OPTIMAL EXPERIMENTAL DESIGN IN THE CONTEXT OF

OBJECTIVE-BASED UNCERTAINTY QUANTIFICATION

A Dissertation

by

ROOZBEH DEHGHANNASIRI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Edward R. Dougherty
Committee Members,	Byung-Jun Yoon
	Xiaoning Qian
	Erchin Serpedin
	Ivan Ivanov
Head of Department,	Miroslav M. Begovic

August 2016

Major Subject: Electrical Engineering

ABSTRACT

In many real-world engineering applications, model uncertainty is inherent. Large-scale dynamical systems cannot be perfectly modeled due to systems complexity, lack of enough training data, perturbation, or noise. Hence, it is often of interest to acquire more data through additional experiments to enhance system model. On the other hand, high cost of experiments and limited operational resources make it necessary to devise a cost-effective plan to conduct experiments. In this dissertation, we are concerned with the problem of prioritizing experiments, called *experimental design*, aimed at uncertainty reduction in dynamical systems. We take an objective-based view where both uncertainty and modeling objective are taken into account for experimental design. To do so, we utilize the concept of mean objective cost of uncertainty to quantify uncertainty.

The first part of this dissertation is devoted to the experimental design for gene regulatory networks. Owing to the complexity of these networks, accurate inference is practically challenging. Moreover, from a translational perspective it is crucial that gene regulatory network uncertainty be quantified and reduced in a manner that pertains to the additional cost of network intervention that it induces. We propose a criterion to rank potential experiments based on the concept of mean objective cost of uncertainty. To lower the computational cost of the experimental design, we also propose a network reduction scheme by introducing a novel cost function that takes into account the disruption in the ranking of potential experiments caused by gene deletion. We investigate the performance of both the optimal and the approximate experimental design methods on synthetic and

real gene regulatory networks.

In the second part, we turn our attention to canonical expansions. Canonical expansions are convenient representations that can facilitate the study of random processes. We discuss objective-based experimental design in the context of canonical expansions for three major applications: filtering, signal detection, and signal compression. We present the general experimental design framework for linear filtering and specifically solve it for Wiener filtering. Then we focus on Karhunen-Loève expansion to study experimental design for signal detection and signal compression applications when the noise variance and the signal covariance matrix are unknown, respectively. In particular, we find the closed-form solution for the intrinsically Bayesian robust Karhunen-Loève compression which is required for the experimental design in the case of signal compression.

To my beloved parents, Atifeh and Hamid,

and

To my best friend, my brother, Razi

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Edward R. Dougherty, for his guidance, support, and encouragement throughout my Ph.D. studies. I have been most fortunate to have him as my advisor. He has always inspired me with his motivation for math and his insight in tackling the hardest problems. He has changed the way I view science and for this I will be eternally grateful. I would like to thank Professor Byung-Jun Yoon for his invaluable collaboration and fruitful discussions during my first two years at Texas A&M. I would also like to thank Professor Xiaoning Qian, Professor Erchin Serpedin, and Professor Ivan Ivanov for serving on my committee and for all their constructive suggestions.

I am thankful to all my friends and colleagues in the Genomic Signal Processing Lab who made my Ph.D. an unforgettable and pleasant experience.

And finally, my deepest gratitude extends to my mother Atifeh, my father Hamid, and my brother Razi. Without them, I would not have come this far. My appreciation to them for their everlasting love, unwavering dedication, and unconditional support throughout my life is beyond words. I dedicate this thesis to them.

NOMENCLATURE

MOCU	Mean objective cost of uncertainty
IBR	Intrinsically Bayesian robust
MSE	Mean-squared error
MAP	Maximum *a posteriori*
KL	Karhunen-Loève
$\mathfrak{R}(\mathbf{W},i)$	i-th row of matrix $\mathbf{W}$
$\mathfrak{C}(\mathbf{W},j)$	j-th column of matrix $\mathbf{W}$
$\Pr(.)$	Probability operator
$f_X(x)$	Probability density function of X
$f_X(x\|y)$	Conditional density function of X given $Y = y$
$\mathrm{Cov}[\mathbf{X}]$	Covariance matrix of $\mathbf{X}$
$\mathrm{E}_x[g(x)]$	Expectation of $g(x)$ with respect to random variable X
$\mathrm{E}[X\|Y = y]$	Conditional expectation of X given $Y = y$
$\mathrm{E}_{x\|Y=y}[g(x,y)]$	Conditional expectation of $g(x,y)$ relative to X given $Y = y$
$\mathcal{N}(\mu,\Sigma)$	Multivariate Normal distribution, mean μ and covariance Σ
$N(\mathbf{x};\mu,\Sigma)$	Gaussian function, $\frac{1}{\sqrt{(2\pi)^k\|\Sigma\|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu) \right)$
$\mathcal{W}_p(\Sigma,n)$	Wishart distribution, degree of freedom n and scale matrix Σ
$\delta(t)$	Dirac delta function

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Since the earliest days of modern science, it has been recognized that experimental design is critical for the efficient observation of nature. Immanuel Kant, the seminal philosopher of the 18th century, says in the *Critique of Pure Reason*: "It is only when experiment is desired by rational principles that it can have any real utility". Dynamical systems typically involve a large number of variables interacting with each other and therefore often suffer from many uncertain parameters. Moreover, experiments are usually time-consuming and expensive. This brings up the issue of *experimental design* in many branches of science such as signal processing, biological investigations, materials science, etc. The task of experimental design is to evaluate potential experiments to find out which ones are most informative relative to the problem at hand. In other words, experimental design is concerned with maximizing the information content of experiments to be conducted.

1.1 Literature Review on Experimental Design

Experimental design has roots in statistics and machine learning [1, 2, 3]. The first statistical framework for experimental design was proposed by Smith [4]. She was concerned with univariate polynomials and her aim was to find the design which minimizes the maximum variance. In 1943, Wald proposed a sequential hypothesis testing method for linear normal regression based on the D-optimality criterion which seeks to maximize the

determinant of the information matrix [5]. Later Elfving proposed A-optimality criterion aimed at minimizing the average variance of the estimates of the regression parameter [6]. Kiefer utilized the theory of convex optimization for experimental design in the case of linear regression [7]. In 1953, Chernoff proposed locally optimal designs for non-linear models [8]. Experimental design from an information theoretic viewpoint was first proposed by Lindley [9] and further studied in several follow-up works [10, 11, 12, 13, 14]. The basic idea in these methods is to define the information gain $I(\Theta; T_i)$ for experiment T_i, that leads to the estimation of parameter θ_i, as the difference between the entropy before experiment and the conditional entropy after conducting the experiment:

$$I(\Theta; T_i) = H(\Theta) - H(\Theta|T_i)$$

$$= H(\Theta) + \sum_{\theta, \phi} \Pr(\theta, \theta_i = \phi) \log_2 \Pr(\theta|\theta_i = \phi), \qquad (1.1)$$

where $H(\Theta)$ is the model entropy and $\Pr(.)$ is the probability operator. The chosen experiment is the one that maximizes (1.1).

All aforementioned experimental design methods emphasize the application of statistics and focus merely on the model statistical information. To perceive the shortcomings of purely statistical experimental design methods, consider the case that all uncertain parameters are statistically independent and uniformly distributed. In this case, all experiments would possess equal information gain (as defined in (1.1)) and consequently the experimental design does not distinguish between different experiments. From an engineering perspective, the ultimate objective of constructing mathematical models is to design optimal operators. Although statistical information can give us some hints, it does not tell us
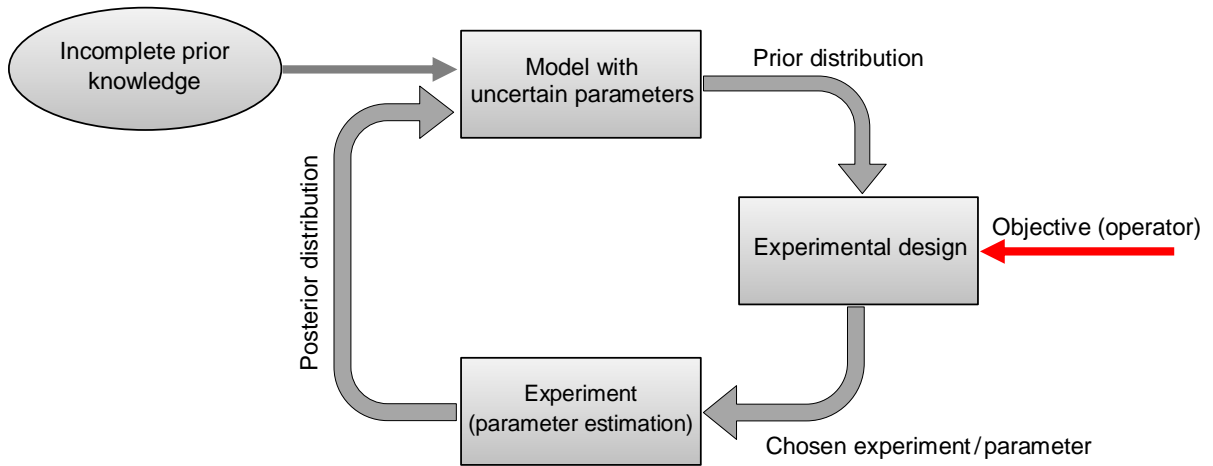
Figure 1.1. The general flowchart of the proposed objective-based experimental design framework.

much about the operator performance in many cases. Therefore, if the ultimate objective is to design operators, the experimental design method aimed at reducing model uncertainty should also consider operator performance.

1.2 Objective-Based Experimental Design

In the absence of model uncertainty, the goal is to design the optimal operator relative to a single perfectly known model. However, when the model knowledge is lacking, the aim becomes to design a robust operator. In this study, we take the viewpoint that experimental design should depend on both the uncertainty and the modeling objective, which is designing operators. Such an approach is called *objective-based* experimental design.

An experimental design scenario begins with a mathematical model possessing unknown parameters. The knowledge regarding unknown parameters is incorporated in the model in terms of a prior distribution. As Figure 1.1 shows, the proposed experimental de-

sign framework uses the prior distribution, which reflects our knowledge about unknown parameters, and takes into account the objective, which is designing robust operators, to find the best experiment that should be conducted first. After conducting the chosen experiment, based on the observed outcome of the experiment, the prior distribution is updated to the posterior distribution. The updated posterior distribution can be used as the new prior distribution and to find the next experiment. We can keep repeating this process.

A crucial step in an objective-based experimental design method, whose aim is to improve operator performance, is the criterion used for robustness definition.

1.3 Robustness Criteria

When it is not realistic to assume that all model parameters are known, it is prudent to design a robust operator by taking into account all possible models consistent with the partial prior knowledge, being called the *uncertainty class*. In signal processing, the design of robust operators goes back to the 1970s, with the goal being to design a linear filter in the presence of an uncertain covariance structure. Qualitatively, an operator is robust if its performance degradation is acceptable relative to all models close to the model for which it has been designed. In fact, robust operator is the best option given the current imperfect state of knowledge regarding the model.

We consider an uncertainty class of models Θ parameterized by the vector of unknown parameters $\theta = \{\theta_1, ..., \theta_k\} \in \Theta$ under the assumption that the true model is given by some specific value of θ. We refer to θ as uncertainty vector. Let ψ denote an operator

such as filter, control, classifier, etc that belongs to a class Ψ of operators such that for each operator $\psi \in \Psi$, if ψ is applied to the model with parameter θ, there is an associated cost $\eta_\theta(\psi)$. The model-specific optimal operator $\psi(\theta)$ for model θ is:

$$\psi(\theta) = \arg\min_{\psi \in \Psi} \eta_\theta(\psi). \tag{1.2}$$

If ψ^Θ denotes the robust operator that has been obtained relative to the uncertainty class Θ, then $\eta_\theta\left(\psi(\Theta)\right) - \eta_\theta\left(\psi(\theta)\right)$ is the increased cost arises from applying the robust intervention ψ^Θ instead of the model-specific optimal operator $\psi(\theta)$ to model θ.

1.3.1 Minimax Approach

Early works on robust operator design were focused on minimax robustness [15, 16, 17, 18, 19], the aim being to find an operator exhibiting the best worst-case performance. Under this approach, the minimax robust operator $\psi^\Theta_{\mathrm{minimax}}$ is the one whose worst performance across the uncertainty class Θ is best among operators in Ψ:

$$\psi^\Theta_{\mathrm{minimax}} = \arg\min_{\psi \in \Psi} \max_{\theta \in \Theta} \eta_\theta(\psi). \tag{1.3}$$

Minimax approach has been used for Wiener filtering [20, 21, 22, 23], Kalman filtering [24, 25], and matched filtering [26, 27] . In [28], it has been shown in the context of game theory, that a minimax robust filter exists if the convexity assumption holds for the uncertainty class . Minimax robustness is a conservative approach and suffers from being overly influenced by models possessing negligible likelihood because it does not take into account the probability mass over the uncertainty class.

1.3.2 Bayesian Approach

The Bayesian framework addresses the problem of outlier models by assuming a prior distribution $f(\theta)$ over the uncertainty class Θ. In this framework, the aim is to find an operator with the best expected performance relative to $f(\theta)$:

$$\psi_{\text{IBR}}^{\Theta} = \arg\min_{\psi \in \Psi} E_\theta \left[\eta_\theta(\psi) \right]. \tag{1.4}$$

Such a robust operator is called an *intrinsically Bayesian robust (IBR)* operator. The IBR operator performs optimally relative to the prior distribution used to represent model with uncertain parameters. It should be recognized that the IBR operator is not guaranteed to perform optimally for each possible model in the uncertainty class. Intuitively, it is expected to perform well over those models, where the prior distribution $f(\theta)$ is concentrated. Designing Bayesian robust operators has been addressed in different engineering applications, including Wiener filtering [29, 30], Kalman filtering [31], classification [32], morphological and binary filtering [33, 34], classification error estimation [35], blind image deconvolution [36], hypothesis testing [37], and control policy design for Markov chains [38].

When it is computationally infeasible to search through the class Ψ to identify an IBR operator, we can confine the search to the set of model-specific optimal interventions $\psi(\theta)$ for models within Θ. We define a *model-constrained Bayesian robust* (MCBR) operator by

$$\psi_{\text{MCBR}}^{\Theta} = \arg\min_{\psi(\phi):\phi \in \Theta} E_\theta \left[\eta_\theta\big(\psi(\phi)\big) \right]. \tag{1.5}$$

6

An MCBR operator is suboptimal relative to an IBR operator, i.e., $\mathrm{E}_{\theta}\left[\eta_{\theta}(\psi_{\mathrm{MCBR}}^{\Theta})\right] \geq \mathrm{E}_{\theta}\left[\eta_{\theta}(\psi_{\mathrm{IBR}}^{\Theta})\right]$. Historically, MCBR filtering goes back to binary filtering [34]. MCBR dynamical intervention for gene regulatory networks was presented in [39].

In this dissertation, we take a Bayesian view to define robustness when dealing with an uncertain dynamical model, be it a gene regulatory network or a canonical expansion.

1.4   Contributions

In this dissertation, we will focus on the problem of experimental design for two major engineering applications: (1) uncertainty reduction in gene regulatory networks models used for designing therapeutic interventions, and (2) uncertainty reduction in canonical expansions used to express random functions in a simpler form.

As remarked earlier, the proposed experimental design framework is objective-based. Therefore, as a first step we need to quantify uncertainty in an objective-based manner. In that regard, we use mean objective cost of uncertainty (MOCU) [40]. MOCU measures the deterioration in the operator performance resulting from the presence of uncertainty. Based on MOCU, we develop an objective-based experimental design method to reduce uncertainty in dynamical models.

In Chapter 2, we provide the proposed experimental design framework based on the concept of MOCU and then utilize it for uncertainty reduction in gene regulatory networks modeled as Boolean networks with perturbation (BNps). In the proposed framework, potential experiments are prioritized based on the MOCU expected to remain after conducting

the experiment. Based on this prioritization, one can select an optimal experiment with the largest potential to reduce the pertinent uncertainty present in the current network model. We demonstrate the effectiveness of the proposed method via extensive simulations based on synthetic and real gene regulatory networks. Work in this chapter is originally from [41].

Chapter 3 addresses computational concerns of the optimal experimental design method for gene regulatory networks. In the process of experimental design, one must find the optimal intervention for every gene regulatory network compatible with the prior knowledge, which can be prohibitively expensive when the size of the network is large. To overcome this difficulty, we propose a computationally efficient experimental design method that incorporates a network reduction scheme by introducing a novel cost function [42]. This cost function takes into account the disruption in the ranking of potential experiments. We then estimate the approximate expected remaining MOCU at a lower computational cost using the reduced networks. Simulation results based on synthetic and real gene regulatory networks show that the proposed approximate method has close performance to that of the optimal method but at lower computational cost. The proposed approximate method also outperforms the random selection policy significantly.

In Chapter 4, we present the general framework for the experimental design in the context of canonical expansions and solve it for two major signal processing problems: optimal linear filtering and signal detection. We note that parameters of the random process appear in the canonical expansion, so that when the expansion is used for operator design,

the uncertainty in these parameters affects the operator objective via the expansion. Hence, optimal experimental design can be approached in terms of the canonical expansion. In particular, we show how experimental design can be used for Wiener filtering.

Having provided a general framework for experimental design in the context of canonical expansions, in Chapter 5, we apply the proposed experimental design framework to Karhunen-Loève (KL) compression to decide which uncertain parameter in the covariance matrix should be determined first to improve the quality of the compressed signal [43]. We find the closed-form solution for the intrinsically Bayesian robust (IBR) KL compression when the covariance matrix is unknown and show that the IBR KL compression can be found in the same form as the ordinary KL compression with the covariance matrix replaced by the expected covariance matrix. We then utilize the expression for IBR KL compression to develop the experimental design framework. Two model assumptions for the covariance matrix are studied: Wishart priors and the blocked covariance model. In particular, we will show how the conditional expectation of each element given the value of another element can be found for the Wishart priors case.

2 OPTIMAL EXPERIMENTAL DESIGN FOR GENE REGULATORY NETWORKS [1]

The main objective in systems biology is to characterize the way genes interact with each other in the context of the mathematical models called gene regulatory networks (GRNs). Gene regulatory network models are increasingly used as a tool to study interactions among genes [44]. Today, of major interest to translational systems biology is to determine beneficial interventions in GRNs for the purpose of identifying potential drug targets. A precondition for using GRNs to design intervention strategies is network identification. Hence, given a model possessing uncertainty, the aim of an experiment is to reduce that uncertainty as it pertains to the intervention objective. Thus, entropy alone is inadequate. One needs a measure that incorporates both the uncertainty and the objective.

From the earliest days of high-throughput gene-expression measurements, the intervention problem has been addressed from two perspectives: (1) dynamical intervention by altering one or more regulatory outputs (expressions) over time [45], and (2) structural intervention via a one-time change of one or more regulatory functions constituting the network [46]. Dynamical intervention interferes with signaling and does not alter network wiring, whereas structural intervention constitutes a one-time alteration of the physical network. Both approaches have mainly developed in the context of probabilistic Boolean

---

networks (PBNs) [47]. Structural intervention, which concerns us here, has been studied from a logical perspective to achieve a desired alteration of the attractor structure of a PBN [48] and in the framework of Markov Chain perturbation theory to derive an altered transition probability matrix that optimally reduces undesirable (pathological) steady-state probability mass [49].

In the vast majority of methods considered for both dynamical and structural intervention, the GRN is assumed to be known, which in the case of Markovian networks means that the transition probability matrix is known. However, given the complex regulatory machinery of the cell and the lack of sufficient data for accurate inference, there is typically significant uncertainty in GRN models. Hence, rather than assume that the model is fully known, it can be beneficial to assume that the true GRN belongs to an *uncertainty class* of networks and the problem is to find a *robust* intervention strategy that is optimal across the uncertainty class. In the case of dynamical intervention in PBNs, robust control policies have been found under two scenarios: (1) no knowledge is assumed concerning the distribution of the networks in the uncertainty class and optimality is defined via a minimax criterion [50]; and (2) there is a prior distribution governing the networks in the uncertainty class and optimality is defined via a Bayesian criterion [39]. Robust design has also been addressed in structural intervention, where one searches for the optimal regulatory function alteration relative to the uncertainty class [40].

It should be recognized that the uncertainty problem is inherent to computational biology owing to the complexity of biological systems and the ubiquity of samples that are

11

small relative to the number of variables. This is why many works analyze gene regulatory networks with uncertainty [51, 52, 53, 54]. In particular, not only is model uncertainty an issue that must be addressed for network intervention (optimal therapy), it is also an issue for biomarker design (optimal diagnosis) and in this context has been treated in the context of uncertainty classes of feature-label distributions [32, 35].

From an experimental perspective, one would like to reduce model uncertainty and thereby improve intervention performance. For smaller uncertainty classes, it is more likely that the performance of a designed robust intervention strategy is close to the performance of the optimal intervention for the actual network. This brings up the issue of experimental design. Experimental design has been utilized in the inference of gene regulatory networks to reduce the entropy of the network model [55, 56, 57, 58]. Here we take the viewpoint that, when designing an intervention strategy we are not so much concerned with reducing model uncertainty from a general perspective, say, entropy; rather, our goal is to reduce uncertainty that will retard the effectiveness of our designed strategy.

Here we present an experimental design method based on the concept of *mean objective cost of uncertainty* (MOCU), introduced in [40]. MOCU is an uncertainty quantification for dynamical models that quantifies the increased cost due to uncertainty, where the cost function depends on ones objective. In the context of controlling GRNs, MOCU measures uncertainty in terms of the differential cost between applying the robust and true-model optimal interventions. According to our proposed method, we conduct experiments to estimate unknown parameters in such a way as to maximize the expected reduction

of MOCU. By computing the expected remaining MOCU after conducting each experiment, we select the experiment that results in the minimum expected remaining MOCU. We desire experiments that estimate these uncertain parameters and would like to know which experiment should be conducted first. To evaluate our proposed experimental design method, we perform simulations on both synthetic and real networks. The simulation results demonstrate the effectiveness of the proposed method.

This chapter is organized as follows. Section 2.1 provides an overview of Boolean networks. Section 2.2 presents our proposed experimental design method based on mean objective cost of uncertainty. In Section 2.3.1, a comprehensive performance analysis of the proposed experimental design method for both synthetic and real networks is given. Finally, we conclude the chapter in Section 2.4.

2.1 Boolean Networks

Boolean networks (BNs) [59] and probabilistic Boolean networks (PBNs) [47] are widely used models for GRNs that have been shown to be effective in capturing these interactions [60, 61, 62, 63, 64, 65, 66]. An n-gene Boolean network (BN) is a pair $(\mathbf{V}, \mathbf{F})$, where $\mathbf{V} = \{X_1, X_2, ..., X_n\}$ is a set of nodes representing the binary expression states of genes and $\mathbf{F} = \{f_1, f_2, ..., f_n\}$ is a set of Boolean functions such that $f_i : \{0,1\}^{k_i} \rightarrow \{0,1\}$ is the Boolean function that determines the expression state of X_i. It is commonplace to refer to gene i as X_i. The binary values $X_i = 0$ and $X_i = 1$ correspond to the gene being turned "off" or "on", respectively. The vector $X(t) = (X_1(t), ..., X_n(t))$ of gene values at time t is

called the *gene activity profile* (GAP). It reflects the "state" of the network at time t. The value of gene i at the next time point, $X_i(t+1) = f_i\big(X_{i_1}(t), X_{i_2}(t), ..., X_{i_{k_i}}(t)\big)$, is determined by the values of k_i predictor genes at time t. In a Boolean network with perturbation (BNp), each gene may randomly flip its value at a given time with a perturbation probability p, independently from other genes. Hence, for a BNp, $X(t+1) = \mathbf{F}(X(t))$ with probability $(1-p)^n$ when there is no perturbation, but $X(t+1)$ may take a different value with probability $1 - (1-p)^n$, when there exists one or more random perturbations. In a BNp, the sequence of states over time can be regarded as a Markov chain with transition probability matrix (TPM) $\mathbf{P} = \big[p_{ij}\big]_{i,j=1}^{2^n}$ where p_{ij} is the probability that state i transitions into state j. Therefore, classical Markov chain theory can be applied for analyzing network dynamics [67, 68]. The general formula of a TPM using Boolean functions and perturbation probability has been derived in [50]. When $p > 0$, the resulting Markov chain is ergodic, irreducible, and possesses a steady-state distribution (SSD) $\pi^T = \pi^T \mathbf{P}$, where the k-th element, π_k, of the column vector π corresponds to the steady-state probability of state k and T denotes the transpose operator. A probabilistic Boolean network (PBN) contains a set of m constituent BNps, called contexts. The choice of which context is chosen at each time step is governed by network selection probabilities $c_1, c_2, ..., c_m$. A PBN switches its context for the next transition according to a network switching probability q. If $q = 1$, then PBN changes its context at every time point and is called an instantaneously random PBN. When $q < 1$, the network is a context-sensitive PBN.

A key objective in modeling gene regulatory networks is to design therapeutic net-

work interventions based on long-run network behavior to avoid abnormal phenotypes such as cancer. The long-run behavior of a GRN is characterized by its SSD and hence the goal of beneficial interventions is to change the SSD of network [69]. In the context of translational genomics, the state space of a network can typically be partitioned into undesirable states (U), corresponding to abnormal (disease) phenotypes, and desirable states (D), corresponding to normal (healthy) phenotypes. The goal in controlling GRNs via interventions is to beneficially alter the dynamics of network to decrease the probability that the network will enter the set U of undesirable states. In other words, intervention aims at minimizing the overall steady-state probability mass $\pi_U = \sum_{i \in U} \pi_i$ in undesirable states. There are two basic categories of intervention approaches: structural interventions and dynamical interventions. Structural interventions [40, 46, 48, 49, 70] alter the long-run behavior of a network via a one-time change of the underlying network structure (wiring). After applying this type of interventions to the network, the regulatory functions of the network and consequently the state transitions of the underlying Markov chain are altered. Dynamical interventions [39, 45, 50, 71, 72, 73] utilize Markov decision theory to flip (or not flip) the value of a certain gene called control gene at each time instant. These interventions typically involve stationary control policies to make a decision at each time.

We restrict ourselves to Boolean networks in this study; however, it should be recognized that the proposed experimental design method is fairly general and can be applied to different models and applications in a straightforward manner.

2.2 MOCU-Based Experimental Design

Let $\theta = (\theta_1, \theta_2, ..., \theta_k)$ be a vector of parameters that characterizes the gene regulatory network. We assume that θ is uncertain and belongs to an uncertainty class Θ of possible networks. For any $\theta \in \Theta$, let $\eta_\theta(\psi)$ be the cost of applying the intervention $\psi \in \Psi$, a class of potential interventions, to the network defined by the uncertainty vector θ. For instance, $\eta_\theta(\psi)$ might be the steady-state probability mass in undesirable states after applying the intervention. Let $\psi(\theta) \in \Psi$ denote an optimal intervention relative to η_θ, meaning that $\eta_\theta(\psi(\theta)) \leq \eta_\theta(\psi)$ for any $\psi \in \Psi$. $\psi(\theta)$ is an optimal intervention for the network with uncertainty vector θ.

An *intrinsically Bayesian robust* (IBR) intervention is defined as

$$\psi_{\text{IBR}}^\Theta = \underset{\psi \in \Psi}{\arg\min}\ \mathrm{E}_\theta \left[\eta_\theta(\psi) \right] \tag{2.1}$$

[40]. The expectation $\mathrm{E}_\theta [\bullet]$ is taken over the probability distribution $f(\theta)$ of θ.

The *mean objective cost of uncertainty* (MOCU) relative to an uncertainty class Θ of networks and a class Ψ of interventions is defined as

$$\mathrm{M}_\Psi(\Theta) = \mathrm{E}_\theta \left[\eta_\theta \left(\psi_{\text{IBR}}^\Theta \right) - \eta_\theta \left(\psi(\theta) \right) \right] \tag{2.2}$$

[40]. MOCU is the expected cost increase that results from applying a robust intervention over all networks in Θ instead of the optimal intervention for the true network, which is unknown.

We also define a *model-constrained Bayesian robust* (MCBR) intervention by

$$\psi_{\text{MCBR}}^\Theta = \underset{\psi(\phi):\phi \in \Theta}{\arg\min}\ \mathrm{E}_\theta \left[\eta_\theta \left(\psi(\phi) \right) \right]. \tag{2.3}$$

16

Empirical results in [40] indicate that, at least for binary PBNs with up to ten genes, the MCBR structural intervention provides an extremely accurate approximation of the IBR structural intervention. Since the large number of MOCU computations required for the simulations performed in the current study would be computationally prohibitive using IBR intervention, we employ MCBR intervention. Using the MCBR intervention, rather than the IBR intervention, we can obtain an approximation of the true MOCU in (2.2) by replacing the optimal IBR operator, $\psi^{\Theta}_{\text{IBR}}$, by the optimal MCBR operator, $\psi^{\Theta}_{\text{MCBR}}$. In what follows, we will refer to the approximate MOCU computed based on MCBR intervention as MOCU. We will denote an MCBR intervention $\psi^{\Theta}_{\text{MCBR}}$.

Consider a GRN possessing k uncertain parameters $\theta_1, \theta_2, ..., \theta_k$. Suppose there exists a corresponding set of k experiments $T_1, T_2, ..., T_k$, where performing experiment T_i would completely determine θ_i such that we would be sufficiently confident about the value of θ_i that we would no longer consider it to be uncertain. In practice, more than one actual experiment might be needed to be conducted for the true estimation of an uncertain parameter but we can consider these experiments collectively as one experiment for our analysis. For simplicity, let us assume that θ_i is a binary variable and that experiment T_i can determine whether $\theta_i = 0$ or $\theta_i = 1$. Our aim is to decide which experiment T_i among the k potential experiments should be conducted first in order to optimally reduce the uncertainty based on a single experiment. Let $\theta | \theta_i = \bar{\theta}_i$ be the *conditional uncertainty vector* composed of all uncertain parameters other than θ_i, with $\theta_i = \bar{\theta}_i$, and let $\Theta | (\theta_i = \bar{\theta}_i) = \left\{ \theta \big| \theta \in \Theta, \theta_i = \bar{\theta}_i \right\}$ be the reduced uncertainty class of networks obtained

by assuming that $\theta_i = \bar{\theta}_i$. Let $M_\Psi(\Theta|\theta_i = \bar{\theta}_i)$ be the remaining MOCU given $\theta_i = \bar{\theta}_i$:

$$M_\Psi(\Theta|\theta_i = \bar{\theta}_i) = E_{\theta|\theta_i=\bar{\theta}_i}\left[\eta_{\theta|\theta_i=\bar{\theta}_i}\left(\psi_{\mathrm{MCBR}}^{\Theta|\theta_i=\bar{\theta}_i}\right) - \eta_{\theta|\theta_i=\bar{\theta}_i}\left(\psi(\theta|\theta_i = \bar{\theta}_i)\right)\right], \qquad (2.4)$$

where the expectation is taken over the conditional probability distribution $f(\theta|\theta_i = \bar{\theta}_i)$ of the remaining uncertain parameters given $\theta_i = \bar{\theta}_i$ and $\psi_{\mathrm{MCBR}}^{\Theta|\theta_i=\bar{\theta}_i}$ is the MCBR intervention for the reduced uncertainty class $\Theta|(\theta_i = \bar{\theta}_i)$:

$$\psi_{\mathrm{MCBR}}^{\Theta|\theta_i=\bar{\theta}_i} = \underset{\psi(\tau):\tau\in\Theta|(\theta_i=\bar{\theta}_i)}{\arg\min}\ E_{\theta|\theta_i=\bar{\theta}_i}\left[\eta_{\theta|\theta_i=\bar{\theta}_i}\left(\psi(\tau)\right)\right]. \qquad (2.5)$$

We define the cost function by

$$\eta_{\theta|\theta_i=\bar{\theta}_i}\left(\psi(\tau)\right) = \tilde{\pi}_{U,\theta|\theta_i=\bar{\theta}_i}\left(\psi(\tau)\right), \qquad (2.6)$$

where $\tilde{\pi}_{U,\theta|\theta_i=\bar{\theta}_i}\left(\psi(\tau)\right)$ is the steady-state probability mass in undesirable states after applying intervention $\psi(\tau)$ to the network defined by the uncertainty vector $\theta|\theta_i = \bar{\theta}_i$ in the reduced uncertainty class $\Theta|\theta_i = \bar{\theta}_i$. We define the expected remaining MOCU after determining the value of θ_i via experiment T_i by

$$M_\Psi(\Theta;\theta_i) = E_{\bar{\theta}_i}\left[M_\Psi(\Theta|\theta_i = \bar{\theta}_i)\right], \qquad (2.7)$$

where the expectation is taken over the marginal probability density function, $f(\bar{\theta}_i)$, for the uncertain parameter θ_i. In order to optimally reduce the uncertainty in the current uncertainty class Θ, we should select the experiment T_{i^*} such that

$$i^* = \underset{i\in 1,2,...,k}{\arg\min}\ M_\Psi(\Theta;\theta_i), \qquad (2.8)$$

since T_{i^*} is expected to minimize the remaining MOCU by determining the value of the parameter θ_{i^*}.

To calculate $M_\Psi(\Theta;\theta_i)$, we need to define the class of interventions. We focus on

18

the structural intervention method proposed in [49]. In [49], intervention is performed via a rank-1 function perturbation such that the relation between the transition probability matrices of the original and perturbed networks is $\tilde{\mathbf{P}} = \mathbf{P} + \mathbf{a}\mathbf{b}^T$, where $\tilde{\mathbf{P}}$ is the transition probability matrix after perturbation and $\mathbf{a}\mathbf{b}^T$ is the rank-1 perturbation matrix, $\mathbf{a}$ and $\mathbf{b}$ being two arbitrary vectors, and $\mathbf{b}^T\mathbf{e} = 0$ for $\mathbf{e}$ (all unity column vector). Single-gene perturbation is a special case of a rank-1 function perturbation in which the output state for only one input state changes and the output states of other states remain unchanged. We consider single-gene perturbations for the class Ψ of structural interventions. Let $\tilde{\mathbf{F}} = \{\tilde{f}_1, \tilde{f}_2, ..., \tilde{f}_n\}$ be the list of Boolean functions for the perturbed BNp. The structural intervention for input state j solely changes the output state for input state j and leaves the rest unaltered: $v = \tilde{\mathbf{F}}(u) \neq \mathbf{F}(u) = w$ and $\tilde{\mathbf{F}}(i) = \mathbf{F}(i)$ for $i \neq u$. The transition probability matrix $\tilde{\mathbf{P}}$ of the perturbed network will be identical to the transitional probability matrix $\mathbf{P}$ of the original network, except for $\tilde{p}_{uw} = p_{uw} - (1-p)^n$ and $\tilde{p}_{uv} = p_{uv} + (1-p)^n$. The SSD of the perturbed BNp can be obtained by

$$\tilde{\pi}_i(u,v) = \pi_i + \frac{(1-p)^n \pi_u (z_{vi} - z_{wi})}{1 - (1-p)^n (z_{vu} - z_{wu})}, \tag{2.9}$$

where π_i is the steady-state probability for the state i, $z_{vi}, z_{wi}, z_{vu}, z_{wu}$ are elements of the fundamental matrix of the BNp, and $\tilde{\pi}_i(u,v)$ is the perturbed steady-state probability for state i after applying the aforementioned intervention [49]. The fundamental matrix of a BNp can be computed as $\mathbf{Z} = [\mathbf{I} - \mathbf{P} + \mathbf{e}\pi^T]^{-1}$, where $\mathbf{I}$ is the $n \times n$ identity matrix and $\mathbf{e}$ is the all unity column vector. Let $\tilde{\pi}_{i,\theta}(u,v)$ be the steady-state probability of state i in the network with uncertainty vector θ after intervention (u,v). Then $\tilde{\pi}_{U,\theta}(u,v) = \sum_{i \in U} \tilde{\pi}_{i,\theta}(u,v)$

is the steady-state probability mass in undesirable states after applying the single-gene per-turbation structural intervention. For a BNp defined by a given uncertainty vector θ, the optimal single-gene perturbation structural intervention $\big(u(\theta), v(\theta)\big)$ is the one that mini-mizes $\tilde{\pi}_{U,\theta}(u,v)$:

$$\big(u(\theta), v(\theta)\big) = \underset{u,v \in \{1,2,3,\ldots,2^n\}}{\arg\min} \tilde{\pi}_{U,\theta}(u,v) \tag{2.10}$$

For each network $\theta \in \Theta$, we find the optimal intervention $\psi(\theta) = \big(u(\theta), v(\theta)\big)$. The MCBR intervention $\psi_{\text{MCBR}}^{\Theta} = \big(u_{\text{MCBR}}^{\Theta}, v_{\text{MCBR}}^{\Theta}\big)$ is chosen from the set $\{\psi(\theta), \theta \in \Theta\}$ such that it can minimize the expected error over the uncertainty class as shown in (2.3).

