University of California
Los Angeles

Sensor Placement in Linear Fields: Estimation and Discrimination

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Electrical Engineering

by

Mohamed Nabil Hassan Hajjchehade

2007

The thesis of Mohamed Nabil Hassan Hajjchehade is approved.

William Kaiser

Mark Hansen

Gregory J. Pottie, Committee Chair

University of California, Los Angeles

2007

To Hassan, Alia, Manar, Samer, Hiba and Bassel.

Table of Contents

1 Introduction **1**

1.1 Notation 3

1.1.1 Variables 3

1.1.2 Means and Covariance Matrices 3

1.1.3 Probability Density Functions 3

1.2 Statistical Models 4

1.2.1 The Regression Framework 4

1.2.2 The Gaussian Process Framework 8

2 Related Work **11**

3 Experiment Design for Model Estimation **13**

3.1 Estimation of a Single Model 13

3.1.1 Sensor Placement in the Regression Framework 14

3.1.2 Optimal Sensor Placement in the Gaussian Process Framework 17

3.1.3 Interpretation in terms of Entropies 18

3.2 Estimation of Multiple Models 18

3.2.1 Using the Mean 19

3.2.2 Using Constraints 20

4 Experimental Design for Discrimination between Models . . . **22**

4.1 Discrimination between Two Models 23

4.2 Discrimination between Multiple Models 28

5 Conclusions and Future Work 31

References 34

List of Figures

3.1 Comparison between Random Designs and $D - optimal$ Designs. The Random Design curve is the average of 1000 simulations. . . 16

4.1 Selected Locations for Discrimination between the models. 7 Locations were selected in 8 Iterations 27

4.2 Selected Locations for Discrimination between the models. 7 Locations were selected in 8000 Iterations 27

4.3 Selected Locations for Discrimination between three models. 8 Locations were selected in 8 Iterations 30

4.4 Selected Locations for Discrimination between three models. 11 Locations were selected in 8000 Iterations 30

Acknowledgments

I would like to thank my advisor Professor Greg Pottie for his support, guidance and suggestions. He introduced me to sensor networks and to the modeling problem in sensor networks. I also thank Professor Mark Hansen and Professor William Kaiser for serving on my thesis committee. Part of the work on sensor placement for discriminating between two models presented was a project I did for the Design, Modeling and Analysis for Embedded Sensing class taught in the Winter 2007 quarter by Professor Mark Hansen. His help and suggestions were fundamental for the work presented here. I would like to thank my friends Laura Balzano, Raghavendra Prabhu and Roja Bandari for the very helpful discussions and suggestions.

Abstract of the Thesis

Sensor Placement in Linear Fields: Estimation and Discrimination

by

Mohamed Nabil Hassan Hajjchehade

Master of Science in Electrical Engineering

University of California, Los Angeles, 2007

Professor Gregory J. Pottie, Chair

Many applications in sensor networks require the estimation of spatial environmental phenomena. Finding strategies for sensor placement and data collection to optimize the estimation is a fundamental task for these applications. In this thesis we present two optimization formulations for obtaining good strategies. The first formulation considers the estimation of a spatial field of known model structure using $D-optimal$ design and is based on minimizing the estimation error. This formulation is extended to estimating multiple fields with different model structures simultaneously. The locations for this formulation are found before collecting any data. The second formulation considers discriminating between multiple competing model structures for the field using $T-optimal$ designs introduced in [1]. This formulation is based on minimizing the probability of error in selecting the correct model. The locations for this formulation are found in a sequential algorithm that provides a strategy for sensor placement and data collection. We focus on linear fields, and we provide simulations that describe these solutions and show their benefits.

CHAPTER 1

Introduction

Wireless sensing systems are very useful for applications in which we need to learn about natural phenomena (e.g. temperature, humidity, light, etc.) over spatial and temporal fields. Statistical models are widely used to make sense of the collected data and to answer questions regarding these phenomena such as prediction of the phenomena changes over time or space, effects of some phenomena on other phenomena, or any sort of inference.

One can randomly place sensors, collect measurements and make inferences about the phenomena in question. This work instead focuses on the problem of directing the initial placement of sensors and the collection of measurements, depending on the application and the desired final inferences. We are interested in many aspects of the sensor placement problem. The use or choice of certain statistical models, combined with the desired inferences, leads us to formulate methods or techniques which direct sensor placement for high quality inference results. The methods presented are optimization methods where the resulting placement is optimal in achieving the desired inference. In this thesis we focus on spatial data, linear regression models to represent these data and inferences based on the regression models. We also present some generalizations that lead to the Gaussian process model and the corresponding sensor placement methods.

Justifying the use of certain statistical models is an important open problem, but it is beyond the scope and the goal of this thesis. Nonetheless we stress the

relationship between commonly used statistical models. We refer to a model as a characterization of how unknown quantities (phenomena) are related to known quantities (measurement locations). Two characterizations are widely used: a function where the phenomena are a function of the locations and a probability distribution of the phenomena given the locations. Linear regression belongs to the first type while Gaussian processes fall in the second one. Throughout the analysis in this thesis, we focus on spatial field and we assume that the field does not change rapidly over time. We intend to study and understand the design of sensor placement to optimize the modeling of the spatial field. In future work, we plan to incorporate temporal changes of the field in the design of the placement and the modeling of the spatiotemporal field. We restrict the analysis as well to static nodes.

After describing the notation used in this thesis, we review in Section 1.2 the two frameworks and elaborate on the relationship between them.

The contributions of this thesis are as follows. In Chapter 3, we present the problem of finding sensor placement in order to minimize the error in estimating a linear spatial field using optimal experimental design techniques. We focus on $D-optimal$ design, and we show that it minimizes the entropy of the field given the measurements. We show as well the connection between the Gaussian process framework, used in [4], and the regression framework. We describe three optimization approaches for placing sensors to estimate multiple fields simultaneously. These three approaches can also be used to find placements for robust model estimation when there is uncertainty in the model describing the field. In Chapter 4, we describe a strategy to place sensors and collect data to select a model for the field from a set of models. This strategy ensures a minimum probability of error in the model selection.

1.1 Notation

1.1.1 Variables

Throughout the thesis we use four types of variables: scalar, vector, matrix and random. We employ the following notation: lower case for both scalars and vectors, upper case for matrices and bold for random variables. For example $\boldsymbol{x}$ is a random variable, scalar or vector, and x is a realization of $\boldsymbol{x}$. Similarly, $\boldsymbol{X}$ is a matrix with random entries while X is a realization.

1.1.2 Means and Covariance Matrices

We use $\mu_{\boldsymbol{x}}$ to denote the mean of a random variable $\boldsymbol{x}$. The index refers to the random variable. We use the upper case letter $C_{\boldsymbol{x}}$ to denote a covariance matrix for a random variable $\boldsymbol{x}$. Similarly $C_{\boldsymbol{xy}}$ is the joint covariance matrix of $\boldsymbol{x}$ and $\boldsymbol{y}$ i.e.

$$C_{\boldsymbol{y}} = E\left[(\boldsymbol{x} - \mu_{\boldsymbol{x}})(\boldsymbol{y} - \mu_{\boldsymbol{y}})^T\right] \tag{1.1}$$

and $C_{\boldsymbol{x}|\boldsymbol{y}}$ is the covariance matrix of $\boldsymbol{x}$ given $\boldsymbol{y} = y$ i.e.

$$C_{\boldsymbol{x}|\boldsymbol{y}} = E\left[(\boldsymbol{x} - \mu_{\boldsymbol{x}})(\boldsymbol{x} - \mu_{\boldsymbol{x}})^T|\boldsymbol{y} = y\right] \tag{1.2}$$

1.1.3 Probability Density Functions

We use $f_{\boldsymbol{x}}(x)$ to denote the probability density function (PDF) of the random variable $\boldsymbol{x}$. Similarly $f_{\boldsymbol{xy}}(x, y)$ denotes the joint PDF of $\boldsymbol{x}$ and $\boldsymbol{y}$ and $f_{\boldsymbol{x}|\boldsymbol{y}}(x|y)$ is the PDF of $\boldsymbol{x}$ given $\boldsymbol{y} = y$.

