $[-R, R]$, and are quantized to k bits. The range corresponds to the upper and lower limits of the amplitude range of the sensors. The set Θ_n consists of a finite number of models (a bound on the number of partitions is derived in the Appendix). Assume that $p(n)$ satisfies the summability condition (Kraft inequality)

$$\sum_{\theta \in \Theta_n} e^{-p(n)|\theta|} \leq 1, \quad (3)$$

where again $|\theta|$ denotes the number of squares (alternatively we shall call this the number of leaves in the pruned tree description of the boundary) in the partition θ . It is shown in the Appendix that $p(n) \leq \gamma \log n$ satisfies (3). Let $\hat{\theta}_n$ denote the solution to

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta_n} R(\theta, x) + 2\sigma^2 p(n)|\theta|, \quad (4)$$

where, as before, x denotes the array of measurements at the finest scale $\{x_{i,j}\}$, and $|\theta|$ denotes the number of squares in the partition associated with θ . This

is essentially the same estimator as defined in (2) except that the values of the estimate are quantized in this case.

Let θ_n^* denote the true value of the field at resolution $1/\sqrt{n}$ (i.e., $\theta_n^*(i, j) = E[x_{i,j}]$). Then, applying Theorem 7 in [5], the MSE of the estimator $\hat{\theta}_n$ is bounded above according to

$$\frac{1}{n} \sum_{i,j=1}^{\sqrt{n}} E \left[\left(\hat{\theta}_n(i, j) - \theta_n^*(i, j) \right)^2 \right] \leq \min_{\theta \in \Theta_n} \frac{1}{n} \left\{ 2 \sum_{i,j=1}^{\sqrt{n}} (\theta(i, j) - \theta_n^*(i, j))^2 + 8\sigma^2 p(n) |\theta| \right\} \quad (5)$$

The upper bound involves two terms. The first term, $2 \sum_{i,j=1}^{\sqrt{n}} (\theta(i, j) - \theta_n^*(i, j))^2$, is a bound on the bias or approximation error. The second term, $8\sigma^2 p(n) |\theta|$, is a bound on the variance or estimation error. The bias term, which measures the squared error between the best possible model in our class and the true field, is generally unknown. However, if we make certain assumptions on the smoothness of the boundary, then the rate at which this term decays as function of the partition size $|\theta|$ can be determined.

Assume that the field being sensed is composed of homogeneous regions separated by a one-dimensional boundary. If the boundary is a Lipschitz function [3, 10] or more generally has a box-counting dimension (closely related to Hausdorff dimension) of 1, then by carefully calibrating quantization and penalization as discussed in the Appendix (taking $k \sim 1/4 \log n$ and setting $p(n) = 2/3 \log n$) it follows that

$$\frac{1}{n} \sum_{i,j=1}^{\sqrt{n}} E \left[\left(\hat{\theta}_n(i, j) - \theta_n^*(i, j) \right)^2 \right] \leq O \left(\sqrt{\frac{\log n}{n}} \right). \quad (6)$$

This result shows that the MSE decays to zero at a rate of $\sqrt{\log n/n}$. This rate cannot be significantly improved by any estimator. From [3, 6] we know that for Lipschitz boundaries, the minimax rate is $O(1/\sqrt{n})$, which shows that our estimator is within a square-root of a logarithmic factor of the best possible convergence rate. The minimax rate is the fastest rate of convergence achievable with any estimator (“min”) for the most challenging (“max”) Lipschitz boundary. Faster rates of decay are theoretically possible if one assumes that the boundary is even smoother. As an extreme case, suppose the boundary can be exactly described parametrically (e.g., a line). Then the boundary problem is one of parameter estimation and the rate of convergence is $O(1/n)$. Extensions of our approach are possible which can take advantage of smoother boundaries, which may provide convergence rates approaching the parametric rate. These extensions are part of our ongoing work and will be discussed in Section 6.