2.3 Performance Assessment

2.3.1 Simulation Setup

The simulations involve GRNs with genes regulated according to the commonly used majority vote rule [74]. Regulations in the network are governed by a regulatory matrix $\mathbf{R}$, where R_{ij} represents the regulatory relation from gene j to gene i as follows:

$$R_{ij} = \begin{cases} 1 & \text{the relation from } j \text{ to } i \text{ is activating} \\ -1 & \text{the relation from } j \text{ to } i \text{ is suppressive} \\ 0 & \text{there is no relation from } j \text{ to } i \end{cases} \tag{2.11}$$

A given gene takes the value 1 if the majority of its regulator genes up-regulate it and the value 0 if the majority of the predictor genes down-regulate it; otherwise, it remains

20

unchanged. Under this rule,

$$
X_i(t+1) = f_i\big(X(t)\big) = \begin{cases} 1 & \text{if } \sum_j R_{ij}X_j(t) > 0 \\[2mm] 0 & \text{if } \sum_j R_{ij}X_j(t) < 0 \\[2mm] X_i(t) & \text{if } \sum_j R_{ij}X_j(t) = 0 \end{cases} \tag{2.12}
$$

We assume that for certain gene pairs, we are aware of the existence of regulatory relations based on prior biological knowledge; however, the precise type of regulation (i.e., activating or suppressive) may not be known. Therefore, the uncertain parameters in our simulations would be these regulatory relations. Each uncertain parameter θ_i, corresponding to an uncertain regulatory relation of an unknown type, can take on two different values: 1 for activating regulation and -1 for suppressive regulation. For a network with k uncertain regulations, the uncertainty class Θ contains 2^k potential networks that differ in one or more of these uncertain regulations. The proposed experimental design method is used to decide which uncertain parameter would be better to determine first, or equivalently, which experiment should be conducted first, in order to maximally reduce the uncertainty in the current network model and thereby optimally improve the performance of structural intervention.

After performing the optimal experiment, we are left with a smaller number of uncertain parameters that lead to a reduced uncertainty class of networks. Suppose we have performed an experiment to estimate the parameter θ_i and that the experiment has identified the true value to be $\theta_i = \mu_i$. We denote the reduced uncertainty class as $\Theta|\theta_i = \mu_i$ and the robust intervention for this reduced uncertainty class as $\psi_{\text{MCBR}}^{\Theta|\theta_i=\mu_i}$. An effective experiment selection strategy should allow us to find out the best parameter θ_{i*} to be determined

first, such that on average the optimal robust intervention $\psi_{\text{MCBR}}^{\Theta|\theta_{i^*}=\mu_{i^*}}$ for the reduced uncertainty class $\Theta|\theta_{i^*} = \mu_{i^*}$ would outperform other robust interventions on the true (unknown) network after identifying θ_j ($j \neq i^*$).

To illustrate the proposed experimental design strategy, consider $k = 5$ uncertain parameters in the GRN. Suppose the five potential experiments, each identifying one of the five parameters, $\theta_1, \theta_2, \cdots, \theta_5$, have been ranked to obtain an ordered $\theta_{1'}, \theta_{2'}, \cdots, \theta_{5'}$. Performing the experiment $T_{i'}$ leads to the identification of the unknown parameter $\theta_{i'}$ and results in the expected remaining MOCU $M_\Psi(\Theta; \theta_{i'})$, such that

$$M_\Psi(\Theta; \theta_{1'}) < M_\Psi(\Theta; \theta_{2'}) < \cdots < M_\Psi(\Theta; \theta_{5'}). \tag{2.13}$$

To measure the overall gain for performing the optimal experiment $T_{1'}$ relative to other suboptimal experiments, we define

$$\xi_i = \eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{(i+1)'}=\mu_{(i+1)'}} \right) - \eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{1'}=\mu_{1'}} \right), \tag{2.14}$$

where μ is the vector of true parameter values corresponding to θ. For example, ξ_1 denotes the difference between the cost $\eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{2'}=\mu_{2'}} \right)$ of applying the robust intervention, derived for the reduced uncertainty class that results from conducting the second best experiment $T_{2'}$, to the true network and the cost $\eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{1'}=\mu_{1'}} \right)$ of applying the robust intervention obtained from conducting the optimal experiment $T_{1'}$. ξ_i ($i = 1, 2, ..., k-1$) quantifies the expected benefit of performing the best experiment predicted by the proposed strategy compared to other experiments, in terms of the operational cost that could be further reduced by performing the selected experiment.

2.3.2 Simulations on Synthetic Networks

To evaluate the performance of the proposed experimental design strategy, we have performed simulations based on synthetic BNps. In our simulations, $k = 2, 3, 4, 5$ uncertain parameters are considered, assuming a uniform distribution $f(\theta)$ for all potential networks $\theta \in \Theta$. The analysis can be easily extended to other distributions. To make the simulations computationally tractable, we consider networks with six genes, $X_1, ..., X_6$, where each gene has three predictor genes. To generate a random BNp, we randomly select three predictor genes for each gene with uniform probability and randomly assign 1 (up-regulation) or -1 (down-regulation) to the corresponding entries in the regulatory matrix $\mathbf{R}$. The perturbation probability is set to $p = 0.001$. States for which $X_1 = 1$ are assumed to be undesirable, so that the set of undesirable states is $U = \{32, ..., 63\}$. For a given k, we generate 1,000 synthetic BNps and randomly select 50 different sets of k edges (i.e., regulations) for each network. In each case, the regulatory information of other edges is retained while that of the k selected edges is assumed to be unknown.

From a translational perspective, the salient issue in evaluating an experimental design scheme using synthetic networks is controllability. Unlike real biological networks, which are controllable to a certain extent, many randomly generated networks may not be controllable. In other words, regardless of the intervention applied to the network, the SSD shift that results from the intervention may be negligible. For such networks, the difference between optimal and suboptimal experiments may be insignificant. For this reason, to examine the practical impact of experimental design, we must take controllability into

Table 2.1. The average gain of conducting the optimal experiment predicted by the proposed experimental design strategy in comparison to other suboptimal experiments.

	Average ξ_1	Average ξ_2	Average ξ_3	Average ξ_4
$k = 2$	0.0584	N/A	N/A	N/A
$k = 3$	0.0544	0.0718	N/A	N/A
$k = 4$	0.0545	0.0750	0.0855	N/A
$k = 5$	0.0474	0.0696	0.0803	0.0863

account. In this work, the percentage decrease of total steady-state mass in undesirable states after intervention is used as a measure of controllability:

$$\Delta = \frac{\pi_U - \tilde{\pi}_U(j^*, s^*)}{\pi_U} \times 100\%,$$

where controllable networks have a larger Δ.

Table 2.1 summarizes the average gain of performing the optimal experiment predicted by the proposed strategy over other suboptimal experiments. The average is taken over different sets of uncertain regulations and different networks with $\Delta \geqslant 40\%$. For $k = 2$, we calculate ξ_1; for $k = 3$, we calculate ξ_1 and ξ_2; and so on. As we can see in Table 2.1, the average gain is always positive. The results in Table 2.1 clearly show that the robust intervention derived from the uncertainty class reduced by conducting the optimal experiment outperforms the robust intervention that results from any other suboptimal experiment on average. We can also see that the average ξ_i gets larger for a larger i. For example, for $k = 4$, average $\xi_1 = 0.0545 <$ average $\xi_3 = 0.0855$, which shows that, on average, the gain of determining $\theta_{1'}$ over $\theta_{4'}$ is larger than that of determining $\theta_{1'}$ over $\theta_{2'}$. This demonstrates that $M_\Psi(\Theta, i)$ can serve as an effective measure for prioritizing poten-

Table 2.2. The average gain of conducting the optimal experiment predicted by the proposed experimental design strategy in comparison to a randomly selected experiment.

	Average Gain
$k = 2$	0.0291
$k = 3$	0.0430
$k = 4$	0.0533
$k = 5$	0.0571

tial experiments. Furthermore, this suggests that we could expect larger gains when we compare the optimal experiment with an experiment that has a larger $M_\Psi(\Theta; \theta_i)$.

A salient question is how much we can gain by conducting an optimal experiment predicted by the proposed method over a randomly selected experiment. Since we would normally have to randomly pick an experiment unless there are reasons to prefer a specific experiment over the rest, such comparison would be useful in demonstrating the efficacy of the proposed method in a practical setting. We calculate the average gain of applying the robust intervention derived from the reduced uncertainty class obtained by conducting the optimal experiment instead of the intervention that results from a randomly chosen experiment, for all networks with $\Delta \geqslant 40\%$. The simulation results are shown in Table 2.2. It should be noted that the randomly chosen experiment may be identical to the optimal experiment (in fact, they are identical with probability $1/k$), which is the main reason that the performance gain shown Table 2.2 is typically smaller than the gain shown in Table 2.1. For example, the average ξ_1 in Table 2.1 for $k = 2$ is almost two times the average gain for $k = 2$ in Table 2.2, which is due to the fact that the randomly picked experiment will be identical to the optimal experiment predicted by our method about 50% of the time. We can

also see in Table 2.2 that the average gain increases for a larger k. For example, while the average gain for $k = 2$ is 0.0291, it is 0.0571 when $k = 5$. This implies that the performance gap between optimal and random selection is expected to increase as the uncertainty of the network increases.

As mentioned earlier, previous works for experimental design in gene regulatory networks are based on entropy reduction of the model. In [55], the information gain for each experiment is defined as the difference between the model entropy before experiment and the conditional entropy of conducting the experiment:

$$I(\Theta; T_i) = H(\Theta) - H(\Theta|T_i)$$

$$= H(\Theta) + \sum_{\theta, \phi} \Pr(\theta, \theta_i = \phi) \log_2 \Pr(\theta|\theta_i = \phi), \qquad i = 1, 2, ..., k, \qquad (2.15)$$

where $H(\Theta)$ is the model entropy and $\Pr(.)$ is the probability operator. The chosen experiment according to [55] is the one that maximizes (2.15). In our setting (uniform distribution and independent uncertain parameters), with k uncertain parameters, $H(\Theta)$ would be k and $I(\Theta; T_i)$ would be $k - 1$ for each potential experiment. Therefore, this experimental design scheme does not discriminate between potential experiments and as a result it would perform like a random selection approach. This makes sense because (2.15) only takes into account the stochastic properties of the model without considering the objective. Throughout this chapter, whenever we compare our method with the random experiment strategy, in fact, we are also comparing our method with experimental design methods based on entropy, such as [55].

We have compared $\eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{1'}=\mu_{1'}} \right)$ and $\eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{i'}=\mu_{i'}} \right)$ $(i' \neq 1')$ and measured

the proportion of "success" (predicted optimal experiment $T_{1'}$ outperforms the suboptimal experiment $T_{i'}$), "failure" ($T_{i'}$ outperforms $T_{1'}$), and "tie" ($T_{1'}$ and $T_{i'}$ provide identical intervention performance). These results are summarized in Table 2.3. In this table, $\theta_{1'} \sim \theta_{i'}$ denotes the comparison between $\eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{1'}=\mu_{1'}} \right)$ and $\eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{i'}=\mu_{i'}} \right)$. When comparing $\theta_{1'} \sim \theta_{i'}$, a "tie" means that conducting either of the two experiments results in the same intervention performance after the uncertainty reduction, a "success" means that $\eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{1'}=\mu_{1'}} \right) < \eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{i'}=\mu_{i'}} \right)$, and a "failure" means that $\eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{1'}=\mu_{1'}} \right) > \eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{i'}=\mu_{i'}} \right)$. We can see that the "success" proportion is consistently larger than the "failure" proportion, which explains why the gain in Table 2.1 is always positive. For $k > 2$, the proportion of "failure" decreases and the proportion of "success" increases as we compare $\psi_{\text{MCBR}}^{\Theta|\theta_{1'}=\mu_{1'}}$ with $\psi_{\text{MCBR}}^{\Theta|\theta_{i'}=\mu_{i'}}$, $i' \neq 1'$, for a larger i. Moreover, for $k = 2$, the proportion of "tie" is larger than that for $k > 2$. This is because the size of the uncertainty class of networks is small for $k = 2$ and therefore it is more likely that conducting either experiment yields the same robust intervention.

Table 2.3. The proportion of success, failure, and tie of the optimal experiment predicted by the proposed strategy in comparison to other suboptimal experiments.

	$\theta_{1'} \sim \theta_{2'}$			$\theta_{1'} \sim \theta_{3'}$			$\theta_{1'} \sim \theta_{4'}$			$\theta_{1'} \sim \theta_{5'}$		
	Success	Failure	Tie	Success	Failure	Tie	Success	Failure	Tie	Success	Failure	Tie
$k=2$	38.07%	15.29%	46.64%	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
$k=3$	40.76%	22.32%	36.92%	42.84%	15.30%	41.86%	N/A	N/A	N/A	N/A	N/A	N/A
$k=4$	40.97%	25.82%	33.21%	42.98%	19.21%	37.82%	43.75%	15.95%	40.30%	N/A	N/A	N/A
$k=5$	43.00%	28.76%	28.24%	45.02%	22.62%	32.36%	45.63%	18.32%	36.05%	46.17%	15.96%	37.87%

Table 2.4. The proportion of overall success, overall failure, and overall tie of the optimal experiment predicted by the proposed strategy in comparison to all other suboptimal experiments.

	$\theta_{1'} \sim \theta_{i'}(i \neq 1)$		
	Overall Success	Overall Failure	Overall Tie
$k = 2$	38.07%	15.29%	46.64%
$k = 3$	44.86%	28.33%	26.81%
$k = 4$	44.90%	37.10%	18.00%
$k = 5$	44.71%	43.13%	12.16%

Table 2.4 shows the proportions of "overall success", "overall failure", and "overall tie" for the proposed experimental design strategy. Here, an "overall success" means that $\eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{1'}=\mu_{1'}} \right) \leqslant \eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{i'}=\mu_{i'}} \right)$ for all $i' \neq 1'$ (except in the case that $\eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{1'}=\mu_{1'}} \right) = \eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{i'}=\mu_{i'}} \right)$ for all $i' \neq 1'$). An "overall tie" means that $\eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{1'}=\mu_{1'}} \right) = \eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{i'}=\mu_{i'}} \right)$ for all $i' \neq 1'$. Finally, an "overall failure" means that $\eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{1'}=\mu_{1'}} \right) > \eta_\mu \left(\psi_{\text{MCBR}}^{\Theta|\theta_{i'}=\mu_{i'}} \right)$ for at least one $i' \neq 1'$. As this table shows, the proportion of "overall success" is larger than that of "overall failure" for all k. The proportion of "tie" decreases with increasing k, as the size of the uncertainty class of networks increases. While the proportion of "overall tie" decreases with increasing k, the proportion of "overall failure" increases. This is intuitive, since by increasing the number of uncertain regulations k, it becomes more difficult to have an "overall success", since $\psi_{\text{MCBR}}^{\Theta|\theta_{1'}=\mu_{1'}}$ has to outperform all other robust interventions, whose number increases with k.

Now, let us consider the difference between the expected remaining MOCU of the optimal experiment and that of a suboptimal experiment:

$$\Delta \text{MOCU} = M_\Psi(\Theta; \theta_{i'}) - M_\Psi(\Theta; \theta_{1'})$$

29
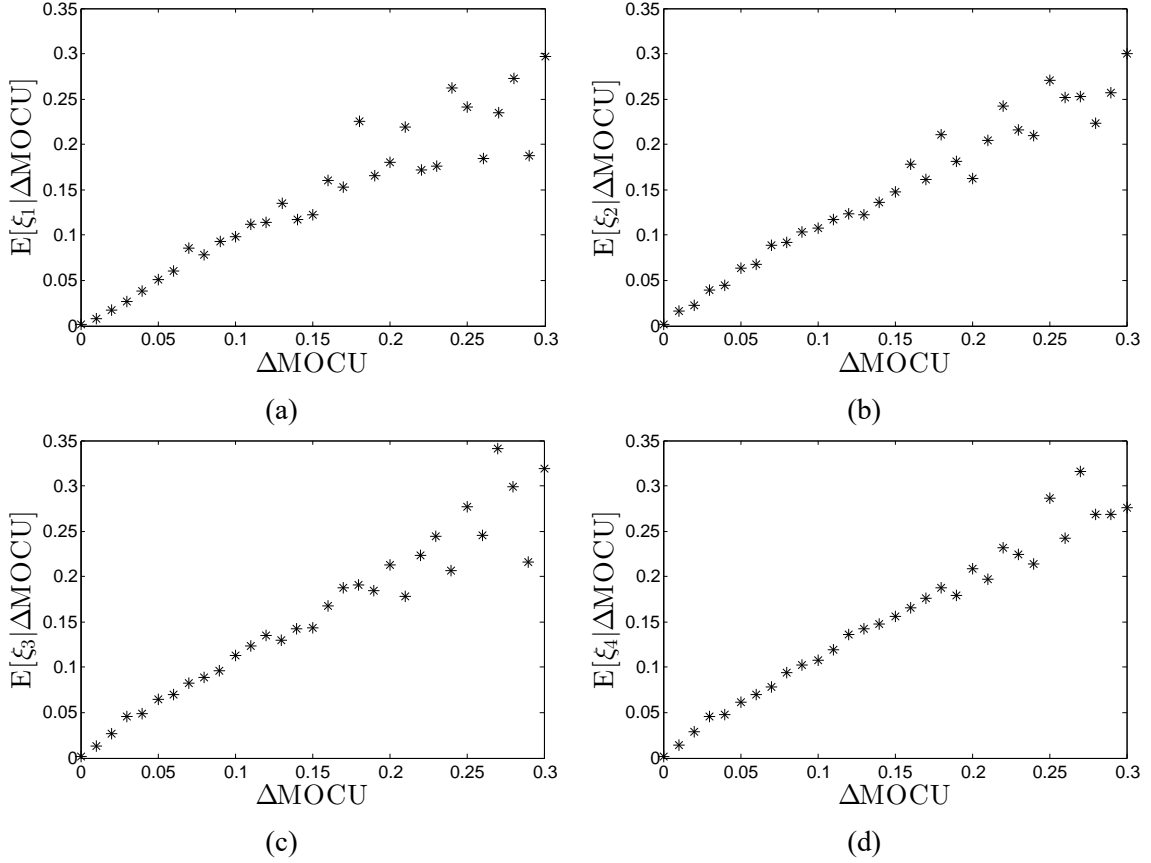
Figure 2.1. The empirical conditional expectation of the gain $\mathrm{E}\big[\xi_i|\Delta\mathrm{MOCU}\big]$ given the difference in MOCU between the optimal and suboptimal experiments. Synthetic BNps with five uncertain regulations are considered. (a) $\theta_{1'} \sim \theta_{2'}$. (b) $\theta_{1'} \sim \theta_{3'}$. (c) $\theta_{1'} \sim \theta_{4'}$. (d) $\theta_{1'} \sim \theta_{5'}$.

for $i' \neq 1'$. Figure 2.1 shows the empirical conditional expectation, $\mathrm{E}\big[\xi_i|\Delta\mathrm{MOCU}\big]$, of ξ_i ($i = 1, 2, 3, 4$) given $\Delta\mathrm{MOCU}$ estimated based on all random networks with $\Delta \geqslant 40\%$. The average gain is positive for all $\Delta\mathrm{MOCU}$. This shows that, on average, the robust interventions obtained by conducting the optimal experiments predicted by our proposed method outperform the robust interventions obtained from other suboptimal experiments when applied to the true network. Moreover, as $\Delta\mathrm{MOCU}$ increases, the average gain increases in a more or less linearly proportional manner. Another interesting observation is that
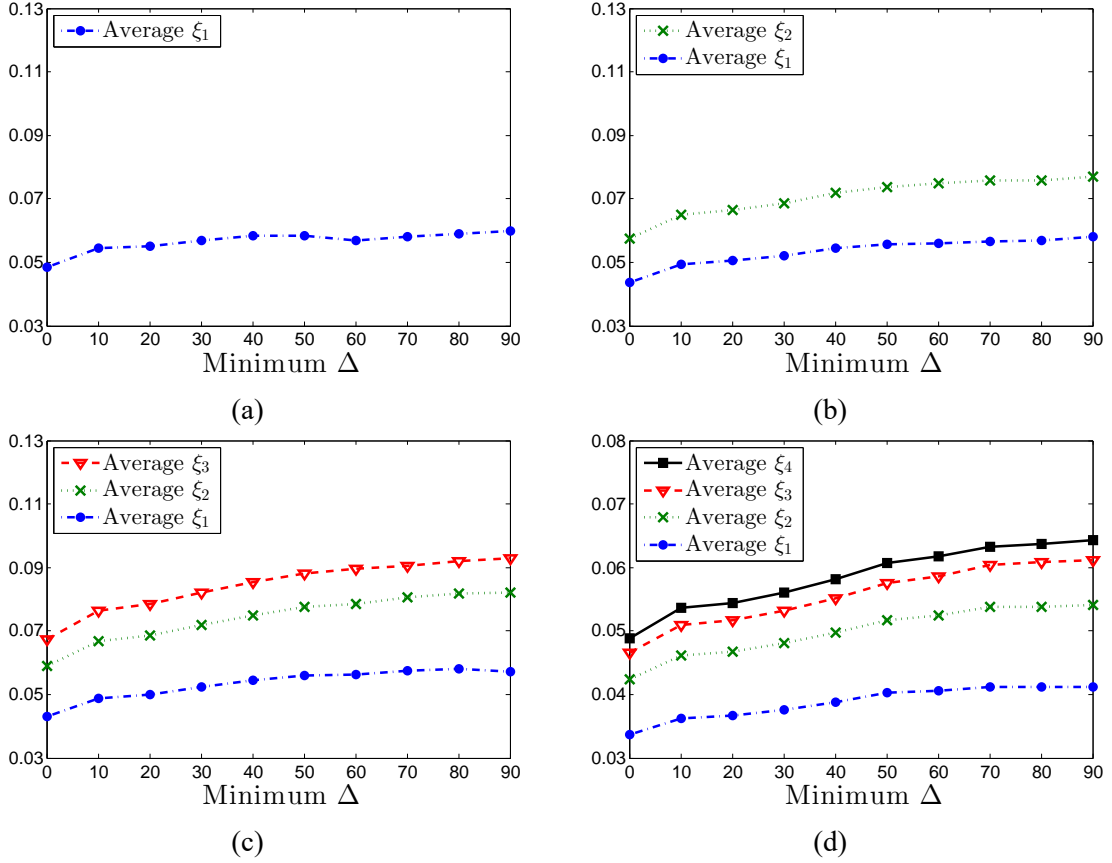
Figure 2.2. Effect of the controllability of the synthetic BNp on the average performance gain of the proposed experimental design method. (a) Networks with 2 uncertain regulations. (b) Networks with 3 uncertain regulations. (c) Networks with 4 uncertain regulations. (d) Networks with 5 uncertain regulations.

$\mathrm{E}\big[\xi_i|\Delta\mathrm{MOCU}\big]$ does not significantly differ for different i. This result is intuitive, since we expect the gain to depend on the estimated $\Delta\mathrm{MOCU}$, and not the predicted rank of the suboptimal experiment.

To see how the controllability Δ measured in terms of the SSD shift that can be achieved by optimal intervention, affects the average gain of the proposed experimental design strategy, we compute the average ξ_i ($i = 1, 2, 3, 4$) for random networks whose controllability (i.e., Δ) exceeds a certain minimum value, where we consider minimum Δ
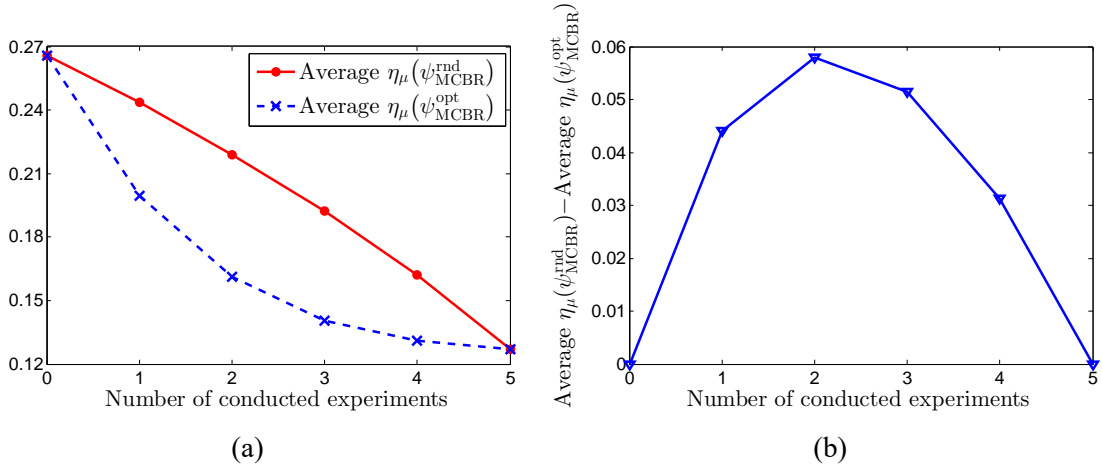
Figure 2.3. Performance comparison based on a sequence of experiments. (a) The average cost of robust intervention after performing the sequence of experiments predicted by the proposed strategy and the average cost after performing randomly selected experiments. (b) The performance difference between the proposed approach and the random selection approach.

ranging between 0% and 90%. According to Figure 2.2, the average gain of ξ_i increases as the minimum Δ increases, regardless of i and the number of uncertain regulations k. For example, for $k = 5$ uncertain regulations in the network, the average gain based on comparing $\theta_{1'}$ to $\theta_{5'}$ is slightly below 0.07 for all networks, but it increases to almost 0.1 when we consider only highly controllable networks with $\Delta \geqslant 90\%$.

Figure 2.3 compares the performance of the proposed experimental design method and that of the random selection approach based on a sequence of experiments. Assuming $k = 5$ uncertain regulations in each network, we perform 5 consecutive experiments until the network does not contain any uncertainty. First, we consider adopting the proposed experimental design strategy, where at each step, we select the optimal experiment predicted by our method, conduct the experiment to reduce the uncertainty class, and repeat this process until the network is fully identified. For comparison, we perform similar simulations

by conducting a randomly selected experiment at each step until there is no uncertainty about the network. In both cases, the network will be fully identified after conducting 5 experiments. To compare the performance of the two approaches, after conducting each experiment, we derive the robust intervention based on the reduced network class, apply it to the true (unknown) network, and measure the cost of intervention (i.e., total steady-state mass in undesirable states). The average performance is estimated based on 1,000 synthetic BNps and 50 different sets of uncertain regulations for each of these networks. Let $\psi_{\text{MCBR}}^{\text{opt}}$ denote the robust intervention obtained by taking the proposed strategy and let $\psi_{\text{MCBR}}^{\text{rnd}}$ denote the robust intervention obtained by performing randomly selected experiments. As seen in Figure 2.3(a), the curves corresponding to these two methods begin and end at the same average cost, but the curve that corresponds to the proposed experimental design strategy drops much more sharply at the beginning compared to the random selection approach. This clearly demonstrates the effectiveness of the proposed method in reducing the network uncertainty. Figure 2.3(b) plots the difference between the average $\eta_\mu(\psi_{\text{MCBR}}^{\text{rnd}})$ and the average $\eta_\mu(\psi_{\text{MCBR}}^{\text{opt}})$. In both figures, the performance difference is especially large for the first few experiments and $\eta_\mu(\psi_{\text{MCBR}}^{\text{opt}})$ quickly approaches the minimum cost attained by the optimal intervention. This fast convergence is important, considering the difficulty of performing a large number of experiments in real applications.

2.3.3 Performance Evaluation Based on the Mammalian Cell Cycle Network

In this section, we evaluate the performance of the proposed experimental design strategy based on the mammalian cell cycle network. The cell cycle involves a sequence

33
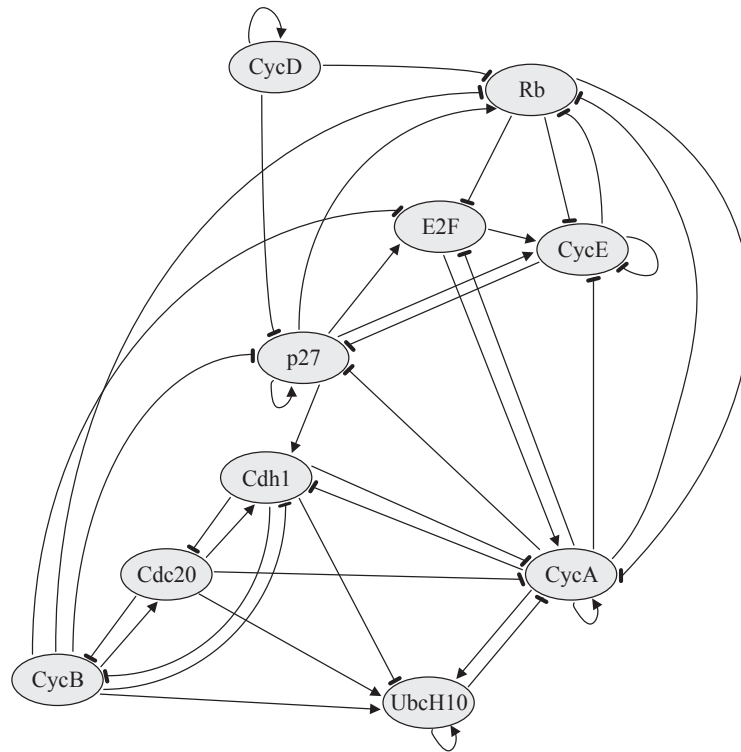
Figure 2.4. A gene regulatory network model of the mammalian cell cycle. Normal arrows represent activating regulations and blunt arrows represent suppressive regulations.

of events resulting in the duplication and division of the cell. It occurs in response to growth factors and, under normal conditions, it is a tightly controlled process. A regulatory model for the mammalian cell cycle is proposed in [66]. This model contains 10 genes: CycD, Rb, p27, E2F, CycE, CycA, Cdc20, Cdh1, UbcH10, and CycB. We represent this gene regulatory network by a BNp, where the perturbation probability is set to $p = 0.001$ and genes are numbered in the previous order. The regulatory model for this network is shown in Figure 2.4. The blunt arrows represent suppressive regulations and the normal arrows represent activating regulations. The cell cycle in mammals is controlled via extra-cellular stimuli. Positive stimuli activate Cyclin D (CycD) in the cell leading to cell division. CycD
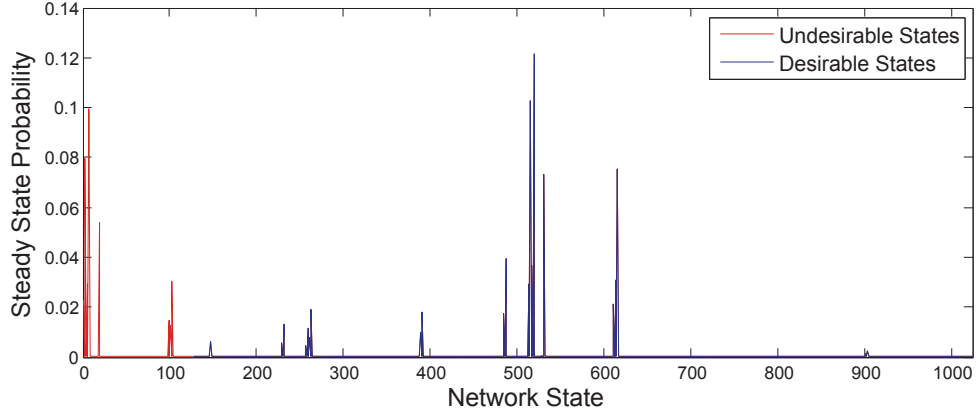
Figure 2.5. The steady state distribution of the mammalian cell cycle network modeled by a BNp with perturbation probability $p = 0.001$.

inactivates Rb protein, a tumor suppressor, via phosphorylation. When gene p27 and either CycE or CycA are active, the cell cycle stops, because Rb can be expressed even in the presence of cyclins. States in which the cell cycle continues even in the absence of stimuli are associated with cancerous phenotypes. For this reason, we regard states with down-regulated CycD, Rb, and p27 ($X_1 = X_2 = X_3 = 0$) as undesirable states, representing cancerous phenotypes. The network SSD is shown in Figure 2.5, where the total undesirable steady-state mass is $\pi_U = 0.3461$ without intervention. Suppose we want to reduce the steady-state probability mass of the set of undesirable states, $U = \{0, ..., 127\}$, via structural intervention. The optimal intervention is to change the transition from the input state 0000000111 to the output state 1110001011 by perturbing the regulatory function such that $\tilde{\mathbf{F}}(0000000111) = 1110001011$. For all other states, their output states remain unchanged after the intervention.

To evaluate the proposed experimental design method based on the given network, we again assume that $k \, (= 2, 3, 4, 5)$ regulations are unknown. For each k, we randomly

35

Table 2.5. The average gain of conducting the optimal experiment predicted by the proposed experimental design strategy in comparison to other suboptimal experiments. The 10-gene mammalian cell cycle network with k unknown regulations are considered.

	Average ξ_1	Average ξ_2	Average ξ_3	Average ξ_4
$k = 2$	0.0208	N/A	N/A	N/A
$k = 3$	0.0207	0.0261	N/A	N/A
$k = 4$	0.0217	0.0337	0.0379	N/A
$k = 5$	0.0365	0.0389	0.0395	0.0425

select 50 different sets of k regulations from the network, for which we assume their regulatory information is not known, and apply the experimental design strategy to predict the optimal experiment to be performed. Table 2.5 summarizes the average gain of the predicted optimal experiment over other suboptimal experiments for different values of k. The average gain is positive in all cases, as in our simulations based on synthetic BNps. Furthermore, the average gain ξ_i increases with i. For example, when $k = 5$,

$$\text{average } \xi_4 > \text{average } \xi_3 > \text{average } \xi_2 > \text{average } \xi_1.$$

Table 2.6 shows the proportion of "success", "failure", and "tie" for applying the proposed experimental design strategy. The results based on the mammalian cell cycle network are consistent with the results obtained from the synthetic networks. The "success" rate is consistently and significantly higher than the "failure" rate in all cases, thereby demonstrating the effectiveness of the proposed method. The proportion of "success" increases when we compare the optimal experiment with an experiment with larger $M_\Psi(\Theta, i')$ (i.e., for larger i'), which shows that the MOCU provides a sound mathematical basis for predicting the effectiveness of potential experiments.

Table 2.6. The proportion of success, failure, and tie of the optimal experiment predicted by the proposed strategy in comparison to other suboptimal experiments. The 10-gene mammalian cell cycle network with k unknown regulations are considered.

	$\theta_{1'} \sim \theta_{2'}$			$\theta_{1'} \sim \theta_{3'}$			$\theta_{1'} \sim \theta_{4'}$			$\theta_{1'} \sim \theta_{5'}$		
	Success	Failure	Tie	Success	Failure	Tie	Success	Failure	Tie	Success	Failure	Tie
$k=2$	40.00%	24.00%	36.00%	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
$k=3$	52.00%	30.00%	18.00%	54.00%	26.00%	20.00%	N/A	N/A	N/A	N/A	N/A	N/A
$k=4$	48.00%	24.00%	28.00%	56.00%	16.00%	28.00%	60.00%	12.00%	28.00%	N/A	N/A	N/A
$k=5$	56.00%	26.00%	18.00%	60.00%	20.00%	20.00%	60.00%	14.00%	26.00%	68.00%	10.00%	22.00%
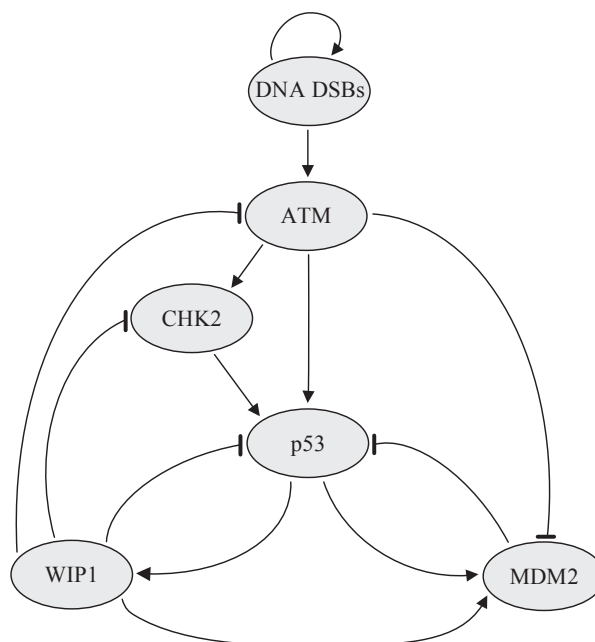
Figure 2.6. A gene regulatory model for the p53 network. Normal arrows represent activating regulations and blunt arrows represent suppressive regulations.