For Gaussian random variables, we use the familiar notation

$$\boldsymbol{x} \sim \mathcal{N}(\mu_{\boldsymbol{x}}, C_{\boldsymbol{x}}) \tag{1.3}$$

to mean that $\boldsymbol{x}$ is a Gaussian random variable with mean $\mu_{\boldsymbol{x}}$ and covariance matrix $C_{\boldsymbol{x}}$.

1.2 Statistical Models

There is a connection between the Gaussian process framework and linear regression in Gaussian noise framework. Both formulations have been extensively studied and widely used to represent data. This connection results from the relationship between the least squares criterion and Maximum Likelihood (ML) estimation under normal errors model. We think that understanding the connection is important for the problem of sensor placement, since it relates different approaches used to solve this problem. In Section 1.2.1 we review the linear regression framework under Gaussian noise. In Section 1.2.2 we review the Gaussian processes framework.

1.2.1 The Regression Framework

Consider the 1-dimensional polynomial regression framework where a polynomial model is used to fit the data and make inferences:

$$\boldsymbol{y}_i = a_0 + a_1 x_i + \cdots + a_n x_i^n + \boldsymbol{e}_i, \quad i = 1, \ldots, m \tag{1.4}$$

where x_i is the measurement location i, $\boldsymbol{e}_i$ is measurement noise at the measurement location i and a_i are unknown parameters. Letting ^{T} represent the transpose operator, we can rewrite (1.4) in vector form

$$\boldsymbol{y}_i = a^T v_i + \boldsymbol{e}_i, \quad i = 1, \ldots, m \tag{1.5}$$

where $a = [\ a_0 \quad a_1 \quad \ldots \quad a_n\ ]^T$ and $v_i = [\ 1 \quad x_i \quad \ldots \quad x_i^n\ ]^T$. We assume that $v_1, \ldots, v_m$ span $\mathbb{R}^{n+1}$. Furthermore if we collect $\boldsymbol{y}_i$ in an m-dimensional vector

called, $\boldsymbol{y}$ we get:

$$\boldsymbol{y} = X^T a + \boldsymbol{e}, \tag{1.6}$$

where X is the matrix whose columns are the v_i vectors. X is called the *design matrix*. We assume that $\boldsymbol{e} \sim \mathcal{N}(0, C_{\boldsymbol{e}})$.

The maximum likelihood estimate of a is found by:

$$\max_a \quad f_{\boldsymbol{y}}(y|a) \tag{1.7}$$

where $f_{\boldsymbol{y}}(y|a)$ is the density function of $\boldsymbol{y}$ parameterized by a. In this Gaussian setup, $f_{\boldsymbol{y}}(y|a)$ is a parameterized Gaussian density function of the form

$$f_{\boldsymbol{y}}(y|a) = \frac{1}{(2\pi)^{n/2}|C_{\boldsymbol{y}}|^{1/2}} exp\left(-\frac{1}{2}(y-\mu)^T C_{\boldsymbol{y}}^{-1}(y-\mu)\right) \tag{1.8}$$

where $\mu = X^T a$ and $C_{\boldsymbol{y}} = C_{\boldsymbol{e}}$.

The ML problem (1.7) can be solved by solving the following least-squares problem

$$\min_a \quad \|(y - X^T a)^T C_{\boldsymbol{e}}^{-1}(y - X^T a)\|_2^2 \tag{1.9}$$

The solution is given by:

$$\begin{aligned} \hat{\boldsymbol{a}} &= \left(\textstyle\sum_{i=1}^m v_i C_{\boldsymbol{e}}^{-1} v_i^T\right)^{-1} \textstyle\sum_{i=1}^m v_i C_{\boldsymbol{e}}^{-1} \boldsymbol{y}_i \\ &= (X C_{\boldsymbol{e}}^{-1} X^T)^{-1} X C_{\boldsymbol{e}}^{-1} \boldsymbol{y} \end{aligned} \tag{1.10}$$

The resulting estimation error, $\tilde{\boldsymbol{a}} = \hat{\boldsymbol{a}} - a$, is a Gaussian vector with mean zero and covariance matrix

$$\begin{aligned} C_{\tilde{\boldsymbol{a}}} &= \left(\textstyle\sum_{i=1}^m v_i C_{\boldsymbol{e}}^{-1} v_i^T\right)^{-1} \\ &= (X C_{\boldsymbol{e}}^{-1} X^T)^{-1} \end{aligned} \tag{1.11}$$

In the discussion above, we assumed implicitly that the regression parameters are modeled as a deterministic vector a i.e. unknown constants. Now we consider the case when we model these regression parameters as a random vector $\boldsymbol{a} \sim \mathcal{N}(\mu_{\boldsymbol{a}}, C_{\boldsymbol{a}})$. The regression equation becomes

$$\boldsymbol{y} = X^T\boldsymbol{a} + \boldsymbol{e} \tag{1.12}$$

and $\boldsymbol{y} \sim \mathcal{N}(X^T\mu_{\boldsymbol{a}}, X^T C_{\boldsymbol{a}} X + C_{\boldsymbol{e}})$.

In this case the maximum likelihood estimation of $\boldsymbol{a}$ given a realization (or measurements) y of $\boldsymbol{y}$ is given by maximizing the joint distribution of $\boldsymbol{a}$ and $\boldsymbol{y}$ when $\boldsymbol{y}$ is equal to the measurement y:

$$\max_a \quad f_{\boldsymbol{y},\boldsymbol{a}}(y, a) \tag{1.13}$$

Using Bayes' theorem, $f_{\boldsymbol{y},\boldsymbol{a}}(y, a) = f_{\boldsymbol{a}|\boldsymbol{y}}(a|y)\dot{f}_{\boldsymbol{y}}(y)$, we see that solving (1.13) is equivalent to solving what is known as the Maximum A Posteriori (MAP) problem

$$\max_a \quad f_{\boldsymbol{a}|\boldsymbol{y}}(a|y) = \frac{1}{(2\pi)^{n/2}|C_{\boldsymbol{a}|\boldsymbol{y}}|^{1/2}} exp\left(-\frac{1}{2}(a-\mu_{\boldsymbol{a}|\boldsymbol{y}})^T C^{-1}_{\boldsymbol{a}|\boldsymbol{y}}(a-\mu_{\boldsymbol{a}|\boldsymbol{y}})\right) \tag{1.14}$$

where [6]

$$\begin{aligned} \mu_{\boldsymbol{a}|\boldsymbol{y}} &= \mu_{\boldsymbol{a}} + C_{\boldsymbol{a}\boldsymbol{y}} C_{\boldsymbol{y}}^{-1}(y - \mu_{\boldsymbol{y}}) \\ &=^1 \mu_{\boldsymbol{a}} + C_{\boldsymbol{a}} X (X^T C_{\boldsymbol{a}} X + C_{\boldsymbol{e}})^{-1}(y - X^T\mu_{\boldsymbol{a}}) \\ &=^2 \mu_{\boldsymbol{a}} + (C^{-1}_{\boldsymbol{a}} + X C^{-1}_{\boldsymbol{e}} X^T)^{-1} X C^{-1}_{\boldsymbol{e}}(y - X^T\mu_{\boldsymbol{a}}) \end{aligned} \tag{1.15}$$

and

$$\begin{aligned} C_{\boldsymbol{a}|\boldsymbol{y}} &= C_{\boldsymbol{a}} - C_{\boldsymbol{a}\boldsymbol{y}} C^{-1}_{\boldsymbol{y}} C_{\boldsymbol{y}\boldsymbol{a}} \\ &=^1 C_{\boldsymbol{a}} - C_{\boldsymbol{a}} X (X^T C_{\boldsymbol{a}} X + C_{\boldsymbol{e}})^{-1} X^T C_{\boldsymbol{a}} \\ &=^2 (C^{-1}_{\boldsymbol{a}} + X C^{-1}_{\boldsymbol{e}} X^T)^{-1} \end{aligned} \tag{1.16}$$

[1] $C_{\boldsymbol{a}\boldsymbol{y}} = C^T_{\boldsymbol{y}\boldsymbol{a}} = C_{\boldsymbol{a}} X$ and $C_{\boldsymbol{y}} = X^T C_{\boldsymbol{a}} X + C_{\boldsymbol{e}}$.

