To the Graduate Council:

I am submitting herewith a dissertation written by Deborah Morren Flanagan entitled "Optimal Monitoring Systems Using Statistical Experimental Design." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Management Science.

Mandyam M. Srinivasan, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor and
Dean of The Graduate School

OPTIMAL MONITORING SYSTEMS USING STATISTICAL EXPERIMENTAL DESIGN

A Dissertation

Presented for the

Doctor of Philosophy Degree

The University of Tennessee, Knoxville

Deborah Morren Flanagan

December 1997

Dedication

To my husband, George, and our children, Theresa and Christopher.

Acknowledgments

Many people have contributed to my doctoral program process in Management Science at the University of Tennessee.

I would like to thank the faculty of the Management Science, Statistics, and Mathematics departments for their interest, help, and encouragement. I am grateful to my Dissertation Committee for their patience and helpful criticisms and suggestions that improved this document. I am particularly grateful to my committee chair, Prof. Mandyam Srinivasan, and dissertation co-director, Dr. Valerii Fedorov.

I am grateful to Lockheed Martin Energy Research Corporation at Oak Ridge National Laboratory (ORNL) for support of this work through the Laboratory Directed Research and Development program. The support and assistance of my ORNL colleagues, staff and management, during my coursework and especially in support of this dissertation has been invaluable. A special thanks is extended to Stelmo Poteet for her patience and dedication in the preparation of this manuscript.

I am grateful to my parents, Mr. and Mrs. John E. Morren, Sr., whose loving support made all the difference in my reaching this point.

Finally, I am most grateful to have such a wonderful, loving, and patient family as my husband, George, and our children, Theresa and Christopher.

Abstract

A statistical experimental design methodology is demonstrated for determining the optimal number and placement of data collection devices of a monitoring system for a computer network such as the Internet or its subsystems. The approach is based on Mercer's expansion of covariance kernels in the case of continuous controlled variables and on the direct use of convex design theory for discrete designs sets with spatial covariance structure. The merit of the methodology using Mercer's kernel expansion is illustrated with some examples based on covariance structures generated by the Brownian bridge model. The results of this dissertation establish the optimality conditions for the D- and linear criteria in the presence of a spatial covariance structure and discrete variable space. Associated algorithms and computer software are developed to implement the theory. In addition to constructing optimal designs, the computer network examples exhibit the ability of the algorithms to suggest and estimate the efficiency of monitoring system designs which are less than optimal but which may be attractive according to some operational, policy, or cost perspectives.

The applicability of the methods to computer network monitoring is demonstrated with simple network graph examples for single- and multiple-host monitoring systems and with a test case from the Department of Energy (DOE) Energy Sciences Network (ESnet).

The simple network graph examples utilize the D-optimality and minimax

criteria for linear models both with and without a prior information structure. The analysis is repeated for the case when the variance is dependent upon the route selected. The examples show that the transition from single-host to multiple-host monitoring systems produces a significant gain with respect to efficiency and cost reduction.

The ESnet test case is created using an Oak Ridge National Laboratory host computer that interrogates 39 sites of the ESnet host community. The developed methodology for constructing experimental designs for D- and linear optimality in the presence of a spatial covariance structure produces an optimal monitoring system with 12 sites, which is compared with a few other designs.

Contents

List of Tables

List of Figures

Nomenclature

$\in$, $\notin$	belongs to, does not belong to
$\subseteq$, $\subset$	is contained in, is a subset of
R^n	n-dimensional Euclidean space
Z^+	set of positive integers
R^+, R^-	set of positive (negative) real numbers
$\| \vec{x} \|$	norm or length of a vector
$\|D\|$	determinant of matrix D
tr	trace of a matrix; sum of the diagonal matrix elements
I	identity matrix
M	information matrix
$M(\xi)$	information matrix of design ϵ
ξ	experimental design defined by specific support points x_i and weights p_i
$s.t.$	such that
D	variance-covariance matrix; $D = M^{-1}$
$inv\ M$	inverse of matrix M; M^{-1}
$\sum_{i=1}^{n}$	sum over n items
M^T	transpose of matrix M
arg	argument of
min	minimum
max	maximum
Ψ	optimality function
Ξ	set of all probability measures
λ_i	eigenvalue i of a matrix
φ_i	eigenfunction corresponding to eigenvalue λ_i
$\phi(x, \xi)$	sensitivity function

Chapter 1

Statistical Design of Optimal Experiments for Computer Network Performance Evaluation

1.1 Introduction

This dissertation concerns the applicability and extension of optimal experimental design methodologies to the problem of developing a theoretical basis for structuring monitoring systems to evaluate the performance of large computer networks such as the Internet. Monitoring system structure, here, refers to the optimum placement of data collection devices on the system being monitored. This section describes the scope by providing a discussion of the motivation, background infor-

mation on existing network performance research, historical remarks on statistical design, and focus areas.

1.2 Motivation

The motivation is multifaceted. The idea of applying experimental design methodologies to develop a monitoring system is relatively novel even though it does not originate with this work. The application to large computer networks, however, is novel. The concerns and issues of developing a design for an experiment closely parallel those of developing even a reasonable monitoring system. One objective of a monitoring system should always be to collect as little data as necessary to be able to monitor specific parameters of the system with respect to assigned targets and objectives. This implies a purposeful monitoring where each piece of data has a reason to be collected, a question that it will help to address. When a computer network system as large and complex as the Internet is the monitoring subject, providing an optimal observing system becomes even more important. Internet monitoring will be a continuing effort of (1) focusing on monitoring objectives and (2) optimal data collection for analysis of those objectives. The theoretical component of this dissertation considers extending optimal design theory methodologies so that the suspected spatial correlation and non-homogeneous variance structure of computer network traffic data can be accommodated. Finally, algo-

rithms and computer programs need to be developed so that the test results can be extended in the future to construct and implement a monitoring system for a computer network such as the Internet. Each of these motivation components is discussed in further detail.

Monitoring systems are needed to provide different kinds of information about a physical operating system. Metrics are required (1) to determine performance of routine maintenance and problem troubleshooting and (2) to predict performance trends, develop simulations, and evaluate operating system planning. The metrics may characterize the operating system status, routes, capacity, resources, packet losses, response time, or intrusions, to name a few examples (c.f. Paxson (1997) and Cottrell et al. (1997) for computer network metric examples). A monitoring system may be implemented for non-industrial areas such as seismology, climactic, and environmental evaluation and intrusion monitoring. These applications might, for example, consider the optimal placement of sensors that would provide data on the status of a spatial area.

Many data collection decisions must be made by the developers of a monitoring system. These decisions include but are not limited to the following:

1. The type of data collection instrument (hardware or software) to be used.

2. The amount of data to be collected in a given time period, how large a subset of the available data to collect during the period, the length of the period,

and the frequency of data collection.

3. The determination of the data to be collected, i.e., response measures and important influencing factors.

4. How to minimize operating system or spatial area interruption during data collection.

5. The placement of sensors and data collection devices.

6. Which data will be retained and how long (i.e., data storage and retention issues).

7. Quantification of the objectives and the formulation of optimality criteria.

8. The analysis costs of experiments.

We would like to emphasize that this research is focused on the monitoring of networks. We do not consider the very interesting and challenging "active" experiments in which a network's characteristics or regimes may be varied to get information about its behavior at some specific conditions to validate or test such models.

The optimal sizing and structure of a monitoring system with respect to the number and placement of data collection devices can address issues (2) through (8). The optimal number and location of data collection devices can be determined via statistical experimental designs based on optimal design theory.

Optimal design theory methodologies start with a set of factors that have a potentially important impact on the response measure of interest (i.e., the performance metric). The entire feasible set of combinations of factor levels is used to select the minimum set of combinations that will provide the most cost effective information about the response variable (i.e., the performance metrics). In this manner, the minimum number of data collection devices and locations (i.e., specific combinations of factor levels to be monitored) is constructed by the method. Including or deleting locations (i.e., specific factor combinations) from the monitoring system are constraints which can also be optimized. Chapter 1.3 describes the place of optimal design methodologies within the realm of other statistical experimental design methodologies.

The Internet is a very large and complex computer network system with many performance monitoring needs. The network has experienced great growth and is expected to continue. With the large number of users and host computers, it is necessary to provide a reliable and timely system. All of the performance metrics and data collection decisions discussed above are relevant to the Internet. Research on computer network performance metrics and monitoring for the Internet has not included optimizing the network monitoring system or developing a specific monitoring system, as shown in Chapter 1.2.

Different types of data require different model assumptions and, therefore, different theoretical approaches. Computer network traffic has been shown to

exhibit unusual patterns with respect to correlation and variance non-homogeneity that require different assumptions, modeling theory, and response surface design of experiments.

An example of monitoring system development using experimental design is Cheng (1991), who proposes a sampling system for federal highway vehicle traffic data using a variation of response surface design methodology. The problem and approach are different from the computer network application and the approach of this dissertation.

Computer software for generating response surface design of experiments is currently available from various vendors such as the Statistical Analysis System (SAS) Inc. (see SAS (1995)). While this software handles most standard linear model designs, it does not cover the optimal design theory methods needed for our research. In particular, even the most extensive package SAS does has no features for correlated observations, does not allow the computation of optimal designs when prior information is available, and uses algorithms which frequently do not guarantee global optimality. Thus, computer programs which can satisfy the theoretical extensions of this dissertation are necessary, and their development is a part of this dissertation.

1.3 Internet Computer Network Performance Metric Research Background

The research needed to lay the foundations of performance monitoring for the Internet is in its early stages. This section provides highlights from some example research literature on important Internet issues.

"Towards a Framework for Defining Internet Performance Metrics," (Paxson (1996)) describes a format for establishing an international framework for Internet performance diagnosis and evaluation. The article describes and recommends a standardized method of considering and developing metrics and incorporates the work of the Benchmarking Methodology Group of the Internet Engineering Task Force.

"ESnet's Internet Monitoring Activities," (Cottrell et al. (1997)) reports on a study by the ESnet Site Coordinating Committee to determine the impact of the Internet on the reliability and timeliness of network connections to national research sites (e.g., the national laboratories) and international partners. This article describes some current monitoring activities completed at some locations and recommends expanded monitoring to additional locations. Specific metrics to be monitored include delays and throughput using "*ping*" (the Packet Internet Groper) data.

There is a problem with "burstiness", an oscillation between periods of smooth

traffic levels and periods of traffic bursts. Authors have observed such burstiness patterns in computer network traffic data and noted that the traffic displayed a self-similar behavior, which is characterized by long term data dependencies. Briefly, processes with self-similar behavior have long range dependencies, which means that the autocorrelation function decays hyperbolically as the lag increases. While the lag correlations are all individually small, their cumulative effect is important and has features very different from processes with conventional short-range dependencies. Hurst (c.f. Hurst (1951, 1956)) uses his own "R/S-Statistic" to represent many time series as $E[R(n)/S(n)] \simeq cn^H$, as $n \to \infty$, where n is the number of observations; $R(n)$ is the re-scaled, adjusted range; $S(n)$ is the sample standard derivation; and c is a finite constant greater than zero. H is the Hurst parameter, which equals .5 for time series with short-range dependencies and is greater than .5 for long-range dependencies.

Willinger, et al. (1996) provide an excellent bibliographical guide to studies of self-similar network traffic modeling and analysis and related areas such as Poisson modeling, traffic measurement, Brownian motion, Markov processes, and ARIMA time series models.

Downing, Fedorov, et al. (1997) use the concepts of non-homogeneity, compoundness, and double stochasticity of Poisson processes to explain behavioral patterns of the long-memory representation of computer network traffic.

"On the Self-Similar Nature of the Ethernet Traffic," (Leland et al. (1994))

provides detailed theory and analysis of the self-similar nature of some network traffic data. Their study also shows that their analytical tools are cumbersome and dependent on graphs. Additionally, prediction methods for self-similar processes are in the early stages and do not exhibit a desirable level of prediction accuracy.

"STOW Traffic and Related Issues," (Ngyuen and Batsell (1996)) analyzes traffic data from a demonstration of the Synthetic Theater of War-Europe (STOW-E) for the Distribution Interactive System (DIS) at the U. S. Naval Research Laboratory. Analysis of the data shows that a possible mixture of Poisson and self-similar processes exists and recommends a further analysis and study of the nature of such computer network traffic data.

Additional information appears in the literature supporting the idea of self-similar properties for network traffic data. Floyd and Paxson (1994) document a thorough study of the failure of the Poisson model to represent the behavior of traffic data.

Another important area in Internet performance metrics is the idea of developing improved pricing schemes. Pricing also involves some level of monitoring.

In summary, a number of important research issues are being addressed; however, no information appears in the literature with respect to the specifics of developing a large scale computer network monitoring system or of optimizing such a system.

1.4 General Comments on Statistical Methods of Experimental Design

A number of statistical methodologies have been developed for constructing experimental designs. In general, experimental design methods suggest an appropriate sample size (number of design points, experimental units, treatments, experimental runs, etc.) based on prior knowledge about the experimental uncertainties that might be expected and the class of models which may be used to make inferences about the explored systems. The controlled factors (independent variables) represent either qualitative or quantitative variables. Methodologies for standard linear models published by Fisher (1947), Box and Draper (1987), Box, Hunter, and Hunter (1978), Karlin and Studden (1966), Fedorov (1972), Puckelsheim (1993), and many others handle such models as the factorial, fractional factorial, composite, or blocked designs, to name a few. A number of software packages exist which can analyze such standard designs. The corresponding mathematical techniques (c.f. Box and Draper (1987)) work well for simple multi-dimensional linear models like simple linear regression or first order polynomials with possible interaction terms. New techniques are needed, however, for irregular operating regions and more complicated models that allow non-homogeneous covariance structures, and they started to emerge in the early 1970's.

How "good" is an experimental design? That is, how well does it provide a

mathematical environment for making a precise and accurate test of hypotheses from the experiment? Were more or fewer data collection points needed? Which ones? What effect can additional points have on the information that this experimental design will provide? Optimal design theory addresses these questions.

Standard experimental design methods for linear models (regression or response surface) are a subset of optimal design methodologies, which are used as a benchmark for determining how "good" designs are. Optimal design theory methodologies, then, apply to the assumptions and models of standard experimental design methods for linear models but also apply to the following situations:

1. not all factor level combinations are feasible,

2. the design space of the independent variables is irregular, and

3. the experimental design model is non-standard or non-linear.

Optimal design theory methods consider all possible (i.e., feasible) combinations of factor levels in order to select the minimum set of factor combination levels that will provide the most cost effective information about the measured response. Optimal design theory methodology also provides weights that represent the number of repeated observations to be taken at each factor level combination. Including or deleting particular factor level combinations can also be evaluated.

The assumptions of homogeneous variances and no correlation of observational errors are very important to statistical experimental design methodologies. The

assumptions made about the experimental data and model demand special approaches for correlations between observations and heterogeneous variances. Correlations can occur for a variety of reasons, including easily-traced "physical" causes like passing the same routes in the traffic problem. Most frequently a common cause factor leads to the correlations of various observations. The term common cause refers to the occurrence of an event which impacts or influences items which otherwise may be assumed to operate or exist independently: from a reliability perspective, for example, a common power source for multiple subsystems or components may be a common cause factor in that the failure of the power source may cause all the subsystems to fail. These problems were identified at the early stages of development in the theory of optimal design of experiments leading to the concept of orthogonal designs.

The correlation and variance problems are specific to a type of data, the model of the response function, and other assumptions, such as the criteria of optimality, characteristics of the operability region, and cost-type constraints. The application of optimal design theory methodologies to computer networks, then, requires the development of a model for the correlation structure of a test case. Then the design theory extensions and supporting algorithms can be developed. Chapter 2 provides the details of optimal design theory needed for this dissertation.

1.5 Study Focus

The thrust areas and novel contributions of this dissertation are:

1. Extensions of optimal design theory for experiments with prior information, correlated observations, and non-homogeneous variances are developed. The proposed approach is essentially based on Mercer's expansion of covariance kernels in the case of continuous controlled variables (Chapter 3) and on the direct use of convex design theory for discrete design sets (Chapter 3). The theoretical results in Chapters 3.2 and 3.3 are developed in collaboration with V. V. Fedorov.

2. A first order algorithm and computer codes to implement the theory of Chapter 2 and theory extensions of Chapter 3 are developed. Four algorithms that construct optimal designs for (1) D- and linear criteria for continuous and discrete designs and (2) D- and linear criteria for designs with covariance structure were developed in FORTRAN. Software documentation appears in Appendices A and B.

3. The illustration of theoretical results with some classical covariance structures in Chapter 3 is provided. These are numerical examples of the theoretical results using well-known problems which are frequently used for benchmarking.

4. The application of the developed methodology to test cases using delay time data from the Department of Energy (DOE) Energy Sciences Network (Esnet) and a few showcase examples based on widely-used exchange type and steepest descent algorithms are completed. These applications include problem and model development, data collection and preparation for analysis, and computer processing and analysis. The examples allow the methodology to exhibit the theoretical features of optimal design and other features important to practitioners, such as "what if" analyses of alternate optimal or less-than-optimal monitoring system structures.

Chapter 2

Convex Design Theory

This section provides theoretical background for the developed extensions of convex design theory. The background discussion starts with concepts of the linear model and information matrices in Chapter 2.1 and proceeds with the criteria of optimality and main properties of optimal designs for standard linear models in Chapter 2.2. Optimal design in the presence of prior information is presented in Chapter 2.3, and numerical methods for optimal design construction appear in Chapter 2.4.

2.1 The Linear Regression Model and the Information Matrix

The standard linear model considered is

$$y_i = \theta^T f(x_i) + \varepsilon_i, \tag{2.1}$$

where i = 1, ..., n, i.e., we have n response measurements y_i given p independent variables or factors $x_i^T = (x_{1i}, x_{2i}, \ldots, x_{pi})$. The model contains m unknown parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$. Errors ε_i are assumed to be random and $E(\varepsilon_i) = 0$. The variance $\sigma^2 \equiv E(\varepsilon_i^2)$ of the random errors is homogeneous across all levels of the independent variables, and $E(\varepsilon_i \varepsilon_{i'}) = \sigma^2 \delta_{ii'}$. It is crucial for many results that the basis functions $f(x) = (f_1(x), \ldots, f_m(x))^T$ are known.

The least squares estimator for θ is

$$\hat{\theta} = \arg\min_{\theta} v(\theta), \tag{2.2}$$

where

$$v(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{r_i} \left(y_{ij} - \theta^T f(x_i) \right)^2,$$

and the second subscript indicates the possibility of repeated observations. The notation "arg $\min_\theta$" means a solution to the optimization problem $\min_\theta v(\theta)$. The

least squares estimator coincides with the best linear unbiased estimator as proved by the Gauss-Markov Theorem (c.f. Rao (1973)). Noting that $v(\theta)$ is a convex function of θ and setting the derivatives of $v(\theta)$ with respect to θ equal to zero, we can derive (c.f. Rao (1973), Chapter 4a; Seber and Wild (1989), Chapter 12) that

$$\hat{\theta}M = Y, \tag{2.3}$$

where

$$M = \frac{\sum_{i=1}^{n} r_i f(x_i) f^T(x_i)}{\sigma^2}, \tag{2.4}$$

and

$$Y = \frac{\sum_{i=1}^{n} r_i \bar{y}_i f(x_i)}{\sigma^2}, \quad \bar{y}_i = \sum_{j=1}^{r_i} \frac{y_{ij}}{r_i}. \tag{2.5}$$

Matrix M is called the information matrix. If M is regular, i.e., its determinant is greater than zero, then

$$\hat{\theta} = M^{-1}Y. \tag{2.6}$$

The variance of $\hat{\theta}, Var\{\hat{\theta}\}$, is found by using

$$Var\{LZ\} = L\, Var\{Z\} L^T,$$

where L is a given matrix and Z is a random vector. Indeed from (2.6) it follows

that

$$Var\{\hat{\theta}\} = M^{-1}Var\{Y\}M^{-1} = M^{-1} = D,$$

where we have used $Var\{Y\} = M$ (see also (2.5)), which comes from the fact that

$$E\left[\left(Y_\alpha - E(Y_\alpha)\right)\left(Y_\beta - E(Y_\beta)\right)\right] = E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma^{-4}\varepsilon_i\varepsilon_j f_\alpha(x_i)f_\beta(x_j)\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{ij}\sigma^{-2}f_\alpha(x_i)f_\beta(x_j) = \sum_{i=1}^{n}\sigma^{-2}f_\alpha(x_i)f_\beta(x_i),$$

for the covariance of the α-th and β-th components of the $m \times 1$ vector Y with an introduction of the factor r_i in the case of repeated observations, i.e., in general,

$$\sum_{j=1}^{r_i} E\left[\left(Y_{\alpha j} - E(Y_{\alpha j})\right)\left(Y_{\beta j} - E(Y_{\beta j})\right)\right] = \sum_{i=1}^{n} r_i\sigma^{-2}f_\alpha(x_i)f_\beta(x_i).$$

The diagonal elements of D are variances of the parameters' estimators, and the off-diagonal elements of D are their covariances. Generally, the information matrix is calculated directly from the design points and then inverted to obtain D. Note that neither M nor D depend directly on the results of observations y_i. The variance σ^2 is either known or assumed; however, whether σ^2 is known or unknown is not important at this stage because the results do not depend on σ^2.

Both Fedorov (1972) and Puckelsheim (1993) discuss many additional properties of the information matrix M. Among them are these very important proper-

ties:

1. The information matrix is symmetric and non-negative definite.

2. The information matrix is additive and is the sum of the individual information matrices for each observation (design point).

Let the collection

$$\xi = \{x_i, p_i\}_1^n,\ p_i = r_i/N, N = \sum_{i=1}^{n} r_i \tag{2.7}$$

be called the design of an experiment. We can introduce the normalized information matrix

$$M(\xi) = \sigma^2 N^{-1} M = \sigma^2 \sum_{i=1}^{n} p_i f(x_i) f^T(x_i). \tag{2.8}$$

If the set of designs is extended to the set of all possible probability measures $\xi(dx)$ on a compact (for simplicity) set X, then the normalized information matrix is defined as

$$M(\xi) = \int_X f(x) f^T(x) \xi(dx), \int_X \xi(dx) = 1. \tag{2.9}$$

Unless it would otherwise be ambiguous, we omit the word "normalized".

Let M_1, and M_2 be the information matrices of two experiments made according to designs ξ_1 and ξ_2, respectively. Then, because of the additivity of M we can write

$$M = M_1 + M_2 \Leftrightarrow NM(\xi) = N_1 M(\xi_1) + N_2 M(\xi_2),$$

and dividing both sites by N we have

$$M(\xi) = (1-\alpha)M(\xi_1) + \alpha M(\xi_2),$$

where

$$\alpha = \frac{N_2}{N_1 + N_2}, \quad \text{and } \xi = (1-\alpha)\,\xi_1 + \alpha\,\xi_2.$$

Thus, the set of normalized information matrices is convex (see more details in Chapter 2.2). Note that additivity is a property of information matrices only with uncorrelated observations. Calculations and proofs are simpler because of it.

If there exists prior information about the estimated parameters which can be expressed in terms of a prior distribution with mean θ_0 and variance-covariance matrix D_0, then the combined information matrix is the following sum:

$$M_{tot} = M + M_0, \tag{2.10}$$

where $M_0 = D_0^{-1}$. The variance-covariance matrix of the parameter estimators is calculated as $D = M_{tot}^{-1}$. The updated estimator for θ may be presented as

$$\hat{\theta}_{tot} = (M + M_0)^{-1}(M\hat{\theta} + M_0\theta_0), \tag{2.11}$$

where $\hat{\theta}$ is defined by (2.6).

Formula (2.10) contains a hint of how to regularize the estimation problem

when the original information matrix is singular: one has to use an "uncorrected" matrix

$$M(c) = M + cM_0,\ M_0 > 0,\ \ c > 0$$

where frequently $M_0 = I$, and to analyze the behavior of $M^{-1}(c)$ as a function of c. This approach is convenient when a linear combination $\gamma = L\theta$ can be uniquely estimated for a given M. In this case (c.f., Albert (1972))

$$\text{Var}\ \hat{\gamma} = \lim_{c \to 0} L(M + cM_0)^{-1} L^T,$$

where $\hat{\gamma}$ is the best linear unbiased estimator of γ.

A number of authors discuss theory and applications for the varied uses of prior information in experimental design; among them are Atkinson et al. (1982), Atkinson and Donev (1992), Cook and Nachtsheim (1982), Fedorov and Atkinson (1988), Fedorov and Hackl (1997), Khuri (1984), Pilz (1983, 1989), and Puckelsheim (1993), Puckelsheim (1993) also provides a summary of authors publishing in the Bayesian area, and Pilz (1991) provides many references and detailed discussions of the Bayesian model.

2.2 Elements of Convex Design Theory

More about information matrices. Convex design theory provides the basis for

construction of experimental designs which are optimal with respect to some criterion, for example, $\ln|D|$ (D-criterion) or tr $AD, A \geq 0$ (linear criterion). The properties of the information matrix M, the variance-covariance matrix D, the inter-relationships of M and D, and the means of mathematically manipulating functions of M or D are the subjects of convex design theory. The theory is derived from such mathematical areas as the theory of convex sets, matrix algebra, and optimization theory.

To formulate results we need the following definitions and a simple, but important theorem.

Definition 2.1. The set S is *convex* if there exists a point $u \in S$ such that $u = \alpha u_1 + (1-\alpha)u_2$, with $u_i \in S$, $0 \leq \alpha \leq 1$.

Points, straight lines, and spheres are some examples of convex sets.

Definition 2.2 The set S^* of points $= \{u : u = \sum_{i=1}^{k} \alpha_i u_i,\ u_i \in S,$ $\alpha_i \geq 0,\ \sum_{i=1}^{k} \alpha_i = 1\}$ is called the *convex hull* of the set S.

The properties of the information matrix are conveniently summarized in the following theorem (c.f. Cook and Fedorov (1995), Fedorov (1997)):

Theorem 1 *1. For any design ξ the information matrix $M(\xi)$ is a symmetric non-negative definite matrix.*

2. The matrix $M(\xi)$ is singular (i.e., $|M(\xi)| = 0$), if the supporting set of the design ξ contains less than m points (m is the number of unknown parame-

ters).

3. *The set of matrices $M(\xi)$, corresponding to all possible designs, is convex.*

4. *For any design ξ the matrix $M(\xi)$ can be represented in the form*

$$M(\xi) = \sum_{i=1}^{n_0} p_i f(x_i) f^T(x_i),$$

where

$$n_0 \leq [(m+1)m/2] + 1, \quad 0 \leq p_i \leq 1, \sum_{i=1}^{n} p_i = 1.$$

Statements (1) through (3) follow directly from the definition of the information matrix. Statement (4) follows from Caratheodory's theorem, which says that any element of the convex hull may be represented as a convex linear combination of not more than $n_0 = (\dim u + 1)$ elements from S (see Definition 2.2). Any information matrix because of its symmetry is completely defined by $m(m+1)/2$ elements. Combining this result with the fact that

$$M(\xi) = \sum_{i=1}^{n} p_i M(x_i), \quad M(x) = f(x) f^T(x),$$

we arrive at the concluding statement.