2.3.4 Performance Evaluation Based on the Pathways Involving p53 Gene

We now investigate performance of the proposed experimental design method on a p53 network [75]. p53 is a tumor suppressor gene which plays a major role in DNA damage regulation and programmed cell death (apoptosis). It has been observed that p53 is mutated in 30-50% of commonly occurring human cancers [76]. Under normal conditions, the expression level of p53 remains low via the control of MDM2, an oncogene that is often highly expressed in tumor cells. When DNA damage occurs, p53 is up-regulated and either activates other genes involved in DNA repair or it initiates apoptosis. Figure 2.6 shows key pathways that involve the regulation of p53 (see [77] for a detailed dynamical analysis of a very similar network). In the model of Figure 2.6, when a DNA double strand break occurs,

Table 2.7. The average gain of conducting the optimal experiment predicted by the proposed experimental design strategy in comparison to other suboptimal experiments. The 6-gene p53 network with k unknown regulations is considered.

	Average ξ_1	Average ξ_2	Average ξ_3	Average ξ_4
$k = 2$	0.0386	N/A	N/A	N/A
$k = 3$	0.0466	0.0434	N/A	N/A
$k = 4$	0.0343	0.0489	0.0657	N/A
$k = 5$	0.0387	0.0597	0.0622	0.0632

DNA DSBs becomes 1. The model contains five genes: MDM2, p53, WIP1, CHK2, and

ATM.

Like the mammalian cell cycle network, we model the p53 network as a BNp with

perturbation probability $p = 0.001$ and 6 nodes: X_1 (DNA DSBs), X_2 (MDM2), X_3 (p53),

X_4 (WIP1), X_5 (CHK2), and X_6 (ATM). The presence of DNA damage ($X_1 = 1$), perma-

nently up-regulated MDM2 ($X_2 = 1$) and permanently down-regulated p53 ($X_3 = 0$) would

result in an abundance of cancerous cells. For example, TCGA studies on 138 patients with

glioblastoma (a kind of brain tumor) have shown that 32% and 12% of them had mutated

p53 and MDM2 genes, respectively. Therefore, states with $X_1 = 1$, $X_2 = 1$, and $X_3 = 0$

are considered as the undesirable states; i.e., $U = \{48, ..., 55\}$. The steady-state proba-

bility mass of undesirable states before and after optimal structural intervention is 0.3478

and 0.0289, respectively. Our simulations use the same settings as for the mammalian cell

cycle network analysis.

Table 2.7 shows the average gain of conducting optimal experiments instead of other

suboptimal experiments. The average gain is always positive and in most cases average

ξ_i increases with i. However, there is an anomaly for $k = 3$, where average ξ_1 is larger than average ξ_2. Because the average results are obtained based on a single network and 50 different selections of uncertain parameters, we should expect such occurrences since we are not averaging over a large set of simulations as with the synthetic networks. Table 2.8 evaluates the performance of the predicted optimal experiments in terms of percentages of "success", "failure", and "tie". The "success" percentage is always larger than the "failure" percentage and it becomes larger when we compare the optimal experiment with an experiment corresponding to a larger i'. Again, there are a few anomalies, such as the decrease of "success" percentage for $k = 3$ and $k = 4$ when the optimal experiment is compared against the second and third optimal experiments – again not surprising given the small number of observations.

2.4 Discussion

Prioritization of potential experiments is of great practical import in systems biology and translational medicine. In this work, we have proposed a novel framework for evaluating the expected impact of a potential experiment in reducing the amount of uncertainty present in a dynamic network model. We estimate the mean objective cost of uncertainty expected to remain after conducting a specific experiment and select the one expected to optimally reduce network uncertainty. Extensive simulations based on both synthetic and actual networks show that the proposed experimental design strategy significantly outperforms random selection.

Table 2.8. The proportion of success, failure, and tie of the optimal experiment predicted by the proposed strategy in comparison to other suboptimal experiments. The 6-gene p53 network with k unknown regulations is considered.

	$\theta_{1'} \sim \theta_{2'}$			$\theta_{1'} \sim \theta_{3'}$			$\theta_{1'} \sim \theta_{4'}$			$\theta_{1'} \sim \theta_{5'}$		
	Success	Failure	Tie	Success	Failure	Tie	Success	Failure	Tie	Success	Failure	Tie
$k = 2$	26.00%	8.00%	66.00%	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
$k = 3$	34.00%	6.00%	60.00%	30.00%	4.00%	66.00%	N/A	N/A	N/A	N/A	N/A	N/A
$k = 4$	46.00%	24.00%	30.00%	44.00%	16.00%	40.00%	52.00%	14.00%	34.00%	N/A	N/A	N/A
$k = 5$	62.00%	18.00%	20.00%	62.00%	8.00%	30.00%	64.00%	8.00%	28.00%	66.00%	4.00%	30.00%

The proposed experimental design method is objective-based and herein our objective is network intervention. Therefore, the computational burden of the design method is mainly based on the associated network intervention strategy. A salient issue for network intervention methods is their inherent computational complexity [78, 79, 80]. The complexity of network intervention grows exponentially with network size. Computational complexity for experimental design is much greater because we need to find the optimal intervention for every potential network inside the uncertainty class.

Finally, it is worth noting that the problem considered in this work bears conceptual similarity to the online learning problems that have been gaining broad interest in recent years. In online learning, sequential measurements are made, one at a time, to improve an uncertain model. The online knowledge gradient (KG) algorithm is an interesting example that deals with a general class of such online learning problems [81]. It is assumed that one of M alternatives can be measured at each time step, which yields a random reward with an unknown mean and known variance (corresponding to measurement error). The main goal is to make sequential measurements that will maximize the expected total reward to be collected over a time period. To achieve this goal, in every time step, one tries to identify the optimal KG policy that will allow one to choose the single best measurement (among the M available alternatives) that is expected to bring forth the largest improvement. The alternative measurements (or rewards) are typically assumed to be independent Gaussian random variables, but one can incorporate prior beliefs about the measurements and their correlations into the problem via their joint distribution. Although the online

learning problem and the aforementioned KG algorithm bear some conceptual similarities to the sequential experimental design problem considered here and our MOCU-based strategy, there are critical differences. For example, our approach does not require direct modeling of the distribution of the reward (or cost). Instead, we focus on the uncertainty regarding the underlying network as it pertains to the cost of the operation of interest. Even though our ultimate goal is minimizing the cost, it is indirectly attained by optimally improving our knowledge regarding the network in a way that is pertinent to the operation (and its cost) to be performed based on the network.

3   APPROXIMATE EXPERIMENTAL DESIGN FOR GENE REGULATORY NETWORKS [1]

Although, the experimental design method proposed in the previous chapter is optimal, it is computationally expensive. Since our final objective is to improve the performance of the therapeutic interventions, the optimal experimental design method involves finding optimal interventions for all networks which are compatible with the prior knowledge. Finding optimal interventions is computationally expensive whose complexity grows exponentially with the number of genes in network. Therefore, the computational complexity of finding optimal experiment can be prohibitively high for large networks. Thus, it is inevitable to construct a smaller network via deleting some genes from the original large size network and then estimate the optimal interventions using the resulting reduced network. Generally the goal in network reduction methods is to produce networks of smaller size while the dynamical behavior of the original network is preserved. There have been some efforts for network reduction to reduce the complexity of designing interventions [79, 80, 82].

In this chapter, we propose a novel cost function for the gene deletion process which takes into account the disruption in the order of potential experiments when they are ranked according to the proposed experimental design method. Since experiments are ranked based upon the expected remaining MOCU or the MOCU that is expected to remain af-

---

ter performing the experiment, we desire that the network reduction step has a low effect on the expected remaining MOCU corresponding to the potential experiments. When the gene (or genes) suggested by the cost function are deleted from network, the optimal (and robust) interventions are estimated using the reduced networks and then they are used for calculating expected remaining MOCU for prioritizing potential experiments. We show the effectiveness of our proposed cost effective experimental design method through simulations on synthetic and real networks. The simulation results verify that our method can perform comparable to the optimal experimental design method with much lower computations.

MOCU-based optimal experimental design is very general and does not even require a Markovian network. As we will see, finding the best gene to delete is also very general; however, once the genes are deleted, the regulatory structure of the original network must be mapped onto a corresponding regulatory structure on the reduced network, an optimal intervention must be found on the reduced network, and that intervention must be induced to the full network. Reduction and inducement are nontrivial and depend on the nature of the regulatory structure. The problem has been addressed for Boolean networks in [82], to which we refer, and a theoretical analysis is given in [79], where it is noted that the methodology applies to probabilistic Boolean networks (PBNs) [47] by applying the reduction to each constituent BN of the PBN. Moreover, whereas we will restrict intervention to rank-one perturbations [49], which provide a one-time alteration of the regulatory logic, the reduction-inducement paradigm applies to other forms of intervention [79, 82].

3.1   Approximate Experimental Design Method

We briefly recall the optimal experimental design method proposed in the previous chapter.

Suppose there are k uncertain parameters $\theta = (\theta_1, \theta_2, ..., \theta_k)$, such that $\theta \in \Theta$, and that there are k potential experiments T_1, T_2, ..., T_k corresponding to k uncertain parameters. It is assumed that experiment T_i, which might be a complex experiment consisting of several sub-experiments, fully identifies θ_i. The goal of experimental design is to find out which experiment T_i, $1 \leq i \leq k$ should be conducted first, or how to rank potential experiments effectively. In the last chapter, we showed that the optimal experiment T_{i^*} to be conducted first is obtained as:

$$i^* = \underset{i=1,2,...,k}{\arg\min}\ \mathrm{M}_\Psi(\Theta; \theta_i), \tag{3.1}$$

where $\mathrm{M}_\Psi(\Theta; \theta_i)$ is the expected remaining MOCU after conducting experiment T_i which can be obtained as

$$\mathrm{M}_\Psi(\Theta; \theta_i) = \mathrm{E}_{\bar{\theta}_i}\left[\mathrm{M}_\Psi(\Theta | \theta_i = \bar{\theta}_i)\right] \tag{3.2}$$

such that

$$\mathrm{M}_\Psi(\Theta | \theta_i = \bar{\theta}_i) = \mathrm{E}_{\theta | \theta_i = \bar{\theta}_i}\left[\eta_{\theta | \theta_i = \bar{\theta}_i}\left(\psi_{\mathrm{IBR}}^{\Theta | \theta_i = \bar{\theta}_i}\right) - \eta_{\theta | \theta_i = \bar{\theta}_i}\left(\psi(\theta | \theta_i = \bar{\theta}_i)\right)\right], \tag{3.3}$$

According to (3.2), calculating the expected remaining MOCU requires finding the optimal intervention $\psi(\theta)$ for each $\theta \in \Theta$ and the robust intervention $\psi_{\mathrm{IBR}}^{\Theta | \theta_i = \bar{\theta}_i}$ for each possible remaining uncertainty class. The complexity of finding optimal interventions grows exponentially with network size n. For finding an optimal single-gene structural

46

intervention, we need to search among all possible $2^n \times 2^n$ state pairs and calculate the new steady-state probability $\tilde{\pi}_i$ for each state i in the set of undesirable states U. Thus, the complexity is $\mathcal{O}(2^{3n})$. This heavy computational cost motivates us to reduce the size of network in order to reduce the complexity of finding optimal interventions, thereby reducing the complexity of the experimental design.

Assuming that gene g is deleted from a network with regulatory function $\mathbf{F}$, we define a regulatory function $\mathbf{F}_{\text{red}}$ for the reduced network. Doing this for each network with uncertainty vector θ in Θ produces the uncertainty class, Θ^g, of reduced networks via the mapping $\theta \to \theta^g$.

To approximate the optimal intervention for a network in Θ, we use the corresponding network in Θ^g, find the optimal intervention for the reduced network $\psi(\theta^g)$, and then induce the intervention to the original network in Θ. This approximate optimal intervention denoted by $\psi(\theta; g)$ is called the *induced optimal intervention*. Also, to find the induced robust intervention, $\psi_{\text{IBR}}^{\Theta}(\text{ind}; g)$, for Θ, first we find the robust intervention, $\psi_{\text{IBR}}^{\Theta^g}$, for Θ^g using (2.1) and then find the induced robust intervention $\psi_{\text{IBR}}^{\Theta}(\text{ind}; g)$ from $\psi_{\text{IBR}}^{\Theta^g}$.

As illustrated in Figure 3.1, in the proposed approximate experimental design method, we find the best gene g^* for deletion via a novel cost function $c(g)$ and then obtain the induced optimal and robust interventions needed for the MOCU calculations in the experimental design step by inducing interventions from uncertainty class of reduced networks Θ^{g^*} to the original uncertainty class Θ.

We now aim to find a gene whose deletion results in minimum degradation in the
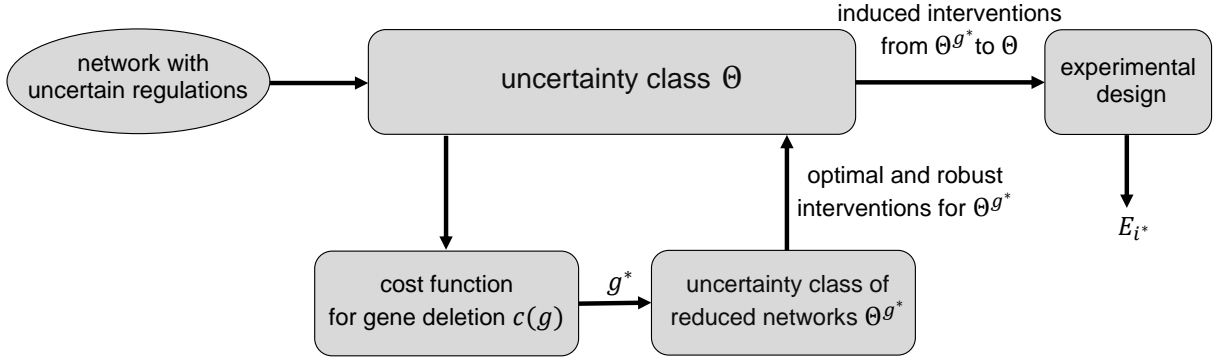
Figure 3.1. An illustrative view of the general approach of the proposed approximate experimental design method.

experimental design process. Keeping in mind that the experimental design is based on the expected remaining MOCU for potential experiments, let $M_{\Psi}^{g}(\Theta|\theta_i = \bar{\theta}_i)$ be the remaining MOCU when uncertain parameter i has value $\bar{\theta}_i$ and we delete gene g,

$$M_{\Psi}^{g}(\Theta|\theta_i = \bar{\theta}_i) = E_{\theta|\theta_i=\bar{\theta}_i}\left[\eta_{\theta|\theta_i=\bar{\theta}_i}\left(\psi_{\text{IBR}}^{\Theta|\theta_i=\bar{\theta}_i}(\text{ind};g)\right) - \eta_{\theta|\theta_i=\bar{\theta}_i}\left(\psi(\theta|\theta_i = \bar{\theta}_i)\right)\right]. \quad (3.4)$$

We define the cost of deleting gene g by

$$c(g) = \sum_{i=1}^{k}\left|M_{\Psi}^{g}(\Theta;\theta_i) - M_{\Psi}(\Theta;\theta_i)\right|, \quad (3.5)$$

where

$$M_{\Psi}^{g}(\Theta;\theta_i) = E_{\bar{\theta}_i}\left[M_{\Psi}^{g}(\Theta|\theta_i = \bar{\theta}_i)\right]. \quad (3.6)$$

The gene g^* minimizing the cost function in (3.5) is selected for deletion:

$$g^* = \underset{g\in 1,2,...,n}{\arg\min}\, c(g). \quad (3.7)$$

The intuition behind this cost function is that our choice of optimal experiment is based upon the expected remaining MOCU corresponding to each experiment. Therefore, we desire that the network reduction step has minimum effect on these quantities. Deleting

48

genes increases the inherent uncertainty of the network because the induced robust intervention cannot perform better than the original robust intervention on average. We want to reduce this increase in the uncertainty of model caused by network reduction. Since

$$\mathrm{E}_{\theta|\theta_i=\bar{\theta}_i}\left[\eta_{\theta|\theta_i=\bar{\theta}_i}\left(\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\bar{\theta}_i}(\mathrm{ind};g)\right)\right] \geq \mathrm{E}_{\theta|\theta_i=\bar{\theta}_i}\left[\eta_{\theta|\theta_i=\bar{\theta}_i}\left(\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\bar{\theta}_i}\right)\right], \qquad (3.8)$$

$\mathrm{M}_{\Psi}^{g}(\Theta;\theta_i) \geq \mathrm{M}_{\Psi}(\Theta;\theta_i)$. Hence, we can omit the absolute value operator in (3.5) to obtain

$$g^* = \arg\min_{g} \sum_{i=1}^{k} \left(\mathrm{M}_{\Psi}^{g}(\Theta;\theta_i) - \mathrm{M}_{\Psi}(\Theta;\theta_i)\right)$$

$$= \arg\min_{g} \sum_{i=1}^{k} \mathrm{M}_{\Psi}^{g}(\Theta;\theta_i), \qquad (3.9)$$

where the second equality follows from the fact that $\mathrm{M}_{\Psi}(\theta;\theta_i)$ does not depend on the gene being deleted. Expanding $\mathrm{M}_{\Psi}^{g}(\Theta;\theta_i)$ yields

$$g^* = \arg\min_{g} \left\{ \sum_{i=1}^{k} \mathrm{E}_{\bar{\theta}_i} \left[\mathrm{E}_{\theta|\theta_i=\bar{\theta}_i} \left[\eta_{\theta|\theta_i=\bar{\theta}_i}\left(\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\bar{\theta}_i}(\mathrm{ind};g)\right) - \eta_{\theta|\theta_i=\bar{\theta}_i}\left(\psi(\theta|\theta_i=\bar{\theta}_i)\right)\right]\right] \right\}$$

$$= \arg\min_{g} \sum_{i=1}^{k} \mathrm{E}_{\bar{\theta}_i} \left[\mathrm{E}_{\theta|\theta_i=\bar{\theta}_i} \left[\eta_{\theta|\theta_i=\bar{\theta}_i}\left(\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\bar{\theta}_i}(\mathrm{ind};g)\right)\right]\right]. \qquad (3.10)$$

The minimization problem in (3.10) is equivalent to the one in (3.7). Based on the cost function in (3.10), for each gene g, we find the expected performance of the induced robust intervention $\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\bar{\theta}_i}(\mathrm{ind};g)$ across the remaining uncertainty class $\Theta|\theta_i=\bar{\theta}_i$, then take the expectation of this average performance relative to the marginal distribution of the uncertain parameter θ_i, and finally sum all values found for the k uncertain parameters.

After removing gene g^*, we find the expected remaining MOCU corresponding to each experiment using equation (3.2) by replacing $\psi(\theta|\theta_i=\bar{\theta}_i)$ with $\psi^{\mathrm{ind}}(\theta|\theta_i=\bar{\theta}_i;g^*)$ and $\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\bar{\theta}_i}$ with $\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\bar{\theta}_i}(\mathrm{ind};g^*)$. An abstract form of the proposed experimental design method has been given in Algorithm 1. A step by step toy example illustrating Algorithm

49

Algorithm 1 Approximate experimental design

1: **input**: Θ, Ψ, $f(\theta)$, $\theta = (\theta_1, ..., \theta_k)$
2: **output**: $T_{i^*}, i^* \in \{1.,,,.,k\}$: the estimated optimal experiment to be conducted first
3: **for** $g = 1 : n$ **do**
4:     $cost(g) \leftarrow 0$
5:     **for** $i = 1 : k$ **do**
6:         **for all** $\bar{\theta}_i$ **do**
7:             build remaining uncertainty class of reduced networks $\Theta^g | \theta_i = \bar{\theta}_i$
8:             compute conditional density function $f(\theta | \theta_i = \bar{\theta}_i)$
9:             find induced robust intervention $\psi_{\text{IBR}}^{\Theta | \theta_i = \bar{\theta}_i}(\text{ind}; g)$
10:             $h_g(\theta_i) \leftarrow \mathrm{E}_{\theta | \theta_i = \bar{\theta}_i} \left[\eta_{\theta | \theta_i = \bar{\theta}_i} \left(\psi_{\text{IBR}}^{\Theta | \theta_i = \bar{\theta}_i}(\text{ind}; g) \right) \right]$
11:         $cost(g) \leftarrow cost(g) + \mathrm{E}_{\theta_i}[h_g(\theta_i)]$
12: $g^* \leftarrow \underset{g=1,2,...,n}{\arg\min} \, cost(g)$
13: **for** $i = 1 : k$ **do**
14:     **for all** $\bar{\theta}_i$ **do**
15:         build remaining uncertainty class $\Theta | \theta_i = \bar{\theta}_i$
16:         compute conditional density function $f(\theta | \theta_i = \bar{\theta}_i)$
17:         compute $\mathrm{M}_{\Psi}^{g^*}(\Theta | \theta_i = \bar{\theta}_i)$ via (3.3) using $\psi_{\text{IBR}}^{\Theta | \theta_i = \bar{\theta}_i}(\text{ind}; g^*)$ and $\psi^{\text{ind}}(\theta | \theta_i = \bar{\theta}_i; g^*)$
18:     $\mathrm{M}_{\Psi}^{g^*}(\Theta; \theta_i) \leftarrow \mathrm{E}_{\theta_i} \left[\mathrm{M}_{\Psi}^{g^*}(\Theta | \theta_i = \bar{\theta}_i) \right]$
19: $i^* \leftarrow \underset{i=1,2,...,k}{\arg\min} \, \mathrm{M}_{\Psi}^{g^*}(\Theta; \theta_i)$
20: **return** i^*
---

1 is also provided in Appendix A.

    This procedure for estimating optimal experiment via deleting one gene can be easily

extended to the deletion of two or more genes. For example, to delete two genes, we need

to evaluate the cost function in (3.10) for all possible two-gene combinations and delete

the pair whose cost is minimum.

3.1.1   Reduction Mappings and Induced Interventions

    If we want to delete gene g from network, we need to find the regulatory function

$\mathbf{F}_{\text{red}}$ for the reduced network. Following [82], every two states of the original network that

differ only in the value of gene g can be collapsed to find the transition rule of the reduced

---
Algorithm 2 Finding induced optimal interventions

1: **input**: $\psi(\theta^g) = (\hat{u}, \hat{v})$, π
2: **output**: $\psi^{\text{ind}}(\theta; g) = (u_g^{\text{ind}}, v_g^{\text{ind}})$
3: $\tilde{u}_1^g \leftarrow$ place 1 in the g-th coordinate of $\hat{u}$
4: $\tilde{u}_0^g \leftarrow$ place 0 in the g-th coordinate of $\hat{u}$
5: **if** $\pi_{\tilde{u}_1^g} \geq \pi_{\tilde{u}_0^g}$ **then**
6: $\quad u_g^{\text{ind}} \leftarrow \tilde{u}_1^g$
7: **else**
8: $\quad u_g^{\text{ind}} \leftarrow \tilde{u}_0^g$
9: $\tilde{v}_1^g \leftarrow$ place 1 in the g-th coordinate of $\hat{v}$
10: $\tilde{v}_0^g \leftarrow$ place 0 in the g-th coordinate of $\hat{v}$
11: **if** $\pi_{\tilde{v}_1^g} \geq \pi_{\tilde{v}_0^g}$ **then**
12: $\quad v_g^{\text{ind}} \leftarrow \tilde{v}_1^g$
13: **else**
14: $\quad v_g^{\text{ind}} \leftarrow \tilde{v}_0^g$
---

network. Let s_1^g and s_0^g be two states with value 1 and 0 for gene g, respectively, and identical values for other genes. State $\overline{s^g}$ can be obtained from either s_1^g or s_0^g by removing the value of gene g. If for the original network, the transition rules for these two states are $\mathbf{F}(s_1^g) = p$ and $\mathbf{F}(s_0^g) = q$, then for the reduced network, $\mathbf{F}_{\text{red}}(\overline{s^g}) = \overline{p^g}$ if $\pi_{s_1^g} > \pi_{s_0^g}$ and otherwise $\mathbf{F}_{\text{red}}(\overline{s^g}) = \overline{q^g}$, where $\overline{p^g}$ and $\overline{q^g}$ are found from states p and q via removing the value of gene g, respectively. Following this procedure, we find the regulatory function $\mathbf{F}_{\text{red}}$ for all states in the reduced network.

As illustrated in Algorithm 2, we find the induced optimal intervention from the optimal intervention for the reduced network. Suppose that the optimal intervention for the reduced network θ^g is $\psi(\theta^g) = (\hat{u}, \hat{v})$. The two corresponding states to $\hat{u}$ in the original network are $\tilde{u}_1^g$ and $\tilde{u}_0^g$, which are found by placing 1 and 0 in the g-th coordinate of $\hat{u}$, respectively. Similarly, there are two states $\tilde{v}_1^g$ and $\tilde{v}_0^g$ in the original network corresponding to state $\hat{v}$. The induced optimal intervention for the original network is

---

Algorithm 3 Finding induced robust interventions

1: **input**: $\psi_{\text{IBR}}^{\Theta^g} = (\hat{u}, \hat{v})$, $\pi(\theta) \, \forall \, \theta \in \Theta$
2: **output**: $\psi_{\text{IBR}}^{\Theta}(\text{ind}; g) = (u_g^{\text{ind}}, v_g^{\text{ind}})$
3: $\pi \leftarrow \text{E}_\theta[\pi(\theta)]$
4: $\tilde{u}_1^g \leftarrow$ place 1 in the g-th coordinate of $\hat{u}$
5: $\tilde{u}_0^g \leftarrow$ place 0 in the g-th coordinate of $\hat{u}$
6: **if** $\pi_{\tilde{u}_1^g} \geq \pi_{\tilde{u}_0^g}$ **then**
7: $\qquad u_g^{\text{ind}} \leftarrow \tilde{u}_1^g$
8: **else**
9: $\qquad u_g^{\text{ind}} \leftarrow \tilde{u}_0^g$
10: $\tilde{v}_1^g \leftarrow$ place 1 in the g-th coordinate of $\hat{v}$
11: $\tilde{v}_0^g \leftarrow$ place 0 in the g-th coordinate of $\hat{v}$
12: **if** $\pi_{\tilde{v}_1^g} \geq \pi_{\tilde{v}_0^g}$ **then**
13: $\qquad v_g^{\text{ind}} \leftarrow \tilde{v}_1^g$
14: **else**
15: $\qquad v_g^{\text{ind}} \leftarrow \tilde{v}_0^g$

$\psi^{\text{ind}}(\theta; g) = (u_g^{\text{ind}}, v_g^{\text{ind}})$, where u_g^{ind} is the one among $\tilde{u}_1^g$ and $\tilde{u}_0^g$ having larger steady-state probability in the original network and v_g^{ind} is the one among $\tilde{v}_1^g$ and $\tilde{v}_0^g$ with larger steady-state probability in the original network.

Analogous to the induced optimal intervention, the induced robust intervention $\psi_{\text{IBR}}^{\Theta}(\text{ind}; g)$ is found from the robust intervention $\psi_{\text{IBR}}^{\Theta^g}$ according to Algorithm 3; however, here we choose the two states possessing larger expected steady-state probability across Θ using the expected SSD, $\pi(\Theta) = \text{E}_\theta[\pi(\theta)]$, where $\pi(\theta)$ is the SSD of the network with uncertainty vector θ in uncertainty class Θ. We can use this procedure to find the induced robust intervention for each remaining uncertainty class $\Theta_{i,\phi}$.

3.1.2 Preliminary Gene Elimination via the Coefficient of Determination

To further reduce the computational cost of the experimental design, we utilize the coefficient of determination (CoD) [83] to eliminate some genes from the optimization

problem without evaluating the cost function and then search among the remaining genes for choosing genes to be removed using the cost function (3.10). The CoD measures the strength of relationship between a target gene Y and a vector X of predictor genes as the difference between the error of the best estimation of gene Y in the absence of other genes and in the presence of genes in X. The CoD is between 0 and 1 and a larger CoD means a stronger connection between the target and predictor genes, in our case the target gene being the aim of intervention. We use the intuition that genes possessing large CoD in relation to the target gene are not likely among the genes that should be deleted because they have strong connection to the target gene. The CoD of the target gene Y, relative to a vector $X = (X_1, ..., X_m)$ of predictor genes is defined by

$$\text{CoD}_X(Y) = \frac{\varepsilon_Y - \varepsilon_{X,Y}}{\varepsilon_Y} \tag{3.11}$$

where ε_Y is the error of the best estimation of Y without any predictors,

$$\varepsilon_Y = \min[\text{Pr}(Y = 0), \text{Pr}(Y = 1)], \tag{3.12}$$

and $\varepsilon_{X,Y}$ is the error of the optimal estimation of Y upon observing X. By assuming that the value of the binary vector X of predictor genes changes from 1 to 2^m, $\varepsilon_{X,Y}$ can be calculated by

$$\varepsilon_{X,Y} = \sum_{j=1}^{2^m} \text{Pr}(X = j) \min \left[\text{Pr}(Y = 0 | X = j), \text{Pr}(Y = 1 | X = j) \right]. \tag{3.13}$$

If $\text{CoD}_X(Y; \theta)$ denotes the CoD of Y relative to X in a network with uncertainty vector θ, then given the uncertainty class Θ the expected CoD of Y relative to X is given

by

$$\mathrm{CoD}_X(Y;\Theta) = \mathrm{E}_\theta\left[\mathrm{CoD}_X(Y;\theta)\right]. \qquad (3.14)$$

Genes possessing strong connection with the target gene in terms of $\mathrm{CoD}_X(Y;\Theta)$ are not considered for deletion. When excluding genes using the CoD it is important to recognize the possibility of intrinsic multivariate prediction [84], where a set of genes may have low individual CoDs with respect to the target gene but may have significant CoD when used together for multivariate prediction. First we calculate $\mathrm{CoD}_X(Y;\Theta)$ for all 3-gene combinations and pick the one with largest CoD. We compute CoD for 3-gene predictors because it has been shown in [59] that the average connectivity of the model cannot be too high providing that the model is not chaotic and it is commonplace to assume 3-gene predictivity in BNs. If we want to exclude less than 3 genes from the search space, then among the 3-gene combination with the largest expected CoD, we choose those genes that have larger expected individual CoD. If we want to exclude more than 3 genes, then in addition to the three genes in the combination with the largest CoD, we choose those genes in the 3-gene combination with the second largest CoD that have larger expected individual CoD and do not belong to the first 3-gene combination. We repeat this process until we reach the desired number of genes to exclude.

If there are initially n genes and we want to delete 3 genes, then we need to evaluate cost function (3.10) for all $C(n,3)$ 3-gene combinations, where $C(n,k)$ denotes the number of combinations of n objects taken k at a time; however, if we exclude s genes from search space then the number of evaluations of (3.10) decreases to $C(n-s,3)$.

Having performed the CoD-based exclusion process and excluded s genes, $X_1', X_2', ..., X_s'$, we search for the genes to be deleted using the cost function in (3.10) among the remaining genes, $\{X_1, X_2, ..., X_n\} - \{X_1', X_2', ..., X_s'\}$.

3.1.3 Computational Complexity Analysis

The first step for the optimal experimental design is estimating optimal interventions $\psi(\theta)$ for each network in Θ. We also need to compute the robust intervention $\psi_{\text{IBR}}^{\Theta|\theta_i=\bar{\theta}_i}$ for each possible remaining uncertainty class $\Theta|\theta_i = \bar{\theta}_i$. Most of the computations are devoted to this step. Finding robust interventions does not require additional calculations because we can store the error of each intervention $\psi \in \Psi$ for the network θ when finding optimal interventions and later use these errors to find robust interventions. Therefore, complexity analysis requires computing the complexity of estimating the optimal interventions.

With n genes that take on binary expression levels, the network has 2^n states. Finding an optimal single-gene function intervention requires searching among all possible 2^{2n} state pairs (u, v) according to (2.9). Assuming without loss of generality that states 2^{n-1} to 2^n are undesirable, (2.9) must be evaluated 2^{n-1} times for each state pair. Thus, the complexity of finding the optimal intervention $\psi(\theta)$ is $\mathcal{O}(2^{3n})$. If there are k uncertain parameters and each can take on l different values, then the uncertainty class Θ contains l^k different networks for which an optimal intervention must be found. Hence, the complexity of the optimal experimental design method is $\mathcal{O}(l^k \times 2^{3n})$.

To analyze the complexity of the proposed approximate method, suppose p genes are to be deleted. Then the cost function in (3.10) must be evaluated for all $C(n-1, p)$

p-gene combinations, $n-1$ instead of n because the target gene cannot be deleted. The complexity of finding an induced optimal intervention for each network after deleting p genes is $\mathcal{O}(2^{3(n-p)})$. Therefore, the complexity of the approximate method is $\mathcal{O}(C(n-1,p) \times l^k \times 2^{3n-3p})$. For large n, it is possible that for small p the complexity of the approximate method can exceed that of the original method; however, by deleting more genes the complexity of the approximate method drops sharply because by deleting each additional gene the complexity of estimating the optimal intervention decreases by eight-fold.

By incorporating the CoD-based gene exclusion step in the approximate method and excluding s genes we are able to decrease the number of p-gene combinations from $C(n-1,p)$ to $C(n-s-1,p)$, which reduces the complexity of the approximate method to $\mathcal{O}(C(n-s-1,p) \times l^k \times 2^{3n-3p})$. Define the *computational gain* λ by

$$\lambda = \frac{l^k \times 2^{3n}}{C(n-s-1,p) \times l^k \times 2^{3n-3p}} = \frac{2^{3p}}{C(n-s-1,p)}, \qquad (3.15)$$

which is the ratio of the complexity of the optimal design method to the complexity of the approximate method when deleting p genes using the cost function in (3.10) and excluding s genes from the search space using the CoD-based gene exclusion step.

Figure 3.2 shows the computational gain λ when deleting p genes and excluding s genes from the search space for network size of 10 and 15. Note that for large n, if we delete very few genes the complexity might exceed that of the optimal experimental design method but as more genes are deleted the complexity of the approximate method becomes much smaller. For example when $n = 15$, searching over all genes and deleting
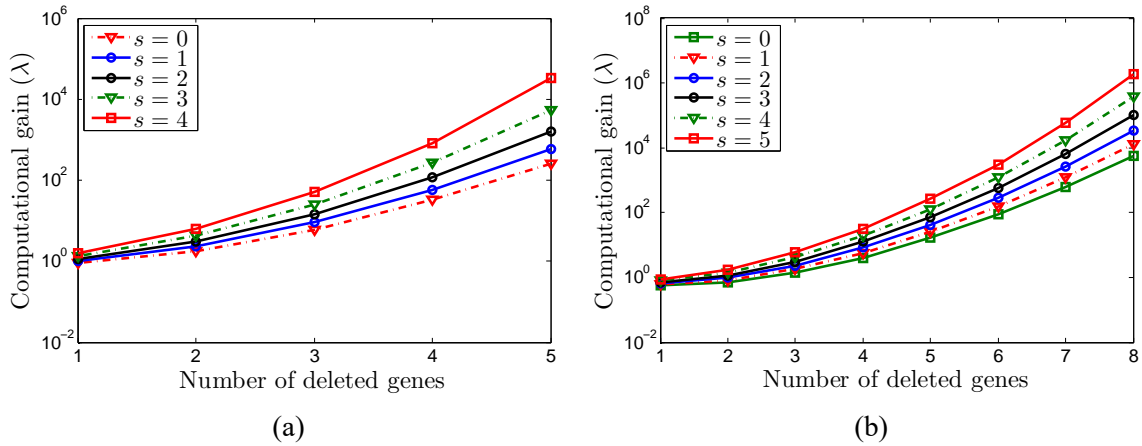
Figure 3.2. Computational gain (λ) of using the proposed approximate method. Different number of genes are deleted and s genes are excluded from the search space. (a) Networks with $n = 10$ genes. (b) Networks with $n = 15$ genes.