[2] Using the Sherman–Morrison–Woodbury formula.

Since $f_{\boldsymbol{a}|\boldsymbol{y}}(a|y)$ is a Gaussian distribution, the solution, $\hat{\boldsymbol{a}}$, of (1.14) is given by $\mu_{\boldsymbol{a}|\boldsymbol{y}}$. The estimation error, $\tilde{\boldsymbol{a}} = \boldsymbol{a} - \hat{\boldsymbol{a}}$, is a Gaussian random variable with mean *zero* and covariance matrix $C_{\tilde{\boldsymbol{a}}} = C_{\boldsymbol{a}|\boldsymbol{y}}$.

For the sake of completeness, we would like to add that the MAP problem, (1.14), is equivalent to the Minimum Mean Square Error ($MMSE$) problem

$$\min_{a} \quad E(\tilde{a}\tilde{a}^T) \tag{1.17}$$

where the $MMSE$ is equal to $C_{\boldsymbol{a}|\boldsymbol{y}}$ given in (1.16), and is achieved by $\mu_{\boldsymbol{a}|\boldsymbol{y}}$ given in (1.15).

In this section we presented two views of the regression framework, the deterministic view where we consider the regression parameter as an unknown constant a and the random view where we consider the regression parameter as a random variable $\boldsymbol{a}$. In both views, the goal is to estimate the regression parameter by $\hat{\boldsymbol{a}}$ in order to estimate the phenomena by $X^T\hat{\boldsymbol{a}}$. Next we relate the two views and then in Section 1.2.2 we relate the Gaussian process framework to the regression framework with random parameters.

Examining the equations of the random view (1.15) and (1.16), we can loosely argue that if we assume that $\mu_{\boldsymbol{a}} = 0$ and let $C_{\boldsymbol{a}}$ go to $'\infty'$ then $C_{\boldsymbol{a}}^{-1}$ would loosely go to $'zero'$ and we will get the following limits

$$\begin{aligned} \mu_{\boldsymbol{a}|\boldsymbol{y}} &\longrightarrow \quad (XC_{\boldsymbol{e}}^{-1}X^T)^{-1}XC_{\boldsymbol{e}}^{-1}y \\ C_{\boldsymbol{a}|\boldsymbol{y}} &\longrightarrow \qquad (XC_{\boldsymbol{e}}^{-1}X^T)^{-1} \end{aligned} \tag{1.18}$$

which coincide with the equations of the deterministic view (1.10) and (1.11) respectively. In other words, increasing $C_{\boldsymbol{a}}$ corresponds to making the distribution of $\boldsymbol{a}$ wider and to increasing the uncertainty in $\boldsymbol{a}$. So letting $C_{\boldsymbol{a}}$ go to $'\infty'$ makes the Gaussian distribution flat and $\boldsymbol{a}$ an unknown constant a. We can also interpret this relationship from the Bayesian perspective; modeling the regression

parameter as a random variable $\boldsymbol{a}$ is a way of incorporating our prior information (or beliefs) about this parameter.

1.2.2 The Gaussian Process Framework

In this section we review the Gaussian process framework in modeling spatial fields (or phenomena). In this framework a Gaussian random variable is associated with the phenomenon at each location and for any set of locations a Gaussian random vector is used to model the phenomenon. The measurement vector will be modeled as follows

$$\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{e} \tag{1.19}$$

with $\boldsymbol{g} \sim \mathcal{N}(\mu_{\boldsymbol{g}}, C_{\boldsymbol{g}})$ and $\boldsymbol{e} \sim \mathcal{N}(0, C_{\boldsymbol{e}})$.

The measurement vector $\boldsymbol{y}$ is a Gaussian random vector with mean $\mu_{\boldsymbol{g}}$ and covariance $C_{\boldsymbol{g}} + C_{\boldsymbol{e}}$. Now in this framework, the maximum likelihood estimation of the phenomenon $\boldsymbol{g}$ is given by maximizing the joint distribution of $\boldsymbol{g}$ and $\boldsymbol{y}$ when $\boldsymbol{y}$ is equal to its realization y

$$\max_{g} \quad f_{\boldsymbol{y},\boldsymbol{g}}(y, g) \tag{1.20}$$

In a similar manner as in the regression framework, (1.20) is equivalent to

$$\max_{g} \quad f_{\boldsymbol{g}|\boldsymbol{y}}(g|y) \tag{1.21}$$

where $f_{\boldsymbol{g}|\boldsymbol{y}}(g|y)$ is a Gaussian distribution with mean [6]

$$\begin{aligned} \mu_{\boldsymbol{g}|\boldsymbol{y}} &= \mu_{\boldsymbol{g}} + C_{\boldsymbol{g}\boldsymbol{y}} C_{\boldsymbol{y}}^{-1}(y - \mu_{\boldsymbol{y}}) \\ &=^{1} \mu_{\boldsymbol{g}} + C_{\boldsymbol{g}}(C_{\boldsymbol{g}} + C_{\boldsymbol{e}})^{-1}(y - \mu_{\boldsymbol{g}}) \end{aligned} \tag{1.22}$$

and covariance matrix

$$\begin{aligned} C_{\boldsymbol{g}|\boldsymbol{y}} &= C_{\boldsymbol{g}} - C_{\boldsymbol{g}\boldsymbol{y}} C_{\boldsymbol{y}}^{-1} C_{\boldsymbol{y}\boldsymbol{g}} \\ &=^{1} C_{\boldsymbol{g}} - C_{\boldsymbol{g}}(C_{\boldsymbol{g}} + C_{\boldsymbol{e}})^{-1} C_{\boldsymbol{g}} \\ &=^{2} (C_{\boldsymbol{g}}^{-1} + C_{\boldsymbol{e}}^{-1})^{-1} \end{aligned} \tag{1.23}$$

If we collect a set of measurements y at certain locations $\{x_1, x_2, \dots\}$, (1.22) would define the maximum likelihood estimate of the phenomena $\boldsymbol{g}$ and (1.23) would define the uncertainty in this estimate. The computation of (1.22) and (1.23) requires the knowledge of $\mu_{\boldsymbol{g}}$ and $C_{\boldsymbol{g}}$. Usually in this framework, the mean $\mu_{\boldsymbol{g}}$ is estimated from the data as a sample mean, and the covariance matrix $C_{\boldsymbol{g}}$ is estimated using a symmetric positive definite kernel function $K(.,.)$, i.e. the ij-th entry of $C_{\boldsymbol{g}}$ is $K(x_i, x_j)$. Often it is assumed that the covariance between any two locations is only a function of their distance. A widely used kernel function is $K(x_1, x_2) = exp\left(-\frac{\|x_1 - x_2\|_2^2}{h^2}\right)$, as in [4], where h is a scaling factor.