Properties of optimal designs. Let the function $\Psi(M)$ be the "criterion of optimality" i.e., it quantitatively describes the objective(s) of a practitioner. A

design ξ^* is called Ψ- optimal if

$$\xi^* = \arg\min_{\xi} \Psi[M(\xi)], \quad \int_X \xi(dx) = 1. \tag{2.12}$$

That is, ξ^* is the design that minimizes the function $\Psi[M(\xi)]$ over the set Ξ of all possible design measures $\xi(dx)$ defined on X.

Assume that

1. the design set X is compact;

   That is, the set X is closed and bounded, has a finite subcover for every open cover, to name a few properties of compact sets.

2. the basis functions $f(x)$ are continuous functions in $X, f \in R^m$;

3. $\Psi(M) \leq \Psi(M+\Delta), M \geq 0, \ \Delta \geq 0$, where M and Δ are nonnegative definite, and Ψ is convex; the above relationship promises that if non-negative matrix Δ is added to M, then Ψ does not decrease.

4. there exists a constant q such that

$$\{\xi : \Psi[M(\xi)] \leq q < \infty\} = \Xi(q) \neq \phi; \quad \Xi(q) \subseteq \Xi$$

   That is, the set of designs with a finite value of the objective function is non-empty.

5. for any $\xi \in \Xi(q)$ and $\overline{\xi} \in \Xi(q)$,

$$\Psi\left[(1-\alpha)M(\xi) + \alpha M(\overline{\xi})\right] = \Psi\left[M(\xi)\right] + \alpha \int \psi(x,\xi)\overline{\xi}(dx)$$

$$+\tau(\alpha, \xi, \overline{\xi}),$$

$$\text{where } \lim_{\alpha \to 0} \frac{\tau(\alpha,\ \xi,\ \overline{\xi})}{\alpha} = 0.$$

The above assumption means the existence of the directional derivative

$$\lim_{x \to 0^+} \frac{1}{\alpha}[\Psi\{(1-\alpha)M(\xi) + \alpha M(\overline{\xi})\} - \Psi\{M(\xi)\}],$$

which has a rather special form. However, this assumption is fulfilled for most optimality criteria used in practice (see Table 2.1).

The previous assumptions are needed in order to prove the following optimality theorem.

Theorem 2 *1. For any optimal design there exists a design with the same information matrix which contains not more than $m(m+1)/2$ supporting points.*

2. A necessary and sufficient condition for a design ξ^* to be optimal is

$$\min_x \psi(x, \xi^*) \geq 0,$$

3. *The set of optimal designs is convex.*

4. *The function $\psi(x, \xi^*)$ achieves zero almost everywhere in the set of support points for the design measure ξ^*.*

The first statement of the theorem provides an upper bound on the number of design points in the information matrix M of an optimal design. Complemented with the last statement of Theorem 2, statement 1 allows us to find an optimal design with a relatively small number of supporting points. For instance, if an optimal design is constructed and contains more design points than this upper bound allows, then another design exists which has $m(m+1)/2$ design points or fewer. The second statement on the necessary and sufficient condition for optimality is used for analytical derivation of optimal designs, which are constructed using some intuitive approaches, and for verification of the optimality of designs. Statement 2 is also crucial for the development of numerical algorithms. The third statement helps to construct optimal designs using, for instance, symmetry properties of basis functions, optimality criterion, and the design set X. The fourth statement is used to verify that the design points and weights constructed by an algorithm form an optimal design solution.

Various versions of this theorem were proved by Fedorov and Malyutov (1972) and Whittle (1973). Additional comments appear in Puckelsheim (1993), Fedorov (1997), Atkinson and Fedorov (1984), and Puckelsheim (1993). In Table 2.1 we

give functions $\phi(x, \xi)$ and constants C for the most popular criteria. Criteria other than the D- and linear criteria are not within the scope of this dissertation.

The sensitivity function

$$\phi(x, \xi) = c - \psi(x, \xi)$$

shows how much the existing design measure added to a point x can improve the design ξ. For the D-criterion, the sensitivity function coincides with the variance of the estimated response, i.e.,

$$\phi(x, \xi) = d(x, \xi) = f^T(x) D(\xi) f(x).$$

This fact leads to the most-quoted equivalence theorem by Kiefer and Wolfowitz (Kiefer (1959)), which is a particular case of Theorem 2.

Theorem 3 *The following three designs are equivalent for the D criterion:*

1. *the design* ξ^* *minimizes* $|D|$ *(maximizes* $|M|$*);*

2. *the design* ξ^* *minimizes* $max_x d(x, \xi)$*;*

3. $max_x d(x, \xi^*) = m$.

This theorem proves that finding a design which satisfies any one of the three stated optimality conditions will automatically satisfy the remaining two. This

Table 2.1: Sensitivity Function $\phi(x,\xi)$ and Constants C for Various Optimality Criteria $\Psi(\xi)$.

$\Psi(\xi)$	$\phi(x,\xi)$	C
$\log\|D\|$, $D = M^{-1}$	$d(x) = f^T(x)Df(x)$	m
$\log\|D_\ell\|$ *	$d(x) - d_k(x)$ $d_k(x) = f_k^T(x)D_k f_k(x)$, $f^T(x) = (f_\ell^T(x), f_k^T(x))$	k
$trAD$, $A \geq 0$	$f^T(x)DADf(x)$	$trAD$
$d(x_0)$	$d^2(x,x_0)$ $d(x,x_0) = f^T(x)Df(x_0)$	$d(x_0)$
$\int_Z d(x)dx$	$\int_Z d^2(x,z)dz$	$\int_Z d(x)dx$
trD^γ	$f^T(x)D^{\gamma+1}f(x)$	trD^γ
$\lambda_{\min} = \lambda_{\min}(M)$ $= \lambda_{\max}^{-1}(D)$	$\sum_{i=1}^{\alpha} \pi_i (f^T(x)P_i)^2$ $\gamma_{\min} P_i = MP_i$, α is a multiple of $\lambda_{\min}$, $\sum_{i=1}^{\alpha} \pi_i = 1$, $0 \leq \pi_i \leq 1$	$\lambda_{\min}$

* D_ℓ is a submatrix of D corresponding to ℓ parameters, $k = m - \ell$.

means that solving the problem of finding the design that minimizes the maximum standardized variance $d(x,\xi) = f^T(x)D(\xi)f(x)$ is equivalent to finding the design which minimizes $|D|$. We want to emphasize that this is true only for cases when observational errors are homogeneous. If $\sigma^2(x) = E[\varepsilon^2(x)]$ is not constant, then $d(x,\xi)$ must be replaced by $\sigma^{-2}(x)d(x,\xi^*)$ in statements 2 and 3.

2.3 Optimal Design in the Presence of Prior Information

In constructing optimal designs for large networks, we experienced a serious problem with ill-conditioning of the information matrices. To overcome that difficulty, there exist two approaches: using some prior information or regularizing the corre-

sponding optimization problem (c.f. Fedorov (1997)). From a formal perspective, both approaches are very similar and are based on adding some positive definite matrix to the original information matrix. In what follows (see Chapter 4) we prefer to work in the framework of an approach based on the use of prior information and, therefore, will restrict this section to that discussion.

Let prior information be described by a prior distribution with a given θ_0 and dispersion matrix D_0, which is assumed to be regular (see also Chapter 2.1). Combining this information with the information gained in an experiment made according to design ξ, we have the normalized version of (2.11).

$$\hat{\theta}_{tot} = M_{tot}^{-1}(\xi)\left[M(\xi)\hat{\theta} + M_0\theta_0\right], \tag{2.13}$$

and

$$Var\{\hat{\theta}_{tot}\} = \sigma^2 N^{-1}(M(\xi) + M_0), \tag{2.14}$$

where

$$M_{tot}(\xi) = M(\xi) + M_0, \text{ and } M_0 = \sigma^{-2}ND_0^{-1}.$$

Note that the normalized total information matrix depends now on σ^2 and N.

An optimal design is now defined as

$$\xi_B^* = \arg\min_{\xi} \Psi\left[M_{tot}(\xi)\right], \tag{2.15}$$

and is frequently referred to as a Bayesian optimal design. Optimization problem (2.15) may be considered a very special case of (2.12) if one notices that

$$\Psi\left[M_{tot}(\xi)\right] = \Psi\left[M(\xi) + M_0\right] = \Psi'\left[M(\xi)\right]. \tag{2.16}$$

We distinguish this case from others for two reasons. First, the explicit handling of prior information leads to relatively simple and sound formulas. Second, we needed to develop numerical algorithms for this case, and there are no known sources of computer software that allow the construction of optimal designs with prior information.

To develop the necessary results, we needed to find the sensitivity functions for the criteria described in (2.16) and then to apply Theorem 2.

Let us start with the D-criterion, i.e.,

$$\Psi\left[M_{tot}(\xi)\right] = -\ln\left|M(\xi) + M_0\right|.$$

Noting that

$$\lim_{\alpha\to 0}\frac{\partial}{\partial\alpha}\ln\left|M\left[(1-\alpha)\xi + \alpha\bar{\xi}\right] + M_0\right|$$

$$= tr\left[M(\xi) + M_0\right]^{-1}\left[M(\bar{\xi}) - M(\xi)\right]$$

$$= \int_X f^T(x)\left[M(\xi) + M_0\right]^{-1} f(x)\bar{\xi}(dx) - tr\left[M(\xi) + M_0\right]^{-1} M(\xi), \tag{2.17}$$

we conclude that

$$\psi(x,\xi) = trM_{tot}^{-1}(\xi)M(\xi) - d_{tot}(x,\xi), \tag{2.18}$$

where

$$\phi(x,\xi) = d_{tot}(x,\xi) = f^T(x)\left[M(\xi) + M_0\right]^{-1} f(x).$$

The latter function is the variance of the best linear unbiased estimator of the response function if prior information is included. Thus, observations must be allocated at points where this variance might be greatest. The first term in (2.18) equals m if $M_0 = 0$.

Similarly for the linear criterion, when $\Psi[M_{tot}(\xi)] = trAM_{tot}^{-1}(\xi)$, we can derive that

$$\lim_{\alpha\to 0}\frac{\partial}{\partial\alpha} trA\left[M[(1-\alpha)\xi + \alpha\bar{\xi}] + M_0\right]^{-1}$$

$$= trAM_{tot}^{-1}(\xi)M(\xi)M_{tot}^{-1}(\xi) - \int_X f(x)^T M_{tot}(\xi)AM_{tot}(\xi)f(x)\bar{\xi}(dx). \tag{2.19}$$

Thus,

$$\psi(x,\xi) = trAM_{tot}^{-1}(\xi)M(\xi)M_{tot}^{-1}(\xi) - f^T(x)M_{tot}^{-1}(\xi)AM_{tot}^{-1}(\xi)f(x) \tag{2.20}$$

or

$$\phi(x,\xi) = f^T(x)M_{tot}^{-1}(\xi)AM_{tot}^{-1}(\xi)f(x). \tag{2.21}$$

Function (2.21) does not have so clear an explanation as the sensitivity function

for the D-criterion. As will be seen subsequently, both functions are essential for the development of numerical search procedures for constructing optimal designs.

Using (2.20) and (2.21) it is possible to trace what kind of changes to make in Table 2.1 for criteria which include prior information. This dissertation is confined to the D- and linear optimality criteria.

2.4 Numerical Methods of Optimal Design Construction

The definition of the sensitivity function hints about how to improve an existing current design ξ. Let there be a design ξ_s that must be improved. For instance, we may select a design

$$\overline{\xi} = (1-\alpha)\xi_s + \alpha\xi', \;\; 0 < \alpha < 1.$$

It is expedient to select a "correcting" design ξ' such that

$$\xi' = \arg\min_{\xi} \Psi\left[M(1-\alpha)\xi_s + \alpha\xi\right]. \tag{2.22}$$

Noting that (see assumption 5 from Chapter 2.2):

$$\Psi\left[M(1-\alpha)\xi_s + \alpha\xi\right] \simeq \Psi\left[M(\xi_s)\right] + \alpha \int \psi(x,\xi_s)\xi(dx),$$

we may replace (2.22) by

$$\xi' = \arg\min_{\xi} \int \psi(x, \xi_s)\xi(dx) \tag{2.23}$$

or

$$\xi' = \arg\min_{\xi} \int \phi(x, \xi_s)\xi(dx) \tag{2.24}$$

when α is small enough.

Both (2.23) and (2.24) have a very simple solution. Namely ξ' must be atomized at the point x_s

$$\begin{aligned} x_s &= \arg\min_{x \in X} \psi(x, \xi_s) \\ &= \arg\min_{x \in X} \phi(x, \xi_s). \end{aligned} \tag{2.25}$$

Note that in general a few different solutions may exist for the above problem.

Thus, we have to increase weights p_i at points x_i where $\phi(x, \xi)$ reaches its maximum and decrease weights at points (if there are any) where $\phi(x, \xi)$ is smallest. If we repeat the procedure sufficiently many times we may get close to an optimal design. For a more accurate statement, see Theorem 4. More formally, the iterative procedure may be described as follows:

1. There is a design $\xi_s \in \Xi(q)$. Find

$$x_s = \arg\max\{\phi(x^+, \xi_s),\ \phi(x^-, \xi_s)\},$$

where

$$x^+ = \arg\max_{x \in X} \phi(x, \xi_s)$$

and

$$x^- = \arg\min_{x \in X_s} \phi(x, \xi_s).$$

The set X_s consists of all points with non-zero weights of design ξ_s, i.e., $X_s =$ supp ξ_s.

2. Choose $0 < \alpha_s < 1$ and construct

$$\xi_{s+1} = (1 - \alpha_s)\xi_s + \alpha_s \xi(x_s),$$

where $\xi(x_s)$ is a design measure atomized at x_s.

Authors such as Atkinson and Donev (1992), Fedorov (1972), and Silvey (1980) discuss how to select a sequence α_s to guarantee the convergence of $\Psi[M(\xi_s)]$ to $\min_\xi \Psi[M(\xi)]$, i.e., to guarantee weak convergence of $\{\xi_s\}$. The following three rules are most popular:

$$\lim_{s\to\infty} \alpha_s = 0, \sum_{s=0}^{\infty} \alpha_s = \infty; \tag{2.26}$$

$$\alpha_s = \arg\min_{\alpha} \Psi\left[M((1-\alpha)\xi_\alpha + \alpha\xi(x_s))\right]; \tag{2.27}$$

$$\alpha_s = \begin{cases} \alpha_{s-1}, & \text{if } \Psi\left[M((1-\alpha_{s-1})\xi_s + \alpha_{s-1}\xi(x_s))\right] < \Psi\left[M(\xi_s)\right]), \\ \alpha_{s-1}/\gamma, & \text{otherwise, with } \gamma > 1. \end{cases} \tag{2.28}$$

An extensive discussion of the general properties of the iterative procedure can be found in Fedorov (1972, 1997) and Atkinson and Donev (1992). The same authors also provide detailed recommendations on computer programs to implement the algorithms.

The following result follows from the general results presented by Fedorov (1972) in Chapter 3.2 and combined with the specific properties of the optimality criteria introduced in Chapter 2.2.

Theorem 4 *For any sequence from (2.26), the iterative procedure is weakly converging, i.e.,*

$$\lim_{s\to\infty} \Psi\left[M(\xi_s) + M_0\right] = \min_{\xi} \Psi\left[M(\xi) + M_0\right]. \tag{2.29}$$

The proof for cases (2.27) and (2.28) is based on the monotonicity and boundness of the sequence $\{\Psi[M(\xi_s)]\}$, which result in convergence to some Ψ^*, and subsequent application of Theorem 2.

For case (2.16) the proof is slightly more complicated (compare with Wu and

Wynn (1978)) but is still mainly based on the use of Theorem 2.

Note that Theorem 4 is stronger than the corresponding results from Fedorov (1997), Chapter 3.2, where for some sequences $\{\alpha_s\}$ there exists a possibility that

$$\lim_{s\to\infty} \Psi\left[M(\xi_s)\right] \to \infty. \tag{2.30}$$

This can be explained by the presence of the prior information matrix M_0, which regularizes the optimization problem by excluding singular sequences which lead to (2.30).

In computer implementations of the iterative procedure in this dissertation, we also used the fact that the set of candidate points or design set X for computer networks is a discrete set: it is a collection of sites (nodes) or routes (edges) to be monitored or observed. The intermediate optimization problems (see stage 1 of the iterative procedure) are rather simple and fast. This is especially true in cases when it is possible to compute the functions $f(x)$ for all $x \in X$ and store these values in memory.

In all cases, we recommend starting with $\alpha_0 = N^{-1}$, where N is a number of observations which the practitioner considers reasonable. If rule (2.28) is used, then the iterative procedure coincides as long as α_0 remains constant with the so-called exchange algorithm (c.f. Mitchell (1974)). We recommend selecting $\gamma > 0$ so that γ/α_0 is integer. The quantity γ can be changed with s, but it is important

that $\alpha_{s-1}/\gamma_s \to 0$. More details are given in Appendix A.

Chapter 3

Optimal Design with Correlated Observations

3.1 Existing Results for Optimal Design with Correlated Observations

The problem of correlated observations can effect the optimality and efficiency of a design (see Fedorov (1996), Kiefer and Wynn (1984), and Silvey 1980)). Fedorov (1996) provides an extensive account of research in this area. This section highlights only those issues of design with correlated observations that are relevant to the properties of network performance data and the solution algorithms proposed.

An important difference between optimal design theory for the standard (i.e., with uncorrelated observations) linear model and theory for a model with cor-

related observations is that the information matrix is guaranteed to be additive only for the former case. The additivity of the non-normalized information matrix leads to the convexity of the normalized information matrix $\underline{M}(\xi)$ with respect to $\xi \in \Xi$ and, therefore, meaningful convexity. The optimization problem,

$$\xi^* = \arg\min_{\xi} \Psi[M(\xi)], \tag{3.1}$$

can then be introduced.

If the covariance functions

$$V(x, x') = E[(y(x) - \theta^T f(x))(y(x') - \theta^T f(x'))] \tag{3.2}$$

or the covariance matrix $V(\xi_N)$ are given, then the least squares estimator for the parameters in the case of correlated observations is

$$\hat{\theta} = \underline{M}^{-1}(\xi_N) F(\xi_N) V^{-1}(\xi_N) Y, \tag{3.3}$$

where

$$\underline{M}(\xi_N) = F(\xi_N) V^{-1}(\xi_N) F^T(\xi_N), \tag{3.4}$$

$$V_{ij} = V(x_i, x_j),$$

$$F(\xi_N) = (f(x_1), ..., f(x_N)), \text{ and } Y^T = (y_1, ..., y_N).$$

As before, N is the number of observations. Note that unlike (2.4), (3.4) is not the sum of the information matrices of individual observations.

Note that if $x \in X$ and the design set X is compact, then (with the exception of degenerate cases)

$$\lim_{N\to\infty} N^{-1}\underline{M}(\xi_N) = 0. \tag{3.5}$$

Thus, we cannot use the concept of the normalized information matrix as it was used in Chapter 2. Instead of a relatively simple optimization problem (2.4) we now solve the discrete optimization problem

$$\xi_N^* = \arg\min_{\xi_N} \Psi[M(\xi_N)], \tag{3.6}$$

which cannot be directly embedded in convex optimization theory. Non-additivity of the information matrix is the main cause. Additionally, the set of all possible discrete designs ξ_N is not convex, i.e., the convex combination of ξ_{1N} and ξ_{2N}

$$\xi = (1-\alpha)\xi_{1N} + \alpha\xi_{2N}$$

does not lead to a discrete design with N points. So, either we pursue (3.6) with all its difficulties or we develop new optimization ideas that can fit in the framework of convex optimization theory. The latter is the subject of this dissertation.

A few efforts to solve (3.6) or some approximate versions are discussed briefly.

Brimkulov et al. (1986) proposed the use of a straightforward generalization of the iterative procedures from Chapter 2.4 to construct ξ_N^*. To accomplish that they derived the following recursion formula for the information matrix $\underline{M}$:

$$\underline{M}(\xi_{N+1}) = \underline{M}(\xi_N) + w(x, \xi_N)\varphi(x, \xi_N)\varphi^T(x, \xi_N), \tag{3.7}$$

where

$$w^{-1}(x, \xi_N) = V(x, x) - V^T(x, \xi_N)V^{-1}(\xi_N)V(x, \xi_N) \tag{3.8}$$

and

$$\varphi(x, \xi_N) = f(x) - F(\xi_N)V^{-1}(\xi_N)V(x, \xi_N). \tag{3.9}$$

A recursive formula for $|\underline{M}(\xi_N)|$ is

$$|\underline{M}(\xi_{N+1})| = |\underline{M}(\xi_N)|[1 + w(x, \xi_N)\varphi^T(x, \xi_N)\underline{M}^{-1}(\xi_N)\varphi(x, \xi_N)]. \tag{3.10}$$

Then, in an iterative procedure for D-optimal design construction, the point x_{N+1}^+ to be added to the set ℓ of design points is

$$x_{N+1}^+ = \arg\min_{x \in X}[w(x, \xi_N)\varphi^T(x, \xi_N)\underline{M}^{-1}(\xi_N)\varphi(x, \xi_N)]. \tag{3.11}$$

The point x_{N+1}^{-} to be deleted from the set of design points is

$$x_{N+1}^{-} = \arg\min_{\ell}[w(x_\ell, \xi_{N+1})\phi^T(x_\ell, \xi_{N+1})\underline{M}^{-1}(\xi_{N+1})\phi(x_\ell, \xi_{N+1})]. \qquad (3.12)$$

There is no published information about the proof of convergence of this iterative procedure to a D-optimal design. There are a few applied studies, however, in which it is shown that the procedure frequently leads to "good" designs. For example, Rabinowitz and Steinberg (1990) study the problem of selecting sites for a seismographic network using this exchange-type algorithm. They constructed relatively efficient designs that showed improvement over the use of D-optimal designs which were constructed without covariance structure. A number of authors have studied autoregressive models of correlated observational errors, including Kiefer and Wynn (1984) and Bickel and Herzberg (1979). Autoregressive examples are not within the scope of this dissertation and are not considered further here.

Sacks and Ylvisaker (1966, 1968, 1970a, 1970b) suggest a very interesting and mathematically elegant approach for correlated observations which is based on theory close to classical integral approximation theory. Let $X = \{x : a \leq x \leq b\}$. Following spatial sampling theory ideas (for example, Matèrn (1986)), Sacks and

Ylvisaker consider these discrete designs

$$\xi_N = \{x : x_i = \arg[\int_a^x h(x)dx = \frac{i-1}{N-1}]\}, \tag{3.13}$$

where $h(x)$ is a density function which must be selected to guarantee the asymptotic optimality of ξ_N $(N \to \infty)$ in the sense of some optimality criteria. For instance, for the D-criterion one has to try to maximize

$$\lim_{N\to\infty} \frac{\ln|\underline{M}(\xi_N)| - \ln|\underline{M}|}{\max_{\xi_N}|\underline{M}(\xi_N)| - \ln|\underline{M}|}, \tag{3.14}$$

or, equivalently,

$$h^*(x) = \arg\max_{h(x)} \lim_{N\to\infty} \frac{\ln|\underline{M}(\xi_N)| - \ln|\underline{M}|}{\max_{h(x)}|\underline{M}(\xi_N)| - \ln|\underline{M}|}, \tag{3.15}$$

where the information matrix $\underline{M}$ corresponds to the case with "continuous" observations, i.e., with observations collected at all of $x \in X$. Obviously, $\ln|\underline{M}|$ is the maximal information which may be gained in the considered problem.

Various generalizations and discussions of Sacks and Ylvisaker's ideas can be found in Benhenni and Cambanis (1992), Cambanis (1985), Cressie (1991), Fedorov (1996), and Wahba (1974). Unfortunately, we cannot apply the above results to the monitoring of computer networks because they are essentially based on the fact that x is a continuous variable and that distance may be introduced

in X. It is also important that the latter may be very close to zero for some neighboring points.

The results reported in the next section elaborate and further develop an idea briefly discussed in Fedorov (1996); see also Fedorov (1989) for earlier similar attempts. This idea is based on partitioning the random error into two components: "intrinsic" or "process" randomness and "observational" randomness:

$$\varepsilon(x) = u(x) + \varepsilon(x), \tag{3.16}$$

where the random variable $u(x)$ models the random behavior of the observed system, while $\varepsilon(x)$ describes various uncertainties of the observational process. It is assumed that the covariance between two observations $y(x)$ and $y(x')$, with $x \neq x'$, is explained entirely through $u(x)$ and $u(x')$, i.e.,

$$\text{Cov}\,[y(x), y(x')] = E[(y(x) - \theta^T f(x))(y(x') - \theta^T f(x'))] \tag{3.17}$$

$$= K(x, x') \tag{3.18}$$

$$= E[u(x)u(x')], \tag{3.19}$$

where,

$$E[u(x)] = E[\varepsilon(x)] = 0. \tag{3.20}$$

If $x = x'$, then

$$E[(y(x) - \theta^T f(x))^2] = K(x,x) + \sigma^2,$$

where $\sigma^2 = E[\varepsilon^2(x)]$. We assume that the variance of the observational errors is homogeneous. Otherwise, a simple transformation replaces a non-homogeneous case by a homogeneous one (compare with the comments to (2.1)). We use

$$V(x,x') = K(x,x') + \sigma^2 \delta_{xx'}, \tag{3.21}$$

where $\delta_{xx'} = 1, \text{if } x = x'$, and $\delta_{xx'} = 0$, otherwise. This notation makes results comparable to those considered in the beginning of this section.