Table 3.1. Comparing the approximate processing times (in seconds) of the optimal and approximate experimental design methods when p genes are deleted and s genes are excluded for networks of size n with 4 uncertain regulations.

	$n = 10$		$n = 11$		$n = 12$	
Optimal	468		4846		60169	
Approximate	$s = 0$	$s = 2$	$s = 0$	$s = 2$	$s = 0$	$s = 2$
$p = 3$	81.71	39.64	830	407	9795	5026
$p = 4$	23.61	12.98	215	93	2355	1057
$p = 5$	8.68	8.36	58	35	450	181

1, 2, and 3 genes, $\lambda = 0.5741$, $\lambda = 0.7$, and $\lambda = 1.4$, respectively, but for $p > 3$, λ grows rapidly, reaching $\lambda \approx 600$ when deleting 7 genes. Greater computational gain results from excluding some genes using the CoD-based step. For instance, excluding 3 genes from the search space results in $\lambda = 1.16$ and $\lambda \approx 6350$ when deleting 2 and 7 genes, respectively.

Table 3.1 shows the approximate processing times for performing the optimal and proposed experimental design methods for networks of different size with 4 unknown regu-

lations. Simulations have been run on a machine with 8GB of RAM and Intel(R) Core(TM) i7 CPU, 3.1 GHz. The run times grow exponentially as the number of genes increases. This table clearly suggests that the optimal experimental design method can be applicable to networks of at most $n = 11$ genes but using the proposed approximate experimental design method we can still increase the number of genes in the network. For example, for $n = 12$ genes, optimal experimental design takes around 17 hours to complete but when we use the proposed method and delete 5 genes it takes around 8 minutes without gene exclusion and 3 minutes with 2-gene exclusion – a significant saving in processing time. Note that the ratios between times in Table 3.1 do not exactly follow the computational gain in (19), especially when the size of the reduced network is very small, because the times in the table also include the time required for calculating the SSD, TPM, and fundamental matrices for original networks in Θ.

3.2   Simulation Results

This section evaluates the performance of the approximate method for both synthetic and real GRNs where the majority vote rule is used as the transition rule. Majority vote rule [74, 85, 86, 87] is popular in systems biology, especially when we are interested in the overall dynamics of the network. For example, majority vote is used in [86] to model the dynamics of yeast cell-cycle network. For the majority vote rule, a regulatory matrix $\mathbf{R}$ is defined component-wise by

$$R_{ij} = \begin{cases} 1 & \text{gene } j \text{ activates gene } i \\ -1 & \text{gene } j \text{ suppresses gene } i \\ 0 & \text{no relation from gene } j \text{ to gene } i \end{cases}.$$

According to this rule, gene i takes value 1 if the number of genes that are ON and activate it is more than the number of genes that are ON and suppress it:

$$X_i(t+1) = f_i\big(X(t)\big)$$

$$= \begin{cases} 1 & \text{if } \sum_j R_{ij} X_j(t) > 0 \\ 0 & \text{if } \sum_j R_{ij} X_j(t) < 0 \\ X_i(t) & \text{if } \sum_j R_{ij} X_j(t) = 0 \end{cases}$$

Uncertainty is introduced by assuming that the exact values of some of the nonzero components of **R** are unknown; that is, for some regulations it is not known whether they are activating or suppressive. Each uncertain parameter θ_i can be -1 or 1. Conducting experiment T_i determines the value of parameter θ_i. Let $\mu = (\mu_1, ..., \mu_k)$ denote the true value for the uncertainty vector $\theta = (\theta_1, \theta_2, ..., \theta_k)$. Conducting experiment T_i results in a remaining uncertainty class $\Theta|\theta_i = \mu_i$ consisting of networks with $\theta_i = \mu_i$ and other uncertain parameters being -1 or 1. For $\Theta|\theta_i = \mu_i$ we can determine a robust intervention $\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\mu_i}$. We evaluate the effectiveness of experiment T_i in terms of the error of the resulting robust intervention obtained after experiment on the underlying true network, $\eta_\mu\big(\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\mu_i}\big)$. We define the gain of conducting the chosen experiment T_{i*} over a random experiment T_{rnd} (chosen randomly without using any experimental design) by

$$\xi = \eta_\mu\big(\psi_{\mathrm{IBR}}^{\Theta|\theta_{\mathrm{rnd}}=\mu_{\mathrm{rnd}}}\big) - \eta_\mu\big(\psi_{\mathrm{IBR}}^{\Theta|\theta_{i*}=\mu_{i*}}\big). \tag{3.16}$$

If $\xi > 0$, then the chosen experiment outperforms the random experiment; if $\xi < 0$, then the random experiment outperforms the chosen experiment; and if $\xi = 0$, then they perform the same.

3.2.1 Simulations on Synthetic Examples

For the performance evaluation based on synthetic BNps, we generated 1000 networks randomly and chose 50 different sets of k regulations in each to be unknown – 50000 simulations in total. We assigned 3 random predictor genes to each gene where each one can be randomly activating or suppressive. The gene perturbation probability was set to 0.001. Without loss of generality, we assume that states with up-regulated X_1 are undesirable. We removed the regulatory type of those regulations that have been assumed to be uncertain and retained other regulatory information of the network. We assume that all uncertain parameters are independent from each other and have uniform marginal distribution. The analysis can be easily extended to other distributions. Because X_1 is the target gene, it was excluded from the reduction process. Hence, we look for the best p-gene set to be deleted among $\{X_2, ..., X_n\}$.

Figure 3.3 shows the average gain ξ for networks with $n = 7$ genes and $k = 2, 3, 4, 5$ uncertain regulations. For each k, we delete 1, 2, and 3 genes. Given the deletion of p genes, to evaluate the effectiveness of the proposed cost function in (3.10), we rank all p-gene combinations based on this cost function and compare the performance of the proposed approximate method when deleting each of these sets. For example in Figure 3.3(a), there are 6 different choices for a single gene to be deleted or in Figure 3.3(b) there
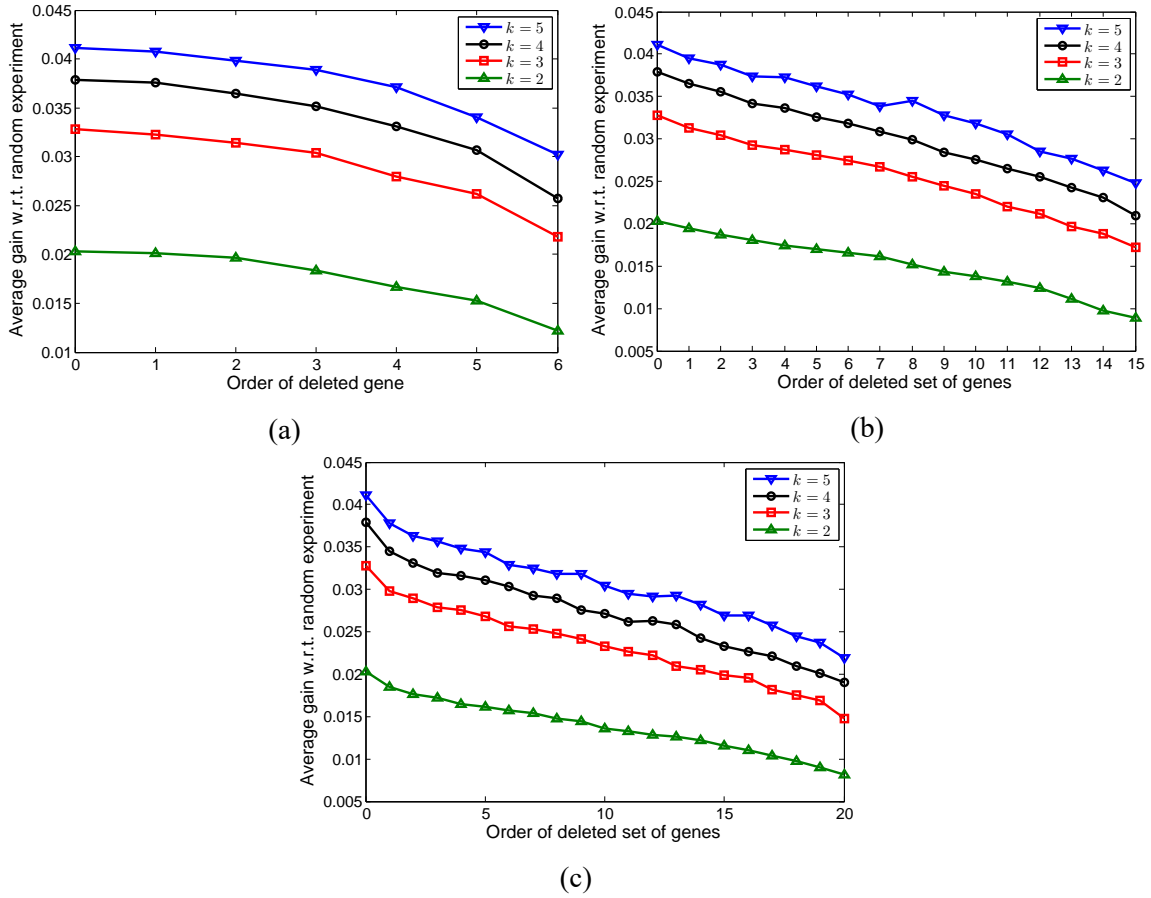
Figure 3.3. Evaluating the effectiveness of the proposed cost function for 7-gene networks with k uncertain regulations. The average gain of conducting the chosen experiments by the proposed approximate method with respect to the random experiments when deleting different genes is shown. (a) Deleting one gene. (b) Deleting two genes. (c) Deleting three genes.

are $C(6,2) = 15$ different selections for two genes to be deleted. In all subfigures in 3.3, the average gain when the order of the deleted set is 0 corresponds to optimal experimental design. This figure shows that for different number of uncertain regulations and different number of deleted genes, deleting those sets that correspond to a lower cost function results in larger average ξ. Denoting average ξ by $\bar{\xi}$, for $k = 5$, where $\bar{\xi} = 0.0411$ for the optimal method, if we delete the gene with minimum cost, then $\bar{\xi} = 0.0408$, but if we delete the

61
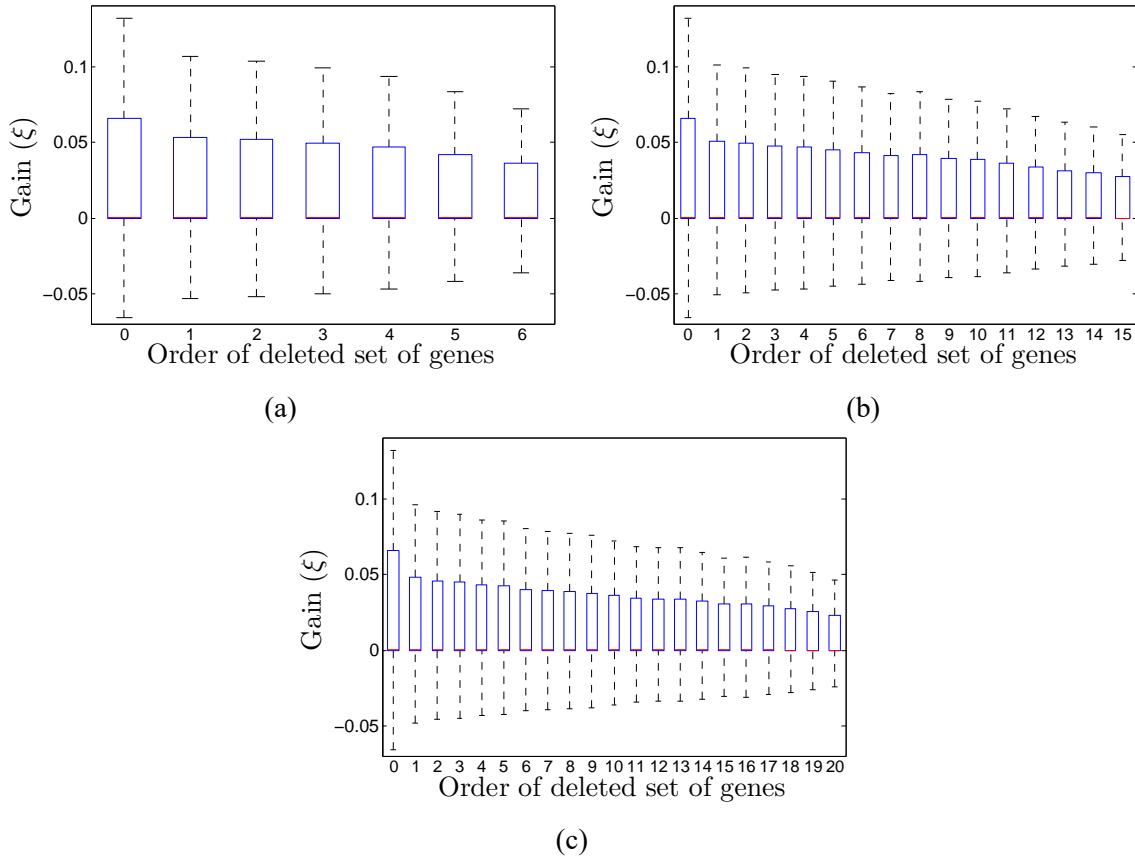
Figure 3.4. The box plot of the gain of conducting the chosen experiment by the proposed approximate method with respect to the random experiment when deleting different genes. 7-gene networks with 5 uncertain regulations are considered. (a) Deleting one gene. (b) Deleting two genes. (c) Deleting three genes.

gene with maximum cost, then $\bar{\bar{\xi}} = 0.0302$. When deleting two genes, corresponding to the best pair of genes (corresponding to the minimum cost) $\bar{\bar{\xi}} = 0.0395$ but for the pair corresponding to the largest cost (15th set) $\bar{\bar{\xi}} = 0.0248$. When deleting three genes, for the best set of deleted genes $\bar{\bar{\xi}} = 0.0378$ and for the worst set $\bar{\bar{\xi}} = 0.0219$.

Figure 3.4 provides the box plots for 7-gene networks possessing 5 uncertain regulations when 1, 2, and 3 genes are deleted. The box extends from the first quartile (25th percentile) to the third quartile (75th percentile) of the data. The lines extending vertically

from the box are called "whiskers". Herein we set whisker length to the interquartile range (distance between the first and third quartiles). The red line in the box represents the median. Note that in the given box plots median and first quartile might not be distinguishable as they are very close to each other but in fact they have different values. The number on the x-axis is the ranking of the set of deleted genes, running from the minimum cost of deletion on the left to the maximum cost of deletion on the right. For optimal experimental design the first quartile, median and third quartile are -1.57×10^{-5}, 5.38×10^{-5}, and 0.662, respectively. For approximate experimental design, as we delete gene(s) whose corresponding cost function is larger, the first quartile, median, and third quartile decrease. For example, in case of deleting 3 genes, if we delete the set of genes corresponding to the minimum cost of deletion the first quartile, median, and third quartile are -3.53×10^{-5}, 1.51×10^{-6}, and 0.048, respectively but if we delete the set of genes with the maximum of deletion cost the first quartile, median, and third quartile would be -0.00055, 0, and 0.023 respectively. These box plots indicate the promising performance of the proposed cost function because the boxes cover larger values when we delete set of genes possessing smaller cost function.

Figure 3.5 shows performance evaluation for 8-gene networks with $k = 4$ uncertain regulations, deleting up to four genes from the original networks. Again, this figure verifies the promising performance of the proposed cost function. It can be observed that when gene sets possessing larger cost are deleted, the resulting average gain decreases. For example, when we delete 4 genes $\bar{\xi} = 0.0390$ for the optimal method and $\bar{\xi}$ for the approximate
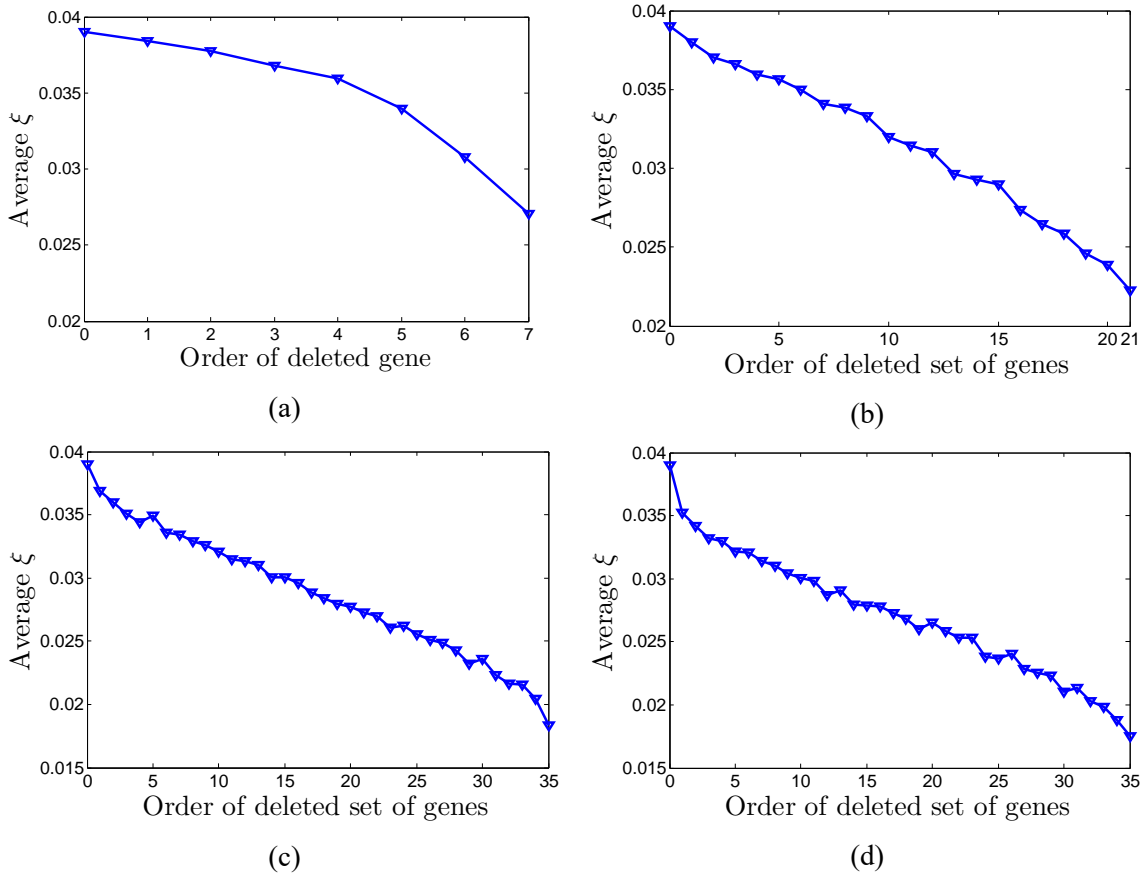
Figure 3.5. Evaluating the effectiveness of the proposed cost function for 8-gene networks with 4 uncertain regulations. The average gain of conducting the chosen experiments by the proposed approximate method with respect to the random experiments is shown. (a) Deleting one gene. (b) Deleting two genes. (c) Deleting three genes. (d) Deleting four genes.

method decreases from 0.0352 to 0.0175 if we delete the 35th set of 4 genes according to the cost function instead of the first set. To consider larger networks, we generated 100 random 13-gene networks and in each chose one set of 4 regulations to be unknown. We used the approximate experimental design method and deleted 5 genes. For this size of network it is not possible to perform the optimal experimental design method or compute original optimal and robust interventions to calculate the gain of the chosen experiment over a
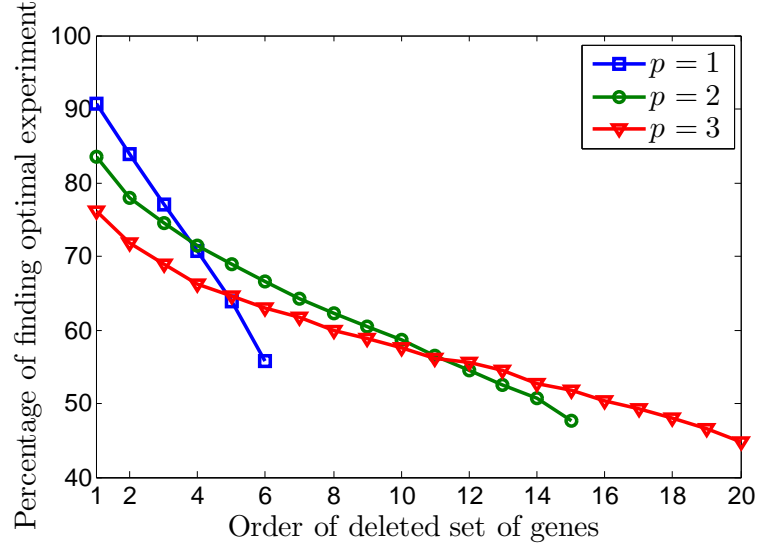
Figure 3.6. Percentage of finding the same experiment as the optimal method. Different p genes are deleted for 7-gene networks possessing $k = 4$ uncertain regulations. Gene sets with larger order have higher cost function.

randomly selected experiment. Hence, we use the induced robust intervention obtained by deleting the set of 5 genes having the minimum cost of deletion as in (14). Therefore, here gain ξ is defined as

$$\xi = \eta_\mu \left(\psi_{\text{IBR}}^{\Theta|\theta_{\text{rnd}}=\mu_{\text{rnd}}} (\text{ind}; g^*) \right) - \eta_\mu \left(\psi_{\text{IBR}}^{\Theta|\theta_{i*}=\mu_{i*}} (\text{ind}; g^*) \right)$$

where g^* is the set of 5 genes with minimum cost of deletion. For this set of simulations, the average gain $\bar{\xi}$ is 0.0192. Note that here the average gain might not be very accurate owing to the small number of simulations. The approximate run time for each simulation was around 5700 seconds.

We now evaluate the proportion of times that we obtain the optimal experiment found by when using the approximate method. Figure 3.6 shows the percentage of finding the optimal experiment when using the approximate method and deleting different number of

genes from 7-gene networks. In this figure, there are 6, 15, and 20 values for deleting 1, 2, or 3 genes, respectively. We observe that deleting the set of genes corresponding to the minimum of the cost function yields the highest likelihood of obtaining an optimal experiment, which is what we would hope for from an efficient approximate method. According to Figure 3.6, when we delete the gene which attains the minimum cost, 90.72% of the simulations yield an optimal experiment, whereas this percentage is 55.73% when deleting the gene with the largest value of the cost function. Similar behavior is observed when deleting 2 or 3 genes. A salient reason that the largest average gain of the approximate method over random experiments is when we delete genes corresponding to the minimum cost function is that it is more likely to get an optimal experiment.

An issue that arises when evaluating experimental design on synthetic networks, as opposed to real biological networks, which typically manifest substantial controllability on account of their need to maintain functionality within changing contexts, a large portion of randomly generated networks might not be controllable and therefore not be responsive to intervention. Hence, intervention has negligible effect on their SSDs and including them in the analysis masks the effect of optimality. To address this issue, we define *controllability* Δ as the percentage decrease of undesirable probability mass after applying optimal intervention:

$$\Delta = \frac{\pi_U - \tilde{\pi}_U}{\pi_U} \times 100\%, \tag{3.17}$$

where π_U and $\tilde{\pi}_U$ are the undesirable probability masses before and after applying optimal intervention to the network, respectively. A larger Δ means that a network is more
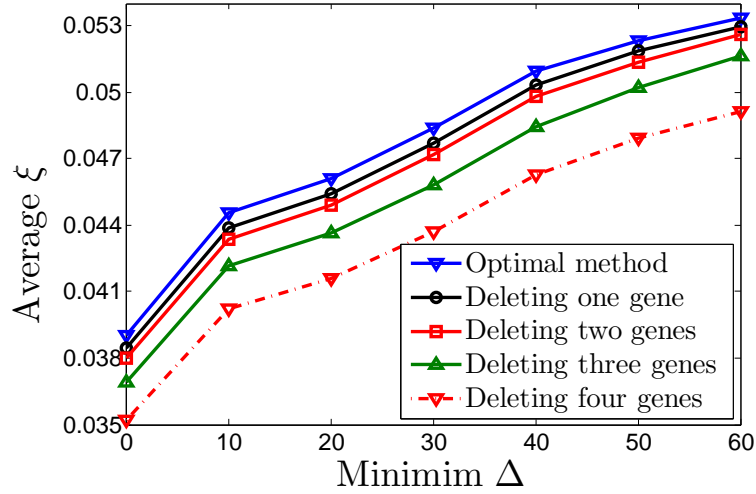
Figure 3.7. Effect of controllability Δ on the performance of the experimental design method. Optimal and approximate experimental design methods when deleting p genes are considered for networks with $n = 8$ genes and $k = 4$ uncertain regulations.

controllable. Figure 3.7 considers the effect of controllability on the performance of experimental design when networks have $n = 8$ genes and $k = 4$ uncertain regulations. The figure shows the average gain $\bar{\xi}$ for the optimal method and the proposed approximate method for networks possessing controllability greater than a given threshold. We observe that $\bar{\xi}$ increases when networks are more controllable, regardless of the number of genes deleted from network. Note that as controllability increases, the difference between the performance of different methods increases. For example, for all networks the average gain for the optimal method and the proposed method when deleting one, two, three, and four genes is 0.0390, 0.0384, 0.0380, 0.0369, and 0.0352, respectively; but for networks with $\Delta \geq 40\%$ the average gains are 0.0509, 0.0503, 0.0498, 0.0484, and 0.0463, respectively.

To evaluate the effectiveness of the CoD-based gene exclusion algorithm, we com-
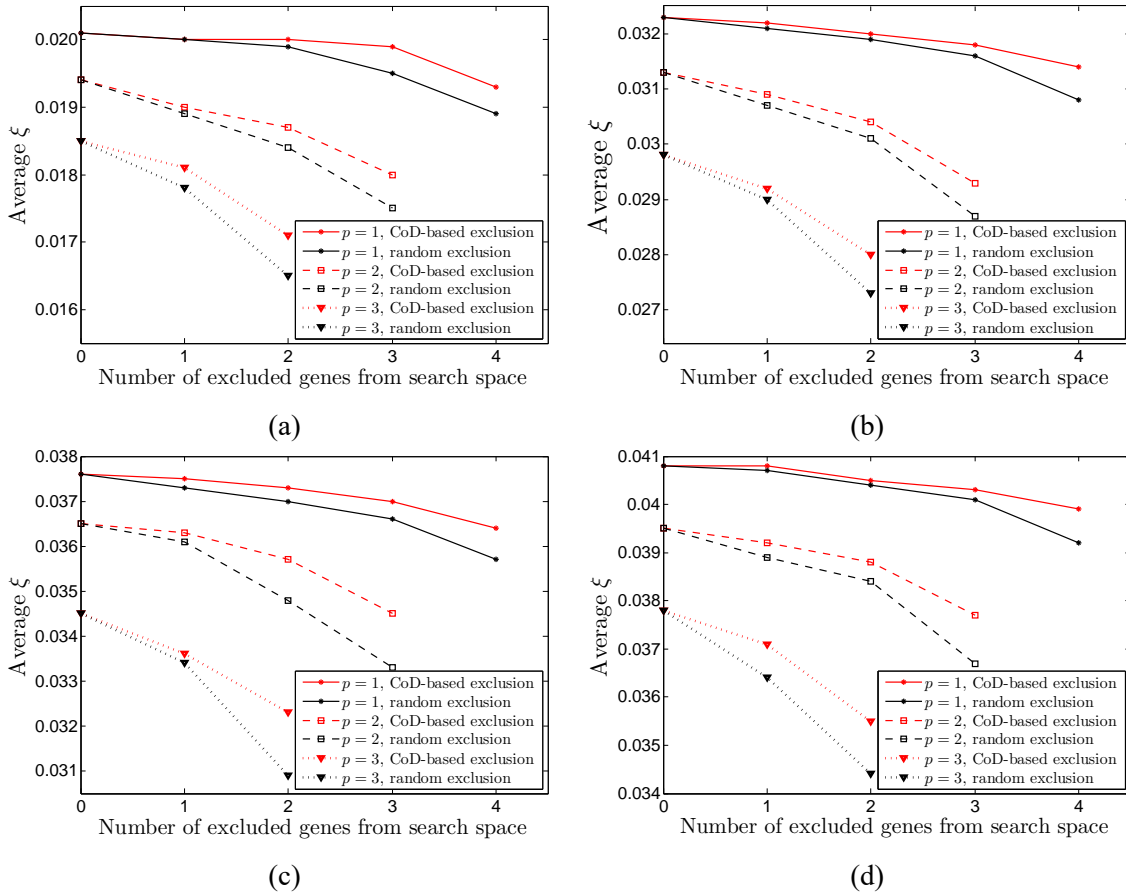
Figure 3.8. Performance evaluation of the CoD-based gene exclusion scheme for 7-gene networks. The average gain of the proposed method over the random experiments when p genes are deleted and different number of genes are excluded from the search space is shown. (a) $k = 2$ uncertain regulations. (b) $k = 3$ uncertain regulations. (c) $k = 4$ uncertain regulations. (d) $k = 5$ uncertain regulations.

pare the average gain of the approximate method when excluding genes from the search space using the CoD-based exclusion algorithm against the average gain when excluding randomly selected genes from the search space. Figure 3.8 shows the average gain $\bar{\xi}$ for networks with $n = 7$ genes and $k = 2, 3, 4, 5$ uncertain regulations. For deleting p genes, we exclude up to $6 - p - 1$ genes from the search space so that for the largest number of genes excluded, the search space contains at least $p + 1$ genes. For example, when deleting

$p = 1$ gene, we exclude 1, 2, 3, and 4 genes; when deleting $p = 2$ genes we exclude 1, 2, and 3 genes; and so on. For each number of uncertain regulations, we observe that the average gain when excluding genes using the CoD-based algorithm is always larger than random gene exclusion, regardless of the number of deleted genes. For example, when $k = 5$, for deleting one gene and excluding 1, 2, 3, and 4 genes randomly, $\bar{\xi} = 0.0407$, 0.0404, 0.0401, and 0.0392 respectively but using the CoD-based scheme and excluding the same number of genes, $\bar{\xi} = 0.0408$, 0.0405, 0.0403, and 0.0399 respectively. If we delete three genes $\bar{\xi} = 0.0378$ without gene exclusion, and if we exclude 1 and 2 genes, then $\bar{\xi} = 0.0364$ and $\bar{\xi} = 0.0344$, respectively, when we exclude genes randomly and $\bar{\xi} = 0.0371$ and $\bar{\xi} = 0.0355$, respectively, when we exclude genes based on CoD . Note that when deleting more genes, the difference between random exclusion and CoD-based exclusion increases because as more genes are deleted, exclusion has a larger impact on the number of candidate sets for evaluating the cost function. For example, when deleting 1 gene, if we exclude one gene, then the number of candidate sets decreases from 6 to 5, but when deleting 3 genes, if we exclude one gene, then the number of candidate sets decreases from $C(6,3) = 20$ to $C(5,3) = 10$.

In Figure 3.9, we also show the box plot for the gain of conducting the chosen experiment if we delete 3 genes and exclude genes from the search space either randomly or via the proposed CoD-based method for 7-gene networks possessing 5 uncertain regulations. We observe that the first quartile, median, and third quartiles are higher when excluding genes using CoD. For example, when randomly excluding 2 genes from the search
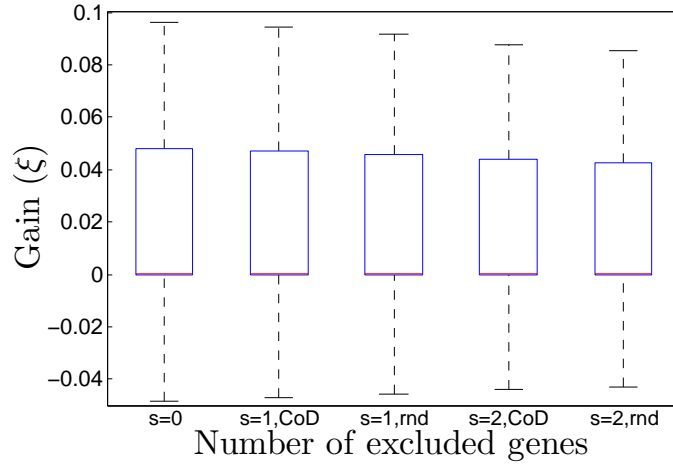
Figure 3.9. The box plot of the gain with respect to the random experiment when *s* genes are excluded randomly or using the proposed CoD-based procedure. 7-gene networks with 5 uncertain regulations are considered.

space, the first quartile, median, and third quartile are -6.64×10^{-5}, 3.33×10^{-7}, and 0.043, respectively; however, when excluding genes using CoD they are -5.62×10^{-5}, 6.04×10^{-7}, and 0.044, respectively.

Figure 3.10 is similar to Figure 3.8 except that it is for 8-gene networks with 4 uncertain regulations. The approximate method is applied deleting 1, 2, 3, and 4 genes. For each number of deleted genes, average ξ is computed for random and CoD-based exclusion.

Table 3.2 lists the percentage that the optimal experiment is found using the approximate method when deleting *p* genes and excluding *s* genes from the search space randomly or according to the CoD-based algorithm. Results are tabulated for 7-gene networks with $k = 2, 3, 4, 5$ uncertain regulations and 8-gene networks with $k = 4$ uncertain regulations. Note that if we are interested in deleting *p* genes, $p + 1$ genes should remain in the search space after the gene exclusion step. For example, for $p = 2$ we exclude up to $s = 3$ genes and for $p = 3$ we exclude up to $s = 2$ genes from the search space. We use N/A in the

70
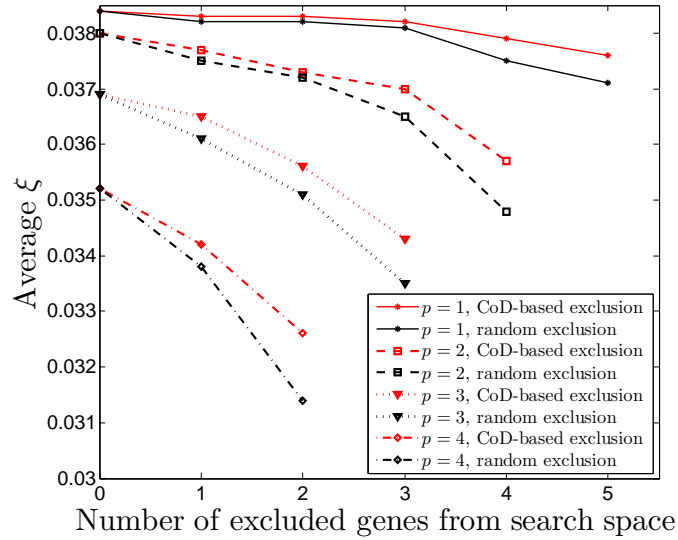
Figure 3.10. Performance evaluation of the CoD-based gene exclusion algorithm for 8-gene networks with $k = 4$ uncertain regulations. The average gain of the proposed approximate experimental design with respect to the random experiments when p genes are deleted and different number of genes are excluded from the search space is shown.

table for those pairs of p and s which are not applicable. We observe that the likelihood of obtaining the optimal experiment is larger when we exclude genes according to the CoD-based algorithm rather than excluding them randomly. A larger proportion of experiments found by the approximate method via excluding genes based on CoD agree with the optimal method. These tables demonstrate the effectiveness of reducing the number of candidate gene sets for the optimization problem by excluding genes based on the CoD.