We consider the assumption of a specific covariance matrix (or kernel function) in the Gaussian process framework to be equivalent to the assumption of a specific model in the regression framework. Indeed, we can see that if we assume that $C_{\boldsymbol{g}} = X^T C_{\boldsymbol{a}} X$, then we can relate $C_{\boldsymbol{g}|\boldsymbol{y}}$ to $C_{X^T\boldsymbol{a}|\boldsymbol{y}}$ of (1.16) as follows

$$\begin{aligned} C_{\boldsymbol{g}|\boldsymbol{y}} &= \left((X^T C_{\boldsymbol{a}} X)^{-1} + C_{\boldsymbol{e}}^{-1}\right)^{-1} \\ &= X^T C_{\boldsymbol{a}} X - X^T C_{\boldsymbol{a}} X (X^T C_{\boldsymbol{a}} X + C_{\boldsymbol{e}})^{-1} X^T C_{\boldsymbol{a}} X \\ &= X^T C_{\boldsymbol{a}|\boldsymbol{y}} X \\ &= C_{(X^T\boldsymbol{a})|\boldsymbol{y}}. \end{aligned} \tag{1.24}$$

[1] $C_{\boldsymbol{g}\boldsymbol{y}} = C_{\boldsymbol{y}\boldsymbol{g}}^T = C_{\boldsymbol{g}}$.

[2] Using the Sherman–Morrison–Woodbury formula.

Both $C_{\boldsymbol{g}|\boldsymbol{y}}$ and $C_{X^T\boldsymbol{a}|\boldsymbol{y}}$ are covariance matrices of the estimated field given the measurements y.

CHAPTER 2

Related Work

The sensor placement problem has been considered by many people. We discuss in this chapter the work done in [4]and [5].

The work in [4] is based on the Gaussian process framework to model the spatial field. As we described above, in this framework a Gaussian random variable is associated with the phenomenon at each location and for any set of locations a Gaussian random vector is used to model the phenomenon at these locations. Given measurements at certain locations, the field value at these locations can be estimated by (1.22) with an error covariance matrix given by (1.23) which depends on the covariance matrix of the Gaussian process. In [4], a pilot deployment is used to collect data and estimate the covariance matrix of the Gaussian process using a kernel function. To relate to our work in this thesis, our assumption of a known regression model for the field in Chapter 3 is similar to assuming a known covariance matrix for the Gaussian process. For this reason we do not need in our work a pilot deployment and the placement we find in Chapter 3 is obtained before any collection of data. Furthermore the sensor placement in [4] is found by maximizing the mutual information between the locations, k, where data are collected and the locations, u, where the process needs to be estimated. In our work the philosophy would be to minimize the entropy of u given k. The entropy is the measure of the uncertainty of u given k while the mutual information is the measure of the reduction of uncertainty in u given k.

The work in [5] considers the sensor placement for maximal coverage in the design space without any knowledge about the model of the field. Mobile sensors are used to travel and collect measurements at these placements. The travel time is taken into consideration in finding the placements. In our work we assume either the model of the field is known (Section 3.1), or the model is one of many defined models (Section 3.2 and Chapter 4).

CHAPTER 3

Experiment Design for Model Estimation

In this chapter we present sensor placement methods based on optimal experiment design for model estimation. We assume that we are restricted to pick the placements from a set of p plausible locations $Z = \{z_1, \ldots, z_p\}$. We assume as well that we want to collect m measurements. In any of the frameworks reviewed above, we can randomly place sensors, collect m measurements and formulate inferences like estimating the parameters of the regression model or estimating the truth g in the Gaussian process framework. Here we ask and answer the question of where to collect the data so that these inferences are done as well as possible or more technically in an optimal fashion. In Section 3.1 we present the basic experiment design problem for estimating a single model and in Section 3.2 we show how an optimal placement can be found to estimate multiple models.

3.1 Estimation of a Single Model

The error covariance matrices $C_{\tilde{\boldsymbol{a}}}$, $C_{\boldsymbol{a}|\boldsymbol{y}}$ and $C_{\boldsymbol{g}|\boldsymbol{y}}$ given in (1.11), (1.16) and (1.23) respectively, do not depend on the measurements $\boldsymbol{y}$ and depend only on the design matrix X. The covariance matrix C reflects the accuracy of the estimation– the smaller (in some sense) the matrix is, the better the estimation accuracy will be.

3.1.1 Sensor Placement in the Regression Framework

Here we show how the sensor placement problem can be formulated in the regression framework in a similar approach as in [3]. The experimental design problem provides us guidance about how to choose from the set $Z = \{z_1, \ldots, z_p\}$ the locations x_j that are maximally informative. The approach we take is to minimize the matrix C in some sense.

Define m_j (≥ 0) to be the number of measurements we collect at the location z_j and w to be the vector with j-th element $w_j = m_j/m$. Note that we might collect multiple measurements at a single location if it is an especially informative location. Since we want to collect m measurements, we get the following equation:

$$\sum_{i=1}^{p} m_j = m \longrightarrow \mathbf{1}^T w = 1$$

where $\mathbf{1}$ is the all-ones vector. The error covariance matrix in (1.11) can now be written as

$$C = \left(\sum_{i=1}^{p} m_j v_j C_{\boldsymbol{e}}^{-1} v_j^T \right)^{-1} = \frac{1}{m} \left(X diag(w) C_{\boldsymbol{e}}^{-1} X^T \right)^{-1}$$

where $diag(w)$ is the diagonal matrix having w as its diagonal. Now we can write the experiment design as an optimization problem as follows:

$$\begin{aligned} \min_{w} \quad & C \\ \text{subject to} \quad & \textstyle\sum_{i=1}^{p} m_j = m \\ & m_j \geq 0 \\ & m_j \text{ is integer} \end{aligned} \tag{3.1}$$

This problem formulation results in an NP-complete problem because of the integer constraint on m_j. We can relax this integer constraint and get an approximate design. To make things simpler and without loss of generality, we assume that $C_{\boldsymbol{e}}^{-1} = I$. Since $\frac{1}{m}$ is a constant we can drop it and the problem becomes

$$\begin{aligned} \min_{w} \quad & C = \left(\sum_{i=1}^{p} w_j z_j z_j^T\right)^{-1} = \left(X diag(w) X^T\right)^{-1} \\ \text{subject to} \quad & \mathbf{1}^T w = 1 \\ & w \geq 0 \end{aligned} \tag{3.2}$$

The minimization (3.2) is a vector optimization problem over the space of positive definite matrices. We will introduce a convex scalarization for the problem that provides us with an ordering over this space. The most common scalarization is to minimize the determinant of the error covariance matrix C. This is called the *D-optimal design*. The problem now becomes a convex problem:

$$\begin{aligned} \min_{w} \quad & \det\left(X diag(w) X^T\right)^{-1} \\ \text{subject to} \quad & \mathbf{1}^T w = 1 \\ & w \geq 0 \end{aligned} \tag{3.3}$$

There are many other scalarizations ($E - optimal$, $A - optimal$, etc.) that minimize different convex functions of the error covariance matrix C. $E - optimal$ minimizes the norm of the error covariance matrix and $A - optimal$ minimizes the trace of the error covariance matrix. An elaborated discussion is presented in [3]. We will focus on the $D - optimal$ design since it can be interpreted as minimizing the entropy of the estimation error as we see in Section 3.1.3.