4 Accuracy-Energy Trade-off

A key characteristic of our proposed method is the explicit consideration of the cost of communication in the construction of the tree describing the boundary. Energy consumption is defined by two communication costs: the cost of communication due to the construction of the tree (*in-network cost*) and the cost of communicating the final boundary estimate (*out-of-network cost*). We will show that the expected number of leaves produced by our algorithm is $O(\sqrt{n})$, and that the in-network and out-of-network energy consumption is proportional to this number. Recall that the rate of decay for the MSE is $\text{MSE} \sim \sqrt{\log n/n}$. Therefore, ignoring the logarithmic factor, the accuracy-energy trade-off required to achieve this optimal MSE is roughly $\text{MSE} \sim 1/\text{Energy}$. Contrast this trade-off with that of a naive approach in which each of the n sensors transmits its data, directly or by multiple hops, to an external point. In this case, the in-network and out-of-network energy costs are $O(n)$, which lead to the trade-off $\text{MSE} \sim 1/\sqrt{\text{Energy}}$, since we know that no estimator exists that can result in an MSE decaying faster than $O(1/\sqrt{n})$. Thus, our proposed hierarchical boundary estimation method offers substantial savings over the naive approach while optimizing the tradeoff between accuracy and complexity of the estimate.

4.1 Out-of-network Communication Cost

It is clear that the out-of-network communication cost is proportional to the final description of the boundary, thus it is of interest to compute the expected size of the tree, or $E[|\hat{\theta}|]$. Each decision in the pruning process is based on comparing the complexity and fitness of an average value to the data in a certain dyadic square to that of the best subpartition model for that square (passed up from the bottom).

An upper bound on $E[|\hat{\theta}|]$ is derived in the Appendix. The upper bound is based on the probability of pruning or not pruning at each node for our hierarchical algorithm. If no boundary is present, then the probability of pruning at each node can be bounded from above by the tail probability of a certain chi-square distribution. The chi-square distribution arises from the assumed Gaussian observation model and the sum-of-squared errors criterion used in pruning. Using another upper bound for the tail probability, we show in the Appendix that if no boundary is present in the square under consideration, and with a penalty $p(n) = 2/3 \log n$, the probability of not pruning tends to zero as n increases. This implies that $E[|\hat{\theta}|] \rightarrow 1$ as $n \rightarrow \infty$. Thus, for large sensor networks, the expected number of leaves (partition pieces) in the case where there is no boundary (simply a homogeneous field) is one.

To consider the inhomogeneous case where a boundary does exist, if the boundary is a Lipschitz function or has a box counting dimension of 1, there exists a pruned RDP with at most $C'\sqrt{n}$ squares (leaves) that includes the $O(\sqrt{n})$ squares of sidelength $1/\sqrt{n}$ that the boundary passes through (see the Appendix for a fuller discussion of this property). Thus an upper bound on the number of leaves required to describe the boundary in the noiseless case is given by $C'\sqrt{n}$.

In the presence of noise, we can use the results above for the homogeneous case to bound the number of spurious leafs due to noise (zero as n grows); as a result, for large sensor networks, we can expect at most $C'\sqrt{n}$ leafs in total. Thus, the expected energy required to transmit the final boundary description is $\text{Energy} = O(\sqrt{n})$.

4.2 In-network Communication Cost

The in-network communication cost is intimately tied to the expected size of the final tree, as this value determines how much pruning will occur. We have seen above that the out-of-network cost is proportional to \sqrt{n} and herein we shall show that the in-network communication cost is also $O(\sqrt{n})$. At each scale $2^j/\sqrt{n}$, $j = 0, \dots, 1/2 \log_2 n - 1$, the hierarchical algorithm passes a certain number of data or averages, n_j , corresponding to the number of squares in the best partition (up to that scale), up the tree to the next scale. We assume that a constant number of bits k , is transmitted per measurement. These $k n_j$ bits must be transmitted approximately $2^j/\sqrt{n}$ meters (assuming the sensor domain is normalized to 1 square meter). Thus, the total in-network communication energy in bit-meters is:

$$\mathcal{E} = k \sum_{j=0}^{1/2 \log_2 n - 1} n_j 2^j / \sqrt{n}.$$

In the naive approach, $n_j = n$ for all j , and therefore $\mathcal{E} \approx kn$. In the hierarchical approach, first consider the case when there is no boundary. We have already seen that in such cases the tree will be pruned at each stage with high probability. Therefore, $n_j = n/4^j$ and $\mathcal{E} \approx 2k\sqrt{n}$. Now if a boundary of length $C\sqrt{n}$ is present, then $n_j \leq n/4^j + C\sqrt{n}$. This produces $\mathcal{E} \leq k(C+2)\sqrt{n}$. Thus, we see that our hierarchical algorithm results in $\mathcal{E} = O(\sqrt{n})$.

5 Simulations

We next present representative simulation results on the efficacy of the proposed boundary estimation algorithm. We considered a host of sensor network densities observing the same phenomenon. Sensor networks of size 4^k for $k = 2, \dots, 8$ distributed over a square meter were considered. The sensors operated in an environment with three different noise levels ($\sigma^2 = 1, 10, 100$). In Figure 2 (a), we see the mean-squared error (MSE) as a function of the network size (which relates directly to density). The MSE is averaged over 50 realizations of the noise. As predicted by the theoretical results, we see the expected decay in MSE. The in-network communication cost as scaled by the distance traveled is provided in Figure 2(b). As predicted, this cost is proportional to \sqrt{n} . Figure 2(c) shows the average size of the boundary estimate (number of leafs) as a function of the network size and a line fit to the data. This plot corresponds to the out-of-network communication costs. We see that the predicted bounds for both costs

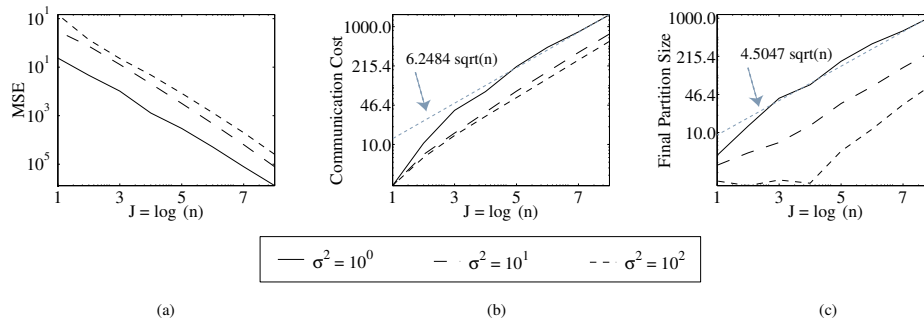


Figure 2: (a) Estimation accuracy as a function of the total number of nodes. (b) In-network communication cost as a function of the total number of nodes. (c) Out-of-network communication cost, $E[|\hat{\theta}|]$, as a function of the total number of nodes.

are in fact conservative, and in practice the constant in $O(\sqrt{n})$ is quite modest (here it is 4 – 6). The final partition size (and hence the communication cost) decreases as the noise variance increases due to the fact that the overall penalty is a function of the noise variance. Thus as the noise variance increases, it is more likely that pruning will occur.

Figure 3 shows single realizations of the boundary estimation process for three resolutions/sensor network densities. The penalty function employed was that derived in the Appendix and we see that the resultant boundary estimates offer the desired tradeoff between accuracy and energy consumption.

6 Conclusions and Ongoing Work

In this work, we have proposed a method for boundary estimation in sensor networks. The boundary estimate is determined via complexity regularization of a hierarchical tree-based estimation method. We demonstrated theoretically that our method nearly achieves the optimal trade-off $\text{MSE} \sim 1/\text{Energy}$, which shows that no other scheme can be devised that will (asymptotically) perform significantly better. Simulation experiments agreed very well with the theoretical predictions. In future work we plan to investigate more sophisticated boundary estimation techniques based on “wedgelets” [3] and “platelets” [10]. These methodologies are also based on hierarchical partitions and trees, but have additional flexibility which allows for a more parsimonious description of smooth boundaries and smooth variations in the mean of homogeneous regions. We are also currently incorporating the effects of imperfect wireless signaling into our theoretical framework and simulation studies. Finally, we are investigating the issue of tracking a slowly time-varying boundary.