Table 3.2. Percentage of finding the same experiment as the optimal method using the proposed approximate method with gene exclusion from the search space

	$n = 7, k = 2$					
	$p = 1$		$p = 2$		$p = 3$	
	Random	CoD	Random	CoD	Random	CoD
$s = 0$	92.59	92.59	87.72	87.72	82.90	82.90
$s = 1$	91.78	91.96	85.85	86.44	80.05	80.85
$s = 2$	90.23	91.07	83.38	84.77	75.97	77.96
$s = 3$	88.31	89.89	79.28	82.01	N/A	N/A
$s = 4$	85.18	87.24	N/A	N/A	N/A	N/A

	$n = 7, k = 3$					
	$p = 1$		$p = 2$		$p = 3$	
	Random	CoD	Random	CoD	Random	CoD
$s = 0$	90.84	90.84	84.19	84.19	77.87	77.87
$s = 1$	89.78	90.07	81.94	82.67	74.46	75.49
$s = 2$	88.19	89.08	78.67	80.42	69.56	71.58
$s = 3$	85.77	87.54	73.95	76.76	N/A	N/A
$s = 4$	82.15	84.79	N/A	N/A	N/A	N/A

	$n = 7, k = 4$					
	$p = 1$		$p = 2$		$p = 3$	
	Random	CoD	Random	CoD	Random	CoD
$s = 0$	90.72	90.72	83.56	83.56	76.17	76.17
$s = 1$	89.65	90.06	81.11	82.00	72.44	73.67
$s = 2$	87.95	89.08	77.71	79.58	67.10	69.80
$s = 3$	85.54	87.57	72.51	76.07	N/A	N/A
$s = 4$	81.81	84.51	N/A	N/A	N/A	N/A

Table 3.2 continued

| | $n = 7, k = 5$ | | | | | |
| | $p = 1$ | | $p = 2$ | | $p = 3$ | |
	Random	CoD	Random	CoD	Random	CoD
$s = 0$	91.28	91.28	83.71	83.71	76.46	76.46
$s = 1$	90.21	90.56	81.32	82.31	72.55	73.86
$s = 2$	88.60	89.63	78.17	79.96	67.09	69.72
$s = 3$	86.33	88.14	72.66	76.26	N/A	N/A
$s = 4$	82.64	85.58	N/A	N/A	N/A	N/A

| | $n = 8, k = 4$ | | | | | | | |
| | $p = 1$ | | $p = 2$ | | $p = 3$ | | $p = 4$ | |
	Random	CoD	Random	CoD	Random	CoD	Random	CoD
$s = 0$	92.43	92.43	86.98	86.98	80.90	80.90	74.97	74.97
$s = 1$	91.69	92.05	85.21	85.94	78.10	79.05	70.95	72.04
$s = 2$	90.59	91.40	82.53	84.13	74.53	76.29	65.81	67.94
$s = 3$	89.03	90.58	79.47	82.00	68.98	71.91	N/A	N/A
$s = 4$	86.83	88.85	74.51	77.93	N/A	N/A	N/A	N/A
$s = 5$	83.09	86.29	N/A	N/A	N/A	N/A	N/A	N/A

We have also evaluated the performance of the approximate experimental design method when a sequence of experiments is conducted. Suppose there are $k = 5$ uncertain regulations and we conduct five experiments to identify all unknown regulations. For each set of unknown regulations, at each step we utilize the experimental design to choose one of the possible experiments, conduct the chosen experiment, and measure the performance (undesirable probability mass after intervention) of the robust intervention obtained after the experiment on the underlying true network. Continuing, for the remaining uncertain
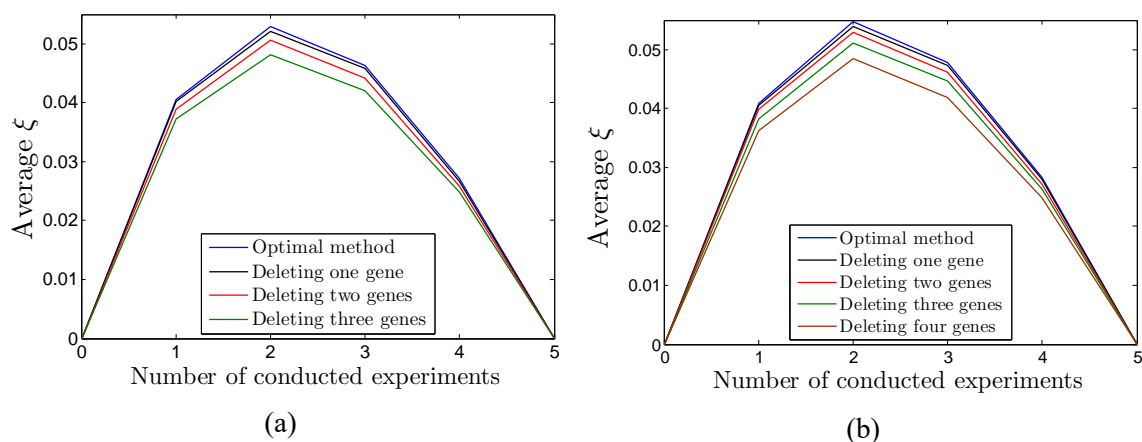
Figure 3.11. Performance comparison based on a sequence of experiments. The average gain for the optimal method and the proposed approximate method when deleting p genes are shown for $k = 5$ uncertain regulations. (a) Networks with $n = 7$ genes. (b) Networks with $n = 8$ genes.

regulations we use the experimental design method and repeat the previous procedure until there is no more unknown regulation remaining in the network. We also do sequential experiments randomly where at each step we choose an experiment randomly, measure the performance of its corresponding robust intervention, and again choose one experiment randomly among the remaining ones. The difference between the undesirable probability mass after applying the robust interventions corresponding to the randomly chosen experiment and the chosen experiment through experimental design at each step is the gain of conducting the chosen experiment at that step. Figure 3.11 shows the average gain over random selection for the optimal method and the approximate design method deleting up to three genes for 7-gene networks and up to four genes for 8-gene networks. The figure indicates that the approximate design method has reliable performance compared to the optimal method. Moreover, similar to the optimal method, the average gain increases
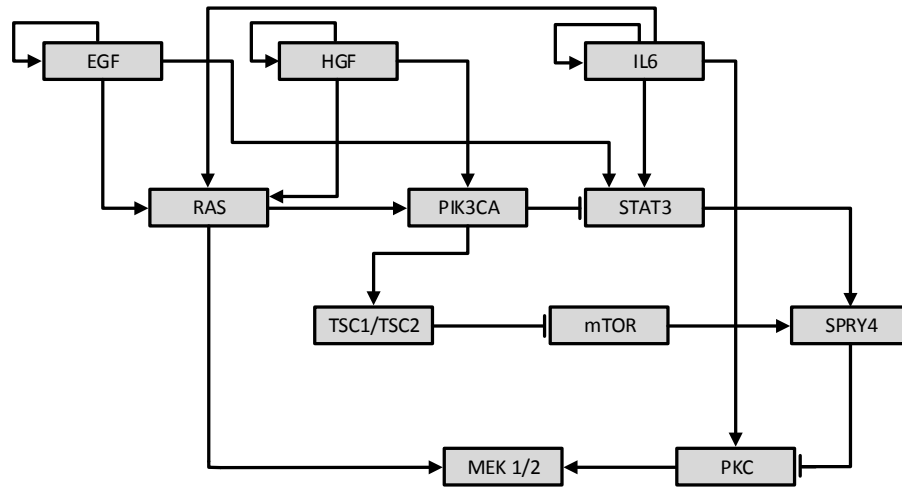
Figure 3.12. The regulatory relations among key genes involved in the colon cancer pathways. Normal arrows represent activating regulations and blunt arrows represent suppressive regulations.

sharply in the beginning for the approximate method. This is very important in real applications owing to the cost and time required for conducting experiments. Note that when we conduct all five experiments the average gain is zero because after five experiments the network is fully known and we can exactly calculate the optimal intervention regardless of the approach taken to choose experiments.

3.2.2 An Example Based on the Colon Cancer Pathway

In this section, we analyze the performance of the proposed experimental design method on the colon cancer pathways used in [88]. We focus on the pathways formed by 11 genes extracted from the complete pathway set, as used in [89]. These are shown in Figure 3.12: STAT3, RAS, IL6, HGF, PIK3CA, EGF, TSC1/TSC2, mTOR, SPRY4, PKC, and MEK 1/2. Normal and blunt arrows represent activating and suppressive regulations, respectively. We modeled the pathways as a BNp with perturbation probability 0.001.

Table 3.3. Performance of the proposed approximate method on the colon cancer pathways when deleting p genes and excluding s genes from the search space.

	$p = 3$		$p = 4$		$p = 5$	
	$s = 0$	s=2	s=0	s=2	s=0	s=2
$\bar{\xi}$	0.0239	0.0235	0.0231	0.0229	0.0206	0.0198

Genes are named as they have been introduced. For example, STAT3 is X_1 and MEK 1/2 is X_{11}.

EGF, HGF, and IL6 are three stimulation factors that carry the external signals gener-ated by neighboring cells to downstream genes and activate downstream pathways. Signal transducers and activators of transcription (STATs) constitute a family of transcription fac-tors that can be activated via extracellular signaling proteins such as cytokins and growth factors. These play a major role in regulating downstream processes such as cell growth, survival, and apoptosis [90]. STAT3 is an oncogene observed to be highly activated in many cancers, in particular, colon cancer [91, 92]. Hence, STAT3 has been recognized as a legitimate target for cancer therapy [93]. We considered states with up-regulated STAT3 ($X_1 = 1$) as undesirable states, so that the set of undesirables states is $U = \{1024, ..., 2047\}$. Before intervention the probability mass of undesirable states π_U is 0.5525. The optimal intervention for this network is transitioning state 11111110101 to state 01011001101; that is, $\tilde{\mathbf{F}}(11111110101) = 01011001101$ for the regulatory function after intervention. The undesirable probability mass after intervention $\tilde{\pi}_U$ is 0.3837.

To evaluate the proposed approximate method, we randomly selected 100 different sets of $k = 4$ regulations and assumed that they are uncertain, meaning that their regulatory

information is unknown. If experiments are chosen according to the optimal experimental design method, then $\bar{\bar{\xi}} = 0.0244$. Table 3.3 compares the average gain $\bar{\bar{\xi}}$ of the experiments chosen by the approximate method when deleting $p = 5$ genes and excluding s genes from the search space using the CoD-based algorithm. The table shows that we can obtain meaningful gain when the approximate experimental design method is used to select the experiment to be conducted first.

3.3 Discussion

We have proposed a computationally effective experimental design method for reducing uncertainty in gene regulatory networks. This method can effectively approximate the optimal experimental design method in the previous chapter, which is based on the mean objective cost of uncertainty (MOCU). To reduce computational complexity, we use network reduction to estimate the optimal and robust interventions needed for finding an optimal experiment. We introduced a novel cost function for gene deletion that takes into account the effect of gene deletion on the ranking of potential experiments. Because potential experiments are ranked based on the MOCU in the proposed objective-based experimental design framework, the proposed cost function is also based on the MOCU. Simulation results on both synthetic and real networks show that while our proposed method can greatly reduce computations, its performance is comparable to the optimal method and much better than random gene deletion. Greater computational reduction is achieved by excluding genes from the search space based on their CoD with the target gene whose

expression the intervention is aimed at altering.

We have assumed a uniform distribution over the uncertainty class. If one has relevant prior knowledge, perhaps it can be used to construct a distribution reflecting it. Care must be taken because concentrating the mass of the distribution in the wrong place can lead to poorer results. In Bayesian terminology, the distribution on the uncertainty class is called a *prior distribution*. Putting a non-uniform prior on Θ does not change the reduction procedure introduced here; however, some calculations are altered by including the weights. Prior construction is a difficult problem and has been considered in the context of gene regulation, but not in the context of network construction. Rather, pathway knowledge has been used to construct prior distributions governing uncertainty classes of feature-label distributions for optimal Bayesian classification [32]. Prior construction is particular to each application, examples being gene/protein signaling pathways in Gaussian and discrete phenotype classifications [89, 94]. Prior construction for uncertainty classes of the kind considered in this chapter constitutes an important issue for future study – and not just in relation to the specific problem considered herein.

4 OPTIMAL EXPERIMENTAL DESIGN IN THE CONTEXT OF CANONICAL EXPANSIONS: FILTERING AND SIGNAL DETECTION

This chapter addresses the following general experimental design problem: given a set of potential experiments, each of which can determine an unknown parameter in the signal model, find an experiment that results in the minimum expected remaining MOCU. This experiment should be the one conducted first. We have previously addressed this problem as it relates to gene regulatory networks, where the network topology is incomplete owing to missing parameters and the aim is to find the experiment to maximally reduce the MOCU.

Herein we address optimal parameter determination in the framework of random processes represented as canonical expansions. We simply note that parameters of the random process appear in the expansion, so that when the expansion is used for operator design, the uncertainty in these parameters affects the operator objective via the expansion. Hence, optimal experimental design can be approached in terms of the expansion. We will discuss canonical expansions in detail in the sequel. In this chapter, we will treat uncertainty quantification, robust operator design, and experimental design for two fundamental signal processing applications of canonical expansions: optimal linear filtering and signal detection.

The remainder of this chapter is organized as follows. Section 4.1 is devoted to a

briefly review of canonical expansions. We review the notion of effective characteristics as a means to robust filtering in Section 4.2. In Section 4.3, we apply the experimental design framework for the general case of robust filtering in the context of canonical expansions. We adopt the experimental design method to Wiener filtering in Section 4.4. Section 4.5 involves applying experimental design for signal detection. Finally, Section 4.6 contains concluding remarks.

4.1 Canonical Expansions

Many problems in engineering and physical science such as data compression, estimation and detection, and control involve random processes, and these can often be more readily solved with the process being expressed as a canonical expansion [95]. Canonical expansions are effective tools for studying random process when we are only concerned with the second-order statistics of the whole process. Consider a random process $X(t)$, indexed by the variable t. The canonical expansion for $X(t)$ is given as

$$X(t) = \mu_X(t) + \sum_{k=1}^{\infty} Z_k x_k(t), \tag{4.1}$$

where $\mu_X(t)$ is the mean function of $X(t)$, $x_k(t)$, $k = 1, 2, ...$ are deterministic functions called coordinate functions, and Z_k, $k = 1, 2, ...$ are uncorrelated zero-mean random variables called coefficients. Also, $Z_k x_k(t)$ are called elementary functions. The sequence of random variables Z_k can be regarded as a discrete white-noise process. Therefore, the sum in (4.1) is also called a discrete white-noise representation for the centered process $X(t) - \mu_X(t)$. Utilizing canonical expansions simplifies the problem of dealing with a set

of correlated random variables defined over a continuous domain t to the problem of dealing with a discrete sequence of uncorrelated random variables which is much easier.

4.1.1 Fourier Representation

The concept of canonical expansions in (4.1) is parallel to the idea of Fourier-series representations as both involve decomposition in terms of the orthogonal functions. Assume that the sequence of random variables $\{Z_i\}$ provides an orthonormal system, being a collection of uncorrelated random variables each having unit variance, for the vector space of zero-mean random variables. According to the projection theorem, a zero-mean random variable X with finite second moment can be decomposed as:

$$X = \sum_{i=1}^{\infty} \mathrm{E}\left[X\overline{Z_i}\right]Z_i,$$

(4.2)

where $\mathrm{E}\left[X\overline{Z_i}\right]$ is called the Fourier coefficient of X relative to Z_i. Now consider a random process $X(t)$ for a fixed time t, then the Fourier representation in terms of m random variables $Z_1, Z_2, ..., Z_m$ is given by:

$$X_m(t) = \sum_{i=1}^{m} \mathrm{E}\left[X(t)\overline{Z_i}\right]Z_i.$$

(4.3)

As t varies, Fourier coefficients $\hat{x}_i(t) = \mathrm{E}\left[X(t)\overline{Z_i}\right]$ can be regarded as deterministic functions of t. If for any t, the Fourier representation converges to $X(t)$ in the mean-square sense, i.e.,

$$\lim_{m\to\infty} \mathrm{E}\left[\left|X(t) - X_m(t)\right|^2\right] = 0$$

(4.4)

then in the spirit of (4.1), the canonical representation for $X(t)$ is

$$X(t) = \sum_{i=1}^{\infty} Z_i\hat{x}_i(t),$$

(4.5)

4.1.2 Karhunen-Loève Expansion

In 1947, Karhunen established the analytical theory of Karhunen-Loève (KL) expansion from the viewpoint of orthogonal functions [96]. Later in 1965, Pugachev put KL expansion into a more general canonical expansion form similar to the idea of Fourier representation [95].

For a zero-mean random process $X(t)$ defined over T with covariance function $K_X(t_1, t_2)$, if there is a weight function $w(t)$ such that

$$\int_T \int_T |K_X(t_1, t_2)|^2 w(t_1) w(t_2) dt_1 dt_2 < \infty, \tag{4.6}$$

then $X(t)$ can be represented in the form of KL canonical expression

$$X(t) = \sum_{i=1}^{\infty} Z_i u_i(t), \tag{4.7}$$

where u_i is the i-th eigenfunction of the covariance function,

$$\int_T K_X(t_1, t_2) u_i(t_1) w(t_2) dt_2 = \lambda_i u_i(t_1), \tag{4.8}$$

for $i = 1, 2, ...$, with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq 0$, and $\{Z_i\}, i = 1, 2, ...$, are the generalized Fourier coefficients of $X(t)$ relative to the set $\{u_i(t)\}$,

$$Z_i = \int_T X(t) \overline{u_i(t)} w(t) dt, \tag{4.9}$$

such that $Var[Z_i] = \lambda_i$. The set of eigenfunctions $\{u_i(t)\}$ forms an orthonormal system on T relative to the weight function $w(t)$:

$$\int_T u_i(t) \overline{u_j(t)} w(t) dt = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}. \tag{4.10}$$

For discrete random processes, the integral in (4.8) changes to a sum:

$$\sum_{n_2=1}^{\infty} K_X(n_1, n_2) u_i(n_2) w(n_2) = \lambda_i u_i(n_1), \tag{4.11}$$

for $i = 1, 2, ...$, if $w(n_i) = 1$ for all n_i and the random process $X(n)$ is defined over N points, then (4.11) takes matrix form

$$\mathbf{Ku} = \lambda \mathbf{u}, \tag{4.12}$$

where $\mathbf{K}$ is the covariance matrix for $X(n)$ and λ and $\mathbf{u} = \left[u(1), ..., u(n) \right]^T$ are the eigenvalue and eigenvector of the covariance matrix $\mathbf{K}$. The KL expansion is

$$X(n) = \sum_{i=1}^{N} Z_i u_i(n), \tag{4.13}$$

where $Z_i = \mathbf{X}^T \mathbf{u}_i$ and $\mathbf{u}_i$ is the eigenvector of the random process covariance matrix $\mathbf{K}$ with corresponding eigenvalue of λ_i. Eigenvectors $\mathbf{u}_i$ build an orthonormal system meaning that $< \mathbf{u}_i, \mathbf{u}_j >= 0$ if $i \neq j$ and $||\mathbf{u}_i|| = 1$.

As will be discussed later, KL expansion is particularly important for data compression. KL expansion can be used to achieve the optimal compression in terms of the mean-squared error.

4.1.3 Integral Canonical Expansion

Specifically, an *integral canonical expansion* of a random process $X(t)$ takes the form

$$X(t) = \mu_X(t) + \int_{\Xi} Z(\xi) x(t, \xi) \, d\xi, \tag{4.14}$$

where $\mu_X(t)$ is the mean of $X(t)$, $Z(\xi)$ is white noise over Ξ (the domain of ξ), and the coordinate functions $x(t, \xi)$ are deterministic. Referring to the integral canonical expansion

83

in (4.14), the covariance function of continuous white noise is the generalized function $K_Z(\xi, \xi') = I(\xi)\delta(\xi - \xi')$, where $I(\xi)$ is the intensity of the white noise, $\delta(\xi)$ is the Dirac delta function, and the theory of integral representation is interpreted in the generalized sense. $X(t)$ has covariance function

$$K_X(t,t') = \int_\Xi I(\xi)x(t,\xi)\overline{x(t',\xi)}\,d\xi, \tag{4.15}$$

and $x(t,\xi) = K_{XZ}(t,\xi)I(\xi)^{-1}$, where $K_{XZ}(t,\xi)$ is the cross-covariance between $X(t)$ and $Z(\xi)$. Integral canonical expansions are formed via a kernel $a(t,\xi)$ by defining

$$Z(\xi) = \int_T X(t)\overline{a(t,\xi)}\,dt. \tag{4.16}$$

Three conditions are necessary and sufficient for a canonical expansion to result [95, 97]:

$$x(t,\xi) = \frac{1}{I(\xi)}\int_T a(s,\xi)K_X(t,s)\,ds, \tag{4.17}$$

$$\int_T \overline{a(t,\xi)}x(t,\xi')\,dt = \delta\left(\xi - \xi'\right), \tag{4.18}$$

$$\int_\Xi x(t,\xi)\overline{a(t',\xi)}\,d\xi = \delta(t - t'). \tag{4.19}$$

The intensity of the white noise is

$$I(\xi) = \int_\Xi \int_T \int_T K_X(t,t')\overline{a(t,\xi)}a(t',\xi')\,dt\,dt'\,d\xi'. \tag{4.20}$$

For simplicity (while not affecting the theory in any consequential manner), we will assume $\mu_X(t) = 0$ and replace the covariance K_X by the auto-correlation R_X without loss of generality.

4.2 Effective Characteristics

The basic filtering problem involves jointly distributed observation and signal random processes, $(X(t), Y(s))$, $t \in T, s \in V$, with T and V being index sets. Optimal signal estimation involves estimating a signal $Y(s)$ at time s via a filter ψ given observations $\{X(t)\}_{t \in T}$. We write $\psi(X)(s)$ to emphasize that $\psi(X)$ is meant to estimate the signal $Y(s)$ at time s. Optimization is relative to a family of filters, $\mathcal{L}$, where a filter $\psi \in \mathcal{L}$ is a mapping $\psi : \mathcal{O} \to \mathbb{C}$ and $\mathcal{O}$ is the space of possible observed signals, each signal being a complex-valued function on T. Performance measurement is relative to a cost, $\eta(Y(s), \psi(X)(s))$, quantifying the cost or error in estimating signal $Y(s)$ by $\psi(X)(s)$. If an *optimal filter* exists for a fixed $s \in V$ (with finite error), then it can be expressed as

$$\widehat{\psi}(X)(s) = \arg\min_{\psi \in \mathcal{L}} \eta(Y(s), \psi(X)(s)), \tag{4.21}$$

where the minimum may be achieved by more than a single $\psi \in \mathcal{L}$. In the case of mean-squared error (MSE), $\eta(Y(s), \psi(X)(s)) = \mathrm{E}\left[\left|Y(s) - \psi(X)(s)\right|^2\right]$ and an optimal filter is referred to as a minimum-mean-square-error (MMSE) filter.

When the statistical model is not known with certainty, we assume that the joint process belongs to an *uncertainty class* of processes and the optimality is defined relative to the uncertainty class. Given an uncertainty class defined by the parameter set Θ, so that each $\theta \in \Theta$ corresponds to a distribution $F_\theta(x, y; t, s)$ in the uncertainty class, a prior distribution $f(\theta)$, a cost function η, and a filter class $\mathcal{L}$, an *intrinsically Bayesian robust*

(IBR) filter in $\mathscr{L}$ relative to $f(\theta)$ is defined by

$$\psi_{\text{IBR}}^{\Theta}(X)(s) = \underset{\psi \in \mathscr{L}}{\arg\min} \, \mathrm{E}_{\theta}\left[\eta\left(Y_{\theta}(s), \psi(X_{\theta})(s)\right)\right]. \qquad (4.22)$$

An IBR filter minimizes the expected cost over the uncertainty class. The adjective "intrinsic" refers to the fact that the $\arg\min$ is taken over all filters in $\mathscr{L}$, as opposed to constraining optimization to filters that are optimal for some $\theta \in \Theta$ [29, 34].

Filter error and optimality often depend on only certain characteristics, a *characteristic* being a deterministic function derived from the joint process. For instance, the MMSE linear filter depends on only second-order moments and can be represented via power spectra. It has been demonstrated how an IBR filter can be expressed in exactly the same closed-form as a model-specific optimal filter with the original characteristics replaced by "effective" characteristics [30].

An observation and signal pair, $\left(X(t), Y(s)\right)$, is *solvable under η and $\mathscr{L}$* if there exists $\widehat{\psi} \in \mathscr{L}$ minimizing $\eta\left(Y(s), \psi(X)(s)\right)$ over all $\psi \in \mathscr{L}$. An observation and signal pair $(X_{\text{eff}}(t), Y_{\text{eff}}(s))$ is an *effective process under η, $\mathscr{L}$ and Θ* if for all $\psi \in \mathscr{L}$, both $\mathrm{E}_{\theta}\left[\eta\left(Y_{\theta}(s), \psi(X_{\theta})(s)\right)\right]$ and $\eta\left(Y_{\text{eff}}(s), \psi(X_{\text{eff}})(s)\right)$ exist and

$$\mathrm{E}_{\theta}\left[\eta\left(Y_{\theta}(s), \psi(X_{\theta})(s)\right)\right] = \eta\left(Y_{\text{eff}}(s), \psi(X_{\text{eff}})(s)\right). \qquad (4.23)$$

Proposition 1. *If there exists a solvable effective process, $(X_{\text{eff}}(t), Y_{\text{eff}}(s))$, with the optimal filter $\widehat{\psi}_{\text{eff}}$, then $\psi_{\text{IBR}}^{\Theta} = \widehat{\psi}_{\text{eff}}$. The effective process may or may not belong to the uncertainty class, Θ, but should be solvable [30].*

IBR filters can be determined in terms of characteristics of the joint random process; in fact, we need only find effective characteristics, not necessarily an effective process.

Moreover, the error of a filter can often be expressed in the form $\mathscr{G}(\omega, \kappa)$, where ω refers to the characteristics of $\big(X(t), Y(s)\big)$ and κ refers to the filter parameters – for instance, in the case of linear filtering ω corresponds to the auto- and cross-correlation functions, and κ to the filter weighting function.

A class, Λ, of process pairs, $(X_\lambda(t), Y_\lambda(s))$, is *reducible under η and $\mathscr{L}$* if there exists a functional $\mathscr{G}$, called a *cost functional*, such that for each $\lambda \in \Lambda$ and $\psi \in \mathscr{L}$,

$$\eta\big(Y_\lambda(s), \psi(X_\lambda)(s)\big) = \mathscr{G}(\omega_\lambda, \kappa_\psi), \tag{4.24}$$

where ω_λ represents a collection of process characteristics relative to $\big(X_\lambda(t), Y_\lambda(s)\big)$ and κ_ψ represents parameters for filter ψ.

A collection of characteristics, ω, is *solvable in the weak sense under cost functional $\mathscr{G}$ and $\mathscr{L}$* if there exists a solution to

$$\widehat{\psi} = \arg\min_{\psi \in \mathscr{L}} \mathscr{G}(\omega, \kappa_\psi). \tag{4.25}$$

Given a set of characteristics, ω, which are solvable in the weak sense, there is an optimal filter, $\widehat{\psi}$ that possesses a functional, $\mathscr{G}(\omega, \kappa_{\widehat{\psi}})$. If Θ is contained in a reducible class, then the characteristic ω_{eff} is said to be an *effective characteristic in the weak sense under cost functional $\mathscr{G}$, $\mathscr{L}$, and Θ* if for all $\psi \in \mathscr{L}$, both $\mathrm{E}_\theta\big[\mathscr{G}(\omega_\theta, \kappa_\psi)\big]$ and $\mathscr{G}(\omega_{\text{eff}}, \kappa_\psi)$ exist and

$$\mathrm{E}_\theta\big[\mathscr{G}(\omega_\theta, \kappa_\psi)\big] = \mathscr{G}(\omega_{\text{eff}}, \kappa_\psi). \tag{4.26}$$

Proposition 2. *If Θ is contained in a reducible class and there exist weak-sense solvable weak-sense effective characteristics, ω_{eff}, with the optimal filter $\widehat{\psi}_{\text{eff}}$, then $\psi_{\text{IBR}}^\Theta = \widehat{\psi}_{\text{eff}}$ [30].*

If there exists an effective process providing the effective characteristics, we say that these are effective in the *strong sense*; otherwise, we say they are effective in the *weak sense*. For filter optimization, weak-sense effective characteristics are sufficient.

4.3 IBR Filtering in the Context of Canonical Expansions

This section reviews the general theory of IBR filtering in the framework of canonical expansions [30] and then introduces experimental design in this framework. In the next section, we apply the general integral-canonical theory to Wiener filtering, in which experimental design applies in the transformed domain.

4.3.1 Linear Filtering

Following [95, 97], optimal linear filtering can be addressed in the framework of canonical expansions. A linear filter takes the form

$$\psi(X)(s) = \int_T g(s,t)X(t)dt. \tag{4.27}$$

We assume that $X(t)$ and $Y(s)$ are zero-mean complex-valued random processes. Optimization involves finding a weighting function, $g(s,t)$, to minimize the MSE. If G is a linear function space on T and, for any $g(s,t) \in G$, the stochastic integral of (4.27) gives a random variable having a finite second moment, then $\widehat{g}(s,t)$ yields the optimal linear estimator of $Y(s)$ based on $X(t)$ if and only if it satisfies the Wiener-Hopf equation,

$$R_{YX}(s,t) = \int_T \widehat{g}(s,u)R_X(u,t)du, \tag{4.28}$$

where R_X is the auto-correlation function for $X(t)$ and R_{YX} is the cross-correlation function between $Y(s)$ and $X(t)$. If the optimal filter exists, then it can be shown via the Wiener-

Hopf equation that

$$\widehat{g}(s,t) = \int_{\Xi} \frac{\overline{a(t,\xi)}}{I(\xi)} R_{YZ}(s,\xi)\, d\xi, \tag{4.29}$$

where

$$R_{YZ}(s,\xi) = \int_{T} R_{YX}(s,u)a(u,\xi)du. \tag{4.30}$$

Plugging $\widehat{g}(s,t)$ in (4.27) and substituting (4.16) into (4.27) yields the MMSE estimate of

$Y(s)$ as:

$$\widehat{Y}(s) = \int_{\Xi} \frac{Z(\xi)R_{YZ}(s,\xi)}{I(\xi)}\, d\xi, \tag{4.31}$$

which possesses the MSE

$$\mathrm{E}\left[|Y(s) - \widehat{Y}(s)|^2\right] = R_Y(s,s) - \int_{\Xi} \frac{|R_{YZ}(s,\xi)|^2}{I(\xi)}\, d\xi. \tag{4.32}$$

4.3.2   IBR Linear Filtering

Now consider designing an IBR linear filter for an uncertainty class of signal models

$\{X_\theta, Y_\theta\}$, $\theta \in \Theta$, and suppose that the estimation is made at time s using observations

at $t \in T$. Let $R_{\Theta,Y}(s,s) = \mathrm{E}_\theta\left[R_{Y_\theta}(s,s)\right]$, $R_{\Theta,X}(t,u) = \mathrm{E}_\theta\left[R_{X_\theta}(t,u)\right]$, and $R_{\Theta,YX}(s,t) =$

$\mathrm{E}_\theta\left[R_{Y_\theta X_\theta}(s,t)\right]$ for all $s \in V$ and $t,u \in T$. It is straightforward to show that $R_{\Theta,X}(t,u)$ is

a valid auto-correlation function and therefore there exists a zero-mean Gaussian process

X_{eff} with the auto-correlation function $R_{\Theta,X}(t,u)$. Similar reasoning shows that there exists

a Gaussian process Y_{eff} with auto-correlation $R_{\Theta,Y}(s,s)$ at s and with the cross-correlation

$R_{\Theta,YX}(s,t)$ at s and all $t \in T$. In the robust model the error is given by

$$\mathrm{E}_\theta\left[\mathrm{E}\left[\left|Y_\theta(s) - \int_T g(s,t)X_\theta(t)dt\right|^2 \Big| \theta\right]\right] = \mathrm{E}_\theta\left[R_{Y_\theta}(s,s)\right] - \int_T \mathrm{E}_\theta\left[R_{Y_\theta X_\theta}(s,t)\right]\overline{g(s,t)}dt$$

$$- \int_T \overline{\mathrm{E}_\theta\left[R_{Y_\theta X_\theta}(s,t)\right]}g(s,t)dt$$

$$+ \int_T \int_T \mathrm{E}_\theta\left[R_{X_\theta}(t,u)\right]g(s,t)\overline{g(s,u)}dtdu$$

$$= R_{\Theta,Y}(s,s) - \int_T R_{\Theta,YX}(s,t)\overline{g(s,t)}dt$$

$$- \int_T \overline{R_{\Theta,YX}(s,t)}g(s,t)dt$$

$$+ \int_T \int_T R_{\Theta,X}(t,u)g(s,t)\overline{g(s,u)}dtdu. \quad (4.33)$$

Thus, (4.23) is satisfied and $(X_{\mathrm{eff}}, Y_{\mathrm{eff}})$ is an effective joint process. All previous equations for characteristics hold except that all characteristics are replaced by the effective characteristics $R_{\Theta,Y}$, $R_{\Theta,X}$, $R_{\Theta,YX}$.

In this effective setting, the three necessary and sufficient conditions for a canonical expansion take the form

$$x_\Theta(t,\xi) = \frac{1}{I_\Theta(\xi)} \int_T a(s,\xi)R_{\Theta,X}(t,s)\,ds, \quad (4.34)$$

$$\int_T \overline{a(t,\xi)}x_\Theta(t,\xi')\,dt = \delta\left(\xi - \xi'\right), \quad (4.35)$$

$$\int_\Xi x_\Theta(t,\xi)\overline{a(t',\xi)}\,d\xi = \delta(t-t'), \quad (4.36)$$

where the intensity of the white noise is given by

$$I_\Theta(\xi) = \int_\Xi \int_T \int_T R_{\Theta,X}(t,t')\overline{a(t,\xi)}a(t',\xi')\,dt\,dt'\,d\xi'. \quad (4.37)$$

If the three conditions hold and the Wiener-Hopf equation is satisfied for the effective

process, then the IBR filter can be found as

$$g_{\text{IBR}}^{\Theta}(s,t) = \int_{\Xi} \frac{\overline{a(t,\xi)}}{I_{\Theta}(\xi)} R_{\Theta,YZ}(s,\xi) \, d\xi, \tag{4.38}$$

where $R_{\Theta,YZ}(s,\xi) = \text{E}_{\theta}[R_{Y_{\theta}Z_{\theta}}(s,\xi)]$. The estimate obtained by applying the IBR filter to process X_{θ} is given by

$$\psi_{\text{IBR}}^{\Theta}(X_{\theta})(s) = \int_{T} g_{\text{IBR}}^{\Theta}(s,t) X_{\theta}(t) dt = \int_{\Xi} \frac{Z_{\theta}(\xi) R_{\Theta,YZ}(s,\xi)}{I_{\Theta}(\xi)} d\xi, \tag{4.39}$$

and the optimal average MSE can be computed using (4.32) and replacing $R_Y(s,s)$ and $R_{YZ}(s,\xi)$ with effective characteristics $R_{\Theta,Y}(s,s)$ and $R_{\Theta,YZ}(s,\xi)$, respectively.

4.3.3 Optimal Experimental Design

We consider an *uncertainty class* of models Θ parameterized by $\theta = \{\theta_1, ..., \theta_k\} \in \Theta$, a class Ψ of operators, and an associated cost $\eta_{\theta}(\psi)$. The optimal operator $\psi(\theta)$ for model θ and the IBR operator for the uncertainty class are obtained according to (1.2) and (1.4), respectively.

Assume that there are k experiments $T_1, ..., T_k$, where conducting each experiment T_i is equivalent to the exact determination of the uncertain parameter θ_i. The question that an effective experimental design should address is: Which experiment should be conducted first or equivalently how experiments should be ranked such that only experiments with high priority are conducted?