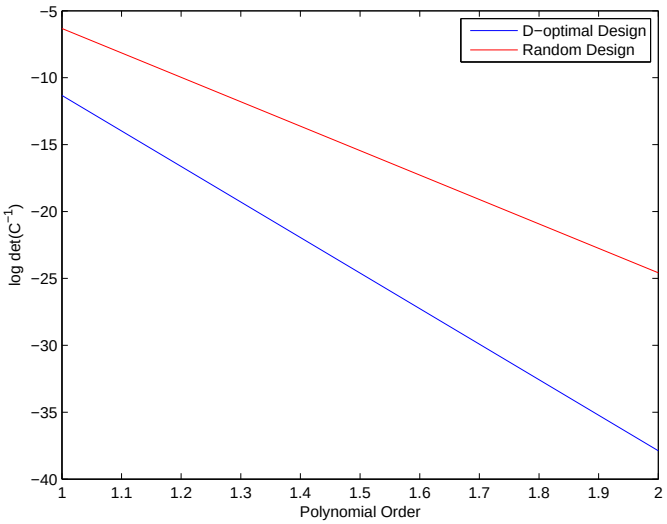

Figure 3.1: Comparison between Random Designs and $D-optimal$ Designs. The Random Design curve is the average of 1000 simulations.

The solution of (3.3), $w_i = m_i/m$, can be interpreted as the relative frequency of choosing the location z_i. The value $w_i \cdot m$ can also be rounded to get integer values for m_i. We notice that this whole formulation does not use the data to find the approximate optimal locations for the sensors. As we said before, it is assuming that the model we choose to fit in (1.4) is the true model!

Why is this formulation useful? The problem (3.3) corresponds to finding the placements to minimize the volume of the confidence ellipsoid in estimating the regression parameters [3]. Here we present figure that shows the benefit in solving the $D-optimal$ sensor placement problem on the regression model estimation. This benefit is in terms of the volume of the estimation confidence ellipsoid. Indeed the volume of the confidence ellipsoid resulting from the $D-optimal$ locations is smaller than the one from random locations as shown in Figure 3.1.

Similarly we can formulate the sensor placement problem for the regression

framework with random parameter $\boldsymbol{a} \sim \mathcal{N}(0, C_{\boldsymbol{a}})$. The problem is

$$\begin{aligned} \min_{w} \quad & \det\left(C_{\boldsymbol{a}}^{-1} + X diag(w) X^T\right)^{-1} \\ \text{subject to} \quad & \mathbf{1}^T w = 1 \\ & w \geq 0 \end{aligned} \tag{3.4}$$

3.1.2 Optimal Sensor Placement in the Gaussian Process Framework

We can ask the question of optimal placement in the Gaussian process framework in a similar manner as in the regression framework. The goal would be to select the locations such that the covariance matrix $C_{\boldsymbol{g}|\boldsymbol{y}}$ is as small (in some sense) as possible. Using the the determinant scalarization we get the following problem

$$\min_{w} \quad \det(C_{\boldsymbol{g}}^{-1} + C_{\boldsymbol{e}}^{-1})^{-1} \tag{3.5}$$

$$\text{subject to} \quad \text{Constraints depending on } C_{\boldsymbol{g}} \tag{3.6}$$

From the discussion at the end of Section 1.2.2, we see that if we assume that $C_{\boldsymbol{g}} = X^T C_{\boldsymbol{a}} X$, (1.23) reduces to (3.4). The constraints in (3.6) reduce as well to the constraints in (3.3) of the regression framework. This relates the sensor placement problems in both frameworks.

3.1.3 Interpretation in terms of Entropies

One could think of (3.6) as an entropy minimization problem. In fact the entropy of $f_{\boldsymbol{g}|\boldsymbol{y}}(g|y)$ is given by

$$\begin{aligned} h(\boldsymbol{g}|\boldsymbol{y}) &= \tfrac{1}{2}\log\left((2\pi e)^n \det(C_{\boldsymbol{g}|\boldsymbol{y}})\right) \\ &= \tfrac{1}{2}n\left(1+\log(2\pi)\right) + \tfrac{1}{2}\log\det(C_{\boldsymbol{g}|\boldsymbol{y}}) \end{aligned} \tag{3.7}$$

We can see that minimizing $h(\boldsymbol{g}|\boldsymbol{y})$ leads to (3.6). This explains what we said in Chapter 2 about the relationship between this work and [4]. We use an entropy minimization approach while [4] uses a mutual information maximization approach.

3.2 Estimation of Multiple Models

In this section we are trying to find designs which are $'$good$'$ for multiple models. This situation arises, for example, when we are deploying nodes that have multiple sensor modalities (e.g. temperature, light, humidity, etc). In this situation, we are dealing with different modalities, each of which has a different model. So the goal of the experiment design is to find locations that result in high accuracy in estimating multiple fields together. This situation arises as well when we are estimating one field with an uncertain model i.e. the field model could be one of multiple models. For example, we want to estimate a temperature field but we are uncertain if we should assume that the field model is a first or a second order polynomial. So the goal is to find locations that result in an estimation robust to the model assumption. We focused as above on polynomial models. We will illustrate the approach in the following example of estimating multiple fields [1]:

[1] We can use the same discussion below to the case of a single field with multiple possible models.

Assume that we have nodes that have two sensing modalities, a temperature sensor and a light sensor. Assume further that the temperature field is a first order polynomial model and the light field is a second order polynomial model:

$$\begin{aligned} \boldsymbol{t}_i &= a_0 + a_1 x_i + \boldsymbol{e}_i, \quad i = 1, \ldots, m \longrightarrow \boldsymbol{t} = X_{\boldsymbol{t}}^T a + \boldsymbol{e} \\ \boldsymbol{l}_i &= b_0 + b_1 x_i + b_2 x_i^2 + \boldsymbol{e}_i, \quad i = 1, \ldots, m \longrightarrow \boldsymbol{l} = X_{\boldsymbol{l}}^T b + \boldsymbol{e} \end{aligned} \tag{3.8}$$

where $\boldsymbol{t}$ refers to temperature and $\boldsymbol{l}$ refers to light. $X_{\boldsymbol{t}}$ is the design matrix corresponding to a first order polynomial, and $X_{\boldsymbol{l}}$ is the design matrix corresponding to a second order polynomial. Note that the example can be easily interpreted as uncertain model assumptions for a single field.

We present two formulations for this problem. The first one minimizes the mean of the estimation errors of the different fields (or models) while the second one minimizes the estimation errors of some fields (or models) with a constraint on the accuracy of estimation of the other fields (or models). We will focus on the example above of two fields but the approach can be easily extended to multiple fields.

3.2.1 Using the Mean

We sought good designs by simply minimizing the mean of the estimation errors corresponding to the two different models. The first approach is to use the arithmetic mean and solve

$$\begin{aligned} \min_{w} \quad & \left(\det \left(X_{\boldsymbol{t}} diag(w) X_{\boldsymbol{t}}^T \right)^{-1} + \det \left(X_{\boldsymbol{l}} diag(w) X_{\boldsymbol{l}}^T \right)^{-1} \right) \\ \text{subject to} \quad & \mathbf{1}^T w = 1 \\ & w \geq 0 \end{aligned} \tag{3.9}$$

Which is a convex problem since the utility function is the addition of two convex functions and the constraints are linear. In case we have more than two fields, we would need to add the utility functions corresponding to the multiple fields. The second approach is to use the geometric mean and solve

$$\begin{aligned} \min_{w} \quad & \left(\det\left(X_{\boldsymbol{t}} diag(w) X_{\boldsymbol{t}}^T\right)^{-1} \cdot \det\left(X_{\boldsymbol{l}} diag(w) X_{\boldsymbol{l}}^T\right)^{-1}\right) \\ \text{subject to} \quad & \mathbf{1}^T w = 1 \\ & w \geq 0 \end{aligned} \tag{3.10}$$

which is also a convex problem. In case of multiple fields, we would need to multiply the utility functions corresponding to the multiple fields.