Fedorov (1996) assumes that

$$K(x,x') = \sum_{\alpha=1}^{p} \lambda_\alpha \varphi_\alpha(x) \varphi_\alpha(x'), \tag{3.22}$$

where $p < \infty$; λ_α is the $\alpha - th$ eigenvalue of $K(x,x')$, and $\varphi(x)$ is the corresponding eigenfunction, i.e.,

$$\lambda_\alpha \varphi_\alpha(x) = \int_Z K(x,x') \varphi_\alpha(x') dx', \tag{3.23}$$

where $X \subset Z$, and Z is some regular region (e.g., interval, square, circle, etc.). In the discussion that follows we prefer to call $K(x,x')$ the "covariance kernel" to avoid confusion with $V(x,x')$, which is named the covariance function. The partition of (3.16) together with (3.21) and (3.22) allows replacing (2.1) by a

regression model with random coefficients (i.e., the second kind regression model):

$$y(x) = \theta_1^T f(x) + \theta_2^T \varphi(x) + \varepsilon(x), \tag{3.24}$$

where all components of θ_1 are deterministic, while

$$E(\theta_2) = 0, \quad E(\theta_2 \theta_2^T) = \Lambda, \text{ and}$$

$$\Lambda_{\alpha\beta} = \delta_{\alpha\beta} \lambda_\alpha.$$

Presentation (3.24) provides the opportunity to use convex design theory as follows. One introduces

$$M_{tot}(\xi) = \begin{pmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{21}(\xi) & M_{22}(\xi) + M_{022} \end{pmatrix}, \tag{3.25}$$

where

$$M_{11}(\xi) = \int_X f(x) f^T(x) \xi(dx),$$

$$M_{12}(\xi) = M_{21}^T(\xi) = \int_X f(x) \varphi^T(x) \xi(dx),$$

$$M_{22}(\xi) = \int_X \varphi(x) \varphi^T(x) \xi(dx),$$

$$M_{022} = \sigma^{-2} N \Lambda.$$

Matrix (3.25) is a replacement of the information matrix $M(\xi)$. Consequently, optimization problem (2.4) is replaced by

$$\xi^* = \arg\min_{\xi} \Psi[M_{tot}(\xi)]. \tag{3.26}$$

Optimization problem (3.26) can be considered in the framework of the convex design theory as it was done by Fedorov (1996).

In the following sections, we expand the results in these directions:

1. evaluate and analyze the role of remainder terms in approximation (3.22);
2. develop numerical procedures for optimization problem (3.26) for continuous and discrete weights.
3. illustrate our findings with some analytical and numerical results for various covariance kernels $K(x, x')$ (e.g., the Brownian bridge and the Poisson kernel); and
4. apply our findings to design a set of optimal monitoring sites related to the ESnet computer network (Chapter 5) and arbitrary computer networks (Chapter 4).

3.2 Optimal Design Based on Mercer's Expansion of Covariance Kernel

Model, estimators, criteria. To make presentation simpler we assume that in model (3.24) we have $\theta_1^T f(x) \equiv 0$ and

$$y(x) = u(x) + \varepsilon(x). \tag{3.27}$$

That is, there is no trend. Our main objective is to estimate the parameters θ_2 or to predict $u(x)$ at a given set of points. The material is presented in the manner of Fedorov and Flanagan (1998).

It is convenient to introduce the following matrices and vectors:

$$\{K(\xi_n)\}_{ii'} = K(x_i, x_{i'}), \tag{3.28}$$

$$\{K(x, \xi)\}_i = K(x, x_i), \tag{3.29}$$

$$Y^T = (y(x_1), ..., y(x_N)).$$

One can check that under model (3.27)

$$E(Y) = 0,\ E(YY^T) = V(\xi_N) = \sigma^2 I + K(\xi_N).$$

Let the prediction of $u(x)$ on a given set X_{pr} be the goal for a practitioner. The estimator

$$\hat{u}(x) = K^T(x, \xi_N)V^{-1}(\xi_N)Y \tag{3.30}$$

minimizes the variance of prediction, and

$$\begin{aligned} Var(u(x) - \hat{u}(x)) &= E[(u(x) - \hat{u}(x))^2] \\ &= K(x,x) - K^T(x, \xi_N)V^{-1}(\xi_N)K(x, \xi_N), \end{aligned} \tag{3.31}$$

given that $E(u(x) - \hat{u}(x)) = 0$, and where expectation is taken with respect to u and ε. Throughout this discussion, the use of any inverse matrix automatically assumes the existence of this matrix.

Without loss of generality we may assume that $E(u(x)) = u_0(x) = 0$ for all x. Otherwise, a simple transformation $u'(x) = u(x) - u_0(x)$ reduces any problem with nonzero $u_0(x)$ to the previous one. It is important that $u_0(x)$ be known.

To check the optimality of (3.30) (compare with Ripley (1981)), let us introduce

$$\tilde{u}(x) = L^T Y,$$

where L is an n x 1 vector, such that

$$E_{u,\varepsilon}\left(L^T Y - u(x)\right) = 0. \tag{3.32}$$

Subscripts u and ε indicate that expectation is taken with respect to both corresponding random variables. By definition

$$\begin{aligned} \mathrm{Var}_{u,\varepsilon}\tilde{u}(x) &= E_{u,\varepsilon}(L^T Y - u(x))^2 \\ &= E_{u,\varepsilon}(L^T Y Y^T L - 2L^T Y u(x) + u^2(x)) \\ &= L^T(\sigma^2 I + K(\xi_N))L - 2L^T K(x,\xi_N) + K(x,x). \end{aligned} \tag{3.33}$$

Minimizing (3.33), i.e., differentiating, with respect to L subject to (3.32), we have that

$$(\sigma^2 I + K(\xi_N))L = K(x,\xi_N). \tag{3.34}$$

We then have $L = (\sigma^2 I + K(\xi_N))^{-1}K(x,\xi_N)$, which leads to (3.30) if the matrix $V(\xi_N) = \sigma^2 I + K(\xi_N)$ is regular. Substituting (3.34) in (3.33) gives (3.31).

The most commonly used objective functions related to (3.30) are

$$\Psi_1(\xi_N) = \max_{x \in X_{pr}} Var\left(u(x) - \hat{u}(x)\right), \tag{3.35}$$

and

$$\Psi_2(\xi_N) = \int_{X_{pr}} Var\left(u(x) - \hat{u}(x)\right) dx, \tag{3.36}$$

where X_{pr} is a set of points where we want to know $u(x)$. If X_{pr} contains a finite number of points, then the integral in (3.36) is to be replaced by a sum.

The design problem may be stated as

$$\xi_N^* = \arg\min_{\xi_N} \Psi(\xi_N). \tag{3.37}$$

Most publications are concerned with the criterion $\Psi_2(\xi_N)$; see Cambanis (1985) and Fedorov (1996) for a survey of main results and further references. Standardly it is assumed that the observational error is negligible, i.e. $\sigma^2 = 0$, and the number of observing points is large enough to use asymptotical ($N \to \infty$) results.

We want to emphasize that it is essential for our approach that $\sigma^2 \neq 0$. On an intuitive level it means that starting with some p

$$\sigma^2 \gg \sum_{\alpha=p+1}^{\infty} \lambda_\alpha \varphi_\alpha(x)\varphi_\alpha(x') \tag{3.38}$$

for all $x \in X$; see also the concluding part of this section for more comments.

If (3.38) is valid, then model (3.24) can be replaced by its approximate version

$$y_i = \theta^T \varphi(x_i) + \varepsilon_i, \tag{3.39}$$

where the vector $\varphi^T(x) = \{\varphi_\alpha(x)\}_1^p$, the parameters θ are random, and

$$E(\theta) = 0, \quad E(\theta\theta^T) = \Lambda, \quad \Lambda_{\alpha\beta} = \lambda_\alpha \delta_{\alpha\beta}.$$

We skip subscript "2" used previously to make the notation simpler. In the frame of (3.39)

$$E(YY^T) = V(\xi_N) = \sigma^2 I + \Phi^T(\xi_N)\Lambda\Phi(\xi_N), \tag{3.40}$$

where $\Phi(\xi_N) = (\varphi(x_1), \ldots, \varphi(x_N))$.

Let us try to predict $u(x) = \theta^T\varphi(x_i)$ using the regression analysis technique. The best linear unbiased estimator for θ is

$$\hat{\theta} = \sigma^{-2}\left(\underline{M}(\xi) + \Lambda^{-1}\right)^{-1}\Phi(\xi_N)Y, \tag{3.41}$$

$$\underline{M}(\xi_N) = \sigma^{-2}\Phi(\xi_N)\Phi^T(\xi_N). \tag{3.42}$$

This estimator minimizes the dispersion matrix of the difference $\hat{\theta} - \theta$, and

$$D(\xi_N) = E\left[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\right] = (\underline{M}(\xi_N) + \Lambda^{-1})^{-1}. \tag{3.43}$$

Unbiasedness of $\hat{\theta}$ means here that $E(\hat{\theta}) = E(\theta) = 0$, where expectation in the first case is taken with respect to u and ε (compare, for instance, with Pilz (1991)).

Let us select $\tilde{u}(x) = \varphi^T(x)\hat{\theta}$ as a predictor for $u(x)$. On an intuitive level it is obvious that $\hat{u}(x)$ and $\tilde{u}(x)$ must coincide in the frame of approximation (3.39).

Indeed, using the identity

$$(A^{-1} + BB^T)^{-1} = A - AB(B^T AB + I)^{-1} B^T A, \tag{3.44}$$

one can verify that

$$\begin{aligned} & Var\left(\tilde{u}(x) - u(x)\right) = \varphi^T(x) D(\xi_N) \varphi(x) \\ = \; & K_p(x,x) - K_p^T(x,\xi_N) \left(\sigma^2 I + K_p(\xi_N)\right)^{-1} K_p(x,\xi_N) \\ = \; & C_p(x,\xi) = Var\left(\hat{u}(x) - u(x)\right) \end{aligned} \tag{3.45}$$

where

$$K_p(x,x') = \sum_{\alpha=1}^{p} \lambda_\alpha \varphi_\alpha(x) \varphi_\alpha(x'),$$

and the matrix $K_p(\xi)$ and the vector $K_p(x,\xi)$ are defined parts of $K(\xi)$ and $K(x,\xi)$, respectively.

Now (3.35) and (3.36) may be rewritten as

$$\Psi_1(\xi_N) = \max_{x \in X_{pr}} \varphi^T(x) D(\xi_N) \varphi(x) \tag{3.46}$$

and

$$\Psi_2(\xi_N) = \mathrm{tr} A D(\xi_N), \quad A = \int_{X_{pr}} \varphi(x) \varphi^T(x) dx. \tag{3.47}$$

Remainders. In (3.45) we neglected the fact that the set of the eigenfunctions

of $K(x,x')$ was truncated and we used only p of them. More thorough analysis shows that

$$
\begin{aligned}
&E\left[(\tilde{u}_p(x)-u(x))^2\right] = C_p(x,\xi_N) + R_p(x,\xi_N),\\
&R_p(x,\xi_N) \;=\; {}_pK(x,x)\\
&+K_p^T(x,\xi_N)\left[\sigma^2 I + K_p(\xi_N)\right]^{-1} {}_pK(\xi_N)\left[\sigma^2 I + K_p(\xi_N)\right]^{-1} K_p(x,\xi_N)\\
&-2{}_pK^T(x,\xi_N)\left[\sigma^2 I + K_p(\xi_N)\right]^{-1} K_p(x,\xi_N). \qquad (3.48)
\end{aligned}
$$

We introduce subscript p to emphasize that a particular structure depends upon the first p eigenvalues and eigenfunctions, such as $\tilde{u}_p(x) = \sum_{\alpha=1}^{p} \lambda_\alpha \varphi_\alpha(x)$. The front subscript p indicates dependence upon the complimentary set of eigenvalues and eigenfunctions, such as ${}_p\tilde{u}(x) = \sum_{\alpha=p+1}^{\infty} \lambda_\alpha \varphi_\alpha(x)$. The terms ${}_pK(x,x)$, ${}_pK(x,\xi_N)$ and ${}_pK(\xi_N)$ are the counterparts of $K_p(x,x), K_p(x,\xi_N)$ and $K_p(\xi_N)$, respectively.

Using (3.48) we can evaluate the remainder terms for various optimality criteria. For instance, for $\Psi_2(\xi_n)$ and $X_{pr} = Z$ we have

$$
\sum_{\alpha=p+1}^{\infty} \lambda_\alpha \le \int_Z R_p(x,\xi_N)dx \le \sum_{\alpha=p+1}^{\infty} \lambda_\alpha \left(1 + C(\xi_N)\right), \qquad (3.49)
$$

and there exists a constant $C > 0$ such that $C(\xi_N) \le C\sigma^{-4}$ for any design ξ_N. To

derive (3.49) one has to use the fact that

$$\int_Z {}_pK(x,x)dx = \sum_{\alpha=p+1}^{\infty} \lambda_\alpha \text{ and } \int_Z K_p(x,x_i)K_p(x,x_j)dx \le q < \infty$$

for all $x_i, x_j \in Z$. Note that the integral of the last term in (3.48) vanishes because of the orthogonality of eigenfunctions. Thus, selecting p sufficiently large we may neglect the remainder terms. Obviously, the larger σ values lead to the smaller admissible p.

Continuous Optimal Designs. In what follows we pursue a very transparent and simple idea. First, we formulate results for optimization problems (3.46) and (3.47) in terms of regression model (3.39), i.e., using our knowledge of Λ and $\varphi(x)$. At the second stage we translate our findings to the language of covariance kernels and best linear predictors.

Let us allow the possibility of repeated observations, and let $N = \sum_{i=1}^{n} r_i$, where r_i is the number of observations at point x_i. That is, for sufficiently large N we can can apply the concept of continuous designs similar to Chapter 2. Let

$$D(\xi) = \left[\sigma^{-2} N M(\xi) + \Lambda^{-1}\right]^{-1}, \tag{3.50}$$

where

$$M(\xi) = \int_X \varphi(x)\varphi^T(x)\xi(dx).$$

Now we can define an optimal design as

$$\xi^* = \arg\min_{\xi} \Psi(\xi), \tag{3.51}$$

It is expedient to note that, in general, the optimal design ξ^* depends upon N and σ^2 similar to some problems in the Bayesian approach to experimental design theory; see Chapter 2.4. When N is fixed, all the results developed in convex design theory may be used almost directly. For instance, for the linear criterion (3.47) we have:

Theorem 1 *The design ξ^* is linear optimal if and only if for all $x \in X$*

$$\phi_1(x, \xi^*) \leq trM(\xi^*)D(\xi^*)AD(\xi^*), \tag{3.52}$$

where $\phi_1(x, \xi) = \varphi^T(x)D(\xi)AD(\xi)\varphi(x)$, and equality takes place at all supporting points of ξ^*.

For criterion (3.46) in the simplest case when $X_{pr} = X$, we have the following analogue to the Kiefer-Wolfowitz theorem (see Chapter 2, Theorem 3).

Theorem 2 *1. The design ξ^* is minimax if and only if for all $x \in X$*

$$\phi_2(x, \xi^*) \leq trM(\xi^*)D(\xi^*), \tag{3.53}$$

where $\phi_2(x, \xi) = \varphi^T(x)D(\xi)\varphi(x)$, and equality takes place at all supporting points

of ξ^*.

2. Minimax designs coincide with D-optimal designs, i.e.

$$\xi^* = \arg\min_{\xi} |D(\xi)| .$$

Let us recall that in the above theorem $D(\xi) = (\sigma^{-2}NM(\xi) + \Lambda^{-1})^{-1}$; this is unlike the classical counterpart where $D(\xi) = M^{-1}(\xi)$. Both theorems are routine results following from convex design theory. Thus, in the framework of approximation (3.39) the design for correlated observations cases can be embedded in the well developed area of convex design theory.

Let us now move in the reverse direction: having the results stated in the above theorems and using presentation (3.40), let us try to re-formulate results in terms of the covariance kernels.

In what follows we consider only designs with a finite number of supporting points. This fact is not very restrictive because for every design with given matrix $M(\xi)$, there exists a design ξ' with a finite number of supporting points and exactly the same matrix $M(\xi')$ (see Chapter 2, Theorem 2). From identity (3.44) it follows that

$$D(\xi) = \Lambda - \Lambda\Phi(\xi)(J(\xi) + K_p(\xi))^{-1}\Phi^T(\xi)\Lambda, \tag{3.54}$$

where

$$J_{ij}^{-1}(\xi) = \sigma^{-2}N_i p_i \delta_{ij} \text{ and } \Phi(\xi) = (\varphi(x_1), \ldots, \varphi(x_n)) .$$

Combining (3.53) and (3.54) and introducing the function

$$C_p(x, x', \xi) = K_p(x, x') - K_p(x, \xi) \left(J(\xi) + K_p(\xi)\right)^{-1} K_p(x', \xi) \tag{3.55}$$

we come to the following result.

Theorem 3 *The design ξ^* minimizes the average variance of prediction if and only if for all $x \in X$*

$$\int_{X_{pr}} C_p^2(x, x', \xi^*) dx' \leq \int_X \int_{X_{pr}} C_p^2(x, x', \xi^*)\, \xi^*(dx) dx', \tag{3.56}$$

and equality holds at all supporting points of ξ^*.

For the minimax case with $X_{pr} = X$ we have the following theorem.

Theorem 4 *1. The design ξ^* is minimax if and only if for all its supporting points x_i^**

$$C_p(x_i^*, \xi^*) = \max_{x \in X} C_p(x, \xi^*). \tag{3.57}$$

2. Minimax designs coincide with D-optimal designs:

$$\xi^* = \arg\max_{\xi} \left|J(\xi) + K_p(\xi)\right| \left|J(\xi)\right|^{-1}. \tag{3.58}$$

To get (3.58) one has to notice that

$$\left|\sigma^{-2}NM(\xi)+\Lambda^{-1}\right| = \frac{|J(\xi)+K_p(\xi)|}{|\Lambda|\,|J(\xi)|}.$$

Optimization problem (3.58) may be considered as a maximization of the determinant of the variance-covariance matrix of observations. The idea that it can lead to good prediction was probably first stated by Shewry and Wynn (1987) in a different setting.

Example. Let $K(x,x')$ be the Poisson kernel, i.e.,

$$K(x,x') = \frac{1-\zeta^2}{1-2\zeta\cos 2\pi(x-x')+\zeta^2},$$

$$0 \le x, x' \le 1, 0 < \zeta < 1.$$

It is known (see, for instance, Kanwal (1971)), that

$$\lambda_0 = 1,\ \lambda_{2\alpha-1} = \lambda_{2\alpha} = \zeta^\alpha$$

and

$$\varphi_0(x) = 1,\ \varphi_{2\alpha-1} = \sqrt{2}\cos 2\alpha\pi x,\ \varphi_{2\alpha}(x) = \sqrt{2}\sin 2\alpha\pi x,\ \alpha = 1,2\ldots,$$

where the new index notation is related to what was previously used.

Let us start with the criterion $\Psi(\xi)$ and assume that $X_{pr} = X = Z$ and Z coincides with $[0,1]$. Let us recall that X_{pr} is the set of points where a practitioner is interested in estimation of $u(x)$; X is a set of points where observations can be made; and Z is the set of points used for the definitions of $K(x,x')$ and $\varphi(x)$. One can verify that the utility matrix A, which is defined in (3.47), is the identity matrix for $X_{pr} = [0,1]$. To find an optimal design let us follow a rather standard way in convex design theory: first, we make a guess about that design and then we verify, with the help of Theorem 1 (or Theorem 3), the correctness of this guess.

Noting that the correlation function depends only upon a distance between x and x' it is natural to consider designs with an equidistant allocation of supporting points as candidates for optimal designs. If there are p unknown parameters, i.e., $2\alpha + 1 \leq p$, then we may consider

$$\xi^* = \left\{\mu_i = p^{-1}, x_i = (i-1)/p, \;\; i = 1, \ldots, p\right\}. \tag{3.59}$$

If $x_1, \ldots, x_p$ are equidistant then

$$n^{-1} \sum_{i=1}^{n} \varphi_\beta(x_i)\varphi_{\beta\prime}(x_i) = \delta_{\beta\beta'},$$

$$\beta, \beta' = 0, 1, \ldots, 2\alpha - 1, 2\alpha, \ldots, 2\alpha_0 - 1, 2\alpha_0, \;\; n \geq p = 2\alpha_0 + 1.$$

From that we can conclude that the information matrix coincides with the identity

matrix, and

$$D(\xi^*) = \left[r^{-1}M(\xi^*) + \Lambda^{-1}\right]^{-1} = \begin{pmatrix} \frac{1+r}{r} & 0 & 0 & \dots & 0 \\ 0 & \frac{\zeta+r}{r\zeta} & 0 & \dots & 0 \\ 0 & 0 & \frac{\zeta^2+r}{r\zeta^2} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{\zeta^{\alpha_o}+r}{r\zeta^{\alpha_o}} \end{pmatrix}^{-1},$$

where $r = \sigma^2/N$. Consequently

$$\begin{aligned} \phi(x,\xi^*) \quad &= \varphi^T(x)D(\xi^*)AD(\xi^*)\varphi(x) = \left(\tfrac{r}{1+r}\right)^2 + \\ &+ \left(\tfrac{r\zeta}{\zeta+r}\right)^2 \left(2\cos^2 2\pi x + 2\sin^2 2\pi^{\alpha_0} x\right) + \dots \\ &+ \left(\tfrac{r\zeta^{\alpha_0}}{\zeta^{\alpha_o}+r}\right)^2 \left(2\cos^2 2\pi\alpha_0 x + 2\sin^2 2\pi\alpha_0 x\right) \\ &\equiv \left(\tfrac{r}{1+r}\right)^2 + 2\sum_{\alpha=1}^{\alpha_o} \left(\tfrac{r\zeta^\alpha}{\zeta^\alpha+r}\right)^2, \end{aligned} \tag{3.60}$$

and

$$\begin{aligned} &\mathrm{tr}\, M(\xi^*)D(\xi^*)AD(\xi^*) = \mathrm{tr}D^2(\xi^*) \\ &= \left(\frac{r}{1+r}\right)^2 + 2\sum_{\alpha=1}^{\alpha_0} \left(\frac{r\zeta^\alpha}{\zeta^\alpha+r}\right)^2. \end{aligned} \tag{3.61}$$

From Theorem 1, (3.60) and (3.61), it immediately follows that ξ^* is optimal, i.e., minimizes $\Psi_2(\xi)$. It is interesting to note that the structure of ξ^* does not depend upon N, σ^2 or ζ. If N is a total number of available sensors or sites to

monitor, then it is reasonable to select n equal N. Similar to the classical case, this design is optimal for any α_0 such that $2\alpha_0 + 1 \leq N$.

Similar exercises show that the proposed design ξ^* minimizes $\Psi_1(\xi)$.

3.3 Numerical Algorithms and Examples

Theorems 1 and 2 from the previous section lead immediately to numerical procedures that are similar to those discussed in Chapter 2.4. Actually, the corresponding algorithms are technically identical to algorithms developed for the Bayesian approach. At every *s-th* iteration of these algorithms one has to find either

$$\min_{x \in X_s} \zeta(x, \xi_s) \quad \text{or} \quad \max_{x \in X} \zeta(x, \xi_s), \tag{3.62}$$

where the sensitivity function ζ may coincide with ϕ_1 or ϕ_2, respectively, and $X_s = \text{supp}\, \xi_s$. The algorithm is simple, but it is necessary to know the eigenfunctions $\varphi(x)$.

Theorems 3 and 4 allow the development of numerical procedures that "directly" use the covariance kernel $K_p(x, x')$. For instance, the first order exchange algorithm can be written for the minimax criterion with $X_{pr} = X$ as follows:

Step a. There is a design ξ_s. Find

$$x_s^+ = \arg\max_{x \in X} C_p(x, \xi_s), \tag{3.63}$$

and construct $\xi_s^+ = \xi_s + \alpha_s \delta(x_s^+)$, where $\delta(x)$ is a probability measure atomized at x.

Step b. Find

$$x_s^- = \arg \min_{x \in X_s} C_p(x, \xi_s), \tag{3.64}$$

where $X_s = \operatorname{supp} \xi_s^+$, and construct $\xi_{s+1} = \xi_s - \alpha_s \delta(x_s^-)$.

The changes which must be done in the case of the linear criterion are evident. To guarantee convergence of the above iterative procedure, the sequence $\{\alpha_s\}$ may be chosen similarly to what was discussed in Chapter 2.4.

To use the above numerical procedure one has to find $K_p(x, x')$, and that may be rather difficult. If approximation (3.39) is valid, then $K_p(x, x')$, may be replaced by $K(x, x')$, which in this case is assumed to be known. As soon as it is done, the iterative procedure becomes a simple tool for "optimal" design construction. The word "optimal" is used in quotation marks to emphasize that a rigorous mathematical analysis of the limit behavior of ξ^* defined by (3.50) when $p \to \infty$ is still an open research problem.

If the eigenvalues and eigenfunctions of the covariance kernel are known, then the algorithms based on (3.62) are more practical: we need to operate with matrices of the size $p \times p$. The use of the iterative procedures based on the direct use of variance-covariance matrices leads to the necessity to operate with matrices of the size $n_s \times n_s$, where n_s is the number of support points in ξ_s.

Frequently, intermediate designs ξ_s contain a rather large number of support points, such that $n_s \gg p$, and that makes the latter procedure computationally more demanding than the first one.

Example. Let $K(x, x')$ be the covariance kernel corresponding to the Brownian bridge, i.e.

$$K(x, x') = \min(x, x')\,[1 - \max(x, x')]\,, \quad 0 \leq x,\ x' \leq 1.$$

It is known (see Kanwal (1971)) that

$$\lambda_\alpha = \alpha^{-2}\pi^{-2}, \varphi_\alpha(x) = \sqrt{2}\sin\alpha\pi x, \quad \alpha = 1, 2, \ldots.$$

We selected the minimax criterion with $X_{pr} = X = [0, 1], p = 9$, and initial design ξ_o with $x_i = i/20$, $i = 1, \ldots, 20$, and $\mu_i \equiv 0.05$. The iterative procedure based on (3.62) was run for three different values of $r = \sigma^2/N$: $1, 0.1$ and $r \to 0$. The last case corresponds to the standard D-optimal design. The constructed designs are shown in Fig. 3.1.

Fig. 3.2 and 3.3 present the sensitivity functions of those designs for the second and third case together with the corresponding functions for the initial design. Curiously enough the support points of optimal design (at least in the framework of accuracy which we managed to reach) for case $r \to 0$ coincide with

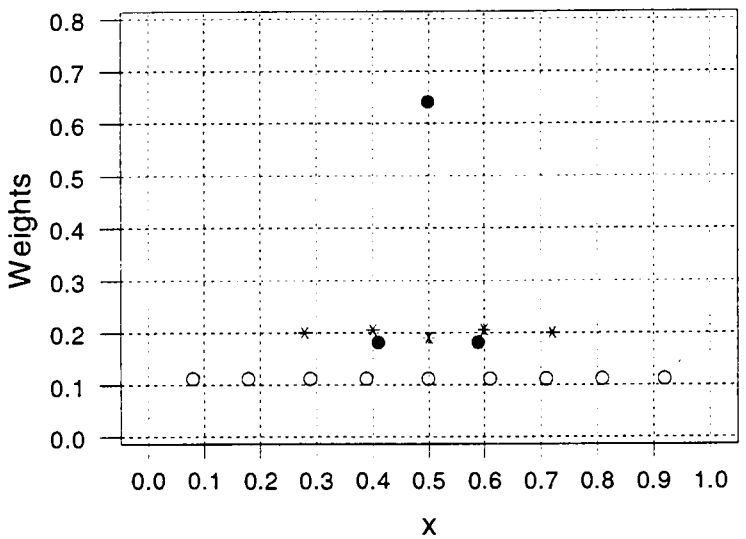

Figure 3.1: Optimal Designs. Solid Circles, Asterisks and Circles Stand for $r = 1, 0.1$ and $r \to 0$, Respectively.