Following discussions in the previous chapters, the experiment T_{i^*} resulting in the minimum expected remaining MOCU should be conducted first:

$$i^* = \arg\min_{i \in 1,...,k} \text{M}_{\Psi}(\Theta; \theta_i). \tag{4.40}$$

91

This reduces to

$$i^* = \underset{i \in 1,\ldots,k}{\arg\min} \, \mathrm{E}_{\bar{\theta}_i} \left[\mathrm{E}_{\theta|\theta_i=\bar{\theta}_i} \left[\eta_{\theta|\theta_i=\bar{\theta}_i} \left(\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\bar{\theta}_i} \right) - \eta_{\theta|\theta_i=\bar{\theta}_i} \left(\psi(\theta|\theta_i = \bar{\theta}_i) \right) \right] \right]$$

$$= \underset{i \in 1,\ldots,k}{\arg\min} \, \left\{ \mathrm{E}_{\bar{\theta}_i} \left[\mathrm{E}_{\theta|\theta_i=\bar{\theta}_i} \left[\eta_{\theta|\theta_i=\bar{\theta}_i} \left(\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\bar{\theta}_i} \right) \right] \right] - \mathrm{E}_{\bar{\theta}_i} \left[\mathrm{E}_{\theta|\theta_i=\bar{\theta}_i} \left[\eta_{\theta|\theta_i=\bar{\theta}_i} \left(\psi(\theta|\theta_i = \bar{\theta}_i) \right) \right] \right] \right\}$$

$$= \underset{i \in 1,\ldots,k}{\arg\min} \, \left\{ \mathrm{E}_{\bar{\theta}_i} \left[\mathrm{E}_{\theta|\theta_i=\bar{\theta}_i} \left[\eta_{\theta|\theta_i=\bar{\theta}_i} \left(\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\bar{\theta}_i} \right) \right] \right] - \mathrm{E}_{\theta} \left[\eta_{\theta} \left(\psi(\theta) \right) \right] \right\}$$

$$= \underset{i \in 1,\ldots,k}{\arg\min} \, \mathrm{E}_{\bar{\theta}_i} \left[\mathrm{E}_{\theta|\theta_i=\bar{\theta}_i} \left[\eta_{\theta|\theta_i=\bar{\theta}_i} \left(\psi_{\mathrm{IBR}}^{\Theta|\theta_i=\bar{\theta}_i} \right) \right] \right], \tag{4.41}$$

where the third equality follows from the law of total expectation and the fourth equality is due to the independence of the second term inside the optimization expression from the variable of optimization. We shall refer to the value being minimized as the *experimental design value* and denote it by $\mathscr{D}(\theta_i)$. We shall refer to the parameter to be determined first according to (4.41) as the *primary parameter*. Regarding experimental design, the uncertain parameter θ_i to determine first (the primary parameter) is, based on (4.41), given by

$$i^* = \underset{i \in 1,\ldots,k}{\arg\min} \, \mathrm{E}_{\bar{\theta}_i} \left[\mathrm{E}_{\theta|\theta_i=\bar{\theta}_i} \left[\eta \left(Y_{\theta|\theta_i=\bar{\theta}_i}(s), \psi_{\mathrm{IBR}}^{\Theta|\theta_i=\bar{\theta}_i} \left(X_{\theta|\theta_i=\bar{\theta}_i} \right)(s) \right) \right] \right]$$

$$= \underset{i \in 1,\ldots,k}{\arg\min} \, \mathrm{E}_{\bar{\theta}_i} \left[R_{\Theta|\theta_i=\bar{\theta}_i,Y}(s,s) - \int_{\Xi} \frac{\left| R_{\Theta|\theta_i=\bar{\theta}_i,YZ}(s,\xi) \right|^2}{I_{\Theta|\theta_i=\bar{\theta}_i}(\xi)} \, d\xi \right]$$

$$= \underset{i \in 1,\ldots,k}{\arg\max} \, \mathrm{E}_{\bar{\theta}_i} \left[\int_{\Xi} \frac{\left| R_{\Theta|\theta_i=\bar{\theta}_i,YZ}(s,\xi) \right|^2}{I_{\Theta|\theta_i=\bar{\theta}_i}(\xi)} \, d\xi \right], \tag{4.42}$$

where $R_{\Theta|\theta_i=\bar{\theta}_i,YZ}$ and $R_{\Theta|\theta_i=\bar{\theta}_i,X}$ can be found by taking the conditional expectation given $\theta_i = \bar{\theta}_i$ and $I_{\Theta|\theta_i=\bar{\theta}_i}(\xi)$ is obtained using (4.37) with $R_{\Theta,X}$ replaced by $R_{\Theta|\theta_i=\bar{\theta}_i,X}$. We illus-

trate (4.42) with a discrete example. A discrete canonical representation takes the form

$$X(t) = \sum_{l=1}^{\infty} Z(l)x_l(t), \qquad (4.43)$$

where

$$Z(l) = \int_T X(t)a_l(t)dt. \qquad (4.44)$$

The three necessary and sufficient conditions for an integral canonical representation are replaced by L^2 orthogonality conditions, the integral representations are replaced by summations, and the white-noise intensity $I(\xi)$ is replaced by the variance of $Z(l)$ [97]. This all holds analogously for robust filter design, with all correlation functions replaced by the effective correlation functions.

Consider the following signal plus noise model parameterized by $\theta = (\theta_1, \theta_2)$:

$$X_\theta(t) = Y_{\theta_1}(t) + N_{\theta_2}(t), \qquad (4.45)$$

where $N_{\theta_2}(t)$ is a white noise process with intensity $\sigma_{\theta_2}^2$; θ_1 parameterizes some feature of $Y(t)$, such as phase, amplitude, frequency, etc.; and that θ_1 and θ_2 are statistically independent. Paralleling the analysis in [30], we obtain

$$R_{\Theta,X}(u,t) = R_{\Theta,Y}(u,t) + \mathrm{E}_{\theta_2}\left[\sigma_{\theta_2}^2\right]\delta(u-t), \qquad (4.46)$$

where $R_{\Theta,X}(u,t) = \mathrm{E}_\theta\left[R_{X_\theta}(u,t)\right]$, $R_{\Theta,Y}(u,t) = \mathrm{E}_{\theta_1}\left[R_{Y_{\theta_1}}(u,t)\right]$, and $\mathrm{E}_{\theta_2}[\sigma_{\theta_2}^2]$ are the effective characteristics. Since noise is uncorrelated, it can be seen that $R_{Y_\theta X_\theta}(s,u) = R_{Y_\theta}(s,u)$. Because we use a discrete canonical expansion, the integral over ξ is replaced by a summation. Also, $a_l(t) = x_l(t)$. Assume that λ_l^Θ and $x_l^\Theta(t)$ are the eigenvalues and the eigenfunctions of $R_{\Theta,Y}$, respectively. Substituting (4.46) in (4.37) and keeping in mind that the

canonical expansion is discrete, we can find I_l^Θ, the discrete form of $I_\Theta(\xi)$, as follows:

$$I_l^\Theta = \sum_{l'=1}^{\infty} \iint_T \left(R_{\Theta,Y}(t,t') + E_{\theta_2}\left[\sigma_{\theta_2}^2\right]\delta(t-t') \right)\overline{x_l^\Theta(t)}x_{l'}^\Theta(t')\,dt\,dt'$$

$$= \sum_{l'=1}^{\infty} \int_T \left(\int_T R_{\Theta,Y}(t,t')x_{l'}^\Theta(t')\,dt' \right)\overline{x_l^\Theta(t)}\,dt + E_{\theta_2}\left[\sigma_{\theta_2}^2\right]\sum_{l'=1}^{\infty}\int_T \left(\int_T \delta(t-t')x_{l'}^\Theta(t')\,dt' \right)\overline{x_l^\Theta(t)}\,dt$$

$$= \sum_{l'=1}^{\infty} \int_T \lambda_{l'}^\Theta x_{l'}^\Theta(t)\overline{x_l^\Theta(t)}\,dt + E_{\theta_2}\left[\sigma_{\theta_2}^2\right]\sum_{l'=1}^{\infty}\int_T x_{l'}^\Theta(t)\overline{x_l^\Theta(t)}\,dt$$

$$= \lambda_l^\Theta + E_{\theta_2}\left[\sigma_{\theta_2}^2\right], \tag{4.47}$$

where the third equality results because $x_{l'}^\Theta(t')$ is an eigenfunction of $R_{\Theta,Y}(t,t')$ and the set

of eigenfunctions $\{x_l^\Theta(t)\}$ forms an orthonormal system on T. Moreover,

$$R_{\Theta,YZ}(s,l) = E_\theta\left[R_{Y_\theta Z_\theta}(s,l)\right] = E_\theta\left[\int_T R_{Y_\theta X_\theta}(s,u)x_l^\theta(u)\,du\right]$$

$$= E_{\theta_1}\left[\int_T R_{Y_{\theta_1}}(s,u)x_l^{\theta_1}(u)\,du\right]$$

$$= E_{\theta_1}\left[\lambda_l^{\theta_1}x_l^{\theta_1}(s)\right], \tag{4.48}$$

where λ_l^θ and $x_l^\theta(t)$ are the eigenvalues and the eigenfunctions of R_{Y_θ}, respectively. Using

(4.39) and substituting (4.47) and (4.48), the IBR filter is given by

$$\psi_{\text{IBR}}^\Theta(X_\theta)(s) = \sum_{l=1}^{\infty} \frac{E_{\theta_1}\left[\lambda_l^{\theta_1}x_l^{\theta_1}(s)\right]}{\lambda_l^\Theta + E_{\theta_2}\left[\sigma_{\theta_2}^2\right]}Z_l^\theta, \tag{4.49}$$

and the expected MSE is

$$E_\theta\left[\eta\left(Y_{\theta_1}(s),\psi_{\text{IBR}}^\Theta(X_\theta)(s)\right)\right] = R_{\Theta,Y}(s,s) - \sum_{l=1}^{\infty} \frac{E_{\theta_1}^2\left[\lambda_l^{\theta_1}x_l^{\theta_1}(s)\right]}{\lambda_l^\Theta + E_{\theta_2}\left[\sigma_{\theta_2}^2\right]}. \tag{4.50}$$

According to (4.42), to determine which parameter, θ_1 or θ_2, should be determined

first, we compare

$$D(\theta_1) = E_{\theta_1}\left[\sum_{l=1}^{\infty} \frac{(\lambda_l^{\theta_1})^2}{\lambda_l^{\theta_1} + E_{\theta_2}\left[\sigma_{\theta_2}^2\right]}\left|x_l^{\theta_1}(s)\right|^2\right], \tag{4.51}$$

and:

$$D(\theta_2) = \mathrm{E}_{\theta_2}\left[\sum_{l=1}^{\infty} \frac{\mathrm{E}_{\theta_1}^2\left[\lambda_l^{\theta_1} x_l^{\theta_1}(s)\right]}{\lambda_l^{\Theta} + \sigma_{\theta_2}^2}\right]. \tag{4.52}$$

If (4.51) > (4.52), then θ_1 is primary; otherwise, θ_2 is primary.

4.4   Wiener Filtering

When processes $X(t)$ and $Y(s)$ are jointly wide-sense stationary (WSS), meaning that their second-order statistics are translation invariant, a closed-form solution for the optimal linear filter can be found. Let $r_X(\tau)$ and $S_X(\omega)$ denote the auto-correlation function and the power spectral density for $X(t)$, respectively, and $r_y(\tau)$ and $S_Y(\omega)$ denote the auto-correlation function and the power spectral density for $Y(t)$, respectively. The cross-correlation function $R_{YX}(t,t')$ is denoted by $r_{YX}(\tau)$, with Fourier transform $S_{YX}(\omega)$. Letting $T = (-\infty,\infty)$ and $a(t,\omega) = e^{j\omega t}$, (4.16) becomes

$$Z(\omega) = \int_{-\infty}^{\infty} X(t)e^{-j\omega t}dt. \tag{4.53}$$

The three necessary and sufficient conditions for a canonical expansion are satisfied and $X(t)$ possesses the integral canonical expansion

$$X(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} Z(\omega)e^{j\omega t}d\omega. \tag{4.54}$$

From (4.20), $Z(\omega)$ has intensity $I(\omega) = 2\pi S_X(\omega)$.

The Wiener-Hopf equation simplifies to

$$r_{YX}(\zeta) = \int_{-\infty}^{\infty} \widehat{g}(\zeta - \tau)r_X(\tau)d\tau. \tag{4.55}$$

Applying the Fourier transform $\mathscr{F}$ with $S_X(\omega) = \mathscr{F}\left[r_X(\tau)\right]$, $S_Y(\omega) = \mathscr{F}\left[r_Y(\tau)\right]$, and

$S_{YX}(\omega) = \mathscr{F}\left[r_{YX}(\tau)\right]$ yields

$$\widehat{G}(\omega) = \frac{S_{YX}(\omega)}{S_X(\omega)}, \tag{4.56}$$

where $\widehat{G}(\omega) = \mathscr{F}\left[\widehat{g}(\tau)\right]$ is the optimal linear (Wiener) filter that is spatially invariant and is called Wiener filter. The MSE of the Wiener filter is:

$$\mathrm{E}\left[\left|Y(s) - \widehat{Y}(s)\right|^2\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S_Y(\omega)S_X(\omega) - \left|S_{YX}(\omega)\right|^2}{S_X(\omega)}\, d\omega. \tag{4.57}$$

With uncertainty, the effective power spectra are $S_{\Theta,X}(\omega) = \mathscr{F}\left[\mathrm{E}_\theta\left[r_{X_\theta}(\tau)\right]\right]$ and $S_{\Theta,YX}(\omega) = \mathscr{F}\left[\mathrm{E}_\theta\left[r_{Y_\theta,X_\theta}(\tau)\right]\right]$. The IBR Wiener filter $G_{\mathrm{IBR}}^{\Theta}(\omega)$ is found by plugging $S_{\Theta,X}(\omega)$ and $S_{\Theta,YX}(\omega)$ in (4.56) [30].

Keeping in mind our aim is to find out which unknown parameter should be estimated first, we rewrite (4.30) as

$$r_{YZ}(s - \omega) = \int_{-\infty}^{\infty} r_{YX}(s - u)e^{j\omega u}\,du = e^{j\omega s}\int_{-\infty}^{\infty} r_{YX}(\tau)e^{-j\omega\tau}\,d\tau = e^{j\omega s}S_{YX}(\omega). \tag{4.58}$$

Therefore, $\left|r_{YZ}(s - \omega)\right|^2 = \left|S_{YX}(\omega)\right|^2$.

Substitution into (4.42) with the corresponding notation changes dictates the primary parameter:

$$\begin{aligned}
i^* &= \arg\max_{i\in 1,\ldots,k} \mathrm{E}_{\bar{\theta}_i}\left[\int_{-\infty}^{\infty} \frac{\left|r_{\Theta|\theta_i=\bar{\theta}_i,YZ}(s - \omega)\right|^2}{\mathrm{E}_{\theta|\theta_i=\bar{\theta}_i}\left[I_{\theta|\theta_i=\bar{\theta}_i}(\xi)\right]}\, d\omega\right] \\
&= \arg\max_{i\in 1,\ldots,k} \mathrm{E}_{\bar{\theta}_i}\left[\int_{-\infty}^{\infty} \frac{\left|S_{\Theta|\theta_i=\bar{\theta}_i,YX}(\omega)\right|^2}{S_{\Theta|\theta_i=\bar{\theta}_i,X}(\omega)}\, d\omega\right],
\end{aligned} \tag{4.59}$$

where $S_{\Theta|\theta_i=\bar{\theta}_i,X}(\omega) = \mathscr{F}\left[r_{\Theta|\theta_i=\bar{\theta}_i,X}(\tau)\right]$ and $S_{\Theta|\theta_i=\bar{\theta}_i,YX} = \mathscr{F}\left[r_{\Theta|\theta_i=\bar{\theta}_i,YX}(\tau)\right]$.

4.4.1 Blurring and Additive Noise

We consider reconstruction of the signal process when it is blurred with another random process $H_\theta(t)$ and corrupted by an additive noise process $N_\theta(t)$:

$$X_\theta(t) = \int_{-\infty}^{\infty} H_\theta(\tau) Y_\theta(t-\tau) \, d\tau + N_\theta(t), \qquad (4.60)$$

where $N_\theta(t)$ is assumed to be uncorrelated to $Y_\theta(t)$ and $H_\theta(t)$ and is white. For WSS $Y_\theta(t)$ and $N_\theta(t)$, the Wiener filter for a fixed θ is given by

$$\widehat{G}(\omega) = \frac{\overline{\mathscr{H}_\theta(\omega)} S_{Y_\theta}(\omega)}{|\mathscr{H}_\theta(\omega)|^2 S_{Y_\theta}(\omega) + S_{N_\theta}(\omega)}, \qquad (4.61)$$

where $\mathscr{H}_\theta(\omega) = \mathscr{F}[H_\theta(t)]$ and $S_{N_\theta}(\omega)$ is the spectral density function of $N_\theta(t)$. Based on (4.57) the MSE is

$$\mathrm{E}\left[|Y_\theta(s) - \widehat{Y}_\theta(s)|^2 \right] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_{Y_\theta} S_{N_\theta}}{|\mathscr{H}_\theta(\omega)|^2 S_{Y_\theta} + S_{N_\theta}} \, d\omega. \qquad (4.62)$$

In terms of effective characteristics, the IBR Wiener filter is given by [30]

$$G_\Theta^{IBR}(\omega) = \frac{S_{\Theta,YX}(\omega)}{S_{\Theta,X}(\omega)}, \qquad (4.63)$$

where

$$S_{\Theta,YX}(\omega) = \overline{\mathrm{E}_\theta\left[\mathscr{H}_\theta(\omega)\right]} S_{\Theta,Y}(\omega),$$

$$S_{\Theta,X}(\omega) = \mathrm{E}_\theta\left[|\mathscr{H}_\theta(\omega)|^2\right] S_{\Theta,Y}(\omega) + S_{\Theta,N}(\omega), \qquad (4.64)$$

and $S_{\Theta,N}(\omega) = \mathscr{F}\left[r_{\Theta,N}(\tau)\right]$ and $S_{\Theta,Y}(\omega) = \mathscr{F}\left[r_{\Theta,Y}(\tau)\right]$. The expected MSE of the Wiener filter is:

$$\mathrm{E}_\theta\left[\eta\left(Y_\theta(s), \psi_{IBR}^\Theta(X_\theta)(s)\right)\right] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{J(\omega)|S_{\Theta,Y}(\omega)|^2 + S_{\Theta,Y}(\omega) S_{\Theta,N(\omega)}}{\mathrm{E}_\theta\left[|\mathscr{H}_\theta(\omega)|^2\right] S_{\Theta,Y}(\omega) + S_{\Theta,N}(\omega)} \, d\omega, \qquad (4.65)$$

where $J(\omega) = \mathrm{E}_\theta\left[|\mathscr{H}_\theta(\omega)|^2\right] - \left|\mathrm{E}_\theta\left[\mathscr{H}_\theta(\omega)\right]\right|^2$.

4.4.2 Gaussian Blurring and Additive White Noise for a WSS Process

Suppose the blurring function $h(n)$ has Gaussian form $h_{\sigma_h^2}(n) = (2\pi\sigma_h^2)^{-\frac{1}{2}}\exp(-\frac{n^2}{2\sigma_h^2})$, the power spectral density for the noise process $N(n)$ is σ_n^2, and the model for the underlying signal process is $Y(n) = 2ku(n) - ku(n+1)$, where $u(n) \sim N(0,1)$. Suppose that $\theta = (\sigma_h^2, \sigma_n^2)$ is unknown and we want to find out which parameter, σ_h^2 or σ_n^2, should be determined first in the following discrete signal observation model:

$$X_\theta(n) = h_{\sigma_h^2}(n) * Y(n) + N_{\sigma_n^2}(n). \tag{4.66}$$

Let $f(\theta) = f(\sigma_h^2)f(\sigma_n^2)$ denote the prior distribution function for θ, where $f(\sigma_h^2)$ and $f(\sigma_n^2)$ denote the marginal priors for σ_h^2 and σ_n^2, respectively.

In order to evaluate the experimental design value in (4.59) for σ_h^2 we use the equation

$$
\begin{aligned}
D(\sigma_h^2) &= \mathrm{E}_{\sigma_h^2}\left[\int_{-\infty}^{\infty} \frac{\left|S_{\Theta|\sigma_h^2,YX}(\omega)\right|^2}{S_{\Theta|\sigma_h^2,X}(\omega)}\,d\omega\right] \\
&= \int\left[\int_{-\infty}^{\infty} \frac{\left|\mathcal{H}_{\sigma_h^2}(\omega)\right|^2 \left|S_Y(\omega)\right|^2}{\left|\mathcal{H}_{\sigma_h^2}(\omega)\right|^2 S_Y(\omega) + \mathrm{E}_{\sigma_n^2}[\sigma_n^2]}\,d\omega\right] f(\sigma_h^2)\,d\sigma_h^2,
\end{aligned}
\tag{4.67}
$$

where $S_Y(\omega)$ can be found using realizations for $Y(n)$. Similarly, the experimental design value according to (4.59) for σ_n^2 is found from

$$
\begin{aligned}
D(\sigma_n^2) &= \mathrm{E}_{\sigma_n^2}\left[\int_{-\infty}^{\infty} \frac{\left|S_{\Theta|\sigma_n^2,YX}(\omega)\right|^2}{S_{\Theta|\sigma_n^2,X}(\omega)}\,d\omega\right] \\
&= \int\left[\int_{-\infty}^{\infty} \frac{\left|\mathrm{E}_{\sigma_h^2}\left[\mathcal{H}_{\sigma_h^2}(\omega)\right]\right|^2 \left|S_Y(\omega)\right|^2}{\mathrm{E}_{\sigma_h^2}\left[\left|\mathcal{H}_{\sigma_h^2}(\omega)\right|^2\right] S_Y(\omega) + \sigma_n^2}\,d\omega\right] f(\sigma_n^2)\,d\sigma_n^2.
\end{aligned}
\tag{4.68}
$$
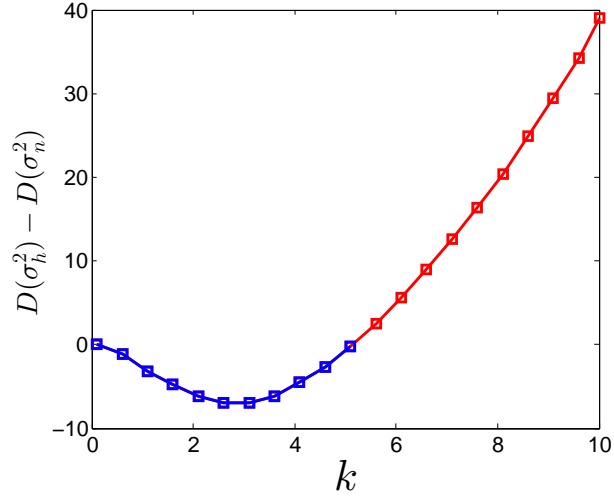
Figure 4.1. The difference between the experimental design values corresponding to the determination of unknown parameters σ_n^2 and σ_h^2 for different k.

The primary parameter is chosen according to

$$\theta_{i^*} = \begin{cases} \sigma_h^2 & \text{if } (4.67) > (4.68) \\ \\ \sigma_n^2 & \text{otherwise} \end{cases} . \qquad (4.69)$$

Assuming that $\sigma_n^2 \in [0,3]$ and $\sigma_h^2 \in [0.5,2]$, Figure 4.1 shows the difference between the experimental design values computed for σ_n^2 and σ_h^2 using equations (4.68) and (4.67), respectively, for different values of k, which sets the amplitude of the process $Y(n)$. In the figure, when the difference is negative, meaning that $D(\sigma_h^2) < D(\sigma_n^2)$, the curve is shown in blue and otherwise it is shown in red. When the signal has low amplitude, σ_n^2 is primary, but as k gets larger, the blurring parameter σ_h^2 becomes primary. This makes sense because we know that, whereas for low-amplitude processes the additive noise is more important, for high-amplitude signals the blurring function plays a major role in signal reconstruction.

Figure 4.2 shows the performance of the IBR Wiener filter designed after determining each uncertain parameter σ_h^2 and σ_n^2 over the uncertainty class for each possible pair
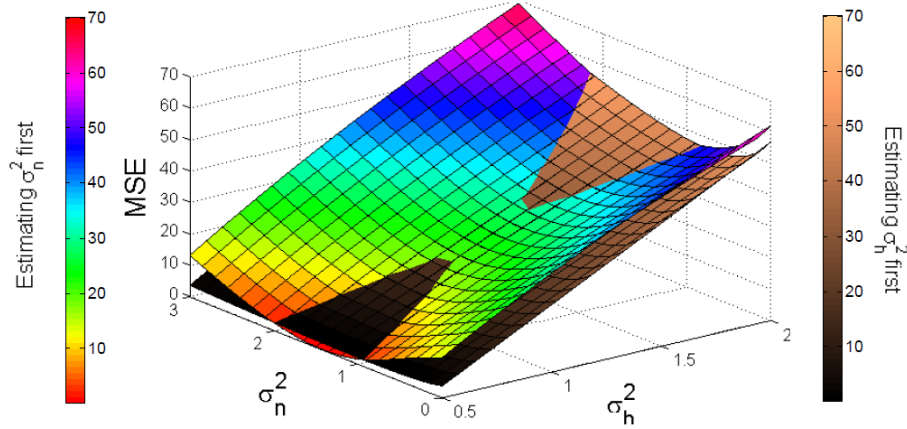
Figure 4.2. The MSE of the IBR filter obtained after determining one unknown parameter over the uncertainty class for k = 8 when σ_h^2 or σ_n^2 is determined first.

(σ_h^2, σ_n^2) with $k = 8$, i.e., $Y(n) = 16u(n) - 8u(n+1)$. Keep in mind that, although the experimental design procedure selects σ_h^2 as primary, for some states the designed IBR filter may perform better by determining σ_n^2 first. For example for $1 \leq \sigma_n^2 \leq 2$ and $0.5 \leq \sigma_h^2 \leq 1$ determining σ_n^2 results in the lower MSE. The point is that the designed IBR filter obtained upon determining σ_h^2 first performs better on average with respect to the uncertainty class.

4.4.3 Gaussian Blurring and Additive White Noise for a Random Phase Signal with Unknown Parameters

A random phase process is of the form $Y(n;A,f_c) = A\cos(2\pi f_c n + \Phi)$, where the amplitude A and frequency f_c are fixed and the phase random variable Φ is uniformly distributed over the interval $[0, 2\pi)$. A random phase signal is WSS. The power spectral density $S_Y(\omega;A,f_c)$ of $Y(n;A,f_c)$ with N samples is $\frac{A^2 N}{4}\delta(f \pm f_c)$. For this signal process, we assume the signal observation model

$$X(n) = h_{\sigma_h^2}(n) * Y(n;A,f_c) + N_{\sigma_n^2}(n). \tag{4.70}$$

100

Suppose $\theta = [A, f_c, \sigma_h^2, \sigma_n^2]$ is unknown and the intent is to find the parameter to determine first. We use (4.59) for each parameter and then determine the parameter having the maximum experimental design value $\mathscr{D}$.

$\mathscr{D}(\sigma_h^2)$ is found using (4.67), where $S_Y(\omega)$ is replaced by $S_{\Theta|\sigma_h^2,Y}(\omega)$ which, on account of the independence of A and f_c from σ_h^2, can be obtained as follows:

$$S_{\Theta|\sigma_h^2,Y}(\omega) = \mathscr{F}\left[E_{\theta|\sigma_h^2}\left[r_{Y_{\theta|\sigma_h^2}}(\tau;A,f_c)\right]\right] = \mathscr{F}\left[E_{A,f_c}\left[r_Y(\tau;A,f_c)\right]\right]$$

$$= E_{A,f_c}\left[\mathscr{F}\left[r_Y(\tau;A,f_c)\right]\right] = E_{A,f_c}\left[S_Y(\omega;A,f_c)\right].$$

(4.71)

Note that interchanging the Fourier transform and expectation integrals is assumed to be justified.

We use (4.68) to calculate $\mathscr{D}(\sigma_n^2)$, where $S_{\Theta|\sigma_n^2,Y}(\omega)$ is similarly found by using (4.71). For the amplitude,

$$\mathscr{D}(A) = E_A\left[\int_{-\infty}^{\infty} \frac{\left|S_{\Theta|A,YX}(\omega)\right|^2}{S_{\Theta|A,X}(\omega)}\,d\omega\right]$$

$$= \int\left[\int_{-\infty}^{\infty} \frac{\left|E_{\sigma_h^2}\left[\mathscr{H}_{\sigma_h^2}(\omega)\right]\right|^2 \left|S_{\Theta|A,Y}(\omega)\right|^2}{E_{\sigma_h^2}\left[\left|\mathscr{H}_{\sigma_h^2}(\omega)\right|^2\right]S_{\Theta|A,Y}(\omega) + E_{\sigma_n^2}\left[\sigma_n^2\right]}\,d\omega\right]f(A)dA,$$

(4.72)

where

$$S_{\Theta|A,Y}(\omega) = \mathscr{F}\left[E_{\theta|A}\left[r_{Y_{\theta|A}}(\tau;A,f_c)\right]\right] = \mathscr{F}\left[E_{f_c}\left[r_Y(\tau;A,f_c)\right]\right]$$

$$= E_{f_c}\left[\mathscr{F}\left[r_Y(\tau;A,f_c)\right]\right] = E_{f_c}\left[S_Y(\omega;A,f_c)\right].$$

(4.73)

Finally,

$$\mathcal{D}(f_c) = \mathrm{E}_{f_c}\left[\int_{-\infty}^{\infty} \frac{\left|S_{\Theta|f_c,YX}(\omega)\right|^2}{S_{\Theta|f_c,X}(\omega)} d\omega\right]$$

$$= \int \left[\int_{-\infty}^{\infty} \frac{\left|\mathrm{E}_{\sigma_h^2}\left[\mathscr{H}_{\sigma_h^2}(\omega)\right]\right|^2 \left|S_{\Theta|f_c,Y}(\omega)\right|^2}{\mathrm{E}_{\sigma_h^2}\left[\left|\mathscr{H}_{\sigma_h^2}(\omega)\right|^2\right] S_{\Theta|f_c,Y}(\omega) + \mathrm{E}_{\sigma_n^2}\left[\sigma_n^2\right]} d\omega\right] f(f_c) df_c. \qquad (4.74)$$

The parameter with maximum experimental design value is determined first. Note that $S_{\Theta|f_c,X}(\omega)$ can be found similarly to (4.73).

To analyze experimental-design performance, we assign some intervals to the uncertain parameters as follows: $\sigma_n^2 \in [0.1, \sigma_{n\max}^2]$, $\sigma_h^2 \in [0.5, \sigma_{h\max}^2]$, $A \in [5, A_{\max}]$, and $f_c \in [0.1, f_{c\max}]$. The nominal values for $\sigma_{n\max}^2$, $\sigma_{h\max}^2$, $A_{\max}$, and $f_{c\max}$ are 1, 4, 10, and 0.15, respectively. Figure 4.3 shows the uncertain parameter to be determined first. The experimental design values for σ_n^2, σ_h^2, A, and f_c are computed using equations (4.68), (4.67), (4.72), and (4.74), respectively. For example, in Figure 4.3 (a), we consider the uncertainty interval of σ_n^2 as $[0.1, \sigma_{n\max}^2]$, $0.5 \leq \sigma_{n\max}^2 \leq 8$. When the interval is small, σ_h^2 is primary, but as the interval gets larger, f_c becomes primary. In Figure 4.3 (d), when the interval of f_c is small, σ_h^2 is primary, but for large uncertainty intervals of f_c, the primary parameter is f_c. Generally, we observe that for different intervals of uncertain parameters, the primary parameter is either frequency f_c or the blurring function parameter σ_h^2.

We now consider experimental-design performance when a sequence of experiments is conducted. For determining all unknown parameters in the signal observation model (4.70), we need to conduct four experiments. For the first experiment, we select the primary parameter using the prior distributions for the parameters. When the first experiment is
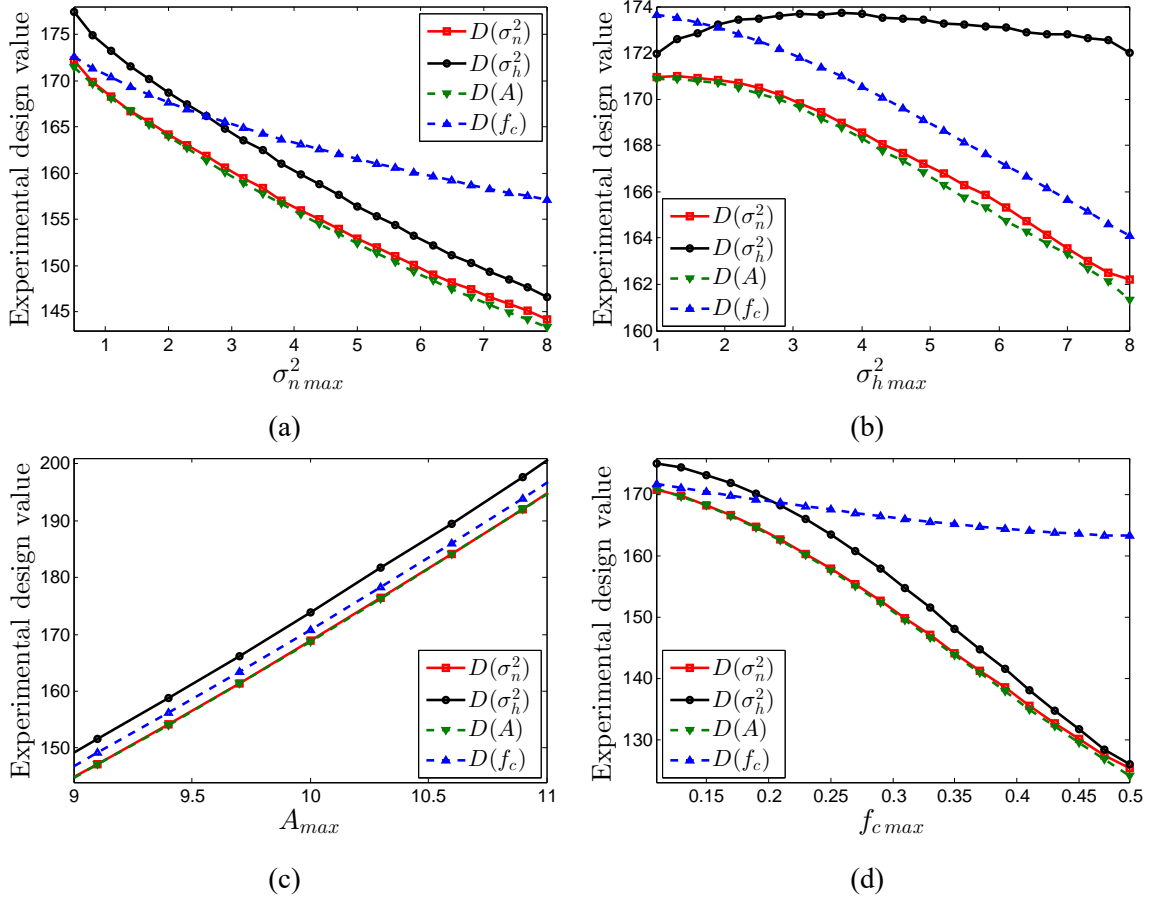
Figure 4.3. Prioritizing the determination of uncertain parameters when the interval of one of the uncertain parameters changes and the other intervals are according to the nominal intervals. The parameter with the maximum experimental design value is the primary parameter. (a) The interval of σ_n^2 changes as $[0.1, \sigma_{n\max}^2]$. (b) The interval of σ_h^2 changes as $[0.1, \sigma_{h\max}^2]$. (c) The interval of A changes as $[5, A_{\max}]$. (d) The interval of f_c changes as $[0.1, f_{c\max}]$.

done, we put the true value of the determined parameter in the signal observation model.

Then, using the updated signal observation model, which has fewer unknown parameters,

the primary parameter among the remaining unknown parameters is found. The procedure

is repeated until all parameters are determined.

To evaluate the performance of the selected experiment at each step, after performing
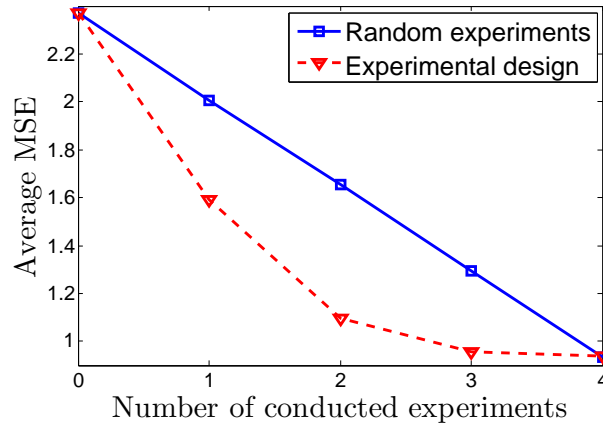
Figure 4.4. The average MSE obtained after performing each experiment in a sequence of experiments for the signal observation model with four unknown parameters. Results are shown when experiments are chosen randomly or based on the proposed experimental design method.

the experiment of interest and incorporating the value of the corresponding parameter in the model (4.70), we find the IBR Wiener filter for the new uncertainty class and calculate its MSE relative to the underlying true model. For simulations, we assume nominal intervals for the uncertain parameters as we had considered for the single experiment case and report the average MSE over 10000 different assumed true models. Figure 4.4 shows the average MSE after conducting different numbers of experiments both when they are chosen randomly and when they are based on experimental design. According to the figure, experimental design achieves much faster decrease in average MSE than random selection. Note that both curves in the figure begin from the same point and reach the same point because initially no experiment has been done and at the end all experiments have been done (and the true model is found regardless of the order of the experiments).