The solution of (3.10) or (3.11) provides good estimation of both the temperature and the light fields. It performs better than the optimal solution assuming the temperature field model in estimating the light field, and it performs better than the optimal solution assuming the light field model in estimating the temperature field. An investigation of the benefits of one approach versus the other is not done here but it is planned as a future work.

3.2.2 Using Constraints

In this section we find designs that minimize the estimation error of one model while ensuring a certain accuracy in the estimation of the other model. This formulation is based on (3.3) with an additional constraint that the resulting estimation error of the second model is less than a specified value. The problem

becomes

$$\min_{w} \quad \left(\det\left(X_{\boldsymbol{t}} diag(w) X_{\boldsymbol{t}}^T\right)^{-1}\right)$$

$$\text{subject to} \quad \det\left(X_{\boldsymbol{l}} diag(w) X_{\boldsymbol{l}}^T\right)^{-1} \leq f \tag{3.11}$$

$$\mathbf{1}^T w = 1$$

$$w \geq 0$$

which is a convex problem. As above, in this formulation the solution provides good estimation of both the temperature and the light fields. The estimation error of the temperature field is minimized and the estimation error of the light field is below a user-defined value.

The benefits of these three approaches are that they ensure good estimation accuracy of multiple fields (or models). This will help in the design of node placements in case of multiple fields or uncertain model assumptions. In the example of a single field with multiple possible model assumptions, these approaches present a limitation that they do not scale with increasing the number of possible model assumptions. In the example of multiple fields, the limitation above is not present since a small finite number of modalities is usually present on the sensor node. But another drawback is present; these three formulations assume implicitly that the different modality fields are independent while usually the environmental fields are correlated and we can benefit from the correlations.

CHAPTER 4

Experimental Design for Discrimination between Models

In Chapter 3 we used $D-optimal$ experiment design to find the locations, from a set of locations Z, where we should collect measurements for optimizing the spatial field estimation by a linear regression model. As mentioned above, it was assumed that the model structure, the polynomial order in our example, is known. In this chapter we add some uncertainty to the problem. We look at the situation when the field is one of multiple regression models and we want to design a strategy to determine which of the models is true. A common method is to randomly place sensors, collect measurements, perform a likelihood test and pick the model that has the maximum likelihood given the measurements collected. But we can find strategies for placements and data collection to optimize the likelihood test. This falls into the fundamental experimental procedure where one has a prior set of plausible models for the field and resorts to guided data collection to pick the model, from the set, that best describes the field. It turns out that the data collection should be done in iterations with the estimation of the parameters of the regression models. In Section (4.1) we present the case of two plausible models and in Section (4.2) we present the case of multiple models.

4.1 Discrimination between Two Models

In this section we show how to find the optimal sensor placements for discriminating between two competing regression models. The regression models have known structure but unknown parameters. We do the following in this order: we derive the likelihood test then we formulate the sensor placement problem for optimizing the likelihood test and finally we present the algorithm of [1] that solves the optimization problem.

The two competing linear regression models are:

$$\begin{aligned} h_1(x,a) &= X_1^T a \\ h_2(x,b) &= X_2^T b. \end{aligned} \tag{4.1}$$

where X_1 and X_2 are the corresponding design matrices.

If m measurements are collected at locations $z_1, \ldots, z_m$ then

$$\boldsymbol{y} = h(z) + \boldsymbol{e}, \tag{4.2}$$

where the phenomenon $h(z)$ is a binary random variable and $\boldsymbol{e} \sim \mathcal{N}(0, C_{\boldsymbol{e}})$ is the measurement noise. Assume that we believe that $h(z)$ is $h_1(z)$ with probability α_1 and $h_2(z,a)$ with probability α_2. We shall assume, without loss of generality, that $C_{\boldsymbol{e}}^{-1} = I$.

The parameters a and b are estimated, as in (1.10), by

$$\hat{a} = (X_1 C_{\boldsymbol{e}}^{-1} X_1^T)^{-1} C_{\boldsymbol{e}}^{-1} X_1 y \tag{4.3}$$

$$\hat{b} = (X_2 C_{\boldsymbol{e}}^{-1} X_2^T)^{-1} C_{\boldsymbol{e}}^{-1} X_2 y. \tag{4.4}$$

The likelihood test is to compare the likelihoods of h_1 and h_2, $P_{\boldsymbol{h}|\boldsymbol{y}} h_1|y$ and $P_{\boldsymbol{h}|\boldsymbol{y}}(h_2|y)$ respectively, and pick the one with the largest value. Using Bayes'

theorem

$$P_{h|y}(h_1|y) = \frac{f_{y|h}(y|h1)P_h(h1)}{f_{\boldsymbol{y}}(y)} = \frac{\alpha_1 f_{y|h}(y|h1)}{f_{\boldsymbol{y}}(y)} \tag{4.5}$$

$$P_{h|y}(h_2|y) = \frac{f_{y|h}(y|h2)P_h(h2)}{f_{\boldsymbol{y}}(y)} = \frac{\alpha_2 f_{y|h}(y|h2)}{f_{\boldsymbol{y}}(y)} \tag{4.6}$$

So the likelihood test reduces to comparing $\alpha_1 f_{y|h}(y|h1)$ and $\alpha_2 f_{y|h}(y|h2)$. The log-likelihood test is

$$\max_i \quad \log(\alpha_i) + \log\left(f_{\boldsymbol{y}|\boldsymbol{h}}(y|h_i)\right) \tag{4.7}$$

where $f\boldsymbol{y}|\boldsymbol{h}(y|h_i) \sim \mathcal{N}(h_i, I)$. This is a binary hypothesis testing problem and the probability of error, P^e, in selecting the model is

$$P^e = Q\left(\frac{\sqrt{\|h_1(x,\hat{a}) - h_2(x,\hat{b}\|_2^2}}{2\sigma}\right) \tag{4.8}$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{(2\pi)}} \exp\left(-\frac{s^2}{2}\right)$

To optimize the discrimination we choose to minimize the probability of error (4.8). Since the function Q is strictly decreasing we find the sensor placement by maximizing the argument of $Q(.)$ in (4.8). Using the vector w as in 3.3, we get the following problem

$$\begin{aligned} \max_w \qquad & \|diag(w)\left(h_1(x,\hat{a}) - h_2(x,\hat{b})\right)\|_2^2 \\ \text{where} \quad & \hat{a} = arg\min_a \|diag(w)\left(y - h_1(x,a)\right)\|_2^2 \\ & \hat{b} = arg\min_b \|diag(w)\left(y - h_2(x,b)\right)\|_2^2 \end{aligned} \tag{4.9}$$

This problem (4.9) was proposed by [1] and the solution $\hat{w}$ was called $T-optimal$ design. As mentioned above the derivation of (4.9) here is more detailed and follows a slightly different approach than the one followed in [1].