Acknowledgments

The authors wish to thank Ms. Rebecca Willett for developing the simulation code for the proposed boundary estimator and for helpful comments on the manuscript, and thank Mr. Rui Castro for his careful reading of the proofs.

7 Appendix

7.1 Number of RDPs in \mathcal{P}

Recall the class \mathcal{P} of RDPs under consideration (all RDPs resulting from pruning P_J , the uniform partition of the unit square into n squares of sidelength $\frac{1}{\sqrt{n}}$). In order to ensure that the Kraft inequality (3) is satisfied, we need to determine how many RDPs there are in \mathcal{P} . More specifically, we will need to know how many partitions there are with exactly ℓ squares/leafs. Notice that since the RDP is based on recursive splits into four, the number of leafs in every partition in \mathcal{P} is of the form $\ell = 3m + 1$, for some integer $0 \leq m \leq (n-1)/3$. The integer m corresponds to the number of recursive splits. For each RDP having $3m + 1$ leafs there is a corresponding partially ordered sequence of m split points (at dyadic positions in the plane). In general, there are $\binom{n}{m} \equiv \frac{n!}{(n-m)!m!}$ possible selections of m points from n (n corresponding to the vertices of the finest resolution partition, P_J). This number is an upper bound on the number of partitions in \mathcal{P} with $\ell = 3m + 1$ leafs (since RDPs can only have dyadic split points).

7.2 Kraft Inequality

Here we show that with k (recall that k is the number of bits employed per transmission) and $p(n)$ properly calibrated, we have

$$\sum_{\theta \in \Theta_n} e^{-p(n)|\theta|} \leq 1 . \quad (7)$$

Let $\Theta_n^{(m)}$ denote the subset of Θ_n consisting of models based on $\ell = 3m + 1$ leaf partitions. Begin by writing

$$\begin{aligned} \sum_{\theta \in \Theta_n} e^{-p(n)|\theta|} &= \sum_{m=0}^{(n-1)/3} \sum_{\theta \in \Theta_n^{(m)}} e^{-(3m+1)p(n)} \\ &\leq \sum_{m=0}^{(n-1)/3} \binom{n}{m} (2^k)^{3m+1} e^{-(3m+1)p(n)} \\ &\leq \sum_{m=0}^{(n-1)/3} \frac{n^m}{m!} (2^k)^{3m+1} e^{-(3m+1)p(n)} \\ &= \sum_{m=0}^{(n-1)/3} \frac{1}{m!} e^{[m \log n + (3m+1) \log(2^k) - (3m+1)p(n)]} . \end{aligned}$$

If $A \equiv m \log n + (3m + 1) \log(2^k) - (3m + 1)p(n) < -1$ (then $e^A < e^{-1}$), then we have

$$\sum_{\theta \in \Theta_n} e^{-p(n)|\theta|} \leq 1/e \sum_{m=0}^{(n-1)/3} \frac{1}{m!} \leq 1.$$

To guarantee $A < -1$, we must have $p(n)$ growing at least like $\log n$. Therefore, set $p(n) = \gamma \log n$, for some $\gamma > 0$. Also, as we will see later in the next section, to guarantee that the quantization of our models is sufficiently fine to contribute a negligible amount to the overall error we must select $2^k \sim n^{1/4}$. With these calibrations we have

$$A = [(7/4 - 3\gamma)m + (1/4 - \gamma)] \log n$$

In order to guarantee that the MSE converges to zero, we will see in the next section that m must be a monotonically increasing function of n . Therefore, for n sufficiently large, the term involving $(\frac{1}{4} - \gamma)$ is negligible, and the condition $A < -1$ is satisfied by $\gamma > 7/12$. We take $\gamma = 2/3$ in practice.