104

4.5   Signal Detection

In this section, we consider signal detection via the Karhunen-Loève canonical expansion. We develop robust signal detection in the presence of uncertainty and show how experimental design can be applied in signal detection when canonical expansions are used. Here optimization will be relative to the criterion of *Maximum A Posteriori* (MAP) probability rather than MSE.

4.5.1   Signal Detection via Karhunen-Loève Expansion

The main concern of signal detection is to analyze a received signal and extract relevant information [98, 99]. In Gaussian signal detection, it is assumed that the processes are Gaussian. We consider optimal signal detection algorithm under the maximum *a posteriori* (MAP) criterion [100, 101] in the framework of the Karhunen-Loève expansion [98]. First we briefly review some essentials on how the signal is estimated via the MAP criterion as given in [98]. Then we compute the maximum *a posteriori* probability obtained by the MAP estimate and subsequently define and compute the MOCU for signal detection. We also show that the MAP estimate of the signal when the noise variance is unknown can be found in a similar manner as when the variance is known.

Following [98], (which can be consulted for more details) suppose signal $X(t)$ has been received during the time interval $[0, T]$ and is of the form

$$X(t) = e[t, Y(t)] + N(t), \qquad (4.75)$$

where $Y(t)$ is the signal, $e[t, Y(t)]$ (a deterministic function of $Y(t)$ and t is the representative

of the modulation scheme, and $N(t)$ is Gaussian noise. Let $R_Y(t, \tau)$, with eigenfunctions $u_i(t)$ and eigenvalues λ_i, be the auto-correlation function for $Y(t)$; and $R_N(t, \tau)$, with eigenfunctions $v_i(t)$ and eigenvalues μ_i, be the auto-correlation function for $N(t)$. Also suppose that Y_i and N_i are the KL expansion coefficients of $Y(t)$ and $N(t)$, respectively:

$$Y(t) = \sum_{i=1}^{\infty} Y_i u_i(t), \tag{4.76}$$

$$N(t) = \sum_{i=1}^{\infty} N_i v_i(t). \tag{4.77}$$

Utilizing the KL expansion, the problem of estimating $Y(t)$ reduces to estimating the corresponding KL coefficients Y_i.

As the coefficients of the KL expansion are independent Gaussian random variables, $\mathbf{Y}_{1:M} = \{Y_1, Y_2, ..., Y_M\}$ and $\mathbf{N}_{1:M} = \{N_1, N_2, ..., N_M\}$ for a fixed M have the following joint probability density functions:

$$f_Y(\mathbf{Y}_{1:M}) = \frac{1}{\sqrt{2\pi}^M \sqrt{\prod_{i=1}^{M} \lambda_i}} \exp\left(-\sum_{i=1}^{M} \frac{Y_i^2}{2\lambda_i}\right), \tag{4.78}$$

$$f_N(\mathbf{N}_{1:M}) = \frac{1}{\sqrt{2\pi}^M \sqrt{\prod_{i=1}^{M} \mu_i}} \exp\left(-\sum_{i=1}^{M} \frac{N_i^2}{2\mu_i}\right), \tag{4.79}$$

If $X_1, X_2, ..., X_M$ are the corresponding KL coefficients of the received signal $X(t)$ and $\mathbf{X}_{1:M} = \{X_1, X_2, ..., X_M\}$, then to calculate the posterior probability of the coefficients Y_i, first we obtain the conditional probability distribution function $f_X(\mathbf{X}_{1:M}|\mathbf{Y}_{1:M})$ as

$$f_X(\mathbf{X}_{1:M}|\mathbf{Y}_{1:M}) = f_N(\mathbf{B}_{1:M}), \tag{4.80}$$

where $\mathbf{B}_{1:M} = \{B_1, B_2, ..., B_M\}$ and

$$B_i = \int_T \left(X(t) - e[t, Y(t)]\right) v_i(t) \, dt. \tag{4.81}$$

The MAP estimate $\hat{\mathbf{Y}}_{1:M}$ of $\mathbf{Y}_{1:M}$ is

$$\hat{\mathbf{Y}}_{1:M} = \arg\max_{\mathbf{Y}_{1:M}} \log\left(f_Y(\mathbf{Y}_{1:M}|\mathbf{X}_{1:M})\right) = \arg\max_{\mathbf{Y}_{1:M}} \log\left(f_Y(\mathbf{Y}_{1:M})f_X(\mathbf{X}_{1:M}|\mathbf{Y}_{1:M})\right) \quad (4.82)$$

$$= \arg\max_{\mathbf{Y}_{1:M}} \left\{ -\sum_{j=1}^{M} \frac{Y_j^2}{\lambda_j} - \sum_{j=1}^{M} \frac{B_j^2}{\mu_j} \right\}. \quad (4.83)$$

To find $\hat{\mathbf{Y}}_{1:M}$, for $i = 1, 2, ...M$, $\frac{\partial}{\partial Y_i}$ is found for the expression inside the optimization in (4.83) and set to 0 to obtain

$$\hat{Y}_i = -\lambda_i \sum_{j=1}^{M} \frac{B_j}{\mu_j} \frac{\partial B_j}{\partial Y_i}\bigg|_{Y_i=\hat{Y}_i}, \quad (4.84)$$

where

$$\frac{\partial B_j}{\partial Y_i}\bigg|_{Y_i=\hat{Y}_i} = -\int_T \frac{\partial e[t,\hat{Y}(t)]}{\partial Y(t)} u_i(t)v_j(t) \, dt. \quad (4.85)$$

Substituting (4.85) in (4.84), the MAP estimate $\hat{Y}_i$ of Y_i is:

$$\hat{Y}_i = \lambda_i \sum_{j=1}^{M} \frac{B_j}{\mu_j} \int_T \frac{\partial e[t,\hat{Y}(t)]}{\partial Y(t)} v_j(t)u_i(t) \, dt. \quad (4.86)$$

Letting $M \to \infty$ yields the MAP estimate of $Y(t)$:

$$\hat{Y}(t) = \sum_{i=1}^{\infty} \hat{Y}_i u_i(t) = \sum_{i=1}^{\infty} \lambda_i \sum_{j=1}^{\infty} \frac{B_j}{\mu_j} \int_T \frac{\partial e[\tau,\hat{Y}(\tau)]}{\partial Y(\tau)} v_j(\tau)u_i(\tau)u_i(t) \, d\tau$$

$$= \sum_{j=1}^{\infty} \frac{B_j}{\mu_j} \int_T \frac{\partial e[\tau,\hat{Y}(\tau)]}{\partial Y(\tau)} v_j(\tau)R_Y(t,\tau) \, d\tau. \quad (4.87)$$

Letting

$$P_N(t,\tau) = \sum_{i=1}^{\infty} \frac{1}{\mu_i} v_i(t)v_i(\tau), \quad (4.88)$$

it can be shown that

$$\int_T R_N(t,t')P_N(t',\tau) \, dt' = \delta(t-\tau). \quad (4.89)$$

Substituting (4.81) in (4.87) and using (4.88), it can be shown that the following set of

integral equations should be solved to find the optimal detector:

$$\hat{Y}(t) = \int_T \frac{\partial e[\tau, \hat{Y}(\tau)]}{\partial Y(\tau)} R_Y(t, \tau) h(\tau) \, d\tau, \tag{4.90}$$

$$h(\tau) = \int_T P_N(\tau, t') \Big(X(t') - e[t', \hat{Y}(t')] \Big) \, dt'. \tag{4.91}$$

Generally there is no straightforward solution for solving (4.90) and (4.91); they have been solved only in some certain cases [100, 101]. For instance, in [100] they have been solved for unity amplitude modulation, $e[t; Y(t)] = Y(t)$, uncorrelated noise samples, $R_N(t, \tau) = \varepsilon_n^2 \delta(t - \tau)$, and exponential form for the signal auto-correlation, $R_Y(t, \tau) = \varepsilon_Y^2 \exp(-k|t - \tau|)$.

From now on to simplify the equations to some extent, we assume that the noise samples are uncorrelated, $R_N(t, \tau) = \varepsilon_n^2 \delta(t - \tau)$. From (4.89) we deduce that

$$P_N(t, \tau) = \frac{1}{\varepsilon_n^2} \delta(t - \tau), \tag{4.92}$$

and (4.91) becomes

$$h(\tau) = \frac{1}{\varepsilon_n^2} \Big(X(\tau) - e[\tau, \hat{Y}(\tau)] \Big). \tag{4.93}$$

Substituting (4.93) in (4.90), the MAP estimate $\hat{Y}(t)$ is obtained by solving the following integral equation:

$$\hat{Y}(t) = \int_T \frac{\partial e[\tau, \hat{Y}(\tau)]}{\partial Y(\tau)} \frac{1}{\varepsilon_n^2} \Big(X(\tau) - e[\tau, \hat{Y}(\tau)] \Big) R_Y(t, \tau) \, d\tau. \tag{4.94}$$

To define MOCU for signal detection in the presence of uncertainty, we need to compute the MAP obtained by $\hat{Y}_i$. Thus, we plug the value found for $\hat{Y}_i$ in (4.86) into the

expression being maximized in (4.83). First we need to compute

$$\sum_{i=1}^{\infty}\frac{\hat{Y}_i^2}{\lambda_i} = \sum_{j=1}^{\infty}\sum_{l=1}^{\infty}\frac{B_jB_l}{\mu_j\mu_l}\iint_T\left[\frac{\partial e[t,\hat{Y}(t)]}{\partial Y(t)}\frac{\partial e[t',\hat{Y}(t')]}{\partial Y(t')}R_Y(t,t')v_j(t)v_l(t')\right]dtdt'. \tag{4.95}$$

We can further simplify (4.95) using the relation

$$\sum_{j=1}^{\infty}\frac{B_j}{\mu_j}v_j(t) = \sum_{j=1}^{\infty}\int_T\left(X(\tau)-e[\tau,\hat{Y}(\tau)]\right)\frac{v_j(\tau)v_j(t)}{\mu_j}d\tau$$

$$= \int_T P_N(t,\tau)\left(X(\tau)-e[\tau,\hat{Y}(\tau)]\right)d\tau. \tag{4.96}$$

If $P_N(t,\tau)$ takes the form as in (4.92), then (4.96) becomes

$$\sum_{j=1}^{\infty}\frac{B_j}{\mu_j}v_j(t) = \frac{1}{\varepsilon_n^2}\left(X(t)-e[t,\hat{Y}(t)]\right). \tag{4.97}$$

Substituting (4.97), (4.95) becomes

$$\sum_{i=1}^{\infty}\frac{\hat{Y}_i^2}{\lambda_i}$$

$$= \frac{1}{\varepsilon_n^4}\iint_T\frac{\partial e[t,\hat{Y}(t)]}{\partial Y(t)}\frac{\partial e[t',\hat{Y}(t')]}{\partial Y(t')}\left(X(t)-e[t,\hat{Y}(t)]\right)\left(X(t')-e[t',\hat{Y}(t')]\right)R_Y(t,t')dtdt'. \tag{4.98}$$

To compute the maximized value in (4.83), we also find

$$\sum_{i=1}^{\infty}\frac{B_i^2}{\mu_i} = \sum_{i=1}^{\infty}\left[\frac{1}{\mu_i}\int_T\left(X(t')-e[t',\hat{Y}(t')]\right)v_i(t')dt'\int_T\left(X(t)-e[t,\hat{Y}(t)]\right)v_i(t)dt\right]$$

$$= \iint_T P_N(t,t')\left(X(t')-e[t',\hat{Y}(t')]\right)\left(X(t)-e[t,\hat{Y}(t)]\right)dtdt'$$

$$= \frac{1}{\varepsilon_n^2}\int_T\left(X(t)-e[t,\hat{Y}(t)]\right)^2dt. \tag{4.99}$$

Using (4.98) and (4.99), the maximized value for (4.83) is

$$
\begin{aligned}
-\sum_{i=1}^{\infty} \frac{\hat{Y}_i^2}{\lambda_i} &- \sum_{i=1}^{\infty} \frac{B_i^2}{\mu_i}\bigg|_{Y_i=\hat{Y}_i} \\
&= -\frac{1}{\varepsilon_n^4} \iint_T \frac{\partial e[t,\hat{Y}(t)]}{\partial Y(t)} \frac{\partial e[t',\hat{Y}(t')]}{\partial Y(t')} \Big(X(t) - e[t,\hat{Y}(t)] \Big) \Big(X(t') - e[t',\hat{Y}(t')] \Big) R_Y(t,t')\, dt dt' \\
&\qquad\qquad\qquad\qquad - \frac{1}{\varepsilon_n^2} \int_T \Big(X(t) - e[t,\hat{Y}(t)] \Big)^2 dt. \qquad (4.100)
\end{aligned}
$$

Now suppose the noise is uncorrelated and its variance is unknown and parameterized by $\theta = \{\theta_1, ..., \theta_k\}$, so that $R_N^{\theta}(t,\tau) = \varepsilon_{\theta}^2 \delta(t-\tau)$, whose eigenfunctions and eigenvalues are denoted by $v_i^{\theta}(t)$ and μ_i^{θ}. We desire to estimate the signal that has the maximum expected *a posteriori* probability with respect to the uncertainty class of noise variance. Keeping in mind that uncertainty only occurs in the noise, the optimization problem in (4.83) should be modified as follows:

$$
\hat{\mathbf{Y}}_{1:M}^{\Theta} = \arg\max_{\mathbf{Y}_{1:M}} \mathrm{E}_{\theta} \left[\log \Big(f_Y(\mathbf{Y}_{1:M}) f_X^{\theta}(\mathbf{X}_{1:M}|\mathbf{Y}_{1:M}) \Big) \right], \qquad (4.101)
$$

where $f_X^{\theta}(\mathbf{X}_{1:M}|\mathbf{Y}_{1:M})$ is the conditional distribution with respect to R_N^{θ}, which can be computed by

$$
f_X^{\theta}(\mathbf{X}_{1:M}|\mathbf{Y}_{1:M}) = f_N^{\theta}(\mathbf{B}_{1:M}^{\theta}), \qquad (4.102)
$$

where

$$
B_i^{\theta} = \int_T \Big(X(t) - e[t,Y(t)] \Big) v_i^{\theta}(t)\, dt, \qquad (4.103)
$$

and

$$
f_N^{\theta}(\mathbf{N}_{1:M}) = \frac{1}{\sqrt{2\pi}^M \sqrt{\prod_{i=1}^{M} \mu_i^{\theta}}} \exp\left(-\sum_{i=1}^{M} \frac{N_i^2}{2\mu_i^{\theta}} \right). \qquad (4.104)
$$

Similar to (4.83),

$$\hat{\mathbf{Y}}^{\Theta}_{1:M} = \underset{\mathbf{Y}_{1:M}}{\arg\max} \left\{ -\sum_{i=1}^{M} \frac{Y_i^2}{\lambda_i} - \mathrm{E}_\theta \left[\sum_{i=1}^{M} \frac{(B_i^\theta)^2}{\mu_i^\theta} \right] \right\}. \tag{4.105}$$

Similar to (4.86), it can be shown that

$$\hat{Y}_i^\Theta = \lambda_i \sum_{j=1}^{M} \int_T \frac{\partial e[t, \hat{Y}^\Theta(t)]}{\partial \hat{Y}^\Theta(t)} \mathrm{E}_\theta \left[\frac{B_j^\theta v_j^\theta(t)}{\mu_j^\theta} \right] u_i(t) \, dt. \tag{4.106}$$

Letting $M \to \infty$, $\hat{Y}^\Theta(t)$ is obtained as

$$\hat{Y}^\Theta(t) = \sum_{i=1}^{\infty} \hat{Y}_i^\Theta u_i(t) = \sum_{i=1}^{\infty} \lambda_i \sum_{j=1}^{\infty} \int_T \frac{\partial e[\tau, \hat{Y}^\Theta(\tau)]}{\partial \hat{Y}^\Theta(\tau)} \mathrm{E}_\theta \left[\frac{B_j^\theta v_j^\theta(\tau)}{\mu_j^\theta} \right] u_i(\tau) u_i(t) \, d\tau. \tag{4.107}$$

The expectation is given by

$$\sum_{j=1}^{\infty} \mathrm{E}_\theta \left[\frac{B_j^\theta v_j^\theta(t)}{\mu_j^\theta} \right] = \int_T \left(X(t') - e[t', \hat{Y}^\Theta(t')] \right) \mathrm{E}_\theta \left[\sum_{j=1}^{\infty} \frac{v_j^\theta(t') v_j^\theta(t)}{\mu_j^\theta} \right] dt'$$

$$= \int_T \left(X(t') - e[t', \hat{Y}^\Theta(t')] \right) \mathrm{E}_\theta \left[P_N^\theta(t, t') \right] dt'$$

$$= \left(X(t) - e[t, \hat{Y}^\Theta(t)] \right) \mathrm{E}_\theta \left[\frac{1}{\varepsilon_\theta^2} \right]. \tag{4.108}$$

Substituting (4.108) in (4.107) yields

$$\hat{Y}^\Theta(t) = \sum_{i=1}^{\infty} \lambda_i \int_T \frac{\partial e[\tau, \hat{Y}^\Theta(\tau)]}{\partial \hat{Y}^\Theta(\tau)} \left(X(t) - e[t, \hat{Y}^\Theta(t)] \right) \mathrm{E}_\theta \left[\frac{1}{\varepsilon_\theta^2} \right] u_i(\tau) u_i(t) \, d\tau$$

$$= \int_T \frac{\partial e[\tau, \hat{Y}^\Theta(\tau)]}{\partial \hat{Y}^\Theta(\tau)} \left(X(t) - e[t, \hat{Y}^\Theta(t)] \right) \mathrm{E}_\theta \left[\frac{1}{\varepsilon_\theta^2} \right] R_Y(t, \tau) \, d\tau. \tag{4.109}$$

Comparing (4.109) with (4.94) shows that finding the estimate with the maximum
expected *a posteriori* probability when the noise variance is unknown reduces to finding
the MAP estimate if the inverse of the noise variance is $\mathrm{E}_\theta \left[\frac{1}{\varepsilon_\theta^2} \right]$. In other words, we have
shown that (4.101) can be solved in the same way that one may solve (4.82).

Now using (4.106), we aim to find the maximum value of the expression being op-
timized in (4.105) as $M \to \infty$. That is, we desire the value of the following expression as

111

$M \to \infty$:

$$-\sum_{i=1}^{M} \frac{(\hat{Y}_i^{\Theta})^2}{\lambda_i} - \mathrm{E}_{\theta}\left[\sum_{i=1}^{M} \frac{(B_i^{\theta})^2}{\mu_i^{\theta}}\bigg|_{Y_i=\hat{Y}_i^{\Theta}}\right]. \tag{4.110}$$

For the first summand in (4.110), substituting (4.106) yields

$$\sum_{i=1}^{M} \frac{(\hat{Y}_i^{\Theta})^2}{\lambda_i} =$$

$$\sum_{i=1}^{M} \lambda_i \sum_{j_1=1}^{M} \sum_{j_2=1}^{M} \iint_{T} \frac{\partial e[t,\hat{Y}^{\Theta}(t)]}{\partial \hat{Y}^{\Theta}(t)} \frac{\partial e[t',\hat{Y}^{\Theta}(t')]}{\partial \hat{Y}^{\Theta}(t')} \mathrm{E}_{\theta}\left[\frac{B_{j_1}^{\theta} v_{j_1}^{\theta}(t)}{\mu_{j_1}^{\theta}}\right] \mathrm{E}_{\theta}\left[\frac{B_{j_2}^{\theta} v_{j_2}^{\theta}(t')}{\mu_{j_2}^{\theta}}\right] u_i(t)u_i(t')dtdt'. \tag{4.111}$$

Using (4.108), the limit of the first summand in (4.110) as $M \to \infty$ is

$$\sum_{i=1}^{\infty} \frac{(\hat{Y}_i^{\Theta})^2}{\lambda_i} = \iint_{T} \left\{ \frac{\partial e[t',\hat{Y}^{\Theta}(t')]}{\partial \hat{Y}^{\Theta}(t')} \frac{\partial e[t,\hat{Y}^{\Theta}(t)]}{\partial \hat{Y}^{\Theta}(t)} \left(X(t') - e[t',\hat{Y}^{\Theta}(t')]\right) \right.$$

$$\left. \times \left(X(t) - e[t,\hat{Y}^{\Theta}(t)]\right) \mathrm{E}_{\theta}^2\left[\frac{1}{\varepsilon_{\theta}^2}\right] R_Y(t,t') \right\} dtdt'. \tag{4.112}$$

Regarding the second term in (4.110),

$$\mathrm{E}_{\theta}\left[\sum_{i=1}^{M} \frac{(B_i^{\theta})^2}{\mu_i^{\theta}}\right] = \iint_{T} \left(X(t) - e[t,\hat{Y}^{\Theta}(t)]\right)\left(X(t') - e[t',\hat{Y}^{\Theta}(t')]\right) \sum_{i=1}^{M} \mathrm{E}_{\theta}\left[\frac{v_i^{\theta}(t)v_i^{\theta}(t')}{\mu_i^{\theta}}\right] dtdt'. \tag{4.113}$$

Since

$$B_i^{\theta} = \int_{T} \left(X(t) - e[t,Y(t)]\right) v_i^{\theta}(t)dt, \tag{4.114}$$

(4.113) with $M \to \infty$ becomes

$$\mathrm{E}_{\theta}\left[\sum_{i=1}^{\infty} \frac{(B_i^{\theta})^2}{\mu_i^{\theta}}\right] = \mathrm{E}_{\theta}\left[\frac{1}{\varepsilon_{\theta}^2}\right] \int_{T} \left(X(t) - e[t,\hat{Y}^{\Theta}(t)]\right)^2 dt. \tag{4.115}$$

Substituting (4.112) and (4.115), and letting $M \to \infty$ in (4.110) yields

$$
-\sum_{i=1}^{\infty} \frac{(Y_i^{\Theta})^2}{\lambda_i} - \mathrm{E}_{\theta} \left[\sum_{i=1}^{\infty} \frac{(B_i^{\theta})^2}{\mu_i^{\theta}} \Big|_{Y_i = \hat{Y}_i^{\Theta}} \right]
$$

$$
= -\mathrm{E}_{\theta}^2 \left[\frac{1}{\varepsilon_{\theta}^2} \right] \iint_T \left\{ \frac{\partial e[t', \hat{Y}^{\Theta}(t')]}{\partial \hat{Y}^{\Theta}(t')} \frac{\partial e[t, \hat{Y}^{\Theta}(t)]}{\partial \hat{Y}^{\Theta}(t)} \right.
$$

$$
\left. \times \left(X(t') - e[t', \hat{Y}^{\Theta}(t')] \right) \left(X(t) - e[t, \hat{Y}^{\Theta}(t)] \right) R_Y(t, t') \right\} dt\, dt'
$$

$$
- \mathrm{E}_{\theta} \left[\frac{1}{\varepsilon_{\theta}^2} \right] \int_T \left(X(t) - e[t, \hat{Y}^{\Theta}(t)] \right)^2 dt. \qquad (4.116)
$$

Observe that (4.116) results from (4.100) by replacing $\hat{Y}(t)$ and $\frac{1}{\varepsilon_n^2}$ with $\hat{Y}^{\Theta}(t)$ and $\mathrm{E}_{\theta} \left[\frac{1}{\varepsilon_{\theta}^2} \right]$. In other words, to find the maximum expected *a posteriori* probability in (4.101), one needs to find the maximum *a posteriori* probability in (4.82) for the case that the inverse of the noise variance is $\mathrm{E}_{\theta} \left[\frac{1}{\varepsilon_{\theta}^2} \right]$.

To define MOCU in the framework of signal detection, we first define a reward function

$$
\zeta_{\theta}(\mathbf{Y}) = -\sum_{i=1}^{\infty} \frac{Y_i^2}{\lambda_i} - \sum_{i=1}^{\infty} \frac{(B_i^{\theta})^2}{\mu_i^{\theta}} \Big|_{Y_i}. \qquad (4.117)
$$

MOCU is then defined by

$$
\mathrm{M}(\Theta) = \mathrm{E}_{\theta} \left[\zeta_{\theta}(\hat{\mathbf{Y}}^{\theta}) - \zeta_{\theta}(\hat{\mathbf{Y}}^{\Theta}) \right], \qquad (4.118)
$$

where $\hat{\mathbf{Y}}^{\theta}$ is the set of model-specific MAP KL coefficients obtained according to (4.86) and $\hat{\mathbf{Y}}^{\Theta}$ is the set of robust MAP KL coefficients obtained according to (4.106). In fact, (4.100) and (4.116) should be used to compute the expectation of the first and the second terms, respectively, in (4.118).

In filtering the goal is to *minimize* an MSE but in signal detection the goal is to *maximize* the *a posteriori* probability. Therefore, while in the MOCU definition in(2.2) we

deduct the performance function (cost function) of the model-specific optimal filter from that of the robust filter, in (4.118) we deduct the performance function (reward function) of the robust estimation from that of the model-specific optimal estimation to calculate MOCU.

Getting back to the problem of experimental design, the primary parameter θ_{i*} is found in the spirit of (4.41):

$$i^* = \arg\max_{i \in 1,\ldots,k} \mathrm{E}_{\bar{\theta}_i}\left[\mathrm{E}_{\theta|\theta_i=\bar{\theta}_i}\left[\zeta_{\theta|\theta_i=\bar{\theta}_i}\left(\hat{\mathbf{Y}}^{\Theta|\theta_i=\bar{\theta}_i}\right)\right]\right]. \tag{4.119}$$

We can compute the inner expectation in (4.119) using (4.116).

In summary, in analogy to the effective structure utilized for filtering uncertain canonical expansions, experimental design can be applied to signal detection in the presence of uncertainty utilizing the same calculational structure as in the certain case. If the optimal signal estimation can be found in the MAP sense, then the MAP estimate in the presence of uncertainty can be solved in exactly the same way. Thus, the experimental design framework can be used in a straightforward manner without having any concern regarding the calculations in the presence of uncertainty to find the robust MAP estimation or the calculations needed for the experimental design process. That having been said, it should be recognized that signal detection using the MAP criterion is not easy and remains an open research area.

4.6 Discussion

This chapter proposed a general methodology for objective-based experimental design in filtering and signal detection. The experimental design takes into account the effect of model uncertainty on the performance of the operator via the mean objective cost of uncertainty (MOCU). MOCU measures the expected performance difference (in a Bayesian setting) in the presence and absence of model uncertainty. In the experimental design method, the parameter possessing the highest impact on the performance of the designed operator is determined first. The optimal objective-based experimental design problem has been addressed for Wiener filtering when some of the parameters of the observation model are unknown and for signal detection when the covariance matrix for the noise process is unknown. In the next chapter, we will consider experimental design for signal compression via the Karhunen-Loève (KL) expansion. We will find the IBR KL compression when the covariance matrix is unknown and then use it to apply the proposed experimental design for signal compression.

5   OPTIMAL EXPERIMENTAL DESIGN IN THE CONTEXT OF CANONICAL EXPANSIONS: KARHUNEN-LOÈVE COMPRESSION

Having laid the theoretical groundwork for canonical-expansion-based experimental design in the previous chapter, here we focus on the role of the Karhunen-Loève (KL) expansion [96] in data compression [102, 103, 104]. For instance, the KL expansion is used for image compression where the statistics of the image pixels are used to generate independent coefficients and then only high energy coefficients are stored or transmitted [105, 106, 107, 108]. In fluid mechanics, flows are often described as nonlinear dynamical processes with infinite dimensions; however, it has been shown that a finite number of parameters can be used for an accurate approximation of a flow [109]. The KL expansion can be used to represent the fluid process via constructing a reduced set of basis elements. This approach has been extensively used in mechanics for complexity reduction [110, 111, 112, 113].

KL compression is based on analyzing the data covariance matrix; however, often the knowledge regarding the actual underlying covariance matrix is not complete. Therefore, to improve the performance of the KL compression, one needs to reduce the uncertainty in the covariance matrix. To decide which uncertain parameter in the covariance matrix should be determined first to improve the quality of the compressed signal, we utilize the experimental design framework proposed in this dissertation.

Uppercase and lowercase letters denote random variables and their realizations, respectively. Bold face letters denote vectors and matrices. For a matrix $\mathbf{W}$, $W(i,j)$ denotes the element of the matrix in row i and column j and $\mathbf{W}^T$, $tr(\mathbf{W})$, and $|\mathbf{W}|$ represent the transpose, trace, and determinant operators, respectively. Also, $\mathfrak{R}(\mathbf{W},i)$ and $\mathfrak{C}(\mathbf{W},j)$ operators extract the i-th row and the j-th column, respectively. For a vector $\mathbf{v}$, $v(i)$ denotes the i-th element of the vector. $\bar{c}$ denotes the complex conjugate of a complex quantity c. Cov$[.]$ and Var$(.)$ denote the covariance and variance of random variables, respectively. $f_X(x)$ and $f_X(x|y)$ denote the marginal probability density of X and the conditional probability density of X given $Y = y$, respectively. Finally, $\mathrm{E}_x[g(x)]$ and $\mathrm{E}[X|Y = y]$ denote the expectation of function $g(x)$ with respect to random variable X and the conditional expectation of random variable X given $Y = y$, respectively.

It has been shown in the previous chapter that the parameter θ_{j*}, known as the *primary parameter*, to be determined first is given by

$$j^* = \underset{j \in 1,\dots,k}{\arg\min} \, \mathrm{E}_{\bar{\theta}_j} \left[\mathrm{E}_{\theta|\theta_j = \bar{\theta}_j} \left[\eta_{\theta|\theta_j = \bar{\theta}_j} \left(\psi_{\mathrm{IBR}}^{\Theta|\theta_j = \bar{\theta}_j} \right) \right] \right], \tag{5.1}$$

where the inner expectation is relative to the conditional distribution function $f_\theta(\theta|\theta_j = \bar{\theta}_j)$, model $(\theta|\theta_j = \bar{\theta}_j)$ is obtained by assigning $\theta_j = \bar{\theta}_j$ in the model with the uncertainty vector θ, and the reduced uncertainty class $\Theta|\theta_j = \bar{\theta}_j$ contains all models resulting from setting $\theta_j = \bar{\theta}_j$.

5.1 Karhunen-Loève Compression

5.1.1 Karhunen-Loève Compression with Known Covariance Matrix

If the random process $X(n)$ is defined over N points, the discrete KL expansion for this random process is given by (4.13). Suppose $m < N$ terms are selected from (4.13) to form the approximation

$$X_m(n) = \sum_{i \in A} Z_i u_i(n), \tag{5.2}$$

where A is the set of indices for the selected terms. The MSE between $X(n)$ and $X_m(n)$ is defined by

$$\text{MSE} < X(n), X_m(n) > = \text{E}\left[\left(X(n) - X_m(n)\right)\left(\overline{X(n) - X_m(n)}\right)\right], \tag{5.3}$$

and can be shown that

$$\text{MSE} < X(n), X_m(n) > = \sum_{i \in A^c} \lambda_i \left|u_i(n)\right|^2, \tag{5.4}$$

where A^c is the set of indices for the removed terms. Considering all $n \in N$, the MSE between X and X_m is defined by

$$\text{MSE} < \mathbf{X}, \mathbf{X}_m > = \sum_{n=1}^{N} \text{MSE} < X(n), X_m(n) > . \tag{5.5}$$

A basic property of the KL expansion is that

$$\text{MSE} < \mathbf{X}, \mathbf{X}_m > = \sum_{i \in A^c} \lambda_i. \tag{5.6}$$

Since $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ...$, given full knowledge of covariance matrix $\mathbf{K}$, the optimal m-term compression is

$$X_m(n) = \sum_{i=1}^{m} Z_i u_i(n). \tag{5.7}$$

Henceforth, the notation $X_m(n)$ will refer to this sum.

5.1.2 Intrinsically Bayesian Robust Karhunen-Loève Compression

Now suppose $\mathbf{K}$ is unknown and Θ is the uncertainty class of all possible covariance matrices $\mathbf{K}^\theta$, $\theta \in \Theta$. We desire a covariance matrix $\mathbf{K}^\Theta$ that can represent the uncertainty class Θ in an effective way, where effectiveness must be defined relative to the uncertainty class and the objective, which is compression.

First we need to show how to compute the MSE between a given process and the compressed process when the compressed process is not obtained necessarily using the covariance matrix of the original process. Suppose the covariance matrix of the random process $X(n)$ is $\mathbf{K}$ and the one used for compression is $\mathbf{K}'$. Hence, the compressed process is

$$X'_m(n) = \sum_{i=1}^{m} Z'_i u'_i(n), \tag{5.8}$$

where $\mathbf{u}'_i$ is the eigenvector of $\mathbf{K}'$ and Z'_i is the generalized Fourier coefficient of $X(n)$ relative to $\mathbf{u}'_i$.

Theorem 1. *If the random processes $X(n)$ and $X'_m(n)$ are defined as in (4.13) and (5.8), respectively, then*

$$\text{MSE} < \mathbf{X}, \mathbf{X}'_m > = \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{m} (\mathbf{u}'_i)^T \overline{\mathbf{K} \mathbf{u}'_i}. \tag{5.9}$$

Proof. Please refer to Appendix B. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Note that as $(\mathbf{u}'_i)^T \overline{\mathbf{K} \mathbf{u}'_i}$ is scalar:

$$(\mathbf{u}'_i)^T \overline{\mathbf{K} \mathbf{u}'_i} = \left((\mathbf{u}'_i)^T \overline{\mathbf{K} \mathbf{u}'_i} \right)^T = (\overline{\mathbf{u}'_i})^T \overline{\mathbf{K}}^T \mathbf{u}'_i = (\overline{\mathbf{u}'_i})^T \mathbf{K} \mathbf{u}'_i$$

Now we utilize (5.9) to find the m-term compression that minimizes the average MSE across the uncertainty class. Let $X^\theta(n)$ denote a random process with fixed covariance matrix $\mathbf{K}^\theta$ whose eigenvalues and eigenvectors are represented by λ_i^θ and $\mathbf{u}_i^\theta$, respectively. The average MSE for the compressed random process $X_m'(n)$ defined in (5.8) is given by

$$\mathrm{E}_\theta\left[\mathrm{MSE} <\mathbf{X}^\theta, \mathbf{X}_m'>\right] = \mathrm{E}_\theta\left[\sum_{i=1}^N \lambda_i^\theta - \sum_{i=1}^m (\overline{\mathbf{u}_i'})^T \mathbf{K}^\theta \mathbf{u}_i'\right]$$

$$= \sum_{i=1}^N \mathrm{E}_\theta\left[\lambda_i^\theta\right] - \sum_{i=1}^m (\overline{\mathbf{u}_i'})^T \mathrm{E}_\theta\left[\mathbf{K}^\theta\right]\mathbf{u}_i'. \tag{5.10}$$

We need to find $\mathbf{u}_i'$ that minimizes (5.10) subject to the constraint $||\mathbf{u}_i'|| = 1$ for $1 \le i \le m$:

$$\underset{\mathbf{u}_i'}{\text{minimize}} \quad \mathrm{E}_\theta\left[\mathrm{MSE} <\mathbf{X}^\theta, \mathbf{X}_m'>\right]$$

$$\text{subject to} \quad ||\mathbf{u}_i'|| = 1, \ i = 1,\dots,m. \tag{5.11}$$

This constrained optimization can be solved via the method of Lagrange multipliers:

$$L(\mathbf{u}_i', \zeta_i) = \sum_{i=1}^N \mathrm{E}_\theta\left[\lambda_i^\theta\right] - \sum_{i=1}^m (\overline{\mathbf{u}_i'})^T \mathrm{E}_\theta[\mathbf{K}^\theta]\mathbf{u}_i' - \sum_{i=1}^m \zeta_i(<\mathbf{u}_i',\mathbf{u}_i'> -1)$$

$$= \sum_{i=1}^N \mathrm{E}_\theta\left[\lambda_i^\theta\right] - \sum_{i=1}^m \sum_{n=1}^N \sum_{n_1=1}^N \mathrm{E}_\theta\left[K^\theta(n,n_1)\right]\overline{u_i'(n)}u_i'(n_1) - \sum_{i=1}^m \zeta_i\left(\sum_{n=1}^N u_i'(n)\overline{u_i'(n)} - 1\right).$$

$$\tag{5.12}$$

The partial derivative of $L(\mathbf{u}_i, \zeta_i)$ relative to $u_i'(n)$ is

$$\frac{\partial L(\mathbf{u}_i, \zeta_i)}{\partial u_i'(n)} = -2\sum_{n_1=1}^N \mathrm{E}_\theta\left[K^\theta(n,n_1)\right]u_i'(n_1) - 2\zeta_i u_i'(n). \tag{5.13}$$

Setting this partial derivative equal to 0 yields the following relation for $\mathbf{u}_i^\Theta$ that minimizes the constrained optimization in (5.11):

$$\sum_{n_1=1}^N \mathrm{E}_\theta\left[K^\theta(n,n_1)\right]u_i^\Theta(n_1) = \zeta_i u_i^\Theta(n). \tag{5.14}$$

The relation in (5.14) shows that the minimizing $\mathbf{u}_i^\Theta$ and the Lagrange multiplier ζ_i are in fact the i-th eigenvector and eigenvalue of the expected covariance matrix $\mathrm{E}_\theta[\mathbf{K}^\theta]$, respec-

tively. Hence, the minimum average MSE is achieved by using the expected covariance matrix $\mathbf{K}^{\Theta} = \mathrm{E}_{\theta}[\mathbf{K}^{\theta}]$. Therefore, the IBR m-term compression when dealing with the unknown covariance matrix is

$$X_m^{\Theta}(n) = \sum_{i=1}^{m} Z_i^{\Theta} u_i^{\Theta}(n), \qquad (5.15)$$

where $\mathbf{u}_i^{\Theta}$ is the eigenvector of $\mathbf{K}^{\Theta}$ and Z_i^{Θ} is the generalized Fourier coefficient relative to $\mathbf{u}_i^{\Theta}$.