We see from (4.9) that we need to know $\hat{a}$ and $\hat{b}$ to find the weights $\hat{w}$ and we need to know $\hat{w}$ to estimate $\hat{a}$ and $\hat{b}$, i.e. on the one hand we need to know the parameters of the models to find the optimal sensor placement and on the other hand we need to collect measurements at certain locations to estimate the parameters of the models. The solution $\hat{w}$ of (4.9) was also given in [1] by the following sequential algorithm

1. Given a design $w(j)$ and a number of measurements N find:

$$\hat{a_j} = arg\min_a \| \left(\boldsymbol{y}_j - h_1(x,a)\right) \|_2^2$$
$$\hat{b_j} = arg\min_b \| \left(\boldsymbol{y}_j - h_2(x,b)\right) \|_2^2$$

2. Add to the design the location z_{j+1} such that:

$$z_{j+1} = arg\max_{z\in Z} \left(h_1(z,\hat{a_j}) - h_2(z,\hat{b_j})\right)^2$$

3. Make a measurement s at z_{j+1} and update

$$w(j+1) = (1 - \tfrac{1}{N+1})w(j) + \tfrac{1}{N+1}\delta(z_{j+1})$$

4. Go back to 1

where $\delta(z_{j+1})$ is a vector with all entries equal to *zero* except the entry corresponding to z_{j+1} is equal to 1. In each iteration the algorithm estimates the parameters, $\hat{a}$ and $\hat{b}$, from the available measurements, picks the point that maximizes the difference between the models, makes a measurement there and updates the weights associated with locations by adding $\frac{1}{N+1}$ weight to the point selected and rescaling so that the weights add up to 1. Details on the convergence of the algorithm to the solution of (4.9) are available in [1]. Finally, given the measurements collected, a likelihood test is used to select the model.

Next we show a simulation we did on a 2-dimensional field. We used the following model to generate data:

$$\boldsymbol{y}_i \;=\; 2 + 3x_i + 0.4v_i + \boldsymbol{e}_i, \quad i = 1, \ldots, m \quad \longrightarrow y = X^T a + \boldsymbol{e} \tag{4.10}$$

where $\boldsymbol{e} \sim \mathcal{N}(0, I)$. We used the algorithm to discriminate between the following models

$$\begin{aligned} h_1(x_i, v_i) &= a_0 + a_1 x_i + a_2 v_i, \quad i = 1, \ldots, m \quad \longrightarrow h_1(x, v, a) = X_1^T a \\ h_2(x_i, v_i) &= b_0 + b_1 x_i^2 + b_2 v_i^2, \quad i = 1, \ldots, m \quad \longrightarrow h_2(x, v, b) = X_2^T b. \end{aligned} \tag{4.11}$$

We choose a set Z with 15 locations as shown in Figure 4.1. Figure 4.1 shows 7 locations selected after 8 iterations, and the likelihood ratio is 4.04. Figure 4.2 shows also 7 locations selected after 8000 iterations, and the likelihood ratio is 14.85. This shows that the algorithm converges well in finding the optimal locations since they were found after 8 iterations and the only change was in the weights of the locations. The weight, m_i/m, reflect the fraction of of measurements that should be collected at the location z_i.

In the case where the two models are nested, for example

$$\begin{aligned} h_1(x_i, v_i) &= a_0 + a_1 x_i + a_2 v_i, \quad i = 1, \ldots, m \quad \longrightarrow h_1(x, v, a) = X_1^T a \\ h_2(x_i, v_i) &= b_0 + b_1 x_i + b_2 v_i + b_3 x_i^2 + b_4 v_i^2, \quad i = 1, \ldots, m \quad \longrightarrow h_2(x, v, b) = X_2^T b, \end{aligned} \tag{4.12}$$

the space of the second model includes the space of the first model. The likelihood of the second model will be, for any design, greater or equal to the likelihood of the first one. So in this case one would be interested in the structure of the model and would add the constraint $b_3^2 + b_4^2 \geq 1$ to step 1 of the algorithm [1].

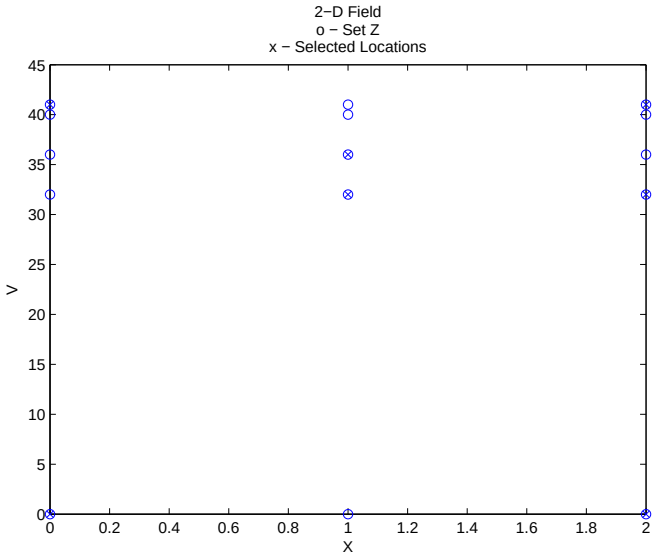

Figure 4.1: Selected Locations for Discrimination between the models. 7 Locations were selected in 8 Iterations

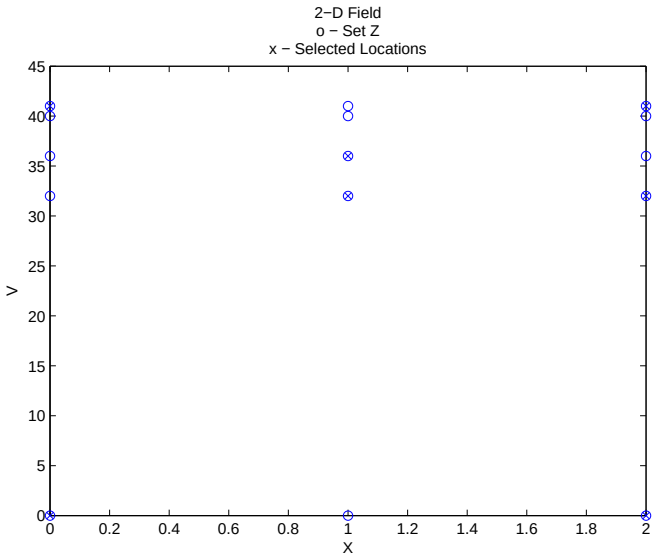

Figure 4.2: Selected Locations for Discrimination between the models. 7 Locations were selected in 8000 Iterations

4.2 Discrimination between Multiple Models

In this section, we show how to extend the work in Section 4.1 to finding a strategy for sensor placements and data collection for discriminating between several competing regression models. As above, the regression models have known structure but unknown parameters. We follow a similar reasoning as above where the strategy is found to minimize the probability of error in selecting the correct model. The probability of error is dominated by the comparison between the likelihood of the true model and the likelihood of the closest model to it. The algorithm is similar to the case of two models. But after step 1, a ranking of the models by goodness of fit is done and then in step 2 we pick the location that maximizes the discrimination between the two best models. Assume we want to discriminate between k models $h_1(z) \dots h_k(z)$. The measurements are modeled as follows

$$\boldsymbol{y} = h(z) + \boldsymbol{e}, \tag{4.13}$$

where $h(z)$ is one of the models.