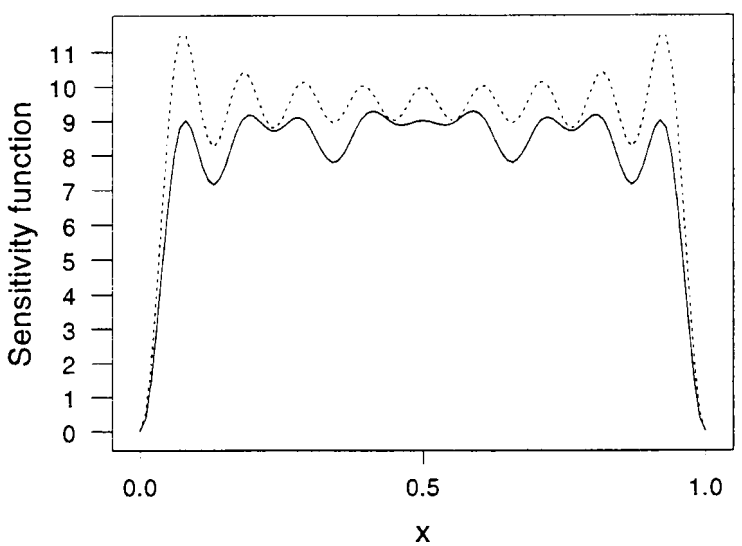

Figure 3.2: Sensitivity Functions for the Initial (dotted lines) and for the Constructed Optimal (solid lines) Design, $r \to 0$.

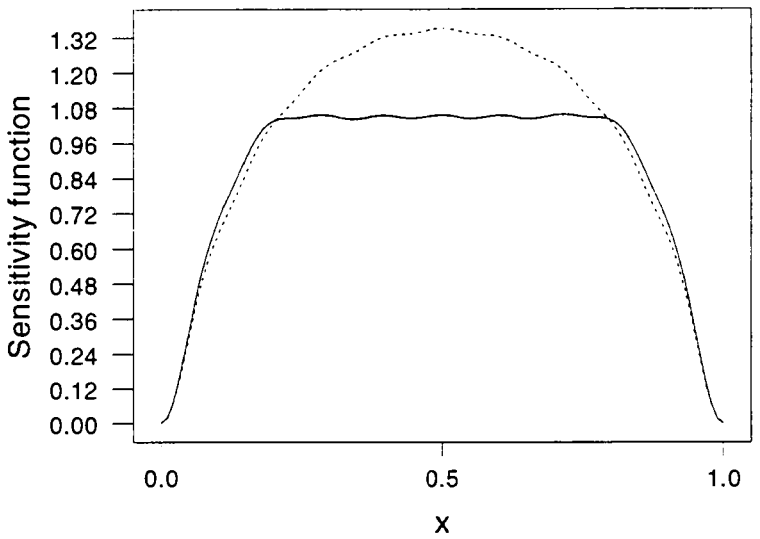

Figure 3.3: Sensitivity Functions for the Initial (dotted lines) and for the Constructed Optimal (solid lines) Design, $r = 0.1$.

points where the function

$$\sum_{\alpha=1}^{n} \sin^2 \alpha\pi x = \frac{n}{2} - \frac{\cos(n+1)\pi x \sin n\pi x}{2 \sin \pi x}, \quad n = 9$$

reaches its local maxima. Note that for $r \to 0$ the computed optimal design is also the standard D-optimal design.

With an increase in r, i.e., with either an increase in the variance σ^2 or with a decrease in the number of available observations N, the number of supporting points n decreases. In other words, if the foreseen observational error becomes large relative to the eigenvalues of a covariance kernel of higher orders, then the optimal allocation of available observations must be mainly oriented towards estimation of the first few parameters of the model.

Obviously, the constructed designs are optimal only in the framework of the selected approximation: we used only the first 9 terms in Mercer's expansion of the covariance kernel. Formulas (3.48) and (3.49) may help to evaluate whether the contribution of the remainder terms is essential. On an intuitive level it is evident that the selected approximation can be acceptable for practical needs: the ratio of the "variance" generated by observational errors to the "variance" generated by the randomness of the first neglected parameter is $\pi^{-2}10^{-2} \cong 0.001$ for $r = 1$.

3.4 Discrete Design Sets

Main assumptions. In many applications, and in computer network monitoring in particular, the design set X may consist of a finite number of points. For instance, X is a collection of sites (nodes) which can be monitored. Similar to Fedorov (1996), we may follow the route which basically coincides with what was discussed in the previous section. First, expand a covariance kernel with respect to its eigenfunctions, and then proceed with discrete versions of Theorems 1 through 4 from Chapter 3.2. In Fedorov and Flanagan (1997), however, we develop a more direct and simpler approach, and that approach is the subject of the discussions that follow.

To emphasize the discrete character of the problem we use "nodes" instead of "points". It also reflects the fact that in the monitoring of computer networks, "points" may be presented by graphs. Whenever it is possible, we try to complement the terminology in design theory with terms related to computer network monitoring.

Similar to (3.27) but in more detail, let the following model be applied

$$y_{jk}(x_i) = u_j(x_i) + \varepsilon_{jk}(x_i), \tag{3.65}$$

where $u_j(x_i)$ describes the i-th node at moment j, and $\varepsilon_{jk}(x_i)$ is the corresponding observational error, $k = 1, \ldots, r_{ij}$. Again, the first term $u(x_i)$ describes the

random behavior of the monitored network, while the second one is related to observational errors or short time disturbances. The same characters are used both for random variables and their realizations. The latter ones are standardly marked by additional indices: i.e., $u(x_i)$ stands for the random variable, and $u_j(x_i)$ is its realization.

Let the vector

$$U = (u(x_1), \ldots, u(x_S))^T \tag{3.66}$$

be observed, and let

$$E_u(U) = U_0, \;\; Var_u(U) = E\left[(U - U_0)(U - U_0)^T\right] = K, \tag{3.67}$$

where the $S \times 1$ vector U_0 and the covariance $S \times S$ matrix K are given. The subscript u (or ε) means that expectation or variance is taken with respect to u (or ε); subscripts $\varepsilon|u$ (or $u|\varepsilon$) are used for conditional expectations. The obvious transform $U \to U - U_0$ zeroes the expectation of U, and, therefore, in what follows we assume that $E_u(U) = 0$. The observational errors $\varepsilon(x_i)$ are assumed to have zero means and to be uncorrelated:

$$E_{\varepsilon|u}\left(\varepsilon_j(x_i)\right) \equiv 0, \quad E_{\varepsilon|u}\left(\varepsilon_j(x_i)\varepsilon_{j'}(x_{i'})\right) \equiv \sigma^2 \delta_{ii'}\delta_{jj'}. \tag{3.68}$$

Introduction of σ^2 depending on x does not lead to any significant changes and

is not considered here. The argument x_i may be considered as a descriptor of site i and might be skipped in this section to simplify notation. However, we will continue to use it to make it easier to bridge our results with standard convex design theory.

Note that in this dissertation we do not use any properties of the set X (compare with Chapters 3.1 and 3.2). For instance, we do not introduce a distance between two points x_i and $x_{i'}$. The concept of a distance is much less natural in communication network measurements than in meteorology or seismology, where the physical distance $\| x_i - x_{i'} \|$ between observing stations may define the behavior of elements $K(x_i, x_{i'})$ of the matrix K as functions of $\| x_i - x_{i'} \|$. Introduction of concepts similar to "distance", such as the number of switches or the complexity of routes between x_i and $x_{i'}$ in communications networks, may lead to more efficient and realistic modeling, but it is beyond the scope of this dissertation.

We assume that the identical (experimental) designs are used for all j:

$$\xi_n = \{p_i, x_i\}_1^n, \ p_i = r_i/N, \ N = \sum_{i=1}^{n} r_i, \ x_i \in X, \ n \leq S.$$

Let $K(\xi_n)$ be a submatrix of K which corresponds to the nodes $x_1, \ldots, x_n$; let $K(x, \xi_n)$ be a column vector of covariances between $u(x)$ and $u(x_1), \ldots, u(x_n)$; let the matrix $W(\xi_n)$ be diagonal with the elements $W_{ii} = N\sigma^{-2} p_i \delta_{ii}$. We also use the matrices $K(X_{pr}, \xi_n) = (K(x_1, \xi_n), \ldots, K(x_q, \xi_q))$, where $x_1, \ldots, x_q \in X_{pr} \subset X$,

and $K(X_{pr})$ is a submatrix of K corresponding to the same nodes.

Estimation and optimality criteria. Let $Y(\xi_n)$ be the vector of averaged observations made according to ξ_n:

$$Y(\xi_n) = \begin{pmatrix} \frac{1}{r_1}\sum_{k=1}^{r_1} y_k(x_1) \\ \vdots \vdots \vdots \\ \frac{1}{r_n}\sum_{k=1}^{r_n} y_k(x_n) \end{pmatrix} = \begin{pmatrix} Y_1 \\ \vdots \vdots \vdots \\ Y_n \end{pmatrix};$$

index j is skipped to emphasize that $Y(\xi_n)$ is a random variable and not a realization. It can be verified by direct minimization that the estimator (compare with (3.30) - (3.34) in Chapter 3.2)

$$\hat{U}(X_{pr}) = K^T(X_{pr}, \xi_n)\left(K(\xi_n) + W^{-1}(\xi_n)\right)^{-1} Y_j(\xi_n) \tag{3.69}$$

minimizes the matrix of expected squared residuals

$$D\left(\xi_n, \tilde{U}(X_{pr})\right) = E_{u,\varepsilon}\left[\left(\tilde{U}(X_{pr}) - U(X_{pr})\right)\left(\tilde{U}(X_{pr}) - U(X_{pr})\right)^T\right]$$

among all linear estimators $\tilde{U}(X_{pr}) = LY(\xi_n)$ such that

$$E_{u,\varepsilon}\left[\tilde{U}(X_{pr}) - U(X_{pr})\right] = 0.$$

In other words

$$D(\xi_n) = D\left(\xi_n, \hat{U}(X_{pr})\right) \leq D\left(\xi_n, \tilde{U}(X_{pr})\right), \tag{3.70}$$

where inequality must be understood in the sense of ordering of nonnegative definite matrices (c.f. Horn and Johnson (1985), Chapter 7.7). From (3.65) and (3.69) it follows that

$$D(\xi_n) = K(X_{pr}) - K^T(X_{pr}, \xi_n)\left(K(\xi_n) + W^{-1}(\xi_n)\right)^{-1} K(X_{pr}, \xi_n) \tag{3.71}$$

The objective of this dissertation is to provide methods which allow the minimization of some given functions of the matrix $D(\xi_n)$, for instance, $trD(\xi_n)$, $\ln|D(\xi_n)|$, $\max_i D_{ii}(\xi_n)$, etc. See Fedorov and Hackl (1997) or Puckelsheim (1993) for details about optimality criteria. Thus, we have to consider the following optimization problem

$$\xi_n^* = \arg\min_{\xi_n} \Psi\left[D(\xi_n)\right], \tag{3.72}$$

where Ψ is a selected objective function (criterion of optimality). In (3.72) the number of nodes (or supporting points) n is fixed. In general, n may be optimized as well.

Properties of optimal designs. Two features of optimization problem (3.72) may cause serious computational hurdles: the weights p_i are discrete, and the

optimal number of supporting points must be found, in general. The problem is simplified both theoretically and numerically if we allow the weights to be continuous, so that $0 \leq p_i \leq 1$, $\sum_{i=1}^{n} p_i = 1$, and make $n = S$. If an optimal n is less than S then some of the weights are zeroes.

Zero weights may necessitate using the limit transition in (3.71). When $n = S$ and X_{pr} coincides with X, then from (3.71) and the identity

$$(A+B)^{-1} = A^{-1} - A^{-1}(A^{-1}+B^{-1})^{-1}A^{-1}$$

it follows that

$$D(\xi_n) = \left(K^{-1} + W(\xi_n)\right)^{-1}. \tag{3.73}$$

The regularity of the matrix K is assumed in the latter formula. If X_{pr} is a subset of X then the covariance matrix (3.70) is a defined submatrix of (3.73).

Using (3.73) we can reformulate the design problem as

$$\xi^* = \arg\min_{\xi} \Psi\left[D(\xi)\right], \tag{3.74}$$

where functions $M(\xi) = D^{-1}(\xi) = K^{-1} + W(\xi)$.

Similar to Chapter 2.2, let us assume that

1. The function $\Psi[D(\xi)]$ is a convex function of ξ and has a directional derivative $\psi(\xi^*, \xi)$ at ξ^* for any $\overline{\xi} = (1-\alpha)\xi^* + \alpha\xi$ and $0 \leq \alpha < 1$.

Theorem 5 *A necessary and sufficient condition for a design ξ^* to be optimal is fulfillment of the inequality*

$$\psi(\xi^*, \xi) \geq 0. \tag{3.75}$$

This result is well known in optimization theory of convex functions and is widely used in experimental design theory (c.f. Cook and Fedorov (1995), Fedorov and Hackl (1997) Chapter 2) (see also Chapter 2.2). The theorem leads to constructive results when some simple presentation of $\psi(\xi^*, \xi)$ exists for a chosen optimality criterion. For instance, for the D-criterion, $\Psi(D) = \ln |D|$, we have

$$\psi(\xi^*, \xi) = trD(\xi^*) \left(W(\xi^*) - W(\xi)\right). \tag{3.76}$$

Noting that

$$tr\ D(\xi)W(\xi) = \sigma^{-2} N \sum_{i=1}^{S} D_{ii}(\xi^*) p_i,$$

and combining (3.75) and (3.76) we have

Theorem 6 *A necessary and sufficient condition for a design ξ^* to be D-optimal is that*

$$\max_i D_{ii}(\xi^*) \leq \frac{\sigma^2}{N} trW(\xi^*)D(\xi^*)$$

and equality holds at all points where $p_i^* > 0$.

A D-optimal design minimizes the maximal variance of prediction:

$$\xi^* = \arg\min_{\xi} \max_{i} D_{ii}(\xi).$$

In this theorem and in what follows $\max_i$ means maximization over all points from X, i.e., $1 \leq i \leq S$. Thus, observations in a D-optimal or minimax design must be placed at points (sites) where prediction might be the worst.

For linear criteria $\Psi(D^{-1}) = trAD$, where $A \geq 0$ is the so-called utility matrix, we have

$$\psi(\xi^*, \xi) = trDAD\left(W(\xi^*) - W(\xi)\right),$$

and the following result holds.

Theorem 7 *A necessary and sufficient condition for a design ξ^* to be linear optimal is that*

$$\max_i \{D(\xi^*)AD(\xi^*)\}_{ii} \leq \frac{\sigma^2}{N} trW(\xi^*)D(\xi^*)AD(\xi^*),$$

and the equality holds at all points where $p_i^* > 0$.

If $A = I$, i.e., the average variance of prediction must be minimized, then the theorem tells us that an optimal design ξ^* allocates observations at sites in which the predicted value $\hat{U}(x^*)$ of $U(x^*)$ might have the greatest average squared covariance with all other $\hat{U}(x), x \in X$.

3.5 First Order Algorithms and Simple Numerical Example

The above theorems help to develop and analyze various first order algorithms for construction of optimal designs. For computer networks the matrices processed during computations have large sizes. It is, therefore, especially important to use recursions which are computationally simple and stable. The most convenient in this sense are algorithms similar to the first order exchange type algorithms (see, for instance, Chapter 2.5, Mitchell (1974), and Fedorov (1997)). The main idea is very simple: at each stage add that new point which improves the current design most and delete from the same (or just corrected) design that point which contributes least. Here we formulate the simplest version of that kind of algorithm for D-criteria.

Let the initial design ξ_0 be such that all weights $p_{0i} = b_i\alpha_0$, where b_i is an integer and $\sum_{i=1}^{S} p_{0i} = 1$. For instance, we may choose $b_i \equiv 1$ and $\alpha_0 = 1/S$.

1. Given ξ_t and $D(\xi_t)$ find

$$a = \arg\max_i D_{ii}(\xi_t). \tag{3.77}$$

Add α_t to the weight of point x_a to construct the design ξ_t^+ and the matrix $D(\xi_t^+)$.

2. Find

$$d = \arg\min_{i \in I_t} D_{ii}(\xi_t^+), \tag{3.78}$$

where I_t is the set of all supporting points of ξ_t, i.e., points with nonzero weights at step t. Delete α_t from the weight of point x_d to modify ξ_t^+ and to construct ξ_{t+1}.

3. If

$$\frac{|D(\xi_{t+1})|}{|D(\xi_t)|} = \frac{1}{(1+\zeta_t D_{aa}(\xi_t))(1-\zeta_t D_{dd}(\xi_t^+))} < 1-\gamma, \tag{3.79}$$

where γ is a small positive number, put $\alpha_{t+1} = \alpha_t$ and go to (a). Otherwise make $\alpha_{t+1} = \alpha_t/2$ and then go to (a).

Computations may be stopped when α_t is sufficiently small.

The choice of p_{0i} and α_0 is a matter of convenience. For instance, the above choice guarantees that no more than α_t^{-1} observations are needed to avoid any "fractional" observation in design ξ_t, which is a frequent case in the continuous design theory setting. Actually in the "classical" version of the exchange algorithm α equals N^{-1}, where N is a preselected number of observations. Unfortunately, in this case the limit design (if it exists) is generally not an optimal one. That is why we introduced the possibility of infinitely reducing the step length α.

The rule for adding and deleting weights becomes obvious if we note that

$$|D(\xi_t^+)| = \frac{|D(\xi_t)|}{1 + \zeta_t D_{aa}(\xi_t)} \tag{3.80}$$

and

$$D(\xi_t^+) = D(\xi_t) - \frac{\zeta_t C^+(\xi_t)}{1 + \zeta_t D_{aa}(\xi_t)}, \tag{3.81}$$

where $C_{ij}^+(\xi_t) = D_{ai}(\xi_t)D_{aj}(\xi_t)$ and $\zeta_t = \sigma^{-2}N\alpha_t$. The above formulae may be derived using the fact that

$$\begin{aligned} D(\xi_t^+) &= \left(K^{-1} + W(\xi_t) + \zeta_t \ell_a \ell_a^T\right)^{-1} \\ &= \left(D^{-1}(\xi_t) + \zeta_t \ell_a \ell_a^T\right)^{-1}, \end{aligned}$$

where $\{\ell_a\}_i = \delta_{ia}$, and identities (3.44) and

$$|A + BB^T| = |A||I + B^T A^{-1} B|.$$

In the versions of (3.80) and (3.81) for the deleting procedure, ζ_t must be replaced by $-\zeta_t$ and ξ_t by ξ_t^+.

Similar to the classical results of experimental design theory (see, for instance, Cook and Fedorov (1995)) the following result may be established.

Theorem 8 *The sequence* $\{|D(\xi_t)|\}$ *converges and*

$$\min_{\xi} |D(\xi)| \leq \lim_{t\to\infty} |D(\xi_s)| \leq (1-\gamma^{-1}) \min_{\xi} |D(\xi)|.$$

The proof is based on the monotonicity of the iterative procedure, convexity of $\ln |D(\xi)|$ as a function of ξ, and Theorem 2.

Note that formulas (3.80), (3.81) and their siblings for the deleting steps are very handy recursions for large size problems.

Example. Following Fedorov and Flanagan (1998), let

$$K_{ij} = 0.004\{\min(i,j)(50 - \max(i,j))\}, i,j = 1,\ldots,49.$$

We computed D-optimal designs for $\sigma^2 N^{-1} = 0.01, 0.1$ and 1. The results (after 100 iterations and $\sigma = 0.001$), are exhibited at Fig. 3.4 and 3.5. The weights were not rounded up or symmetrized with respect to $i = 25$, and that is an obvious step in more practical situations.

In this example we used the iterative procedure without any modification. Actually, we found that some minor amendments improve the rate of convergence significantly. For instance, instead of proceeding with one step forward - one step backward, one can repeat a few steps forward and then exactly the same number of steps backward (in order to keep $\sum_{i=1}^{s} p_i = 1$). Instead of using the constant λ or δ, some slowly diminishing sequences $\{\lambda_t\}$ or $\{\delta_t\}$ do a better job, etc.

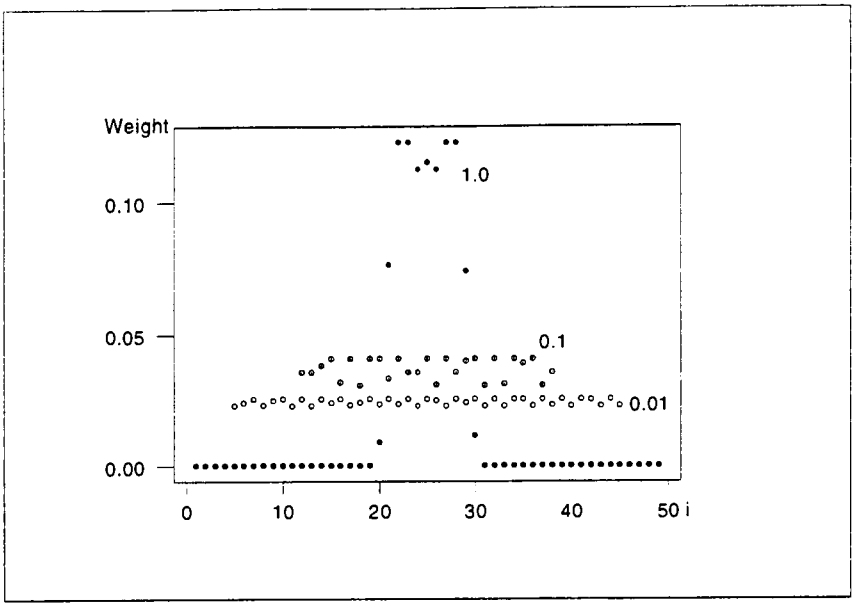

Figure 3.4: Computed Weights for $\sigma^2 N^{-1} = 0.01, 0.1$ and 1

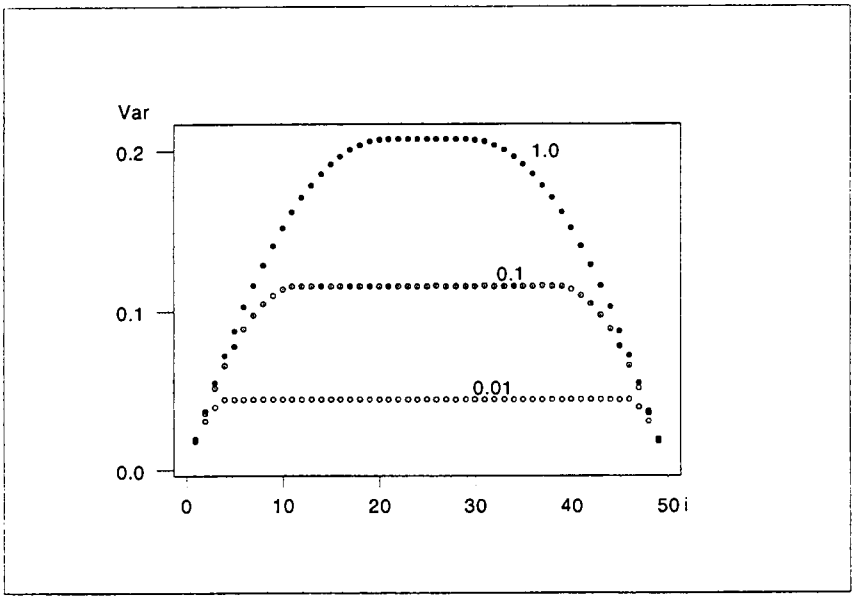

Figure 3.5: Variance of Prediction for $\sigma^2 N^{-1} = 0.01, 0.1$ and 1

As expected, the structure of the optimal designs constructed here and in Chapter 3.2 resemble each other. Indeed, the example considered in this section may be viewed as a discretized version of the limit solution (i.e., $p \to \infty$) for (3.63), (3.64) or directly (3.46). Documentation of the computer program to implement the numerical algorithms appears in Appendix B.

Chapter 4

Optimal Designs on Graphs

4.1 Response Surface Type Models in Network Modeling

As it was mentioned in the introductory chapter, most computer networks can be represented by graphs. Consequently, we may discuss measurements on edges or nodes of the corresponding graphs. These measurements possibly include counts of various types of packets, inter-arrival times, delays, number of lost messages, etc. In an active experiment we may control sample size or select elements of a graph where measuring devices are to be placed. In more complicated cases one can select routes for test-signals (sent packages, files). To discuss the model and the corresponding analysis and design problem, we use two simple graphs. Each node on a graph represents a site which can be treated either as a host, as a recipient or both.

In what follows we construct optimal designs for two response surface models. The first model contains observational errors which have the same variance at all routes, while the second one admits a dependence of this variance upon the selected route. Probably, the closest models were considered by Vardi (1996) who assumed that observations have a Poisson distribution. Our models use only first and second moments without specification of the distribution shape. For definiteness we assume that the delay time is a variable to be observed. Use of this variable also helps to compare results of this chapter with the results of Chapter 5, where some real experiments with delay times are analyzed and designed.

For both models

$$y_i = \sum_{\alpha=1}^{p} \theta_\alpha x_{\alpha i} + \varepsilon_j = \theta^T x_i + \varepsilon_i, \tag{4.1}$$

where

y_i = the delay time for route i, $i = 1, \ldots, S$,

θ^T = $(\theta_1, \ldots, \theta_p)$ is the vector of the p unknown model parameters,

$x_{\alpha i}$ = 1, if the αth edge is included in the route; $x_{\alpha i} = 0$ otherwise,

ε_i = random errors.

We do not use the intercept because the latter corresponds to delays at the host site. Relative to the transmission times (delays) through edges and other

nodes, the host site delays are considered negligible.

In model (4.1) it is assumed that for all i

$$E(\varepsilon_i) = 0, \;\; E(\varepsilon_i^2) = \sigma^2 \text{ and } E(\varepsilon_i \varepsilon_{i'},) = 0, \; i \neq i'. \tag{4.2}$$

Generally, this assumption is not very realistic, and we may want to introduce a more flexible model in which this variance depends upon a chosen route. Model 2 incorporates a variance adjustment as follows:

$$\sigma^2(x_i) = \sum_{\alpha=1}^{S} \sigma_\alpha^2 x_{\alpha i}, \tag{4.3}$$

where σ_α^2 is assumed given.