7.3 Rate of MSE Decay

Consider a complete RDP with m^2 squares of sidelength $1/m$. It is known that if the boundary is a Lipschitz function, or more generally has a box counting dimension of 1, then the boundary passes through $\ell \leq Cm$ of the squares, for some constant $C > 0$ [3, 10, 9]. Furthermore, there exists a pruned RDP with at most $C'm$ leafs, where $C' = 8(C + 2)$, that includes the above ℓ squares of sidelength $1/m$ that contain the boundary [3, 9].

Now consider the upper bound (5), which as stated earlier follows as from an application of Theorem 7 in [5].

$$\begin{aligned} & \frac{1}{n} \sum_{i,j=1}^{\sqrt{n}} E \left[\left(\widehat{\theta}_n(i, j) - \theta_n^*(i, j) \right)^2 \right] \\ & \leq \min_{\theta \in \Theta_n} \frac{1}{n} \left\{ 2 \sum_{i,j=1}^{\sqrt{n}} (\theta(i, j) - \theta_n^*(i, j))^2 + 8p(n)|\theta| \right\} \\ & \leq 2 \int_{[0,1]^2} (\theta - \theta^*)^2 + 8 \frac{7}{12} \frac{\log n}{n} C' m, \end{aligned}$$

where the discretized squared error is bounded by the corresponding continuous counterpart. The squared error $\int_{[0,1]^2} (\theta - \theta^*)^2 \sim \frac{K_1}{m} + \frac{K_2}{\sqrt{n}}$, where the first term is due to the error between the $1/m$ resolution partition along the boundary, and the $1/\sqrt{n}$ term is due to the quantization error overall. Thus, the MSE behaves like

$$\text{MSE} \sim O(1/m) + O(1/\sqrt{n}) + O\left(m \frac{\log n}{n}\right).$$

Taking $m \sim \sqrt{\frac{n}{\log n}}$ produces the desired result: $\text{MSE} \sim O(\sqrt{\log n/n})$.

7.4 Expected Tree Size for Homogeneous Field

We construct an upperbound for $E[|\hat{\theta}|]$ under the assumption of a homogeneous field with no boundary. Let P denote the tree-structured partition associated with $\hat{\theta}$. Note that because P is an RDP it can have $d + 1$ leafs (pieces in the partition), where $d = 3m$, $m = 0, \dots, (n-1)/3$. Therefore, the expected number of leafs is given by

$$E[|\hat{\theta}|] = \sum_{m=0}^{(n-1)/3} (3m + 1) \Pr(|\hat{\theta}| = 3m + 1).$$

The probability $\Pr(|\hat{\theta}| = 3m + 1)$ can be bounded from above by the probability that one of the possible partitions with $3m + 1$ leafs, $m > 0$, is chosen in favor of the trivial partition with just a single leaf. That is, the event that one of the partitions with $3m + 1$ leafs is selected implies that partitions of all other sizes were not selected, including the trivial partition, from which the upper bound follows. This upper bound allows us to bound the expected number of leafs as follows.

$$E[|\hat{\theta}|] \leq \sum_{m=0}^{(n-1)/3} (3m + 1) \#_m p_m,$$

where $\#_m$ denotes the number of different $(3m + 1)$ -leaf partitions, and p_m denotes the probability that a particular $(3m + 1)$ -leaf partition is chosen in favor of the trivial partition (under the homogeneous assumption). The number $\#_m$ can be bounded above by $\binom{n}{m}$, just as in the verification of the Kraft inequality.