The likelihood test is the same as (4.7) where i belongs to $\{1 \dots k\}$. The log-likelihood of the i-th model is given by

$$L_i = -\| (y - h_i(x, a)) \|_2^2 \tag{4.14}$$

The algorithm becomes

1. Given a design $w(j)$ and a number of measurements N, for each model $h_k(x, a)$ find:

$\hat{a_{k_j}} = arg\min_a \| (y_j - h_1(x, a)) \|_2^2$

Rank the models by the goodness of fit, for example $L_u > L_v > ... > L_w$

2. Add to the design the location z_{j+1} such that:

$$z_{j+1} = arg\max_{z \in Z} \left(h_u(z, \hat{a_j}) - h_v(z, \hat{b_j}) \right)^2$$

3. Make a measurement s at z_{j+1} and update

$$w(j+1) = (1 - \tfrac{1}{N+1})w(j) + \tfrac{1}{N+1}\delta(z_{j+1})$$

4. Go back to 1

Next we show a simulation we did on a 2-dimensional field. We used the model in (4.10) to generate the data. We used the algorithm to discriminate between the following models

$$\begin{aligned} h_1(x_i, v_i) &= a_0 + a_1 x_i + a_2 v_i, \quad i = 1, \ldots, m \quad \longrightarrow h_1(x, v, a) = X_1^T a \\ h_2(x_i, v_i) &= b_0 + b_1 x_i^2 + b_2 v_i^2, \quad i = 1, \ldots, m \quad \longrightarrow h_2(x, v, b) = X_2^T b \\ h_3(x_i, v_i) &= c_0 + c_1 x_i^3 + c_2 v_i^3, \quad i = 1, \ldots, m \quad \longrightarrow h_3(x, v, b) = X_3^T c. \end{aligned} \tag{4.15}$$

We chose a set Z of 15 locations as shown in Figure 4.3. Figure 4.3 shows that 8 locations are selected after 8 iterations and the likelihoods are: $L_1 = 0.7627$, $L_2 = 0.0319$ and $L_3 = 2.4e^-4$. Figure 4.4 shows that 11 locations were selected after 8000 iterations and the likelihoods are:$L_1 = 0.9360$, $L_2 = 0.0307$ and $L_3 = 2.5e^-5$. We see here that after 8000 iterations more locations were selected and that the discrimination was only slightly improved.

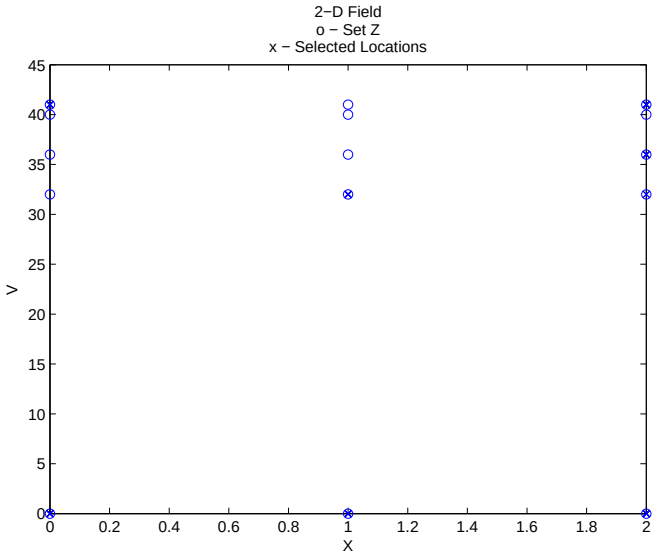

Figure 4.3: Selected Locations for Discrimination between three models. 8 Locations were selected in 8 Iterations

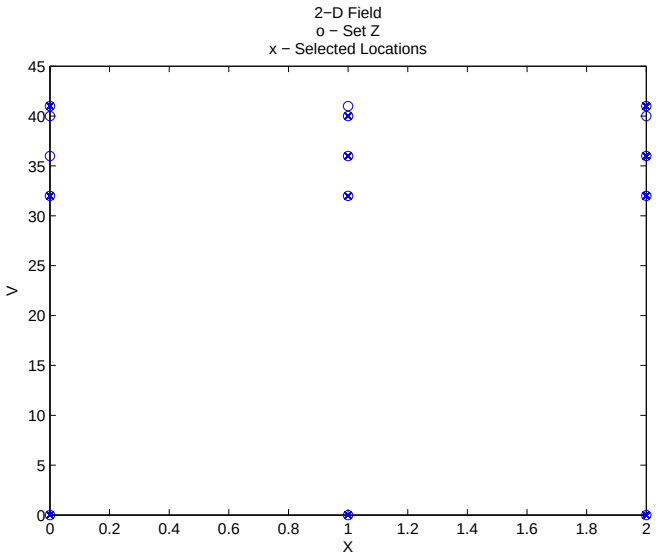

Figure 4.4: Selected Locations for Discrimination between three models. 11 Locations were selected in 8000 Iterations

CHAPTER 5

Conclusions and Future Work

In this thesis we described methods for sensor placement and data collection strategies in linear environmental fields for two purposes; the first is estimation of the field and the second is model selection from a set of models. The methods are based on optimal experimental design techniques. For the first problem, Chapter 3, we showed how to find sensor placement to minimize the error in estimating the field using $D - optimal$ designs. We presented the connection between the regression framework and the Gaussian process framework. The Gaussian process framework reduces to the regression framework if we assume that the Gaussian process has mean $X^T \mu_{\boldsymbol{a}}$ and covariance matrix $X^T C_{\boldsymbol{a}} X$. The sensor placement problem for estimating the process then becomes equivalent to the sensor placement problem for estimating the regression parameters. We showed as well that the problem of $D - optimal$ sensor placement for estimating the field can be interpreted as a minimization of the entropy of the field given the measurements. We presented in simulations the benefit of the experimental design techniques over random placement in reducing the uncertainty in estimating the field model. We presented as well three formulations for optimal sensor placement to estimate multiple fields together. The first formulation is to find the sensor placements which produce the minimum arithmetic mean of the estimation errors of the multiple fields, the second is to find the sensor placements which produce the minimum geometric mean of the estimation errors of the multiple fields, and

the third formulation is to find the sensor placements which produce the minimum mean of the estimation errors of some of the considered fields while ensuring that the estimation errors of the other fields is below some specified value. For the second problem, Chapter 4, we showed how to find sensor placements and data collection strategies for selecting a field model from a finite set of models based on $T - Designs$ of [1] and [2]. We presented a formulation of the problem based on minimization of the probability of error in selecting the field model. The strategy of sensor placement and data collection is sequential.

Many topics for future work can be suggested. We are interested in the following ones. The work of Section 3.2 on estimating multiple modality fields could be extended to incorporate the correlation between the estimated fields. As mentioned above, the formulation presented assumes that the fields corresponding for the different modalities are independent. We expect that incorporating the correlation between the fields will result in a better accuracy in the estimation. Another extension would be to take the work of Section 3.2 on robust estimation of a single field under uncertainty in the structure of the field model and generalize it to uncertainty over a class of models. The work presented in this thesis shows the formulation of the sensor placement problem for robust estimation of the field when the field model belongs to a finite set of models. We are interested in finding sensor placements robust to a class of models, e.g. the class of polynomial models. A similar extension could be made for the work of Chapter 4 to classes of models rather than a finite set of models. Extension of the work of Chapter 4 to different model selection criteria is also interesting. The work presented in this thesis is based on minimizing the probability of error in the selection and assumes that one of the competing models is the correct model. We are interested in the situation when the set of models considered does not include the correct model, and we want to find the closest model to the true one. This situation is

interesting because often the correct field model is not known in advance.

We hope that the methods and discussions presented in this thesis provide a step toward more rigorous strategies of sensor node placements and data collection for reducing the uncertainty and the a priori assumptions in modeling environmental spatial fields.

References

[1] A. Atkinson and V. Fedorov. The design of experiments for discriminating between two rival models. *Biometrika*, 62(1):57 – 70, 1975.

[2] A. Atkinson and V. Fedorov. Optimal design: Experiments for discriminating between several models. *Biometrika*, 62(2):289 – 303, 1975.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[4] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg. Near-optimal sensor placements: Maximizing information while minimizing communication cost. In *Proceedings of Information Processing in Sensor Networks (IPSN)*, April 2006.

[5] M. Rahimi, M. Hansen, W. Kaiser, G. Sukhatme, and D. Estrin. Adaptive sampling for environmental field estimation using robotic sensors. In *Proceedings of The IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005.

[6] C. K. I. Williams. Prediction with gaussian processes: From linear regression to linear prediction and beyond. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 599–622. The MIT Press, 1999.