If there are repeated observations then the second subscript j must be added to the response variable. Similar to the discussion in Chapter 3, we assume that $j = 1, \ldots, r_i$, $\sum_{i=1}^{S} r_i = N$ and $p_i = r_i/N$.

For the first model (c.f. Chapter 3)

$$M(\xi) = \sum_{i=1}^{S} p_i x_i x_i^T, \tag{4.4}$$

while for the second one

$$M(\xi) = \sum_{i=1}^{S} p_i \sigma^{-2}(x_i) x_i x_i^T. \tag{4.5}$$

The simple transform

$$x_i' = \sigma^{-1}(x_i)x_i$$

converts (4.5) into (4.4), and, consequently, the techniques developed in Chapters 2 and 3 for models of type (4.4) may be applied to (4.5).

In what follows all computations are done for the D-criterion. Obviously other criteria may be considered similarly. Our experience shows that even for the simplest graphs it may be difficult to find a non-singular (i.e., $|M(\xi_0)| \neq 0$) initial design ξ_0 to start the iterative procedure. Therefore, in some of our calculations we assumed the existence of prior information with a known variance-covariance matrix D_0 (it can be called the Bayesian approach). Hence, the Bayesian design techniques (see Chapter 2.3) were used in some of our computations. In practice, when no prior information is available the matrix $D_0 \geq 0$ can be considered as a means to create a regular condition. Usually a few designs

$$\xi^*(\gamma) = \arg\min_{\xi} |M(\xi) + \frac{\sigma^2}{N}\gamma D_0^{-1}|$$

with different diminishing γ must be considered to produce the optimal design. In regularized problems D_0 is usually selected to be the identity matrix. Note in the Bayesian case (c.f. Chapter 2.3, see also Fedorov (1997)) that, in general, optimal designs depend upon the ratio σ^2/N.

The analysis process for these examples initially estimates design performance

for the uniform design that includes all candidate design points. For computer network monitoring, such a design represents a monitoring system which collects data at each node. Next, the D-optimal design is constructed and represents the best possible monitoring system (in terms of the minimum variance of prediction) but which unfortunately admits weights p_i which may lead to "fractional" $p_i N$ at some points x_i. The D-optimal design, then, generally does not constitute a monitoring system but becomes a benchmark against which to gauge exact N designs, i.e., containing a given number of observations and with all integer $p_i N$. In some cases, however, an exact design may also be D-optimal. Additional exact designs with fewer design points than the D-optimal design are then constructed based on the D-optimal design. Obviously, the number of observations N cannot be less than the number of model parameters. The additional designs exhibit the capability of the method to accomplish sensitivity analyses with respect to the impact of the placement and number of data collect points on a monitoring network.

The variance of prediction $\max_i d(x_i, \xi)$ for the D-optimal design must be less than the number of parameters m. This requisite variance is reported for each design to prove optimality for the D-optimal design and otherwise to use for efficiency comparisons. For some D-optimal designs, $|\max_i d(x_i, \xi) - m| = \mathcal{L} > 0$, for $\mathcal{L}$ small (less than .05 in these examples). This difference is due to the accuracy of the computer software algorithm.

4.2 Numerical Examples for Two Simple Graphs

To illuminate the main ideas we start with a very simple example and consider the graph with only four nodes and five edges shown in Figure 4.1. The edge numbers appear in bold type in the figure, and the node numbers are circled. Two cases are analyzed:

1. a single-host monitoring system from node 1 and

2. a multi-host monitoring system with all four nodes included in the host set.

All possible routes are used for each case. The single-host monitoring system consists of 9 possible routes, so there are 9 candidate design points. There are 19 routes, and, therefore, there are 19 candidate design points for case (2). The routes for each case appear in Table 4.1. The single-host case includes routes 1 through 9 only.

Experimental designs for a monitoring system appear in Table 4.2 for the single-host case and model (4.1). We observe the following:

1. The exchange-type algorithm designs (see Chapter 2.4) in columns IV and VI are exact designs that are relatively close to D-optimal.

2. If the D-optimal weights from Column II are re-distributed as shown in column III, the new design is only slightly more efficient than the exact design for $N = 6$.

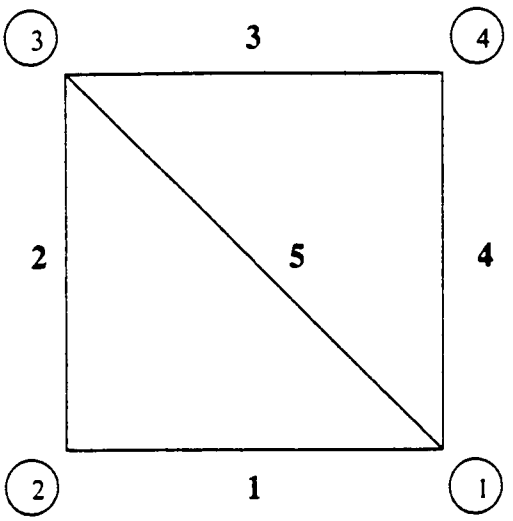

Figure 4.1: Simple Network Graph

Table 4.1: Routes for the Network Problem with 4 Nodes and 5 Edges

Starting-Ending Hosts	Route Number	Edges 1	2	3	4	5
1 - 2	1	0	1	1	1	0
1 - 2	2	1	0	0	0	0
1 - 2	3	0	1	0	0	1
1 - 4	4	0	0	0	1	0
1 - 4	5	0	0	1	0	1
1 - 4	6	1	1	1	0	0
1 - 3	7	0	0	0	0	1
1 - 3	8	1	1	0	0	0
1 - 3	9	0	0	1	1	0
2 - 3	10	0	1	0	0	0
2 - 3	11	1	0	0	0	1
2 - 3	12	1	0	1	1	0
2 - 4	13	1	0	0	1	0
2 - 4	14	0	1	1	0	0
2 - 4	15	0	1	0	1	1
2 - 4	16	1	0	1	0	1
3 - 4	17	0	0	1	0	0
3 - 4	18	0	0	0	1	1
3 - 4	19	1	1	0	1	0

Table 4.2: Experimental Designs for the Single-Host Case of the Network Example with 4 Nodes and 5 Edges

	I	II	III	IV	V	VI
	Uniform	D-optimal Continuous	Rounded 6 points	Exchange-Type 6 points	Exchange-Type 5 points (a)	Exchange-Type 5 points (b)
N	9	9	9	9	9	9
n	9	8	6	6	5	5
p_1	1/9	.14	.15	1/6	0	1/5
p_2	1/9	.14	.15	1/6	1/5	1/5
p_3	1/9	.18	.20	1/6	0	1/5
p_4	1/9	.14	.15	1/6	1/5	1/5
p_5	1/9	.18	.20	1/6	0	1/5
p_6	1/9	.14	.15	1/6	0	0
p_7	1/9	0.00	0.00	0	1/5	0
p_8	1/9	.04	0.00	0	1/5	0
p_9	1/9	.04	0.00	0	1/5	0
$\|D\|$	563	450	463	486	3125	781
$\max_i d(x_i, \xi)$	5.91	5.01	5.83	6.00	15.0	15.0

3. The two designs (Columns V and VI) for $N = 5$ show that the starting solution for Column V produces only a local optimum. This is unusual but may be due to the small ratio of candidate points to the number of parameters (i.e., 9/5). The designs are equivalent with respect to the efficiency in terms of the variance of prediction, $\max_i d(x_i, \xi)$.

4. The optimal exact $N = 5$ design is translated into routes as shown in Figure 4.2. The constructed optimal designs mostly selected routes consisting of single edges.

For the multi-host case and model (4.1), computed experimental designs appear in Table 4.3. The optimal design is translated into routes as shown in Figure 4.3. We observe the following:

1. Columns V and VI show alternate exact designs for $N = 7$.

2. The optimal exact $N = 7$ designs are as efficient as the uniform design with all 19 design points.

3. All designs except for the exact $N = 5$ provide a relatively similar amount of efficiency with respect to $\min_i d(x_i\xi)$.

4. The routes in Table 4.1 that correspond to the multi-host designs in Table 4.3 each contain at least two edges: no single-edged route is part of an optimal design.

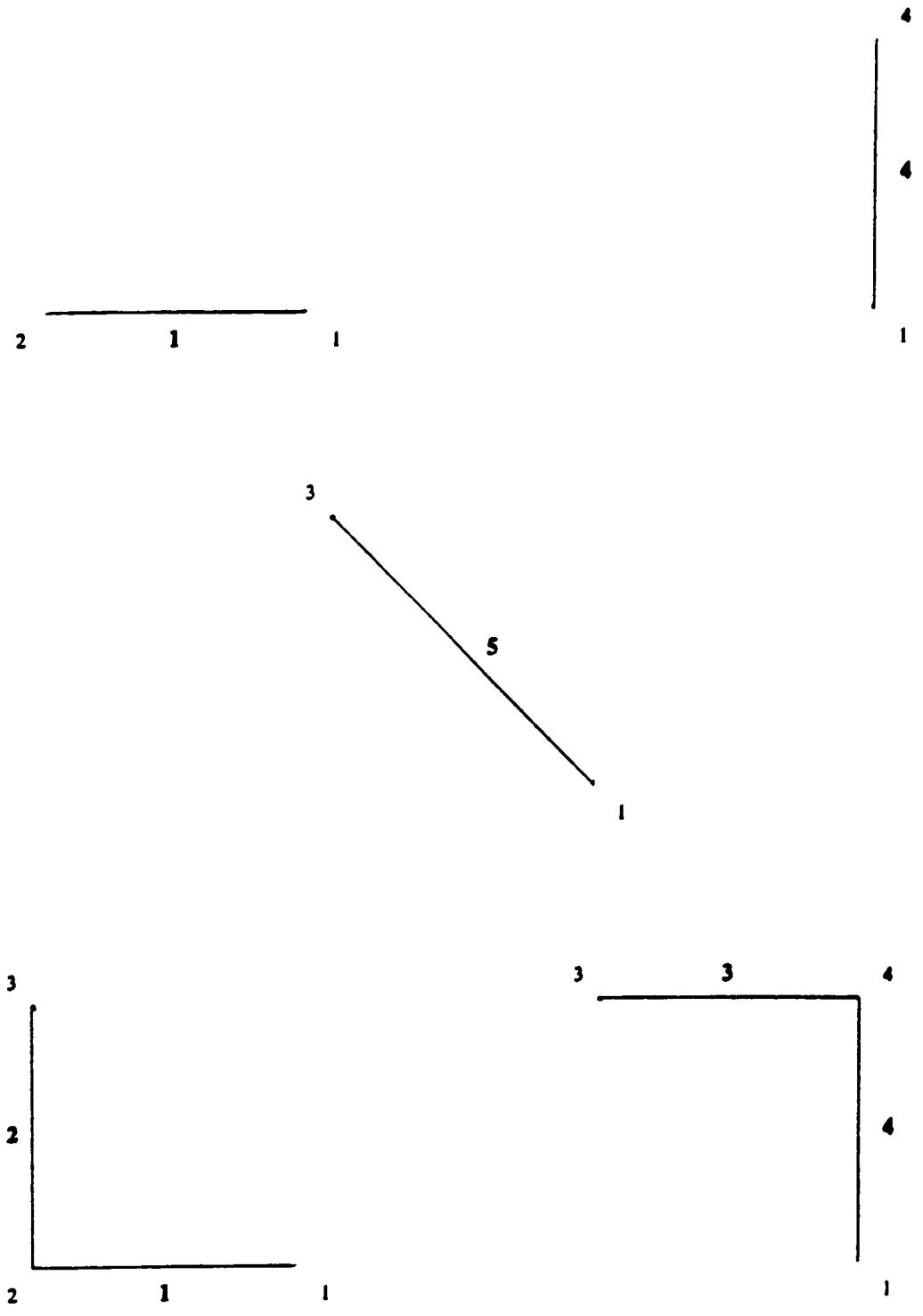

Figure 4.2: Saturated D-optimal Design for a Single-Host Formulation

Table 4.3: Experimental Designs for the Multi-Host Case of the Network Example with 4 Nodes and 5 Edges

	I	II	III	IV	V	VI
	Uniform	D-optimal Continuous	Uniform 14 points	Exchange-Type 5 points	Exchange-Type 7 points (a)	Exchange-Type 7 points (b)
N	19	19	19	19	19	19
n	19	14	14	5	7	7
p_1	1/19	.07	1/14	0	1/7	0
p_2	1/19	0.00	0	0	0	0
p_3	1/19	.05	1/14	0	0	1/7
p_4	1/19	0.00	0	0	0	0
p_5	1/19	.05	1/14	0	1/7	0
p_6	1/19	.07	1/14	0	1/7	1/7
p_7	1/19	0.00	0	0	0	0
p_8	1/19	.06	1/14	1/5	0	0
p_9	1/19	.06	1/14	1/5	1/7	1/7
p_{10}	1/19	0.00	0	0	0	0
p_{11}	1/19	.05	1/14	0	0	0
p_{12}	1/19	.07	1/14	0	0	0
p_{13}	1/19	.06	1/14	0	1/7	0
p_{14}	1/19	.06	1/14	1/5	0	1/7
p_{15}	1/19	.12	1/14	1/5	1/7	1/7
p_{16}	1/19	.12	1/14	1/5	1/7	1/7
p_{17}	1/19	0.00	0	0	0	0
p_{18}	1/19	.05	1/14	0	0	0
p_{19}	1/19	.07	1/14	0	0	1/7
$\|D\|$	193	100	105	195	127	127
$\max_i d(x_i, \xi)$	6.76	5.00	5.95	15.0	6.84	6.84

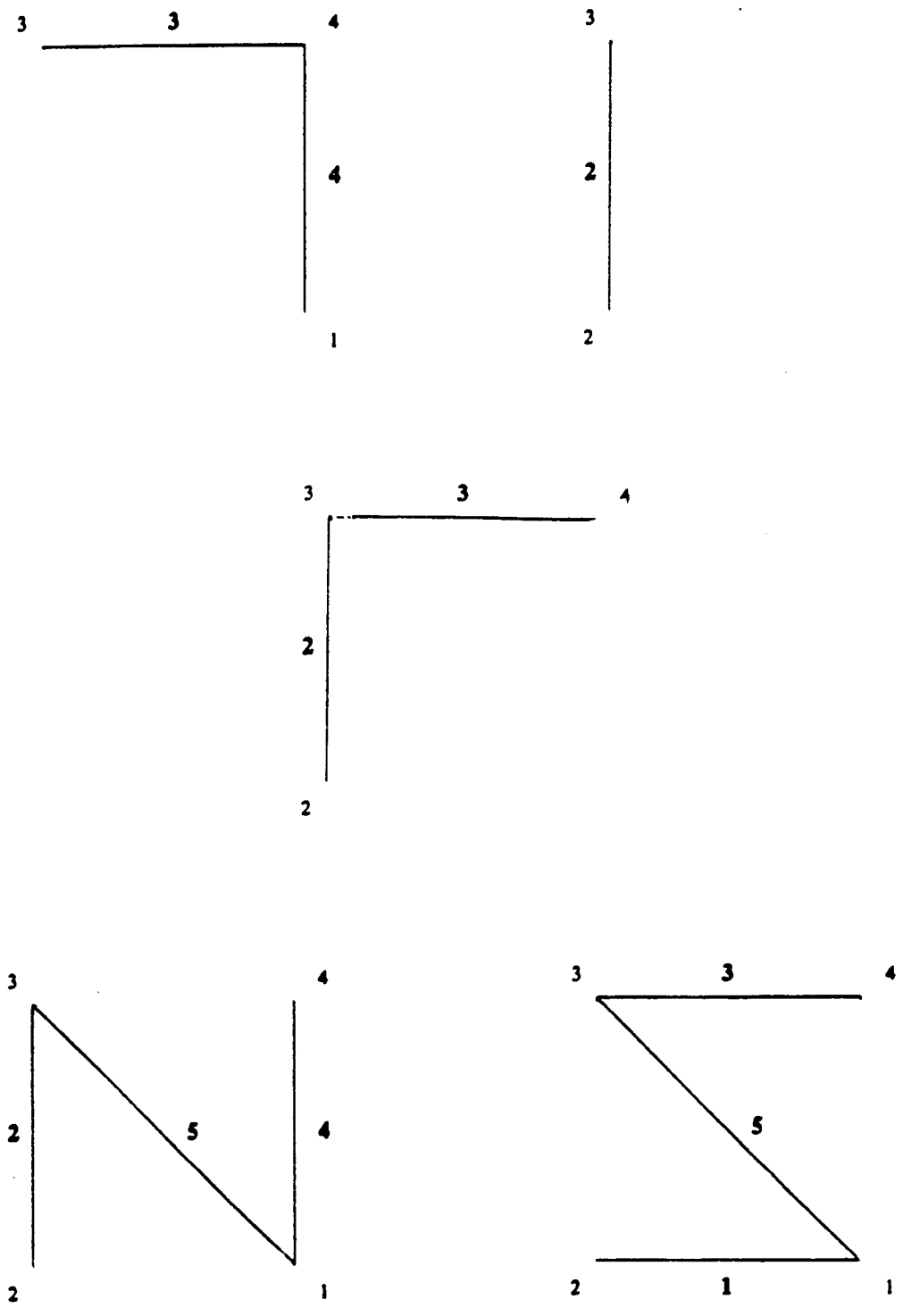

Figure 4.3: Saturated D-optimal Design for a Multi-Host (All Hosts) Formulation

Comparison of the single- and multi-host cases suggests the following:

1. There is no significant difference between the single- and multi-host cases with respect to the efficiency of the designs if the latter one is measured in terms of $\max_i d(x_i, \xi)$. However, in terms of D-optimality, the multi-host designs are much better (3-4 folds improvement): compare the row marked $|D|$ in both tables.

Experimental designs for a monitoring system appear in Table 4.4 for the single-host case and model (4.3). We observe the following:

1. The D-optimal design requires all nine points.

2. Columns IV and V provide alternate optimal exact $N = 5$ designs. There is a considerable loss of efficiency when moving to an $N = 5$ design.

Experimental designs for a monitoring system appear in Table 4.5 for the multi-host case and model (4.3). We observe the following:

1. The exact $N = 5$ design is D-optimal.

2. Columns III and IV are alternate optimal exact designs for $N = 6$.

3. The exact $N = 6$ designs are as efficient as the uniform design with $N = 19$.

4. Using the routes in Table 4.2, it is clear that the D-optimal design (column II) has routes that consist of single edges only.

Table 4.4: Experimental Designs with Variance Adjustment for the Single-Host Case of the Network Example with 4 Nodes and 5 Edges

	I	II	III	IV	V
	Uniform	D-optimal Continuous	Rounded 6 points	Exchange-Type 5 points (a)	Exchange-Type 5 points (b)
N	9	9	9	9	9
n	9	9	6	5	5
p_1	1/9	.07	0	1/5	0
p_2	1/9	.17	1/6	1/5	1/5
p_3	1/9	.14	1/6	1/5	1/5
p_4	1/9	.17	1/6	1/5	1/5
p_5	1/9	.14	1/6	1/5	1/5
p_6	1/9	.07	1/6	0	1/5
p_7	1/9	.12	1/6	0	0
p_8	1/9	.06	0	0	0
p_9	1/9	.06	0	0	0
$\|D\|$	7589	6214	7775	9375	9375
$\max_i d(x_i, \xi)$	6.66	5.00	8.00	10.00	10.00

Table 4.5: Experimental Designs with Variance Adjustment for the Multi-Host Case of the Network Example with 4 Nodes and 5 Edges

	I	II	III	IV
	Uniform	D-optimal	Exchange-Type	
		Continuous	6 points (a)	6 points (b)
N	19	19	19	19
n	19	5	6	6
p_1	1/19	0.00	0	0
p_2	1/19	.20	1/6	1/6
p_3	1/19	0.00	0	0
p_4	1/19	.20	1/6	1/6
p_5	1/19	0.00	0	0
p_6	1/19	0.00	1/6	0
p_7	1/19	.20	0	1/6
p_8	1/19	0.00	0	0
p_9	1/19	0.00	1/6	0
p_{10}	1/19	.20	0	1/6
p_{11}	1/19	0.00	0	0
p_{12}	1/19	0.00	0	0
p_{13}	1/19	0.00	0	0
p_{14}	1/19	0.00	1/6	0
p_{15}	1/19	0.00	0	0
p_{16}	1/19	0.00	0	0
p_{17}	1/19	.20	1/6	1/6
p_{18}	1/19	0.00	0	0
p_{19}	1/19	0.00	0	1/6
$\|D\|$	5251	3125	3888	3888
$\max_i d(x_i, \xi)$	5.99	5.00	6.00	6.00

Comparison of the multi- and single-host cases with variance adjustment (model 4.3) suggests the following:

1. Designs for the multi-host case are only slightly more efficient in terms of the maximal variance of prediction, but similar to model (4.1), the multi-host case exceeds significantly in terms of D-optimality.

Network Graph Example With Eight Nodes and 12 Edges. The second example considers the graph with eight nodes and 12 edges shown in Figure 4.4. Three cases are analyzed: (1) a single-host monitoring system from node 1 with a subset of all the feasible routes, (2) the single-host case with the complete set of feasible routes, and (3) a multi-host system with hosts at nodes 1, 4, and 6 and a nearly complete set of all possible feasible routes. The routes for each case appear in Table 4.6. Case (1) has routes 1 through 25; case (2) has routes 1 through 38; case (3) has all 72.

Experimental designs for the single-host case (1) with model (4.1) appear in Table 4.7. We observe the following: noted:

1. The D-optimal design in column II is for the model without prior information and can be compared directly only with the design without prior information in column I. The D-optimal design is only slightly more efficient than the uniform design with all 25 design points.

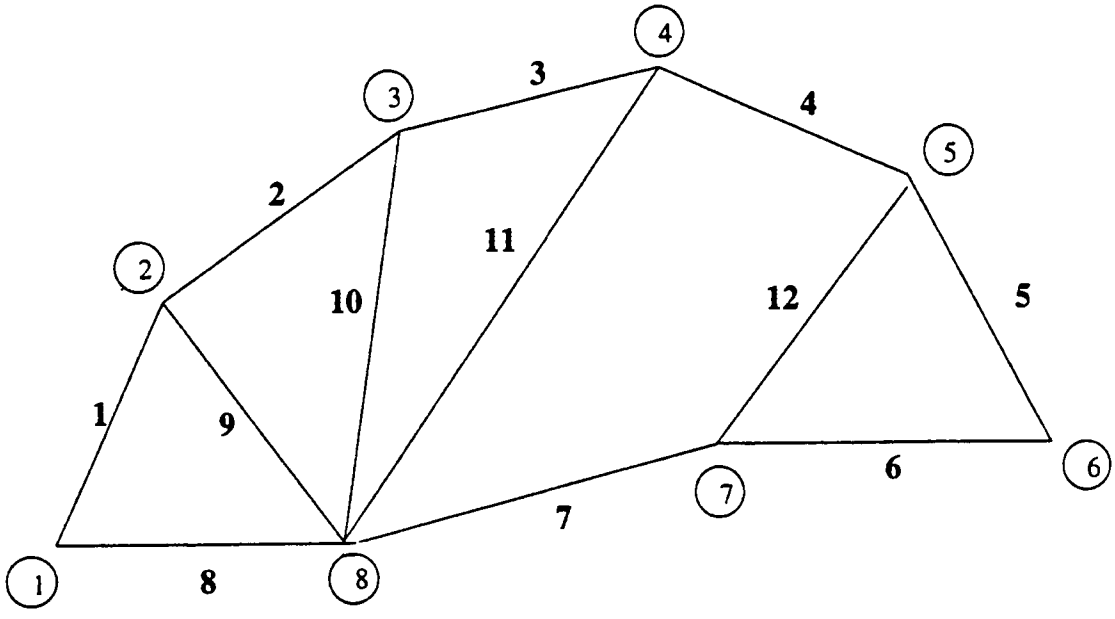

Figure 4.4: Network Graph with 8 Nodes and 12 Edges

Table 4.6: Routes for the Network Example with Eight Nodes and 12 Edges

From-To Hosts	Route Number	Edges												From-To Hosts	Route Number	Edges											
		1	2	3	4	5	6	7	8	9	10	11	12			1	2	3	4	5	6	7	8	9	10	11	12
1 - 3	1	1	1	0	0	0	0	0	0	0	0	0	0	1 - 8	37	1	1	1	0	0	0	0	0	0	0	1	0
1 - 3	2	0	0	0	0	0	0	0	1	0	1	0	0	1 - 8	38	1	1	1	1	0	0	1	0	0	0	0	1
1 - 3	3	0	1	0	0	0	0	0	1	1	0	0	0	4 - 2	39	0	1	1	0	0	0	0	0	0	0	0	0
1 - 6	4	0	0	0	0	0	1	1	1	0	0	0	0	4 - 2	40	0	0	0	0	0	0	0	0	1	0	1	0
1 - 6	5	1	1	1	1	1	0	0	0	0	0	0	0	4 - 2	41	0	1	0	0	0	0	0	0	0	1	1	0
1 - 6	6	0	0	0	0	1	0	1	1	0	0	0	1	4 - 2	42	1	0	0	0	0	0	0	1	0	0	1	0
1 - 2	7	1	0	0	0	0	0	0	0	0	0	0	0	4 - 2	43	0	0	1	0	0	0	0	0	1	1	0	0
1 - 2	8	0	0	0	0	0	0	0	1	1	0	0	0	4 - 2	44	1	0	1	0	0	0	0	1	0	1	0	0
1 - 2	9	0	1	0	0	0	0	0	1	1	1	0	0	4 - 3	45	0	0	1	0	0	0	0	0	0	0	0	0
1 - 4	10	0	0	0	0	0	0	0	1	0	0	1	0	4 - 3	46	0	0	0	0	0	0	0	0	0	1	1	0
1 - 4	11	1	1	1	0	0	0	0	0	0	0	0	0	4 - 3	47	0	1	0	0	0	0	0	0	1	0	1	0
1 - 4	12	1	0	0	0	0	0	0	0	1	0	1	0	4 - 3	48	0	0	0	1	0	0	1	0	0	1	0	1
1 - 5	13	0	0	0	1	0	0	0	1	0	0	1	0	4 - 3	49	0	1	0	1	0	0	1	0	1	0	0	1
1 - 5	14	0	0	0	0	0	0	1	1	0	0	0	1	4 - 3	50	1	1	0	0	0	0	0	1	0	0	1	0
1 - 5	15	0	0	0	0	1	1	1	1	0	0	0	0	4 - 5	51	0	0	0	1	0	0	0	0	0	0	0	0
1 - 7	16	0	0	0	0	0	0	1	1	0	0	0	0	4 - 5	52	0	0	0	0	0	0	1	0	0	0	1	1
1 - 7	17	0	0	0	1	0	0	0	1	0	0	1	1	4 - 5	53	0	0	1	0	0	0	1	0	0	1	0	1
1 - 7	18	1	0	0	0	0	0	1	0	1	0	0	0	4 - 5	54	0	1	1	0	0	0	1	0	1	0	0	1
1 - 8	19	0	0	0	0	0	0	0	1	0	0	0	0	4 - 5	55	0	0	0	0	1	1	1	0	0	0	1	0
1 - 8	20	1	0	0	0	0	0	0	0	1	0	0	0	4 - 6	56	0	0	0	1	1	0	0	0	0	0	0	0
1 - 8	21	1	1	0	0	0	0	0	0	0	1	0	0	4 - 6	57	0	0	0	0	0	1	1	0	0	0	1	0
1 - 6	22	0	0	0	1	1	0	0	1	0	0	1	0	4 - 6	58	0	0	0	0	1	0	1	0	0	0	1	1
1 - 4	23	0	0	0	1	0	0	1	1	0	0	0	1	4 - 6	59	0	0	1	0	0	1	1	0	0	1	0	0
1 - 5	24	1	0	0	1	0	0	0	0	1	0	1	0	4 - 6	60	0	0	0	1	0	1	0	0	0	0	0	1
1 - 5	25	1	1	1	1	0	0	0	0	0	0	0	0	4 - 7	61	0	0	0	1	0	0	0	0	0	0	0	1
1 - 3	26	0	0	1	0	0	0	0	1	0	0	1	0	4 - 7	62	0	0	0	0	0	0	1	0	0	0	1	0
1 - 6	27	0	0	0	1	0	0	0	1	0	0	1	1	4 - 7	63	0	0	0	1	1	1	0	0	0	0	0	0
1 - 2	28	0	1	1	0	0	0	0	1	0	0	1	0	4 - 7	64	0	0	1	0	0	0	1	0	0	1	0	0
1 - 4	29	1	1	0	0	0	0	0	0	0	1	1	0	4 - 7	65	0	1	1	0	0	0	1	0	1	0	0	0
1 - 4	30	0	0	1	0	0	0	0	1	0	1	0	0	4 - 7	66	1	1	1	0	0	0	1	1	0	0	0	0
1 - 5	31	1	0	1	1	0	0	0	0	1	1	0	0	4 - 8	67	0	0	0	0	0	0	0	0	0	0	1	0
1 - 5	32	1	0	0	0	1	1	1	0	1	0	0	0	4 - 8	68	0	0	1	0	0	0	0	0	0	1	0	0
1 - 7	33	1	1	0	0	0	0	1	0	0	1	0	0	4 - 8	69	0	0	0	1	0	0	1	0	0	0	0	1
1 - 7	34	1	1	1	1	0	0	0	0	0	0	0	1	4 - 8	70	0	1	1	0	0	0	0	0	1	0	0	0
1 - 7	35	1	1	1	0	0	0	1	0	0	0	1	0	4 - 8	71	1	1	1	0	0	0	0	1	0	0	0	0
1 - 7	36	1	1	1	1	1	1	0	0	0	0	0	0	4 - 8	72	0	0	0	1	1	1	1	0	0	0	0	0