The probability p_m can be bounded as follows. Note this is the probability of a particular outcome of a comparison of two models. The comparison is made between their respective sum-of-squared errors plus complexity penalty, as given by (2). The single leaf model has a single degree of freedom (mean value of the entire region), and the alternate model, based on the $(3m + 1)$ -leaf has $3m + 1$ degrees of freedom. Thus, under the assumption that the data are i.i.d. zero-mean Gaussian distributed with variance σ^2 , it is easy to verify that the difference between the sum-of-squared errors of the models (single-leaf model sum-of-squares minus $(3m + 1)$ -leaf model sum-of-squares) is distributed as $\sigma^2 W_{3m}$, where W_{3m} is a chi-square distributed random variable with $3m$ degrees of freedom (precisely the difference between the degrees of freedom in the two models). This follows from the fact that the difference of the sum-of-squared errors is equal to the sum-of-squares of an orthogonal projection of the data onto a $3m$ dimensional subspace.

The single-leaf model is rejected if $\sigma^2 W_{3m}$ is greater than the difference between the complexity penalties associated with the two models; that is, if

$$\sigma^2 W_{3m} > (3m + 1)2\sigma^2 p(n) - 2\sigma^2 p(n) = 6m\sigma^2 p(n),$$

where $2\sigma^2 p(n)$ is the penalty associated with each additional leaf in P . According to the MSE analysis in the previous section, we require $p(n) = \gamma \log n$, with

$\gamma > 7/12$. To be concrete, take $\gamma = 2/3$, in which case the rejection of the single-leaf model is equivalent to $W_{3m} > 4m \log n$. The probability of this condition, $p_m = Pr(W_{3m} > 4m \log n)$, is bounded from above using Lemma 1 of Laurent and Massart [7]: If W_d is chi-square distributed with d degrees of freedom, then for $s > 0$

$$Pr(W_d \geq d + s\sqrt{2d} + s^2) \leq e^{-s^2/2}.$$

Making the identification $d + s\sqrt{2d} + s^2 = 4m \log n$ produces the bound

$$p_m = Pr(W_{3m} > 4m \log n) \leq e^{-2m \log n + m\sqrt{3/2(4 \log n - 3/2)}}.$$

Combining the upper bounds above, we have

$$\begin{aligned} E[|\hat{\theta}|] &\leq \sum_{m=0}^{(n-1)/3} (3m+1) \binom{n}{m} e^{-2m \log n + m\sqrt{3/2(4 \log n - 3/2)}}, \\ &= \sum_{m=0}^{(n-1)/3} (3m+1) \binom{n}{m} n^{-m} e^{-m \log n + m\sqrt{3/2(4 \log n - 3/2)}}. \end{aligned}$$

For $n \geq 270$ the exponent $-\log n + \sqrt{3/2(4 \log n - 3/2)} < 0$ and therefore

$$\begin{aligned} E[|\hat{\theta}|] &\leq \sum_{m=0}^{(n-1)/3} (3m+1) \binom{n}{m} n^{-m}, \\ &\leq \sum_{m=0}^{(n-1)/3} (3m+1) \frac{n^m}{m!} n^{-m}, \\ &\leq \sum_{m=0}^{(n-1)/3} (3m+1)/m! < 11. \end{aligned}$$

Furthermore, note that as $n \rightarrow \infty$ the exponent $-\log n + \sqrt{3/2(4 \log n - 3/2)} \rightarrow -\infty$. This fact implies that the factor $e^{-m \log n + m\sqrt{3/2(4 \log n - 3/2)}}$ tends to zero when $m > 0$. Therefore, the expected number of leafs $E[|\hat{\theta}|] \rightarrow 1$ as $n \rightarrow \infty$.

References

- [1] L. Breiman, J. Friedman, R. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth, Belmont, CA, 1983.
- [2] K. Chintalapudi and R. Govindan. Localized edge detection in sensor fields. *University of Southern California, Computer Science Department, Technical Report, 02-773*, 2002. available at <http://www.cs.usc.edu/tech-reports/technical-reports.html>.
- [3] D. Donoho. Wedgelets: Nearly minimax estimation of edges. *Ann. Statist.*, 27:859 – 897, 1999.