Table 4.7: Experimental Designs for the Single-Host Case of the Network Example with 8 Nodes and 12 Edges

	I	II	III	IV	V
	Uniform	*D*-optimal Continuous	*D*-optimal Continuous	Exchange-Type	
	Without, With Prior	Without Prior	With Prior	12 point (a) With Prior	12 point (b) With Prior
N	25	25	25	25	25
n	25	22	22	12	12
p_1	1/25	.05	.04	0	0
p_2	1/25	.06	.06	1/12	1/12
p_3	1/25	.05	.05	1/12	1/12
p_4	1/25	.05	.05	0	1/12
p_5	1/25	.05	.06	1/12	1/12
p_6	1/25	.07	.07	1/12	1/12
p_7	1/25	.04	.02	0	1/12
p_8	1/25	.04	.03	0	0
p_9	1/25	.05	.05	0	0
p_{10}	1/25	.03	.03	0	0
p_{11}	1/25	.05	.05	1/12	1/12
p_{12}	1/25	.04	.04	1/12	0
p_{13}	1/25	0.00	0.00	0	0
p_{14}	1/25	0.00	0.00	0	0
p_{15}	1/25	.05	.05	1/12	0
p_{16}	1/25	.05	.04	0	0
p_{17}	1/25	.05	.06	1/12	1/12
p_{18}	1/25	.04	.05	0	1/12
p_{19}	1/25	0.00	0.00	0	0
p_{20}	1/25	.02	.02	1/12	0
p_{21}	1/25	.05	.06	1/12	1/12
p_{22}	1/25	.06	.06	1/12	1/12
p_{23}	1/25	.05	.05	1/12	1/12
p_{24}	1/25	.03	.04	0	0
p_{25}	1/25	.02	.02	0	0
$\|D\|$	.488 x 10^{11}, .110 x 10^{11}	.291x 10^{11}	.669 x 10^{10}	.271 x 10^{11}	.234 x 10^{11}
$\max_i d(x_i, \xi)$	14.86, 13.62	12.01	12.11	27.45	25.41

2. The designs for the models with prior information in columns III - V can be compared directly. Prior information is needed in order to construct a 12-point design because of singular $M(\xi_0)$. This may be due to the almost 2 to 1 ratio of candidate design points (25) to parameters (12) and the sparcity of the matrix M when only $N = 12$ is considered. The two 12-point designs are considerably less efficient than the D-optimal design.

3. The uniform designs with $N = 25$ are relatively close to the corresponding D-optimal designs for both the model with and the model without prior information.

Experimental designs for the single-host case (2) with model (4.1) appear in Table 4.8. We observe the following:

1. The D-optimal continuous design with 25 design points exhibits not quite twice the efficiency of the uniform continuous design using all 38 points.

2. The exact design for $N = 15$ is as efficient as the uniform design with all 38 candidate points.

3. The exact design for $N = 12$ provides a considerable loss of efficiency.

Experimental designs for the multi-host case (3) with model (4.1) appear in Table 4.9. We observe the following:

1. The D-optimal design with 34 points is more than twice as efficient as the uniform design using all 72 points.

2. The exact $N = 15$ design is as efficient as the uniform design with all 72 points.

3. Moving from an exact $N = 15$ to an exact $N = 12$ design produces a considerable reduction in efficiency.

Comparison across the cases for model (4.1) suggests the following:

1. Increasing the number of single-host candidate points (routes) from 25 (case 1) to 38 (case 2) does not provide an opportunity to improve efficiency or D-optimality.

2. Comparing the single-host (1) case with the multi-host case does not indicate that there is any obvious gain in efficiency of the variance prediction for this network when moving from a single- to a multi-host monitoring environment. There is, however, a significant improvement in D-optimality for the multi-host case.

Experimental designs for the single-host case (1) with model (4.3) appear in Table 4.10. We observe the following:

1. The uniform continuous design with 25 points (column I) is 73% as efficient as the D-optimal design.

Table 4.8: Experimental Designs for the Single-Host (with Additional Routes) of the Network Example with 8 Nodes and 12 Edges

	I	II	III	IV	V	VI
	Uniform	D-optimal Continuous	Rounded 22 points	Exchange-Type 24 points (a)	Exchange-Type 15 points	Exchange-Type 12 points (b)
N	38	38	38	38	38	38
n	38	25	22	24	15	12
p_1	1/38	0.00	0.00	0	0	0
p_2	1/38	0.00	0.00	0	0	0
p_3	1/38	.06	.06	2/24	1/15	1/12
p_4	1/38	.06	.06	1/24	1/15	1/12
p_5	1/38	.03	.03	1/24	0	0
p_6	1/38	.06	.07	2/24	1/15	1/12
p_7	1/38	.07	.07	2/24	1/15	1/12
p_8	1/38	0.00	0.00	0	0	0
p_9	1/38	.03	.03	1/24	0	0
p_{10}	1/38	0.00	0.00	0	0	0
p_{11}	1/38	0.00	0.00	0	0	0
p_{12}	1/38	.02	.02	0	0	0
p_{13}	1/38	0.00	0.00	0	0	0
p_{14}	1/38	0.00	0.00	0	0	0
p_{15}	1/38	0.00	0.00	0	0	0
p_{16}	1/38	0.00	0.00	0	0	0
p_{17}	1/38	.02	.02	0	1/15	0
p_{18}	1/38	.01	0.00	0	0	0
p_{19}	1/38	0.00	0.00	0	0	0
p_{20}	1/38	0.00	0.00	0	0	0
p_{21}	1/38	0.00	0.00	0	0	0
p_{22}	1/38	.06	.07	1/24	1/15	1/12
p_{23}	1/38	.03	.03	1/24	1/15	0
p_{24}	1/38	.03	.03	1/24	0	0
p_{25}	1/38	0.00	0.00	0	0	0
p_{26}	1/38	.03	.03	1/24	0	0
p_{27}	1/38	.03	.03	1/24	0	0
p_{28}	1/38	.01	0.00	0	0	0
p_{29}	1/38	.06	.06	1/24	1/15	1/12
p_{30}	1/38	.05	.05	1/24	1/15	1/12
p_{31}	1/38	.06	.06	1/24	1/15	1/12
p_{32}	1/38	.06	.06	2/24	1/15	1/12
p_{33}	1/38	.04	.04	1/24	1/15	1/12
p_{34}	1/38	.04	.04	1/24	1/15	0
p_{35}	1/38	.06	.06	1/24	1/15	1/12
p_{36}	1/38	.06	.06	1/24	1/15	1/12
p_{37}	1/38	0.00	0.00	0	0	0
p_{38}	1/38	.02	.02	1/24	0	0
$\lvert D \rvert$	$.212 \times 10^{10}$	$.389 \times 10^{9}$	$.404 \times 10^{9}$	$.544 \times 10^{9}$	$.542x\ 10^{9}$	$.284 \times 10^{10}$
$\max_i d(x_i, \xi)$	19.56	12.00	12.41	15.70	19.46	84.00

Table 4.9: Experimental Designs for the Multi-Host Case of the Network Example with 8 Nodes and 12 Edges

	I	II	III	IV
	D-optimal	Uniform	Exchange-Type	
	Continuous	15 point	15 point	12 point (b)
N	72	72	72	72
n	34	15	15	12
p_4	.02	0	0	0
p_5	.03	0	0	0
p_6	.03	1/15	0	0
p_9	.06	1/15	1/15	1/12
p_{12}	.01	0	0	0
p_{22}	.05	1/15	1/15	0
p_{23}	.03	0	0	1/12
p_{24}	.05	1/15	1/15	1/12
p_{26}	.02	0	0	0
p_{27}	.00	0	0	0
p_{28}	.02	0	0	1/12
p_{29}	.02	0	1/15	0
p_{31}	.04	1/15	1/15	0
p_{32}	.06	1/15	1/15	1/12
p_{33}	.04	1/15	0	1/12
p_{34}	.03	0	0	0
p_{35}	.02	0	0	0
p_{36}	.05	1/15	1/15	1/12
p_{41}	.03	0	0	0
p_{42}	.00	0	0	0
p_{43}	.01	0	0	0
p_{44}	.04	1/15	1/15	1/12
p_{48}	.03	1/15	1/15	0
p_{49}	.03	0	0	0
p_{50}	.02	0	0	0
p_{53}	.01	0	0	0
p_{54}	.04	1/15	1/15	1/12
p_{57}	.03	0	1/15	0
p_{58}	.05	1/15	1/15	1/12
p_{59}	.04	1/15	1/15	1/12
p_{60}	.05	1/15	1/15	1/12
p_{65}	.03	0	0	0
p_{66}	.03	1/15	1/15	0
p_{72}	.03	0	0	0
$\lvert D\rvert^a$	.221 x 10^8	.111 x 10^9	.613 x 10^8	.206 x 10^9
$\max_i d(x_i, \xi)^b$	11.95	36.50	22.50	43.50

Table 4.10: Experimental Designs with Variance Adjustment for the Single-Host Case of the Network Example with 8 Nodes and 12 Edges

	I	II	III	IV	V
	Uniform	D-optimal	Exchange-Type		
		Continuous	20 points	12 points (a)	12 points (b)
N	25	25	25	25	25
n	25	24	19	12	12
p_1	1/25	.05	1/20	0	0
p_2	1/25	.06	1/20	1/12	1/12
p_3	1/25	.05	1/20	1/12	1/12
p_4	1/25	.06	1/20	1/12	1/12
p_5	1/25	.03	1/20	0	0
p_6	1/25	.07	2/20	1/12	1/12
p_7	1/25	.06	1/20	1/12	1/12
p_8	1/25	.04	1/20	0	0
p_9	1/25	.01	0	0	0
p_{10}	1/25	.05	1/20	1/12	1/12
p_{11}	1/25	.06	1/20	1/12	1/12
p_{12}	1/25	.04	1/20	0	1/12
p_{13}	1/25	0.00	0	0	0
p_{14}	1/25	.01	0	0	0
p_{15}	1/25	.04	1/20	0	0
p_{16}	1/25	.06	1/20	1/12	1/12
p_{17}	1/25	.05	1.20	0	0
p_{18}	1/25	.04	1/20	1/12	0
p_{19}	1/25	.04	1/20	0	0
p_{20}	1/25	.01	0	0	0
p_{21}	1/25	.04	1/20	1/12	1/12
p_{22}	1/25	.05	1/20	1/12	1/12
p_{23}	1/25	.04	1/20	1/12	1/12
p_{24}	1/25	.02	0	0	0
p_{25}	1/25	.02	0	0	0
$\lvert D \rvert$	.116 x 10^{17}	.747x 10^{16}	.933 x 10^{16}	.308 x 10^{17}	.308 x 10^{17}
$\max_i d(x_i, \xi)$	16.53	12.01	15.27	28.00	28.00

2. The two exact designs for $N = 12$ (columns IV and V) are alternate optimal solutions.

Experimental designs for the single-host case (2) with model (4.3) appear in Table 4.11. We observe the following:

1. For practical purposes, there is little difference in efficiency among the designs in Columns I, IV, and V.

2. There is a considerable loss of efficiency in moving to $N = 12$ design points.

Experimental designs for the multi-host case (3) with model (4.3) appear in Table 4.12. We observe the following:

1. Columns II and III show that there can be a significant difference in efficiency (about 7 times) depending on where observations are taken for a given N.

2. The exact design for N=13 is as efficient as the uniform design with all 72 points. This represents a significant cost savings.

3. There is no significant improvement in moving from an exact design for N= 20 to one for N=25 (columns V and VI).The exact design for N=13 is almost as good as the exact design for N=15.

4. The $N = 20$ exact design is about 75% as efficient as the D-optimal design with 39 design points.

Table 4.11: Experimental Designs with Variance Adjustment for the Single-Host Case (Extended Number of Routes) of the Network Problem with 8 Nodes and 12 Edges

	I	II	III	IV	V	VI
	Uniform	D-optimal Continuous	Rounded 24 Points	Exchange-Type 20 points	Exchange-Type 15 points	Exchange-Type 12 points
N	38	38	38	38	38	38
n	38	27	24	20	15	12
p_1	1/38	.01	0.00	0	0	0
p_2	1/38	.04	.04	1/20	0	0
p_3	1/38	.07	.07	2/20	1/15	1/12
p_4	1/38	.06	.06	1/20	1/15	1/12
p_5	1/38	.03	.03	0	0	0
p_6	1/38	.07	.07	1/20	1/15	1/12
p_7	1/38	.07	.07	1/20	1/15	1/12
p_8	1/38	.01	0.00	0	0	0
p_9	1/38	0.00	0.00	0	0	0
p_{10}	1/38	0.00	0.00	0	0	0
p_{11}	1/38	0.00	0.00	0	0	0
p_{12}	1/38	.04	.04	1/20	1/15	0
p_{13}	1/38	.02	.02	1/20	0	0
p_{14}	1/38	0.00	0.00	0	0	0
p_{15}	1/38	0.00	0.00	0	0	0
p_{16}	1/38	.04	.04	1/20	1/15	1/12
p_{17}	1/38	.02	.02	0	0	0
p_{18}	1/38	.02	.03	0	0	1/12
p_{19}	1/38	.02	.02	0	0	0
p_{20}	1/38	0.00	0.00	0	0	0
p_{21}	1/38	0.00	0.00	0	0	0
p_{22}	1/38	.06	.05	1/20	1/15	1/12
p_{23}	1/38	.03	.03	0	0	0
p_{24}	1/38	0.00	0.00	0	0	0
p_{25}	1/38	0.00	0.00	0	0	0
p_{26}	1/38	.05	.05	1/20	1/15	1/12
p_{27}	1/38	.03	.03	1/20	1/15	1/12
p_{28}	1/38	0.00	0.00	0	0	0
p_{29}	1/38	.05	.05	1/20	1/15	1/12
p_{30}	1/38	.03	.04	1/20	1/15	1/12
p_{31}	1/38	.06	.05	1/20	1/15	0
p_{32}	1/38	.04	.04	1/20	0	0
p_{33}	1/38	.03	.03	1/20	0	0
p_{34}	1/38	.04	.04	1/20	1/15	1/12
p_{35}	1/38	.04	.04	1/20	1/15	0
p_{36}	1/38	.04	.04	1/20	1/15	0
p_{37}	1/38	0.00	0.00	0	0	0
p_{38}	1/38	0.00	0.00	0	0	0
$\lvert D\rvert$	$.374 \times 10^{16}$,	$.130 \times 10^{16}$	$.135 \times 10^{16}$	$.197 \times 10^{16}$	$.261 \times 10^{16}$	$.963 \times 10^{16}$
$\max_i d(x_i, \xi)$	20.75	12.04	13.26	18.50	22.34	36.00

Table 4.12: Experimental Designs with Variance Adjustment for the Multi-Host Case of the Network Example with 8 Nodes and 12 Edges

	I	II	III	IV	V	VI	VII
	D-optimal	Uniform	Exchange-Type				
	Continuous	13 points	13 points	15 points	20 points	25 points	12 points
N	72	72	72	72	72	72	72
n	39	13	13	15	20	24	12
p_1	.02	0	0	0	0	1/25	0
p_2	.04	1/13	1/13	1/15	1/20	1/25	1/12
p_3	.04	1/13	1/13	1/15	1/20	1/25	1/12
p_4	.03	0	1/13	1/15	0	0	0
p_6	.03	1/13	1/13	1/15	1/20	1/25	1/12
p_7	.05	1/13	1/13	1/15	1/20	1/25	1/12
p_8	.02	0	0	0	0	0	0
p_{10}	0.00	0	0	0	0	1/25	1/12
p_{13}	.01	0	0	0	0	0	0
p_{16}	.02	0	0	0	1/20	1/25	0
p_{18}	.02	0	1/13	0	1/20	1/25	0
p_{19}	.01	0	0	0	0	0	0
p_{20}	.01	0	0	0	0	0	0
p_{21}	.01	0	0	0	0	0	0
p_{26}	.03	0	0	0	0	0	0
p_{27}	0.00	0	0	0	0	0	0
p_{32}	.02	0	0	0	0	1/25	0
p_{33}	.01	0	1/13	1/15	1/20	0	0
p_{39}	.04	1/13		1/15	1/20	1/25	0
p_{40}	.04	1/13	0	1/15	1/20	1/25	1/12
p_{41}	.03	0	0	0	1/20	1/25	1/12
p_{42}	.02	0	0	0	1/20	0	0
p_{43}	.04	1/13	1/13	1/15	1/20	1/25	0
p_{45}	.04	1/13	1/13	1/15	1/20	1/25	1/12
p_{46}	.02	0	0	0	0	0	0
p_{47}	0.00	0	0	0	0	0	0
p_{48}	.01	0	0	0	0	1/25	0
p_{49}	.02	0	0	0	0	1/25	0
p_{51}	.05	1/13	1/13	1/15	1/20	1/25	1/12
p_{53}	.01	0	0	0	0	0	0
p_{56}	.05	1/13	0	0	1/20	1/25	0
p_{57}	.03	0	0	0	1/20	1/25	1/12
p_{58}	.04	1/13	0	0	1/20	1/25	0
p_{59}	.02	0	0	0	0	1/25	0
p_{60}	.05	1/13	1/13	1/15	1/20	2/25	1/12
p_{61}	.02	0	0	0	1/20	0	0
p_{62}	.02	0	0	0	0	0	0
p_{63}	.04	1/13	1/13	1/15	1/20	1/25	1/12
p_{64}	.01	0	0	0	0	0	0
p_{67}	.01	0	1/13	1/15	0	1/25	0
p_{69}	0.00	0	0	1/15	0	0	0
$\lvert D\rvert^a$	$.758 \times 10^{14}$	$.755 \times 10^{16}$	$.293 \times 10^{15}$	$.247 \times 10^{15}$	$.113 \times 10^{15}$	$.113 \times 10^{15}$	$.482 \times 10^{15}$
$\max_i d(x_i,\xi)^b$	12.02	173.34	24.87	21.63	16.13	15.73	30.67

[a] $|D| = .338x10^{15}$ for the uniform design with all 72 points.

[b] $\max_i d(x_i,\xi) = 24.73$ for the uniform design with all 72 points.

4.3 Summary and Conclusions

This analysis of two simple network graphs suggests the following:

1. The algorithms can be used successfully to construct optimal designs for the models detailed in Chapter 4.1.

2. The algorithms can suggest and estimate the efficiency of designs which are less than optimal but which may be attractive according to some operational, policy, or cost perspectives.

3. Transition from single-host to multi-host experiments allows us to gain very significantly in terms of D-optimality.

4. In general, the increase of an available number of observations N allows the use of designs where the information yield per observation is higher.

Chapter 5

Problem and Model Description

This chapter provides the results of the application of the methodologies and algorithms developed in Chapters 3.5 and 3.6 to developing a monitoring system for large computer networks.

In most cases a computer network can be represented as a graph with a given number of nodes (vertices or sites) and edges (links, communications channels). Possible objectives of experiment(s) may include evaluating such quantities as the following: delays on a given subset (subset of interest) of edges, processing times at a given subset of nodes, traveling times from one subset of nodes to another, etc. Existing software and hardware allow measuring (see, for example, (Paxson (1997)) for details and a comprehensive list of references) a large variety of network performance indicators; i.e., in general our "measurement" is a vector. Types of measurement strategies may be very different. For instance, a meter can

be installed at any chosen node to measure input and output flows; a measurement software or hardware device can be placed at a host node, and a preselected set of nodes or edges can be monitored; a practitioner can cooperate with others (i.e., there are a few host nodes) to monitor a network. Thus, if we have an opportunity to plan (design) experiments, we may look for the best subset of nodes where devices must be allocated, find the most informative subset of nodes and edges to be monitored by the given host, or in the most general setting, select the most effective team of host nodes (sites) and match them with the sets of nodes and edges to be monitored.

As noted, a monitoring system may have several focus areas and purposes. This chapter does not propose to define the optimal structure of a monitoring system for a few parameters of interest. We consider a single response case to gain a better understanding of the difficulties and benefits that can be encountered when applying experimental design technologies. In particular we start with developing a monitoring system based on delay times. To handle more complicated cases with several responses of interest, either we must (1) introduce a compound objective function (frequently this is a weighted sum of the objective functions each of which targets a single response) or (2) solve constrained optimization problems when one of the most important objective functions is minimized while others are held on or below some prescribed levels. The generalization is straightforward (c.f. Cook and Fedorov (1995), Puckelsheim (1993)), but it is beyond the scope of this

dissertation. It is hoped that this demonstration of what can relatively simply be accomplished will provide justification for this methodology and its extensions to be applied to more host computers, parameters, and performance measures.

The data for this application come from the Energy Sciences Network (ESnet) backbone, a portion of the Internet and an international computer network that is managed and funded by the Department of Energy Office of Energy Research. The ESnet was chosen for this application because of its size and importance to the Department of Energy. Each site on this network is a national laboratory, other national facility, or a national or international partner in the network. The ESnet is engineered, installed, and operated by the staff of the Networking and Telecommunications Department at Lawrence Berkeley National Laboratory (LBNL) located in Berkeley, California. Additional information about the ESnet is available on its web site (http://www.es.net).

In the considered experiment there is only one host computer at Oak Ridge National Laboratory, and there are $S = 39$ nodes (sites) which are of interest. Thus, according to Chapter 3.4, we can use the following model:

$$Y = U(X) + \mathcal{E},$$

$$\text{where } E[U(X)] = U_0, \ Var(U) = K,$$

$$E[\mathcal{E}] = 0, \;\; Var(\mathcal{E}) = \Sigma, \;\; \text{and} \;\; E[\mathcal{E}U^T] = 0, \;\; \sum_{ij} = \sigma^2 \delta_{ij}. \qquad (5.1)$$

Note that unlike Chapter 3.4, the vector U_0, the variance σ^2, and the matrix K are to be estimated from preliminary experiments. We understand that the assumption that all diagonal elements of Σ are equal is very strong, but we stay with it at this exploratory stage.

Another assumption which rigorous mathematical review may criticize is that we assume that the solution of the design problem with U_0, σ, and K replaced by their estimates is close enough to the solution of the design problem with given U_0, σ, and K. This assumption is not very restrictive if preliminary experiments contain a large number of observations.

All computations are completed using the program described in Appendix B and are based on the algorithm developed in Chapter 3.5 (see also Batsell, Fedorov, and Flanagan (1998), and Fedorov and Flanagan (1997)). For estimation of U_0, σ, and K we developed computer software equivalent to the Statistical Analysis System (SAS) software.

5.1 Data Collection

Table 5.1 lists the 39 computer host sites (sites) on the ESnet which were interrogated from a host computer at Oak Ridge National Laboratory (ORNL) to collect network delay times using the Packet Internet Groper (*ping*) network soft-

ware (authored by M. Muss, U.S. Army Ballistic Research Laboratory in 1983); see a rather detailed description in Paxson (1997). The 39 sites were chosen from the ESnet backbone map and the ESnet M-bone schematic: these are both available at the ESnet web site. A larger number of sites were interrogated initially, but sites which did not answer *pings* were dropped from further interrogation.

The *ping* command is directed to a particular host computer site and requests an answer from that site: the answer, if received, notes that the site is active. A computer program written by T. Dunigan at ORNL is used to collect the *ping* data. This program reads a list of sites to *ping* and, unless otherwise specified, *pings* each site three times. The program reports the following information:

1. the round trip times in milliseconds from the ORNL host to the interrogated site;

2. the minimum, average, and maximum delay (round trip) times; and

3. the percent packet loss.

The program also notes when a site is not reachable. The list of sites to be interrogated is then repeated in random order a specified number of times.

To minimize the impact of the time trends on the results, the list of sites was permuted at random for every interrogation session. The data represent 50 sessions of "*pinging*" for the 39 sites from about 11:00 A.M. until 2:00 P.M. E.S.T. one weekday in mid-March 1997. The conclusions from this analysis, then,

Table 5.1: ESnet Sites Interrogated by ORNL

ID	Site Acronym	Site Name
1	JLAB	Thomas Jefferson National Accelerator Facility (Newport News, VA)
2	ARM	Atmospheric Radiation Measurement Project (Lamont, OK)
3	FNAL	Fermi National Accelerator Laboratory (Batavia, IL)
4	SNL	Sandia National Laboratories Albuquerque (Albuquerque, NM)
5	KEK	KEK, Japan
6	NYU*	New York University Courant Institute (New York, NY)
7	MSRI*	Mathematical Sciences Research Institute, Univ. CA (Berkeley, CA)
8	ANL-MR1	Argonne National Laboratory-Main Router 1
9	AMES	AMES Laboratory, Iowa State University (Ames, IA)
10	FSU*	Florida State University (Tallahassee, FL)
11	CIT	California Institute of Technology (Pasadena, CA)
12	MIT*	Massachusetts Institute of Technology (Cambridge, MA)
13	FNAL-MR1	Fermi National Accelerator Laboratory-Main Router 1
14	GAT	General Atomics (San Diego, CA)
15	UTA	University of Texas at Austin (Austin, TX)
16	SRS*	Savannah River Site (Aiken, SC)
17	SLAC	Stanford Linear Accelerator (Stanford, CA)
18	INEL*	Idaho National Engineering Laboratory (Idaho Falls, ID)
19	LLNL	Lawrence Livermore National Laboratory (Livermore, CA)
20	AUCK*	University of Auckland (Auckland, New Zealand)
21	DOE	Department of Energy (Washington, DC)
22	PPPL	Princeton Plasma Physics Laboratory (Princeton, NJ)
23	UTK	University of Tennessee (Knoxville, TN)
24	LANL-MR1	Los Alamos National Laboratory - Main Router 1 (Los Alamos, NM)
25	NASA*	AMES Research Center, NASA (San Francisco, CA)
26	BNL	Brookhaven National Laboratory (Upton, NY)
27	PPPL-local	additional PPPL site
28	CU	Columbia University Academic Information Systems (New York, NY)
29	ANL	Argonne National Laboratory (Argonne, IL)
30	Pro.PPPL	additional PPPL site
31	PNNL*	Pacific Northwest National Laboratory (Richland, WA)
32	OSTI	Office of Scientific and Technical Information (Oak Ridge, TN)
33	NEVIS*	Columbia University Nevis Laboratory (Irvington, NY)
34	LBNL-MR1	Lawrence Berkeley National Laboratory - Main Router 1
35	LLNL-MR2	Lawrence Livermore National Laboratory - Main Router 2
36	NERSC*	National Energy Research Scientific Computing, LBNL (Berkeley, CA)
37	LBNL	Lawrence Berkeley National Laboratory (Berkeley, CA)
38	SNL/LLNL	Sandia National Laboratories at LLNL (Livermore, CA)
39	YALE*	Yale University (New Haven, CT)
40	ORNL	Oak Ridge National Laboratory (Oak Ridge, TN)

* Computer optimal sites.

are limited to an estimated variance-covariance structure calculated in a single day. We understand that these data are too limited to use with confidence to make serious practical inferences, but they are sufficient for a pilot study. For future, more extensive studies we initiated a long term data collection program for the estimation of U_0, σ^2, and K. This program was started on July 16, 1997, and includes interrogation sessions at 11 A.M. and 3 P.M. for weekdays only. The collected data will provide a better understanding of time trends and more accurate estimates of the parameters. This additional data can validate or supplant the variance-covariance estimates of this dissertation.

The data were processed (extraction from the ping text files, deleting outliers, identification of cases with missing values, tabulation) with the help of FORTRAN programs we developed. According to ORNL network experts, the delay times reported by the *ping* program (or even the "traceroute" program) may provide results that are overstated due to the nature of the site routers, which give may frequently give lower priority to processing *ping* requests. The minimum delay time, therefore, is used to represent the normal operating conditions of the network sites. It also helps to reduce the number of cases with missing observations (i.e., it is likely that at least one *ping* out of three will produce a response). The data are screened for missing values (i.e., non-response from a site) or site extreme values. A site extreme value is one that is at least 5 times the next largest value for that site: this definition of an extreme value was chosen in order (1) to remove

extreme observations that likely represented some unusual operating or network condition and (2) to preserve larger response values that likely would contribute to a realistic estimate of the variance. This exclusion of extreme values and missing values is consistent with providing normal network operating conditions.

5.2 Optimal Sites: Results of Computation and Discussion

Appendix C contains the variance-covariance matrix $\hat{K}$ for the ESnet sites. Pairwise covariance calculations allow the use of as much collected data as possible. Note that pairwise estimation of the elements of a variance-covariance matrix, however, does not guarantee positive definiteness. Inverting this $\hat{K}$ resulted in two diagonal elements that were less than zero. Solving for the eigenvalues λ_i of $\hat{K}$ resulted in two small, negative eigenvalues, as expected. Otherwise, the range of the λ_i is $0 < \lambda_i \leq 3651$. In most cases analyzed, the addition of the matrix $\hat{W}$ to $\hat{K}^{-1}$ to form D^{-1} produced a D matrix which was positive definite (all eigenvalues are positive). This allowed computing optimal designs for most cases considered. This exercise gives evidence that the estimation of K is indeed important and can be difficult. Part of the difficulty lies in the necessity to apply rather complicated estimation methods to guarantee the positive-definiteness of $\hat{K}$. Note that this problem is acute for data containing large numbers of missing values. To conserve space, only the diagonal and lower half of the matrix $\hat{K}$ are

printed in Appendix C. The site variances appear in bold within the matrix and are also listed in the first column of the table for easy reference.

Table 5.2 shows various monitoring designs for 10 observations based on σ = 8.0 ms. The value of the standard error σ was estimated through averaging differences between results of neighboring-in-time interrogations over the whole set of interrogations for all sites: this is approximately 8.0 ms. The variances of prediction D_{ii}, the site identification, site acronym, and variances are included in Table 5.2. The $\max_i D_{ii}$ for each design is in bold type. The optimal weights for each site are also reported in Table 5.2. The number of *pings* sent to a particular site must be proportional to the corresponding weight. Obviously, in practice we use "rounded" or uniform weights, which may lead to some increase in the maximal variance of prediction.

No loss of efficiency (i.e., the ratios of $\max_i D_{ii}$) is incurred between the 10- and 12-point uniform designs, and efficiency is only slightly reduced by moving from the D-optimal design to the former two designs. The D-optimal design is nearly four times more efficient than the 39-point design. The $\ln|D|$ is 11.4, 20.2, 11.8, and 11.4 for the D-optimal, 39-point uniform, 12-point uniform, and 10-point uniform designs, respectively. Only the 39-point design, then, shows a significant reduction in D-optimality.

Taking a closer look at the selected design points (sites), it is evident that the selected sites have the highest site variances. This means that the sites with the

Table 5.2: Various Designs to Monitor ESnet Sites ($N = 10, \sigma \simeq 8.0\ ms$)

Site ID	Site	Site Variance	D-optimal Weight	Variance of the Prediction, D_{ii}			
				D-optimal Continuous	Rounded 12 Points	Uniform 39 points	Rounded 10 points
1	JLAB	96.8	.0000	23.7	12.8	18.4	24.2
2	ARM	72.0	.0000	35.1	25.1	25.9	42.0
3	FNAL	90.7	.0000	21.2	8.2	12.8	23.2
4	SNL	8.8	.0000	2.8	1.9	2.5	2.8
5	KEK	56.2	.0000	31.1	18.4	29.6	31.5
6	NYU	1042.7	.0970	58.0	176.7	66.5	56.4
7	MSR1	289.2	.0128	57.9	39.8	34.0	61.2
8	ANL-MR1	100.9	.0000	32.7	19.8	29.3	32.6
9	AMES	83.5	.0000	26.0	11.7	18.9	27.9
10	FSU	1497.7	.1010	58.0	194.8	69.4	58.5
11	CIT	19.3	.0000	8.3	4.9	6.4	8.9
12	MIT	1002.6	.1000	57.8	187.6	68.5	57.9
13	FNAL-MR1	6.1	.0000	5.4	3.8	5.4	5.5
14	GAT	60.5	.0000	21.5	13.2	15.3	23.0
15	UTA	54.0	.0000	17.7	12.3	14.0	18.4
16	SRS	978.5	.0985	57.9	181.3	67.3	57.1
17	SLAC	93.0	.0000	26.0	14.9	15.4	31.8
18	INEL	1537.9	.1010	58.0	194.7	69.3	58.5
19	LLNL	57.8	.0000	27.7	15.5	25.7	28.2
20	AUCK	1967.8	.1036	57.9	**210.8**	**71.3**	60.0
21	DOE	25.3	.0000	12.4	6.9	11.5	12.7
22	PPPL	41.9	.0000	11.4	5.2	7.6	12.2
23	UTK	0.4	.0000	0.3	0.3	0.3	0.3
24	LANL-MR1	51.7	.0000	21.1	10.7	18.5	21.8
25	NASA	949.6	.0978	**58.9**	177.8	66.9	56.8
26	BNL	183.8	.0000	56.7	33.9	49.7	56.8
27	PPPL-local	126.0	.0000	48.2	23.1	37.8	51.3
28	CU	75.3	.0000	23.9	15.7	20.8	23.9
29	ANL	88.3	.0000	22.4	10.1	15.2	24.1
30	Pro.PPPL	35.3	.0000	15.6	9.4	14.7	15.6
31	PNL	365.1	.0853	57.9	130.1	58.8	51.0
32	OSTI	0.4	.0000	0.3	0.3	0.3	0.3
33	NEVIS	402.1	.0809	57.9	70.3	44.1	52.8
34	LBNL-MR1	62.4	.0000	16.4	7.0	11.6	17.4
35	LLNL-MR2	58.9	.0000	25.5	12.7	21.1	26.9
36	NERSC	121.2	.0236	58.0	47.2	36.4	**74.7**
37	LBNL	86.7	.0000	21.3	10.7	16.3	22.0
38	SNL/LLNL	131.8	.0000	43.1	20.9	30.8	46.3
39	YALE	1137.1	.0987	58.0	183.6	65.6	57.3

largest variances (which are actually the routes with the largest variances) will be selected for monitoring in order to minimize the variance. This is consistent with the information matrix concept of gathering the most information at those points where we might observe the greatest variances.

The ESnet map in Figure 5.1 shows the 39 sites interrogated by ORNL (marked with boxes around the site names) and the 12 sites selected as optimal design points. The fact that the number of observations N is 10 and the number of optimal design points is 12 means that fractional observations will be taken at some sites. If the 10-point uniform design is used, then sites MRI and NERSC could be excluded from monitoring. For developing a monitoring plan for *ping* data that originates at ORNL, these results also indicate that ORNL needs to monitor only the 10 - 12 sites (routes) selected as optimal in order to stay knowledgeable about the delay times on those routes on which ORNL network traffic is most likely to experience greater delays.

5.3 Conclusions

Conclusions with respect to optimal design theory and the D-optimal results for the model with correlated errors:

1. The algorithm constructs a continuous D-optimal design which can be used to benchmark alternate discrete and exact designs.

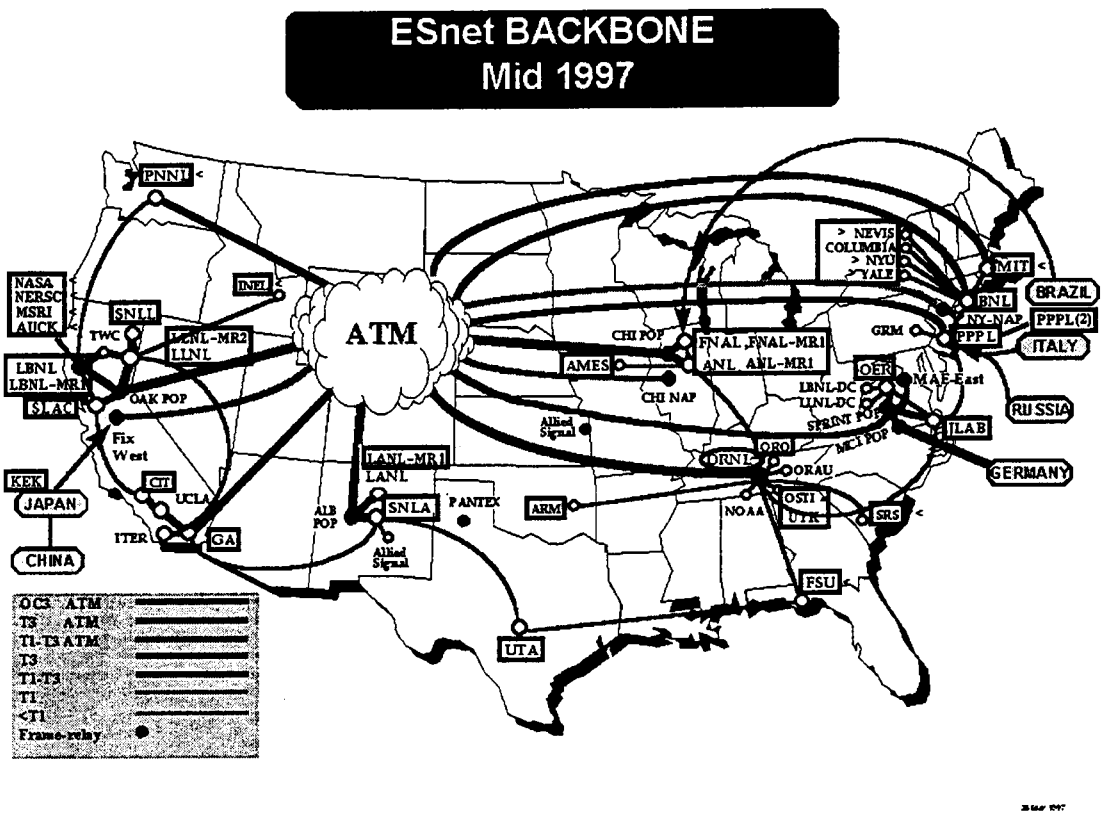

Figure 5.1: ESnet Participants in Optimal Monitoring Sites

2. Each constructed design meets the optimality criteria and is, therefore, globally optimal for the stated ratio σ^2/N. Note that while these ratios were calculated for a specific σ^2 and N, an infinite number of combinations of σ^2 and N produce the identical ratio, and many such combinations could be reasonable for this data.

3. Estimation of the covariance matrix may be difficult, depending on the application. This example demonstrates that a pairwise calculation of the covariances may result in a matrix which is not positive definite. The matrix may be ill-conditioned with some negative eigenvalues and produce some negative diagonal elements, i.e., $\hat{K}_{ii} \leq 0$. The magnitude of such a problem, if it occurs, has to be determined by the experimenter.

4. The algorithm executes rapidly on a SUN workstation and, therefore, would likely execute in even less time on a number of the faster personal computers.

Conclusions with respect to developing a monitoring system for the ESnet and other computer systems:

1. We have to emphasize that the preliminary data collection process is very much defined by the time span in which we are interested. If the time span is a day, i.e., the evolution of network traffic during the daytime is of interest, then observations must be spread throughout the desired portions of the day and perhaps repeated for several consecutive days. If we are

interested in weekly tendencies, then observations must be spread across all seven days with possibly some repeated observations, and so on. These types of experiments should be done infrequently, say once per month for the day-long case. This concept of monitoring through the use of infrequent experiments is analogous to the traditional statistical concept of sampling instead of taking a census. The goal in the monitoring arena is to design a routine activity that will provide almost the same or better information but at a much reduced expense with respect to both measurement and analysis.

2. This dissertation included most of the ESnet national community and provides a good example of the potential of an optimal monitoring system. There is, of course, much more research to be done, but the following statements seem appropriate:

   (a) Monitoring delay times using *ping* data that originates from a single host computer may be accomplished more efficiently when fewer sites are interrogated. Here efficiency is in terms of both the information gained and the costs.

   (b) A multi-host system of interrogation should be considered in order to reduce the monitoring burden on any single site and improve overall efficiency.

(c) The developed methodology can be applied to a more complete specification of the ESnet community with respect to the number of host computers and members to be interrogated.

(d) The methodology and software can be used to construct a variety of designs from which a set of host computers may be selected.

(e) Additional research is required to develop a consistent means of estimating the variance-covariance matrix $\hat{K}$.

Chapter 6

Conclusions and Future Directions

We have developed algorithms and computer software and supporting theory to construct D-optimal and linear optimal designs in the presence of spatial covariance structure to measure and monitor large computer network systems such as the Internet. The study proceeded as follows.

In the light of current levels of traffic on large computer networks such as the Internet and network performance concerns by user communities such as the Department of Energy (DOE), we considered the problem of engineering a monitoring system that could collect performance data as efficiently as possible. This data would, then, be purposefully collected to address specific performance monitoring concerns.

The efficiency concept would consider both minimizing costs and system interruption while maximizing the amount of information obtained. The term "costs" includes such costs as data collection and retention, data analysis and reporting, and the number and placement of data collection devices.

This monitoring problem immediately suggested the application of statistical experimental design methodologies. The nature of network traffic and the network itself, however, presented difficulties with directly applying existing optimal design theory methods. A network is frequently represented as a graph with the points or nodes representing host computers and with the graph edges representing network links between hosts. This is a discrete environment problem with correlated observational errors rather than the continuous one of existing optimal design theory essentially based on the assumption of uncorrelated errors.

While a number of researchers have studied different aspects of optimization in the presence of correlation (see Chapter 3), theory for D- and linear optimal designs and algorithms have not been presented or tested with an application. This dissertation develops and illustrates the required theory (see Chapter 3) with examples that include the Brownian bridge problem (see Chapter 3), simple network graphs (see Chapter 4), and a test case from the DOE Energy Sciences Network (see Chapter 5).

The background of established optimal design theory needed for this dissertation is presented in Chapter 2, which also includes a short survey of numerical

methods and algorithms. Documentation of the computer software developed to implement these methods adapted for the needs of the considered design problems (i.e., inclusion of prior information or regularization of initial information matrices) appears in Appendix A.

The theoretical approach for cases with spatial covariance structure is based on Mercer's expansion of covariance kernels in the case of continuous controlled variables (see Chapter 3) and on the direct use of convex design theory for discrete design sets (see Chapter 3). The merit of the methodology using Mercer's kernel expansion is illustrated with some examples with classical covariance structures. Documentation of the computer software developed to implement these methods appears in Appendix B.

The applicability of the methods to computer network monitoring is demonstrated with simple network graph examples for single- and multiple-host monitoring systems and with a test case from the Department of Energy (DOE) Energy Sciences Network (ESnet).

The simple network graph examples of Chapter 4 utilize the D-optimality criterion for linear models both with and without a prior information structure. The analysis is repeated for the case when the dependence of the variance of the observational errors upon the selected route is allowed. The tables (4.1 - 4.12) report the optimal design and additional designs which attempt to trade off optimality with costs reduction. The examples show that the transition from

single-host to multiple-host monitoring systems produces a significant gain with respect to efficiency and cost reduction or simplicity of a design structure.

The ESnet test case (see Chapter 5) is created using an Oak Ridge National Laboratory host computer that interrogates 39 sites of the ESnet host community. The methodology developed in Chapter 3 for constructing experimental designs for D- and linear optimality in the presence of non-zero covariance produces an optimal monitoring system with 12 sites. The map of the ESnet community in Fig. 5.1 is annotated to indicate the sites interrogated and the sites selected by optimal design. Table 5.2 compares the continuous optimal design with a few other designs with simpler structures. Estimation of the variance-covariance matrix for this example (see estimation discussion in Chapter 5 and the matrix in Appendix C) provided the opportunity to witness the difficulty of such estimation with respect to guaranteeing the positive-definiteness of the matrix.

The results of this dissertation establish the optimality conditions for the D- and linear criterion in the presence of a spatial covariance structure and discrete variable space. Associated algorithms and computer software are developed to implement the theory. In addition to constructing optimal designs, the computer network examples exhibit the ability of the algorithms to suggest and estimate the efficiency of monitoring system designs which are less than optimal but which may be attractive according to some operational, policy, or cost perspectives.

Future directions include both theoretical and research applications to com-

puter network monitoring systems. For example, the traffic on the Internet is a compound process that comprises files of different lengths, different types of transmitting packets (vehicles), and protocols. Typically, a computer network is an open system which has a number of external, uncontrolled variables. Distributions we can observe include some heavy-tailed distributions or mixtures of various distributions for which the second moments frequently are not defined. Therefore, our technique based on the first and second moments must be adapted to these new types of challenging mathematical problems.

An additional applied research direction is the development of methods to estimate the large variance-covariance matrix K for systems such as the ESnet in order to ensure the positive-definiteness of that matrix in the presence of large numbers of missing observations. The extension of the existing method to other metrics is also an open research area.

We would like to emphasize that this dissertation represents the initial stages of the groundwork that must be completed in order to continue building and expanding both convex design theory and computer network performance measurement research and application areas. It is expected that interim points in this continuing development process will likely produce both theoretical and network performance off-shoots that will lead to better understanding of statistical aspects of measuring large computer networks.

Bibliography

[1] Albert, Arthur (1972). *Regression and the Moore-Penrose Pseudoinverse.* Academic Press, New York.

[2] Atkinson, A. C., and V. V. Fedorov (1988). "Optimum Design of Experiments", in S. Kotz and N. I. Johnson (Eds.), *Encyclopedia of Statistics. Supplemental.* Wiley, New York, pp. 107-114.

[3] Batsell, S., V. Fedorov, and D. Flanagan (1998). "Multivariate Prediction: Selection of the Most Informative Components to Measure", Accepted for proceedings of MODA-5 International Workshop. Springer-Verlag, New York.

[4] Benhenni, K., and S. Cambanis (1992). "Sampling Designs for Estimating Integrals of Stochastic Process", *Ann. Statist.*, Vol. 20, pp. 161-196.

[5] Bickel, P. J., and A. M. Herzberg (1979). "Robustness of Design Against Autocorrelation in Time, I", *Ann. Statist.*, Vol. 7, pp. 77-95.

[6] Box, G. E. P., and Norman R. Draper (1987). *Empirical Model-Building and Response Surfaces.* Wiley, New York.

[7] Box, G. E. P., W. G. Hunter and J. S. Hunter (1978). *Statistics for Experiments. An Introduction to Design, Analysis, and Model Building.* Wiley, New York.

[8] Brimkulov, U., G. Krug and V. Savanov (1986). *Design of Experiments for Random Fields and Processes.* Nauka, Moscow.

[9] Brownlee, N. (1995) "New Zealand Experiences with Network Traffic Charging", Presented at the MIT Workshop on Internet Economics, March 1995.

[10] Cambanis, S. (1985). "Sampling Design for Estimating a Random Process", in Rao, C. R. (Ed.) *Handbook of Statistics*, Vol. 5, North-Holland, New York, pp. 337-362.

[11] Cheng, Chih-Hsu (1991). "Optimal Sampling for Traffic Volume Estimation", Ph.D. dissertation, Carlson School of Management, University of Minnesota. December 1991.

[12] Cook, R. D. and C. J. Nachtsheim (1982). "Model Robust, Linear-Optimal Designs", *Technometrics*, Vol. 24, pp. 49-54.

[13] Cottrell, L., G. Haney, T. Healy, L. Logg, D. Martin, L. White and W. Wing (1997). "ESnet's Internet Monitoring Activities", *IEEE Communications*, January 1997.

[14] Cressie, N. A. (1991). *Statistics for Spatial Data.* Wiley, New York.

[15] Downing, D., V. Fedorov, T. Dunigan, and S. Batsell (1997). "Poisson Type Models and Descriptive Statistics of Computer Network Information Flows", Oak Ridge National Laboratory, Oak Ridge, Tennessee, ORNL/TM-1368.

[16] Fedorov, V. V. (1972). *Theory of Optimal Experiments.* Academic Press, New York.

[17] Fedorov, V. V. (1996). "Design of Spatial Experiments: Model Fitting and Prediction", in S. Ghosh and C. R. Rao, (Eds.), *Handbook of Statistics*, Vol. 13., Elsevier Science, New York.

[18] Fedorov, V. V. and A. C. Atkinson (1988). "The Optimum Design of Experiments in the Presence of Uncontrolled Variability and Prior Information,", in Y Dodge, V. V. Fedorov, and H. P. Wynn, (Eds.), *Optimal Design and Analysis of Experiments*, North-Holland, Amsterdam.

[19] Fedorov, V. V. and D. M. Flanagan (1997). "Optimal Monitoring of Computer Networks", Presented at 1997 AMS/SIAM/IMS Joint Summer Re-

search Conference in the Mathematical Sciences, June 28, 1997. To appear in proceedings.

[20] Fedorov, V. V. and D. M. Flanagan (1998). "Optimal Monitoring Network Design Based on Mercer's Expansion of Covariance Kernel", *Journal of Combinatorics and Information Science.* To appear February 1998.

[21] Fedorov, V. V. and A. B. Malyutov (1972). "Design of Experiments in Regression Problems", *Math. Operationsforsch. Statist.*, Vol. 3, pp. 281-321.

[22] Fedorov, V. V. and W. Muller (1989). "Comparison of Two Approaches in the Optimal Design of an Observation Network", *Statistics*, Vol. 19, pp. 339-351.

[23] Fedorov, V. V. and P. Hackl, (1997). *Model-Oriented Design of Experiments.* Springer-Verlag New York, Inc., New York.

[24] Fisher, R.A. (1947). *The Design of Experiments.* 4th Edition, Hafner-Publishing Co., Inc., New York.

[25] Floyd, S. (1994). "Wide Area Traffic: The Failure of Poisson Modeling", SIG COMM 1994 Conference, Lawrence Berkeley National Laboratory, California, 325238, CONF-9408202–2.

[26] Horn, R. A. and C. R. Johnson (1985). *Matrix Analysis.* Cambridge University Press, Cambridge, United Kingdom.

[27] Hurst, H. E. (1951). "Long-term storage capacity of reservoirs", *Transactions of the American Society of Civil Engineers*, Vol. 116, pp. 770-799.

[28] Hurst, H. E. (1956). "Methods of using long-term storage in reservoirs", *Proceedings of the Institution of Civil Engineers, Part I*, Vol. 5, pp. 519-590.