- [4] D. Ganesan, D. Estrin, and J. Heideman. DIMENSIONS: Why do we need a new data handling architecture for sensor networks? In *Proceedings of IEEE/ACM HotNets-I*, Princeton, NJ, October 2002.
- [5] E. Kolaczyk and R. Nowak. Multiscale likelihood analysis and complexity penalized estimation. *Annals of Statistics* (tentatively accepted for publication). Also available at www.ece.rice.edu/~nowak/pubs.html, 2002.
- [6] A. P. Korostelev and A. B. Tsybakov. *Minimax theory of image reconstruction*. Springer-Verlag, New York, 1993.
- [7] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, (5), October 2000.
- [8] Q. Li and A. Barron. Mixture density estimation. In S.A. Solla, T.K. Leen, and K.-R. Müller, editors, *Advances in Neural Information Processing Systems 12*. MIT Press, 2000.
- [9] C. Scott and R. Nowak. Dyadic classification trees via structural risk minimization. In *Proc. Neural Information Processing Systems (NIPS)*, Vancouver, CA, Dec. 2002.
- [10] R. Willett and R. Nowak. Platelets: A multiscale approach to recovering edges and surfaces in photon-limited imaging. *IEEE Trans. Med. Imaging*, to appear in the Special Issue on Wavelets in Medical Imaging, 2003.

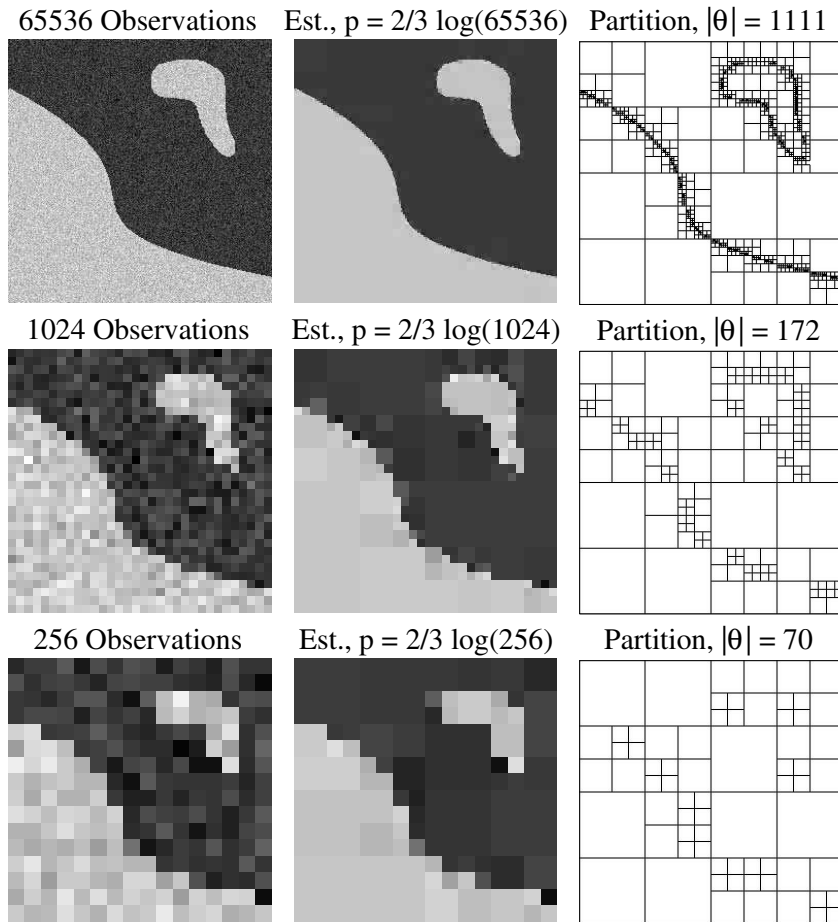


Figure 3: *Effect of sensor network density (resolution) on boundary estimation. Column 1 is the noisy set of measurements, Column 2 is the estimated boundary, and Column 3 is the associated partition.*