[29] Kanwal, R. P. (1971). *Linear Integral Equations, Theory and Technique.* Academic Press, NY.

[30] Khuri, A. I. (1984). "A Note on D-Optimal Designs for Partially Nonlinear Regression Models", *Technometrics*, Vol. 26, pp. 59-61.

[31] Kiefer, J. (1959). "Optimal Experimental Designs", *Journal of the Royal Statistical Society, Series B*, Vol. 21, pp. 272-319.

[32] Kiefer, J. C. and H. P. Wynn (1984). "Optimum and minimax exact treatment designs for one-dimensional auto-regressive errors processes", *Annals of Statistics*, Vol. 12, pp. 431-450.

[33] Leland, W., M. Taqqu, W. Willinger, and D. Wilson (1994). "On the Self-Similar Nature of Ethernet Traffic (Extended Version)", *IEEE/AC Transactions on Networking*, Vol. 2(1), February 1994.

[34] Matèrn, B. (1986). *Spatial Variation, Lecture Notes in Statistics.* Springer-Verlag, New York.

[35] Mitchell, T. (1974). "An Algorithm for Construction of D-optimal Design", *Technometrics*, Vol. 16, pp. 203-210.

[36] Paxson, V. (1995). "Fast Approximation of Self-Similar Network Traffic", Lawrence Berkeley National Laboratory, California, 36750, CONF-9508118-1.

[37] Paxson, Vern (1996). "Towards a Framework for Defining Internet Performance Metrics". *Proceedings of INET '96*, Montreal, June 1996.

[38] Paxson, V. (1997). "Measurement and Analysis of End-to-End Internet Dynamics," Ph.D. thesis, University of California, Berkeley, UCB//CSD-97-945, Lawrence Berkeley National Laboratory, LBNL-40319.

[39] Pilz, J. (1983). *Bayesian Estimation and Design in Linear Models.* Teubner, Leipzig.

[40] Pilz, J. (1989). *Bayesian Estimation and Experimental Design in Linear Regression Models.* Wiley, New York.

[41] Pilz, J. (1991). *Bayesian Estimation and Experimental Design in Linear Regression Models.* Wiley, New York.

[42] Puckelsheim, F.(1993). *Optimal Design of Experiments.* Wiley, New York.

[43] Rabinowitz, N. and D. M. Steinberg (1990). "Optimal Configuration of a Seismographic Network: A Statistical Approach", *Bulletin of the Seismological Society of America*, Vol. 80, pp. 187-196.

[44] Rao, C. R. (1973). *Linear Statistical Inference and its Applications, 2nd Edition.* Wiley, New York.

[45] Ripley, B. D. (1981). *Spatial Statistics.* Wiley, New York.

[46] Sacks, J. and D. Ylvisaker (1966). "Designs for regression problems with correlated errors", *Ann. Math. Statist.,* Vol. 37, pp. 66-89.

[47] Sacks, J. and D. Ylvisaker (1968). "Designs for regression problems with correlated errors, many parameters", *Ann. Math. Statist.,* Vol. 39, pp. 49-69.

[48] Sacks, J. and D. Ylvisaker (1970a). "Designs for regression problems with correlated errors, III", *Ann. Math, Statist.,* Vol. 41, pp. 2057-2074.

[49] Sacks, J. and D. Ylvisaker (1970b). "Statistical designs and integral approximation", Proceedings 12th Bien. Sem. Canad. Math. Cong., *Canad. Math. Cong.,* Montreal, pp. 115-136.

[50] SAS/QC Software: (1995). *Design of Experiments Tools.* SAS Institute, Cary, North Carolina.

[51] Seber, G. A. F. and C. J. Wild (1989). *Nonlinear Regression.* Wiley, New York.

[52] Shewry, M. C., and H. P. Wynn (1987). "Maximum entropy sampling", *Journal of Applied Statistics*, Vol. 14, pp. 165-170.

[53] Silvey, S. D. (1980). *Optimal Design.* Chapman and Hall, London, England.

[54] Vardi, Y. (1996). "Network Tomography: Estimation Source - Destination Traffic From Link Data", *JASA*, 91, pp. 365-377.

[55] Wahba, G. (1974). "Regression Design for Some Equivalence Classes of Kernels", *Ann. Statist.*, Vol. 2, pp. 925-934.

[56] Whittle, P. (1973). "Some General Points in the Theory of Optimal Experimental Design", *J. Roy Statist. Soc.*, Vol, 35, Ser. B, pp. 123-130.

[57] Willinger, W., M. S. Taqqu and A. Erramilli (1996). "A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks", in F. P. Kelly, S. Zachary and I. Ziedins, (Eds.), *Stochastic Networks Theory and Applications*, Clarendon Press, Oxford. pp. 339-366.

[58] Wu, C. F. and H. Wynn (1978). "The Convergence of General Step-Length Algorithms for Regular Optimum Design Criteria", *Ann. Statist.*, Vol . 6, pp. 1273-1285.

Appendices

Appendix A

Description of Computer Software for Constructing D-Optimal and Linear Optimal Experimental Designs for Linear Models

The computer programs that construct D- and linear optimal designs for the continuous and discrete models of Chapters 2 and 3 are briefly described. The program steps, input data and outputs are presented.

Input Data. Up to three input files are needed for this program:

1. the points representing the experimental design space and an initial set of design points (required),

2. either a prior information matrix M_0 or prior variance-covariance matrix D_0 (optional; either M_0 or D_0 is input, not both), and

3. the utility matrix A (required for the linear criterion).

A grid is used to represent the design space in terms of the model. The grid is input as an N x p matrix, where N = the number of grid points and p = the number of parameters of the assumed model. Each row of the matrix is the vector $f(x_i)$, the realization of the model function using point x_i, as described below.

The input data files can be created in any manner that results in data separated by spaces with a marker (e.g., hard return) at the end of each row. The following columns of information are necessary: "weight", and a column for each parameter of the linear model (with a total of p parameters). For example, assuming the second order linear model

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2,$$

the column entries for each design region point are as follows:

1. The first column ("weight") contains a zero for all points which are in the design region but which are not included in the initial set of design points. If a point is part of the initial design set, its weight p_i is nonzero and generally is set to $1/n$, where n = the number of initial design points, with $\sum_{i=1}^{n} p_i = 1$.

2. The second column contains a "1" for the β_0 parameter.

3. The third through $p+2$ columns will contain the function $f_j(x_i)$ corresponding to parameter β_j. For the example model, column three contains all the X_1 values; column four contains all the X_2 values corresponding to each X_1 value; column five contains the values of X_1^2; column six contains the values of X_2^2; and column seven contains the values of X_1 x X_2.

The first line of the input data file must specify the number of variables and the number of estimated parameters for the linear model.

The user is prompted from the program to specify the following:

1. the name of the file which contains the experimental design region and initial set of design points.

2. the maximum number of iterations (the stopping criterion) and the step size α (α is a factor of the initial weight p_i).

3. the number of points k to be added or deleted at each iteration, i.e., k = the length of the excursion.

4. whether a prior information matrix M_0 or prior variance-covariance matrix D_0 will be input. If so, the user is prompted to specify which is being input and whether the matrix will be input from a file or the keyboard.

5. a multiplier for either prior matrix M_0 or D_0. If none is desired, the user should specify any negative number (i.e., multiplier < 0).

Program Steps and Outputs.

1. The initial information matrix M is calculated from information matrix M_d of the initial set of n design points and the optional prior information matrix M_0 (or $M_0 = D_0^{-1}$) input by the user.

2. The initial variance-covariance matrix D is calculated as $D = M^{-1}$. The variance covariance matrices D_0 and D_d are calculated directly from matrices M_0 and M_d, respectively.

3. The initial information is printed to the output file: ("dopt.out" or "linopt.out" for D- and linear criteria, respectively) number of design space points N, number of initial design points n, α, maximum number of iterations, length of excursion k, M_0, M_d, M, D_0, D_d, D, and $|D|$, and tr AD (linear optimal only). If no prior information is provided, the message, "No prior information was input," appears on the monitor, and only matrices M and D are calculated and printed.

4. The program then repeats the following steps until the optimality criterion (i.e., the maximum number of runs) is met:

   (a) The "Forward" procedure sequentially adds k points to the current set of design points. The matrix D is updated after each point x_i is added. The weight p_i of each newly added point is updated (i.e., increased). This process is completed k times.

   (b) The "Backward" procedure sequentially deletes k points from the current set of design points. The matrix D is updated after each point x_i is deleted. The weight p_i of the newly deleted point is updated (i.e., decreased). This process is completed k times.

When the optimality criterion is met (i.e., the maximum number of iterations), the determinant of the optimal matrix D is calculated (and $\mathrm{tr}\, AD$ for the linear optimal criterion), and these results are printed to the respective output files: D, $|D|$, $\mathrm{tr}\, AD$ for linear optimality, the optimal set of design points $\{x_i\}$, and their corresponding weights $\{p_i\}$.

Program Default Values and Notes. Program default values that may need adjustment:

   (a) Maximum number of model parameters ("param" and arrays and matrices in dimension statements) = 30. Before using the program, the user should verify that the matrix dimensions in the print formats are

at least as large as needed.

(b) Maximum number of design space points (arrays and matrices in dimension statements).

(c) Maximum number of excursions (kpt): None.

(d) Maximum number of iterations: None. Change output print statement if > 9999.

(e) The user should verify that the format statements for the determinants are appropriate for the magnitudes expected. If the determinant is too large for the format, the output will display asterisks. If the determinant is too small for the format, the output will display zeroes. Current formats print a double precision format, and $\ln|D|$ is printed to reduce formatting errors.

Additional notes:

(a) The algorithm may use a constant stepsize α or the diminishing α from Chapters 2 and 3.

(b) The optimal solution has to be checked by the user to verify that the optimality criteria have been met.

(c) The user may want to experiment with different values of α and the maximum number of iterations.

(d) The simple subroutines for calculating the matrix determinant and inverse are from Fedorov. Their accuracy has been verified against similar-purpose subroutines by LApack and LINPACK.

Appendix B

Description of Computer Software for Constructing D-Optimal Optimal Experimental Designs with Covariance Structure

The computer programs that construct D-optimal and linear optimal designs for linear models with covariance structure as developed in Chapter 3 are described briefly. The program steps, input data description, and outputs are presented.

This algorithm bears a strong resemblance to the iterative process of the algorithms of Appendix A, except that this algorithm operates on the parameter variance-covariance matrix K only. After input of the matrix K, this algorithm proceeds through a series of forward and backward procedures which increase and decrease, respectively, the weights p_i until a maximum number of iterations have passed. Each forward or backward procedure may be repeated a number k of times (called excursions) before moving to the next procedure.

In the initial stage, $D(\xi_0)$ is computed as $(K^{-1} + W(\xi_0))^{-1}$, where $W(\xi_0) = (N\sigma^{-2})I$. The number N is the number of observations , and σ^2 is the variance of the partitioned error ε.

Within the forward procedure of iteration t, the minimax process selects the maximum variance of prediction D_{ii} (DAD_{ii} for linear optimality, where A is a utility matrix) and increments the weight p_i for that parameter index by the step size alpha. Similarly, the backward procedure selects the minimum variance of prediction D_{ii} (DAD_{ii} for linear optimality) and decreases the weight p_{ii} for that parameter index. The matrix D_t^+ (from the forward procedure) or D_t^- (from the backward procedure) and its determinant are updated at each excursion of the forward and backward procedures as developed in Chapter 3.

Input Data and Program Outputs. Up to three input files are needed for this program:

1. the variance-covariance matrix K or its estimate $\hat{K}$ of the model parameters

(required),

2. the weights p_i for each parameter (required); $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1$, and n = the number of initial design points. An easy initial design includes all candidate points m ($n = m$) with weights $p_i = 1/n$.

3. the utility matrix A (required for the linear criterion only).

The input file for matrix K must have the following format:

1. Line 1 contains the number of model parameters m.

2. Lines 2 through m+1 contain the rows of the matrix K. Each element within a line should be separated by at least one space. Each line is separated by a hard return. No specific numeric format is required.

The input file for the weights p_i will consist of m rows of information and must have the following format:

1. The first entry in each row is an integer number used as an identification number.

2. The second entry is the weight p_i represented as a decimal fraction and separated from the identification number by at least one space.

The utility matrix A has the following format:

1. There are m lines with m entries per line separated by blanks. Each line is separated by a hard return.

The user is prompted (from the monitor) to specify (from the keyboard) each of the following items:

1. the name of the file which contains the variance-covariance matrix K,
2. the name of the file which contains the weights p_i,
3. the name of the file which contains utility matrix A (linear optimality only),
4. the maximum number of iterations (the stopping criterion),
5. the initial step size α_0 (α_0 is a factor of the initial weight p_i), and change values γ separated by a space,
6. the number of points k to be added or deleted at each iteration, i.e., $k =$ the length of the excursion, and
7. the variance σ_ε^2.

Program outputs: all initial input information, D_0, $|D_0|$, D^*, $|D^*|$, initial tr AD (linear only), optimal tr AD (linear only), the maximum variance of the prediction at the last iteration, and optimal design points with corresponding weights p_i. For matrix D, the subscript "0" indicates the initial value, and the superscript "$*$" indicates the value at the last iteration.

Default values and additional notes. Program default values:

1. Maximum number of model parameters ("param" and arrays and matrices in dimension statements) = 30. Before using the program, the user should

verify that the matrix dimensions in the print formats are at least as large as needed.

2. Maximum number of design space points (arrays and matrices in dimension statements).

3. Maximum number of excursions (kpt): None.

4. Maximum number of iterations: None. Change output print statement if > 9999.

5. The user should verify that the format statements for the determinants are appropriate for the magnitudes expected. If the determinant is too large for the format, the output will display asterisks. If the determinant is too small for the format, the output will display zeroes. Current formats print a double precision format, and $\ln |D|$ is printed to reduce formatting errors.

Additional notes:

1. The algorithm may use a constant step size α or the diminishing α from Chapters 2 and 3.

2. The optimal solution has to be checked by the user to verify that the optimality criteria have been met.

3. The user may want to experiment with different values of α and the maximum number of iterations.

4. The simple subroutines for calculating the matrix determinant and inverse are from Fedorov. Their accuracy has been verified against similar-purpose subroutines by LApack and LINPACK.

Appendix C

Variance-Covariance Matrix for Interrogated ESnet Sites

Variance-Covariance Matrix for Interrogated ESnet Sites

Site	Site Variance	JLAB	ARM	FNAL	SNL	KEK
JLAB	96.9	**96.9**				
ARM	72.0	60.0	**72.0**			
FNAL	90.7	89.6	57.7	**90.7**		
SNL	8.8	27.8	16.2	24.4	**8.8**	
KEK	56.2	52.8	42.1	53.0	11.2	**56.2**
NYU	1042.7	-23.8	-45.1	-21.8	-7.1	-17.0
MSRI	289.2	154.4	94.1	159.5	43.0	83.1
ANL-MR1	100.9	96.1	59.8	82.3	28.8	49.6
AMES	83.5	79.9	53.8	87.2	20.3	55.6
FSU	1497.7	10.6	-43.4	2.5	9.9	-30.0
CIT	19.3	32.6	21.8	39.6	8.4	19.7
MIT	1002.6	16.2	-11.3	6.3	11.5	-3.7
FNAL-MR1	6.1	6.0	8.2	6.1	0.0	15.2
GAT	60.5	64.2	38.3	72.2	18.3	29.9
UTA	54.0	-64.4	-36.8	-64.7	-19.1	-29.6
SRS	978.5	-50.3	-54.5	-58.0	-15.4	-46.7
SLAC	93.0	84.0	70.1	87.5	23.2	49.2
INEL	1537.9	-41.2	-41.8	-29.2	-10.0	-53.1
LLNL	57.8	60.0	45.7	58.9	13.8	56.0
AUCK	1967.9	-16.0	-67.2	-36.4	5.1	-68.6
DOE	25.3	38.1	29.9	38.7	8.5	37.0
PPPL	42.0	59.7	38.2	62.9	16.3	35.5
UTK	0.4	-0.5	0.6	0.0	-0.1	-0.6
LANL-MR1	51.8	60.1	44.2	61.3	14.3	51.8
NASA	949.6	-47.7	-74.1	-53.4	-12.1	-43.6
BNL	183.8	130.5	81.0	113.0	39.1	66.6
PPPL-local	126.0	90.1	63.5	103.0	21.5	70.4
CU	75.3	81.6	47.9	71.7	25.5	33.3
ANL	88.3	87.8	56.7	91.0	23.9	53.5
Pro.PPPL	35.3	52.8	36.0	44.4	14.5	37.3
PNNL	365.1	56.9	36.0	54.5	16.0	41.0
OSTI	0.4	-0.8	0.8	-0.8	-0.2	-0.6
NEVIS	402.1	177.8	104.8	172.4	50.7	93.4
LBNL-MR1	62.4	73.5	48.9	75.7	19.5	48.6
LLNL-MR2	59.0	58.7	42.9	67.6	13.4	50.5
NERSC	121.2	65.4	84.1	68.6	17.4	43.8
LBNL	86.7	91.3	58.9	85.6	26.0	52.1
SNL/LLNL	131.8	97.5	63.6	109.2	25.0	64.8
YALE	1137.1	-64.6	-57.6	-60.3	-15.4	-46.8

Variance-Covariance Matrix for Interrogated ESnet Sites (continued)

Site	NYU	MSRI	ANL-MR1	AMES	FSU	CIT
NYU	**1042.7**					
MSRI	-30.3	**289.2**				
ANL-MR1	-26.9	142.0	**100.9**			
AMES	-19.4	144.7	71.1	**83.5**		
FSU	534.4	-6.1	23.5	-9.1	**1497.7**	
CIT	-10.7	65.7	26.6	37.4	-2.5	**19.3**
MIT	97.6	12.4	31.4	-1.0	428.7	-9.8
FNAL-MR1	-3.8	6.8	5.4	8.9	-16.7	2.0
GAT	-15.3	123.7	56.2	65.0	8.3	32.7
UTA	9.6	-112.5	-61.6	-57.0	-14.3	-26.2
SRS	155.2	-92.1	-56.0	-56.8	101.9	-20.3
SLAC	-30.5	147.8	78.1	80.6	-13.5	36.7
INEL	552.6	-27.2	-47.9	-29.0	681.8	-4.9
LLNL	-17.2	94.4	57.4	59.7	-20.6	21.6
AUCK	10.8	-65.9	0.4	-51.3	301.3	-25.6
DOE	-11.5	61.3	35.7	39.7	-21.0	14.8
PPPL	-12.2	105.6	54.8	58.2	2.6	26.4
UTK	-1.8	0.1	-0.7	-0.1	1.0	0.2
LANL-MR1	-17.0	99.6	56.0	61.0	-16.1	23.9
NASA	371.8	-98.3	-44.1	-51.6	502.8	-23.6
BNL	-37.3	195.4	136.0	97.7	29.6	37.3
PPPL-local	-22.5	168.6	77.1	101.4	-21.0	45.8
CU	-26.7	125.9	84.5	59.9	23.0	24.8
ANL	-21.6	153.4	81.5	84.4	2.8	37.4
Pro.PPPL	-10.3	73.9	55.5	41.1	3.8	13.4
PNNL	60.2	89.2	58.0	50.8	37.5	19.9
OSTI	-2.7	-1.5	-0.7	-0.7	1.3	-0.3
NEVIS	-67.2	321.9	170.4	154.8	-12.1	66.6
LBNL-MR1	-18.5	126.7	68.5	71.2	-2.6	30.7
LLNL-MR2	-6.4	109.4	49.9	67.7	-13.2	29.8
NERSC	-45.9	114.0	60.8	64.9	-51.3	29.4
LBNL	-25.5	145.8	90.6	77.0	8.7	31.4
SNL/LLNL	-23.8	183.0	85.0	104.0	-7.8	48.2
YALE	425.1	-60.5	-61.7	-61.7	165.2	-24.9

Variance-Covariance Matrix for Interrogated ESnet Sites (continued)

Site	MIT	FNAL-MR1	GAT	UTA	SRS	SLAC
MIT	**1002.6**					
FNAL-MR1	-3.9	**6.1**				
GAT	-2.2	0.0	**60.5**			
UTA	-1.2	-0.3	-51.1	**54.0**		
SRS	78.9	-6.3	-40.4	50.1	**978.5**	
SLAC	-2.9	5.9	66.4	-61.5	-52.3	**93.0**
INEL	270.6	-18.0	-8.0	6.9	387.8	-19.6
LLNL	3.8	14.2	34.9	-35.0	-47.2	55.3
AUCK	256.7	-14.0	-22.7	36.2	228.3	-65.9
DOE	0.8	9.5	23.3	-22.8	-36.1	36.5
PPPL	5.2	4.1	48.0	-43.1	-37.9	58.3
UTK	3.3	-0.2	0.3	0.6	0.2	0.1
LANL-MR1	5.4	11.9	39.6	-37.4	-41.9	57.0
NASA	231.6	-11.6	-36.2	42.9	188.3	-61.3
BNL	42.1	6.9	78.2	-84.6	-74.7	107.2
PPPL-local	-7.3	13.2	76.2	-64.1	-70.8	95.0
CU	24.2	0.2	53.6	-56.4	-47.0	68.3
ANL	11.0	6.8	68.3	-62.2	-58.0	84.6
Pro.PPPL	16.7	7.8	26.0	-30.3	-36.2	42.2
PNNL	-66.4	6.1	36.2	-40.8	-56.3	50.0
OSTI	0.5	-0.1	-0.5	0.8	2.5	-0.4
NEVIS	31.6	8.7	129.7	-122.4	-116.5	160.1
LBNL-MR1	5.7	7.5	55.1	-50.8	-50.8	70.5
LLNL-MR2	-3.8	10.7	48.3	-40.9	-47.3	62.1
NERSC	-19.6	7.7	49.9	-48.0	-41.8	92.3
LBNL	17.0	6.6	60.9	-61.0	-56.5	81.0
SNL/LLNL	-3.1	9.0	84.1	-72.1	-69.9	100.4
YALE	193.1	-8.1	-41.8	52.4	497.1	-74.3

Variance-Covariance Matrix for Interrogated ESnet Sites (continued)

Site	INEL	LLNL	AUCK	DOE	PPPL	UTK
INEL	**1537.9**					
LLNL	-45.6	**57.8**				
AUCK	414.4	-52.2	**1967.9**			
DOE	-27.4	37.8	-38.1	**25.3**		
PPPL	-20.5	39.4	-26.2	25.9	**42.0**	
UTK	1.0	-0.5	3.3	-0.4	-0.1	**0.4**
LANL-MR1	-36.6	53.8	-48.5	35.5	40.9	-0.4
NASA	474.0	-44.6	198.5	-31.5	-33.5	-0.3
BNL	-63.3	77.4	8.5	48.2	75.2	-0.8
PPPL-local	-33.6	74.0	-82.9	49.8	68.8	0.0
CU	-24.6	40.9	11.7	25.3	47.8	-0.4
ANL	-32.7	59.1	-37.2	38.8	60.7	-0.1
Pro.PPPL	-29.6	40.4	-6.7	25.6	29.6	-0.5
PNNL	-42.4	42.0	240.2	27.3	36.4	-1.5
OSTI	1.8	-0.6	4.0	-0.4	-0.5	0.1
NEVIS	1.8	108.0	-16.5	69.2	114.2	0.5
LBNL-MR1	-31.3	52.9	-37.4	34.7	50.5	-0.2
LLNL-MR2	-22.6	52.2	-66.1	35.2	45.1	0.1
NERSC	-9.8	48.7	-111.0	32.7	45.8	0.4
LBNL	-40.4	58.7	-19.8	37.5	57.0	-0.4
SNL/LLNL	-33.8	70.1	-74.8	46.8	72.9	-0.1
YALE	320.9	-47.9	117.1	-35.6	-40.8	4.2

Variance-Covariance Matrix for Interrogated ESnet Sites (continued)

Site	LANL-MR1	NASA	BNL	PPPL-local	CU	ANL
LANL-MR1	**51.8**					
NASA	-42.9	**949.6**				
BNL	76.0	-58.9	**183.8**			
PPPL-local	75.3	-62.4	106.3	**126.0**		
CU	42.1	-33.0	114.7	63.6	**75.3**	
ANL	60.7	-50.0	111.7	99.7	70.1	**88.3**
Pro.PPPL	37.3	-29.6	74.5	46.8	42.5	44.9
PNNL	41.3	-38.9	79.3	57.4	46.0	54.0
OSTI	-0.5	-3.9	-1.0	-0.9	-0.6	-0.8
NEVIS	110.7	-99.7	232.6	177.2	149.1	167.6
LBNL-MR1	53.3	-43.9	93.6	84.8	57.4	73.8
LLNL-MR2	52.2	-41.2	68.7	85.2	39.4	65.6
NERSC	48.4	-74.3	83.7	78.0	51.9	66.9
LBNL	58.5	-48.8	123.0	87.7	76.4	84.0
SNL/LLNL	73.1	-64.3	117.3	126.3	73.7	105.1
YALE	-44.6	27.1	-84.8	-74.6	-59.9	-60.3

Variance-Covariance Matrix for Interrogated ESnet Sites (continued)

Site	Pro.PPPL	PNNL	OSTI	NEVIS	LBNL-MR1	LLNL-MR2
Pro.PPPL	**35.3**					
PNNL	34.8	**365.1**				
OSTI	-0.5	-1.4	**0.4**			
NEVIS	90.1	99.6	-1.7	**402.1**		
LBNL-MR1	39.4	46.7	-0.7	139.3	**62.4**	
LLNL-MR2	32.2	37.5	-0.6	113.7	56.5	**59.0**
NERSC	35.0	35.6	0.2	122.3	56.6	51.9
LBNL	50.4	56.6	-0.7	167.4	70.6	57.5
SNL/LLNL	47.2	56.4	-1.0	193.3	87.9	83.5
YALE	-36.4	-66.3	1.8	-66.0	-54.5	-48.7

Variance-Covariance Matrix for Interrogated ESnet Sites (continued)

Site	NERSC	LBNL	SNL/LLNL	YALE
NERSC	**121.2**			
LBNL	64.9	**86.7**		
SNL/LLNL	79.5	93.8	**131.8**	
YALE	-86.4	-63.7	-62.4	**1137.1**

Vita

Deborah Flanagan received her Bachelor of Science in Statistics (1975) and Master of Science in Statistics (1981) from the University of Tennessee, Knoxville. She is currently a staff member of Oak Ridge National Laboratory with 20 years of experience in the application to physical sciences areas of a variety of statistical methodologies. Statistical methodology experience includes but is not limited to experimental design, performance metrics, reliability, availability, maintainability, risk analysis, sampling, and Data Quality Objectives for environmental remediation.