When performing KL compression to m terms in the presence of unknown covariance matrix, relative to an uncertainty class Θ of covariance matrices, a class Ψ of m-term compressions, and the cost function η being the MSE between the original and compressed processes as in (5.9), the MOCU is defined similarly to (2.2) by

$$\mathrm{M}_{\Psi}(\Theta) = \mathrm{E}_{\theta}\left[\mathrm{MSE} < \mathbf{X}^{\theta}, \mathbf{X}_m^{\Theta} > - \mathrm{MSE} < \mathbf{X}^{\theta}, \mathbf{X}_m^{\theta} >\right]. \qquad (5.16)$$

5.1.3 Primary Parameters and Optimal Experimental Design

In the spirit of (5.1) and using the relation for MSE in (5.9), the parameter θ_{j*} to be determined first is obtained as

$$
\begin{aligned}
j^* &= \arg\min_{j} \mathrm{E}_{\bar{\theta}_j}\left[\mathrm{E}_{\theta|\theta_j=\bar{\theta}_j}\left[\mathrm{MSE} < \mathbf{X}^{\theta|\theta_j=\bar{\theta}_j}, \mathbf{X}_m^{\Theta|\theta_j=\bar{\theta}_j} >\right]\right] \\
&= \arg\min_{j} \mathrm{E}_{\bar{\theta}_j}\left[\mathrm{E}_{\theta|\theta_j=\bar{\theta}_j}\left[\sum_{i=1}^{N} \lambda_i^{\theta|\theta_j=\bar{\theta}_j} - \sum_{i=1}^{m} (\mathbf{u}_i^{\Theta|\theta_j=\bar{\theta}_j})^T \mathbf{K}^{\theta|\theta_j=\bar{\theta}_j} \bar{\mathbf{u}}_i^{\Theta|\theta_j=\bar{\theta}_j}\right]\right] \\
&= \arg\min_{j} \sum_{i=1}^{N} \mathrm{E}_{\theta}[\lambda_i^{\theta}] - \mathrm{E}_{\bar{\theta}_j}\left[\mathrm{E}_{\theta|\theta_j=\bar{\theta}_j}\left[\sum_{i=1}^{m} (\mathbf{u}_i^{\Theta|\theta_j=\bar{\theta}_j})^T \mathbf{K}^{\theta|\theta_j=\bar{\theta}_j} \bar{\mathbf{u}}_i^{\Theta|\theta_j=\bar{\theta}_j}\right]\right] \\
&= \arg\max_{j} \mathrm{E}_{\bar{\theta}_j}\left[\mathrm{E}_{\theta|\theta_j=\bar{\theta}_j}\left[\sum_{i=1}^{m} (\mathbf{u}_i^{\Theta|\theta_j=\bar{\theta}_j})^T \mathbf{K}^{\theta|\theta_j=\bar{\theta}_j} \bar{\mathbf{u}}_i^{\Theta|\theta_j=\bar{\theta}_j}\right]\right] \\
&= \arg\max_{j} \mathrm{E}_{\bar{\theta}_j}\left[\sum_{i=1}^{m} \lambda_i^{\Theta|\theta_j=\bar{\theta}_j}\right], \qquad (5.17)
\end{aligned}
$$

where $\lambda_i^{\Theta|\theta_j=\bar{\theta}_j}$ is the i-th eigenvalue of $\mathbf{K}^{\Theta|\theta_j=\bar{\theta}_j} = \mathrm{E}_{\theta|\theta_j=\bar{\theta}_j}[\mathbf{K}^{\theta|\theta_j=\bar{\theta}_j}]$ and the last equality holds because $\mathbf{u}_i^{\Theta|\theta_j=\bar{\theta}_j}$ is the eigenvector of the conditional expected covariance matrix $\mathbf{K}^{\Theta|\theta_j=\bar{\theta}_j}$.

As seen from (5.17), to find the primary parameter, for each unknown parameter one needs to find the conditional expectation of the other unknown parameters given each possible value of that parameter. In the next section, we assume that the elements of the covariance matrix are distributed according to the Wishart distribution, which is a widely used prior for the covariance matrix in multivariate signal analysis. We will show how to compute the conditional expectations in a Wishart distribution and then perform experimental design.

5.2 Unknown Covariance Matrix with Wishart Priors

The covariance matrix $\mathbf{K}$ is symmetric and positive semi-definite, meaning that $\mathbf{q}^T\mathbf{Kq} \geq 0$ for all possible vectors $\mathbf{q}$. A suitable prior that can be considered for an unknown covariance matrix is the Wishart distribution. Formally, the Wishart distribution is the distribution of the sample covariance matrix where samples are drawn from a multivariate normal distribution. For a p-variate normal distribution, there are $\frac{p(p+1)}{2}$ random variables in the covariance matrix. The joint distribution of these $\frac{p(p+1)}{2}$ random variables, or equivalently the distribution of the covariance matrix, is called the Wishart distribution and is defined as

$$\mathbf{K} \sim \mathscr{W}_p(\Sigma, n) = \frac{|\mathbf{K}|^{\frac{n-p-1}{2}}}{2^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}\Gamma_p(\frac{n}{2})} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\mathbf{K}\right)\right), \tag{5.18}$$

where $\Gamma_p(.)$ is the multivariate gamma function, $n \geq p$ is called the degree of freedom, and Σ is a $p \times p$ positive definite matrix called the scale matrix. To find the best element of an unknown covariance matrix with Wishart distribution via (5.17), we need to find the conditional expectation of each element of a Wishart distribution given the value of another element. In what follows, we find these conditional expectations with $p = 3$. Later, we propose a recursive algorithm to find the conditional expectations for a matrix of an arbitrary size.

A 4-block representation for $\mathbf{K}$ is given by

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix}, \tag{5.19}$$

where $\mathbf{K}_{ij}$ is a matrix of size $p_i \times p_j$ for $i, j = 1, 2$ and $p_1 + p_2 = p$. Also $\mathbf{K}_{12} = \mathbf{K}_{21}^T$. There is a similar 4-block representation for the scale matrix Σ.

Let

$$\mathbf{K}_{11.2} = \mathbf{K}_{11} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21}, \tag{5.20}$$

and

$$\mathbf{K}_{22.1} = \mathbf{K}_{22} - \mathbf{K}_{21}\mathbf{K}_{11}^{-1}\mathbf{K}_{12}, \tag{5.21}$$

be the Schur complements of $\mathbf{K}_{22}$ and $\mathbf{K}_{11}$, respectively. The following theory is central to our analysis to find the conditional expectations in a Wishart distribution.

Theorem 2. *If* $\mathbf{K} \sim \mathcal{W}_p(\Sigma, n)$ *and* $\mathbf{K}$ *is partitioned according to (5.19), then*

1. $\mathbf{K}_{11.2}$ *is independent from* $\mathbf{K}_{21}$ *and* $\mathbf{K}_{22}$, *and* $\mathbf{K}_{11.2} \sim \mathcal{W}_{p_1}(\Sigma_{11.2}, n - p_2)$.

2. *The conditional distribution of $\mathbf{K}_{21}$ given $\mathbf{K}_{22}$ is a multivariate normal distribution:*

$$\mathbf{K}_{21}|\mathbf{K}_{22} \sim \mathcal{N}(\mathbf{K}_{22}\Sigma', \mathbf{K}_{22} \otimes \Sigma_{11.2}), \qquad (5.22)$$

where $\Sigma' = \Sigma_{22}^{-1}\Sigma_{21}$ and $\otimes$ denotes Kronecker product.

3. $\mathbf{K}_{22}$ *has a Wishart distribution:* $\mathbf{K}_{22} \sim \mathcal{W}_{p_2}(\Sigma_{22}, n)$.

Proof. Please refer to reference [114]. □

Corollary 1. *If $\mathbf{K} \sim \mathcal{W}_p(\Sigma, n)$ and $\mathbf{K}$ is partitioned according to (5.19), then*

1. $\mathbf{K}_{22.1}$ *is independent from $\mathbf{K}_{12}$ and $\mathbf{K}_{11}$, and $\mathbf{K}_{22.1} \sim \mathcal{W}_{p_2}(\Sigma_{22.1}, n - p_1)$.*

2. *The conditional distribution of $\mathbf{K}_{12}$ given $\mathbf{K}_{11}$ is a multivariate normal distribution:*

$$\mathbf{K}_{12}|\mathbf{K}_{11} \sim \mathcal{N}(\mathbf{K}_{11}\Sigma'', \mathbf{K}_{11} \otimes \Sigma_{22.1}), \qquad (5.23)$$

where $\Sigma'' = \Sigma_{11}^{-1}\Sigma_{12}$.

3. $\mathbf{K}_{11}$ *has a Wishart distribution:* $\mathbf{K}_{11} \sim \mathcal{W}_{p_1}(\Sigma_{11}, n)$.

We consider two 4-block representations for matrices $\mathbf{K}$ and Σ. Throughout this section, $\mathbf{K}_{ij}^r$ denotes the block ij in Representation r, where in Representation 1, $p_1 = 1$ and $p_2 = p - 1$, and in Representation 2, $p_1 = p - 1$ and $p_2 = 1$. Also, $\mathbf{K}_{11.2}^r$, $\mathbf{K}_{22.1}^r$, Σ_r', and Σ_r'' are computed with regards to Representation r. Consider a 3×3 covariance matrix

$$\mathbf{K} = \begin{bmatrix} K_1 & K_4 & K_5 \\ K_4 & K_2 & K_6 \\ K_5 & K_6 & K_3 \end{bmatrix}. \qquad (5.24)$$

We can partition $\mathbf{K}$ in two ways. For Representation 1, the blocks are $\mathbf{K}_{11}^1 = [K_1]$, $\mathbf{K}_{12}^1 = [K_4 \ K_5]$, $\mathbf{K}_{21}^1 = \begin{bmatrix} K_4 \\ K_5 \end{bmatrix}$, and $\mathbf{K}_{22}^1 = \begin{bmatrix} K_2 & K_6 \\ K_6 & K_3 \end{bmatrix}$. For Representation 2, the blocks are $\mathbf{K}_{11}^2 =$

$$\begin{bmatrix} K_1 & K_4 \\ K_4 & K_2 \end{bmatrix}, \mathbf{K}_{12}^2 = \begin{bmatrix} K_5 \\ K_6 \end{bmatrix}, \mathbf{K}_{21}^2 = [K_5 \ K_6], \text{ and } \mathbf{K}_{22}^2 = [K_3].$$

5.2.1 Conditional Expectations for a 3×3 Wishart Distribution

In the following several subsections, we find the conditional expectations relative to each element of the covariance matrix using Theorem 2 and Corollary 1.

Conditional expectation given $K_1 = k_1$

Partitioning $\mathbf{K}$ according to Representation 1, it is enough to compute $\mathrm{E}[\mathbf{K}_{12}^1|k_1]$ and $\mathrm{E}[\mathbf{K}_{22}^1|k_1]$. By part 2 of Corollary 1,

$$\mathrm{E}[\mathbf{K}_{12}^1|k_1] = k_1 \Sigma_1''. \tag{5.25}$$

Since $\mathrm{E}[\mathbf{K}_{22.1}^1|k_1] = \mathrm{E}[\mathbf{K}_{22}^1|k_1] - \mathrm{E}[\mathbf{K}_{21}^1(\mathbf{K}_{11}^1)^{-1}\mathbf{K}_{12}^1|k_1]$ and $\mathbf{K}_{11}^1 = K_1$, the conditional expectation of $\mathbf{K}_{22}^1$ given $K_1 = k_1$ is

$$\mathrm{E}[\mathbf{K}_{22}^1|k_1] = \mathrm{E}[\mathbf{K}_{22.1}^1|k_1] + k_1^{-1}\mathrm{E}[\mathbf{K}_{21}^1\mathbf{K}_{12}^1|k_1]. \tag{5.26}$$

As $\mathrm{E}[\mathbf{K}_{21}^1\mathbf{K}_{12}^1|k_1]$ is the conditional auto-correlation matrix of $\mathbf{K}_{12}^1$ given $K_1 = k_1$, using (5.23) we have:

$$\mathrm{E}[\mathbf{K}_{21}^1\mathbf{K}_{12}^1|k_1] = \mathrm{Cov}(\mathbf{K}_{12}^1|k_1) + \mathrm{E}[\mathbf{K}_{12}^1|k_1]^T \mathrm{E}[\mathbf{K}_{12}^1|k_1]$$

$$= k_1 \Sigma_{22.1}^1 + k_1 \left(\Sigma_1''\right)^T k_1 \Sigma_1'' = k_1 \Sigma_{22.1}^1 + k_1^2 (\Sigma_1'')^T \Sigma_1''. \tag{5.27}$$

Using (5.27), and noting that from the first part in Corollary 1, $\mathrm{E}[\mathbf{K}_{22.1}^1|k_1] = \mathrm{E}[\mathbf{K}_{22.1}^1] = (n-1)\Sigma_{22.1}^1$, (5.26) becomes

$$\mathrm{E}[\mathbf{K}_{22}^1|k_1] = n\Sigma_{22.1}^1 + k_1(\Sigma_1'')^T \Sigma_1''. \tag{5.28}$$

Conditional expectation given $K_3 = k_3$

Partitioning **K** according to Representation 2, we need to find the conditional expectations of $\mathbf{K}_{21}^2$ and $\mathbf{K}_{11}^2$ given $K_3 = k_3$, for which we use Theorem 2. Following a similar procedure as in Subsection 5.2.1, we obtain $\mathrm{E}\left[\mathbf{K}_{21}^2 | K_3 = k_3\right]$ using (5.25), where k_1, Σ_{11}^1, and Σ_{12}^1 are replaced by k_3, Σ_{22}^2, and Σ_{21}^2, respectively.

We can also find $\mathrm{E}\left[\mathbf{K}_{11}^2 | K_3 = k_3\right]$ similarly to (5.28), where k_1, $\Sigma_{22.1}^1$, and Σ_1'' are replaced by k_3, $\Sigma_{11.2}^2$, and Σ_2', respectively.

Conditional expectation given $K_2 = k_2$

To compute the conditional expectation given $K_2 = k_2$, we should compute $\mathrm{E}\left[\mathbf{K}_{22}^1 | K_2 = k_2\right]$, $\mathrm{E}\left[\mathbf{K}_{11}^2 | K_2 = k_2\right]$, and $\mathrm{E}\left[K_5 | K_2 = k_2\right]$.

We can compute the conditional expectation $\mathrm{E}\left[\mathbf{K}_{22}^1 | K_2 = k_2\right]$ in a similar way that we computed the conditional expectation of other elements given $K_1 = k_1$ in Subsection 5.2.1, where here we have a matrix of size 2×2 with Wishart distribution $\mathscr{W}_2(\Sigma_{22}^1, n)$ and we want to find the conditional expectation of other elements given the value of the first diagonal element.

We also partition **K** based on Representation 2 and compute the conditional expectation $\mathrm{E}\left[\mathbf{K}_{11}^2 | K_2 = k_2\right]$ in a similar way that we computed the conditional expectation of other elements given $K_3 = k_3$ in Subsection 5.2.1 where here we have a matrix of size 2×2 with Wishart distribution $\mathscr{W}_2(\Sigma_{11}^2, n)$ and we want to find the conditional expectation of other elements given the value of the last diagonal element.

126

We now need to compute $\mathrm{E}\big[K_5|K_2 = k_2\big]$:

$$
\begin{aligned}
\mathrm{E}\big[K_5|K_2 = k_2\big] &= \int k_5 f_{K_5}(k_5|k_2)\,dk_5 \\
&= \int k_5 \int f_{K_5}(k_5|k_1,k_4,k_2) f_{K_1,K_4}(k_1,k_4|k_2)\,dk_1 dk_4 dk_5 \\
&= \int f_{K_1,K_4}(k_1,k_4|k_2)\mathrm{E}\big[K_5|K_1 = k_1, K_4 = k_4, K_2 = k_2\big]\,dk_1\,dk_4 \\
&= \int f_{K_1,K_4}(k_1,k_4|k_2)\Re(\left(\begin{smallmatrix} k_1 & k_4 \\ k_4 & k_2 \end{smallmatrix}\right),1)\Sigma_2''\,dk_1\,dk_4 \\
&= \int f_{K_1,K_4}(k_1,k_4|k_2)\big(k_1\Sigma_2''(1) + k_4\Sigma_2''(2)\big)\,dk_1\,dk_4 \\
&= \Sigma_2''(1)\mathrm{E}\big[K_1|K_2 = k_2\big] + \Sigma_2''(2)\mathrm{E}\big[K_4|K_2 = k_2\big],
\end{aligned}
\tag{5.29}
$$

where we use (5.23) to obtain the fourth equality. Note that $\mathrm{E}\big[K_1|K_2 = k_2\big]$ and $\mathrm{E}\big[K_4|K_2 = k_2\big]$ have already been calculated when we computed $\mathrm{E}\big[\mathbf{K}_{11}^2|K_2 = k_2\big]$.

Conditional expectation given $K_4 = k_4$

It is enough to compute the conditional expectations for $\mathbf{K}_{22}^1$, $\mathbf{K}_{11}^1 = K_1$, and K_5. To compute the conditional expectation given $K_4 = k_4$, first we compute $f_{K_4}(k_4)$ by partitioning $\mathbf{K}$ according to Representation 1 as follows:

$$
f_{K_4}(k_4) = \mathrm{E}_{k_1}\big[f_{K_4}(k_4|k_1)\big] = \mathrm{E}_{k_1}\big[N(k_4;\alpha,\beta)\big],
\tag{5.30}
$$

where $\alpha = k_1\mathfrak{C}(\Sigma_1'',1)$, $\beta = k_1\Sigma_{22.1}^1(1,1)$, and $N(k_4;\alpha,\beta)$ is the value of a multivariate Gaussian function with mean vector α and covariance matrix β at point k_4. Note that one only needs to drop the irrelevant variables from the mean vector and the covariance matrix to obtain the marginal distribution over a subset of multivariate normal random variables. The expectation in (5.30) involves a Gaussian function with respect to a Wishart distribution and cannot be solved analytically. Hence, we compute it numerically by taking

the average of function $N(k_4; \alpha, \beta)$ relative to a Wishart distribution with scale matrix $\Sigma(1,1)$ and n using Monte-Carlo simulations. The numerators in (5.32) and (5.33) will be treated similarly.

If X and Y are two random variables, then the conditional expectation $E[X|y]$ can be found as

$$E[X|Y=y] = \int x \frac{f_{XY}(x,y)}{f_Y(y)} dx = \frac{\int x f_Y(y|x) f_X(x) dx}{f_Y(y)} = \frac{E_x[x f_Y(y|x)]}{f_Y(y)}. \tag{5.31}$$

For computing conditional expectations for $\mathbf{K}_{22}^1$ and K_1, we use the relation in (5.31). Now,

$$E\left[\mathbf{K}_{22}^1 | K_4 = k_4\right] = \frac{E_{\mathbf{K}_{22}^1}\left[\mathbf{K}_{22}^1 f_{K_4}(k_4 | \mathbf{K}_{22}^1)\right]}{f_{K_4}(k_4)} = \frac{E_{\mathbf{K}_{22}^1}\left[\mathbf{K}_{22}^1 N(k_4; \alpha, \beta)\right]}{f_{K_4}(k_4)}, \tag{5.32}$$

where $\alpha = \Re(\mathbf{K}_{22}^1, 1)\Sigma_1'$, $\beta = K_{22}^1(1,1)\Sigma_{11.2}^1$, and the second equality results from using Theorem 2, part 2. Next,

$$E\left[K_1 | K_4 = k_4\right] = \frac{E_{k_1}\left[k_1 f_{K_4}(k_4 | k_1)\right]}{f_{K_4}(k_4)} = \frac{E_{k_1}\left[k_1 N(k_4; \alpha, \beta)\right]}{f_{K_4}(k_4)}, \tag{5.33}$$

where $\alpha = k_1 \mathfrak{C}(\Sigma_1'', 1)$, $\beta = k_1 \Sigma_{22.1}^1(1,1)$, and we use Corollary 1 part 2 to obtain the second equality. Finally,

$$E\left[K_5 | K_4 = k_4\right] = \frac{E_{k_1}\left[\int k_5 f_{K_5, K_4}(k_5, k_4 | k_1) dk_5\right]}{f_{K_4}(k_4)} = \frac{E_{k_1}\left[\int k_5 N([k_4 \ k_5]; \alpha, \beta) dk_5\right]}{f_{K_4}(k_4)}, \tag{5.34}$$

where $\alpha = k_1 \Sigma_1''$ and $\beta = k_1 \Sigma_{22.1}^1$. To compute the numerator in (5.34), we introduce the following lemma, which is the bivariate Gaussian form of the relation $\int x f_{X,Y}(x,y) dx = E[X|y] f_Y(y)$.

$$E[K_5|K_4 = k_4] = \frac{E_{k_1}\left[\exp\left(-\frac{(k_4-k_1\Sigma_1''(1))^2}{2k_1\Sigma_{22.1}^1(1,1)}\right) \frac{k_1\Sigma_1''(2)+\left(k_4-k_1\Sigma_1''(1)\right)\sqrt{\frac{\Sigma_{22.1}^1(2,2)}{\Sigma_{22.1}^1(1,1)}}\frac{\Sigma_{22.1}^1(1,2)}{\sqrt{\Sigma_{22.1}^1(1,1)\Sigma_{22.1}^1(2,2)}}}{\sqrt{2\pi}\sqrt{k_1\Sigma_{22.1}^1(1,1)}}\right]}{f_{K_4}(k_4)}$$

$$= \frac{E_{k_1}\left[\exp\left(-\frac{(k_4-k_1\Sigma_1''(1))^2}{2k_1\Sigma_{22.1}^1(1,1)}\right) \frac{k_1\left(\Sigma_1''(2)-\Sigma_1''(1)\frac{\Sigma_{22.1}^1(1,2)}{\Sigma_{22.1}^1(1,1)}\right)+k_4\frac{\Sigma_{22.1}^1(1,2)}{\Sigma_{22.1}^1(1,1)}}{\sqrt{2\pi}\sqrt{k_1\Sigma_{22.1}^1(1,1)}}\right]}{f_{K_4}(k_4)}. \qquad (5.37)$$

Lemma 1. *If $\sigma_x, \sigma_y > 0$, $\mu_x, \mu_y \in \mathbb{R}$, and $-1 < \rho < 1$ are all constants, then*

$$\int x \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left(\frac{-1}{2(1-\rho^2)}\left[(\frac{x-\mu_x}{\sigma_x})^2 + (\frac{y-\mu_y}{\sigma_y})^2 - 2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y})\right]\right) dx$$

$$= \frac{\mu_x + \frac{y-\mu_y}{\sigma_y}\rho\sigma_x}{\sqrt{2\pi}\sigma_y}\exp\left(-\frac{(y-\mu_y)^2}{2\sigma_y^2}\right). \qquad (5.35)$$

Using Lemma 1, we can show that:

$$\int k_5 N([k_4\ k_5]; k_1\Sigma_1'', k_1\Sigma_{22.1}^1)\, dk_5 = \frac{\mu_{k_5} + \frac{k_4-\mu_{k_4}}{\sigma_{k_4}}\rho\sigma_{k_5}}{\sqrt{2\pi}\sigma_{k_4}}\exp\left(-\frac{(k_4-\mu_{k_4})^2}{2\sigma_{k_4}^2}\right), \qquad (5.36)$$

where $[\mu_{k_4}\ \mu_{k_5}] = k_1\Sigma_1''$, $\sigma_{k_4}^2 = k_1\Sigma_{22.1}^1(1,1)$, $\sigma_{k_5}^2 = k_1\Sigma_{22.1}^1(2,2)$, and $\rho = \frac{k_1\Sigma_{22.1}^1(1,2)}{\sigma_{k_4}\sigma_{k_5}} = \frac{\Sigma_{22.1}^1(1,2)}{\sqrt{\Sigma_{22.1}^1(1,1)\Sigma_{22.1}^1(2,2)}}$. Therefore, in order to compute $E[K_5|K_4 = k_4]$ we can rewrite (5.34) as shown in (5.37).

Conditional expectation given $K_6 = k_6$

This conditional expectation can be computed similarly to the conditional expectation given $K_4 = k_4$; however, here we use Representation 2 for **K** to find the conditional expectations.

Conditional expectation given $K_5 = k_5$

The conditional expectations of $\mathbf{K}_{11}^2$ and $\mathbf{K}_{22}^1$ given $K_5 = k_5$ are needed. First we require $f_{K_5}(k_5)$. By partitioning $\mathbf{K}$ according to Representation 2, we obtain

$$f_{K_5}(k_5) = \mathrm{E}_{k_3}\left[f_{K_5}(k_5|k_3)\right] = \mathrm{E}_{k_3}\left[N(k_5;\alpha,\beta)\right], \qquad (5.38)$$

where $\alpha = k_3\mathfrak{C}(\Sigma_2',1)$, $\beta = k_3\Sigma_{11.2}^2(1,1)$. Now, using the relation in (5.31):

$$\mathrm{E}\left[\mathbf{K}_{11}^2|K_5 = k_5\right] = \frac{\mathrm{E}_{\mathbf{K}_{11}^2}\left[\mathbf{K}_{11}^2 f_{K_5}(k_5|\mathbf{K}_{11}^2)\right]}{f_{K_5}(k_5)} = \frac{\mathrm{E}_{\mathbf{K}_{11}^2}\left[\mathbf{K}_{11}^2 N(k_5;\alpha,\beta)\right]}{f_{K_5}(k_5)}, \qquad (5.39)$$

where $\alpha = \mathfrak{R}(\mathbf{K}_{11}^2,1)\Sigma_2''$ and $\beta = K_{11}^2(1,1)\Sigma_{22.1}^2$.

Next, we partition $\mathbf{K}$ according to Representation 1 and compute $\mathrm{E}\left[\mathbf{K}_{22}^1|K_5 = k_5\right]$ in a similar way:

$$\mathrm{E}\left[\mathbf{K}_{22}^1|K_5 = k_5\right] = \frac{\mathrm{E}_{\mathbf{K}_{22}^1}\left[\mathbf{K}_{22}^1 f_{K_5}(k_5|\mathbf{K}_{22}^1)\right]}{f_{K_5}(k_5)} = \frac{\mathrm{E}_{\mathbf{K}_{22}^1}\left[\mathbf{K}_{22}^1 N(k_5;\alpha,\beta)\right]}{f_{K_5}(k_5)}, \qquad (5.40)$$

where $\alpha = \mathfrak{R}(\mathbf{K}_{22}^1,2)\Sigma_1'$ and $\beta = K_{22}^1(2,2)\Sigma_{11.2}^1$.

5.2.2 Conditional Expectations for an Arbitrary Size Wishart Distribution

In this section, we propose a recursive approach to compute the conditional expectations for a Wishart distribution of size $p \times p$.

Conditional expectation given $K(1,1) = k_{11}$ and $K(p,p) = k_{pp}$

To compute the conditional expectations of other elements given $K(1,1) = k_{11}$, we find the conditional expectation for $\mathbf{K}_{12}^1$ and $\mathbf{K}_{22}^1$ similarly to the 3×3 case in Subsection 5.2.1. The conditional expectations given $K(p,p) = k_{pp}$ are found similarly to the method of Subsection 5.2.1.

Conditional expectation given $K(i,i) = k_{ii}$, $1 < i < p$

To compute the conditional expectations of other elements given $K(i,i) = k_{ii}$, we find the conditional expectations for $\mathbf{K}_{22}^1$, $\mathbf{K}_{11}^2$, and $K(1,p)$.

As $\mathbf{K}_{22}^1$ has a Wishart distribution with the scale matrix Σ_{22}^1, we compute the conditional expectation $\mathrm{E}\left[\mathbf{K}_{22}^1 | K(i,i) = k_{ii}\right]$ recursively, where we need to compute the conditional expectation of a Wishart matrix of size $(p-1) \times (p-1)$ given the value of an element at position $(i-1, i-1)$.

As $\mathbf{K}_{11}^2$ has a Wishart distribution with scale matrix Σ_{11}^2, we compute $\mathrm{E}\left[\mathbf{K}_{11}^2 | K(i,i) = k_{ii}\right]$ recursively, where we need to compute the conditional expectation of a Wishart matrix of size $(p-1) \times (p-1)$ conditioned on the value of an element at position (i,i).

We compute $\mathrm{E}\left[K(1,p) | K(i,i) = k_{ii}\right]$ using Representation 2 as (5.29):

$$\mathrm{E}\left[K(1,p) | K(i,i) = k_{ii}\right] = \sum_{j=1}^{p-1} \Sigma_2''(j) \mathrm{E}\left[K(1,j) | K(i,i) = k_{ii}\right]. \tag{5.41}$$

Conditional expectation given $K(1,i) = k_{1i}$, $2 \le i \le p-1$

To compute the conditional expectations of other elements given $K(1,i) = k_{1i}$, we find the conditional expectations for $\mathbf{K}_{22}^1$, $K(1,1)$, and $K(1,j), j \ne 1$. First, we evaluate

$$f_{K(1,i)}(k_{1i}) = \mathrm{E}_{k_{11}}\left[f_{K(1,i)}(k_{1i}|k_{11})\right] = \mathrm{E}_{k_{11}}\left[N(k_{1i}; \alpha, \beta)\right], \tag{5.42}$$

where $\alpha = k_{11}\mathfrak{C}(\Sigma_1'', i-1)$ and $\beta = k_{11}\Sigma_{22.1}^1(i-1, i-1)$.

Proceeding similarly to (5.32), we compute $\mathrm{E}\left[\mathbf{K}_{22}^1 | K(1,i) = k_{1i}\right]$ as

$$\mathrm{E}\left[\mathbf{K}_{22}^1 | K(1,i) = k_{1i}\right] = \frac{\mathrm{E}_{\mathbf{K}_{22}^1}\left[\mathbf{K}_{22}^1 f_{K(1,i)}(k_{1i}|\mathbf{K}_{22}^1)\right]}{f_{K(1,i)}(k_{1i})} = \frac{\mathrm{E}_{\mathbf{K}_{22}^1}\left[\mathbf{K}_{22}^1 N(k_{1i}; \alpha, \beta)\right]}{f_{K(1,i)}(k_{1i})}, \tag{5.43}$$

where $\alpha = \mathfrak{R}(\mathbf{K}_{22}^1, i-1)\Sigma_1'$ and $\beta = \mathbf{K}_{22}^1(i-1, i-1)\Sigma_{11.2}^1$.

$$\text{E}\big[K(1,j)|K(1,i)=k_{1i}\big]$$

$$= \frac{\text{E}_{k_{11}}\left[\exp\left(-\frac{(k_{1i}-k_{11}\Sigma_1''(i-1))^2}{2k_{11}\Sigma_{22.1}^1(i-1,i-1)}\right)\frac{k_{11}\Sigma_1''(j-1)+\left(k_{1i}-k_{11}\Sigma_1''(i-1)\right)\sqrt{\frac{\Sigma_{22.1}^1(j-1,j-1)}{\Sigma_{22.1}^1(i-1,i-1)}}\frac{\Sigma_{22.1}^1(i-1,j-1)}{\sqrt{\Sigma_{22.1}^1(i-1,i-1)\Sigma_{22.1}^1(j-1,j-1)}}}{\sqrt{2\pi}\sqrt{k_{11}\Sigma_{22.1}^1(i-1,i-1)}}\right]}{f_{K(1,i)}(k_{1i})}$$

$$= \frac{\text{E}_{k_{11}}\left[\exp\left(-\frac{(k_{1i}-k_{11}\Sigma_1''(i-1))^2}{2k_{11}\Sigma_{22.1}^1(i-1,i-1)}\right)\frac{k_{11}\left(\Sigma_1''(j-1)-\Sigma_1''(i-1)\frac{\Sigma_{22.1}^1(i-1,j-1)}{\Sigma_{22.1}^1(i-1,i-1)}\right)+k_{1i}\frac{\Sigma_{22.1}^1(i-1,j-1)}{\Sigma_{22.1}^1(i-1,i-1)}}{\sqrt{2\pi}\sqrt{k_{11}\Sigma_{22.1}^1(i-1,i-1)}}\right]}{f_{K(1,i)}(k_{1i})}. \quad (5.47)$$

Via Representation 1, $\text{E}\big[K(1,1)|K(1,i)=k_{1i}\big]$ is found similarly to (5.33):

$$\text{E}\big[K(1,1)|K(1,i)=k_{1i}\big] = \frac{\text{E}_{k_{11}}\big[k_{11}f_{K(1,i)}(k_{1i}|k_{11})\big]}{f_{K(1,i)}(k_{1i})} = \frac{\text{E}_{k_{11}}\big[k_{11}N\big(k_{1i};\alpha,\beta\big)\big]}{f_{K(1,i)}(k_{1i})}, \quad (5.44)$$

where $\alpha = k_{11}\mathfrak{C}(\Sigma_1'',i-1)$ and $\beta = k_{11}\Sigma_{22.1}^1(i-1,i-1)$.

To find $\text{E}\big[K(1,j),j\neq 1,i|K(1,i)=k_{1i}\big]$, partitioning $\mathbf{K}$ according to Representation

1 yields

$$\text{E}\big[K(1,j)|K(1,i)=k_{1i}\big] = \frac{\text{E}_{k_{11}}\left[\int k_{1j}f(k_{1j},k_{1i}|k_{11})dk_{1j}\right]}{f_{K(1,i)}(k_{1i})} = \frac{\text{E}_{k_{11}}\left[\int k_{1j}N([k_{1i}\ k_{1j}];\alpha,\beta)dk_{1j}\right]}{f_{K(1,i)}(k_{1i})},$$

$$(5.45)$$

where $\alpha = k_{11}[\Sigma_1''(i-1)\ \Sigma_1''(j-1)]$ and

$$\beta = k_{11}\begin{bmatrix} \Sigma_{22.1}^1(i-1,i-1) & \Sigma_{22.1}^1(i-1,j-1) \\ \Sigma_{22.1}^1(i-1,j-1) & \Sigma_{22.1}^1(j-1,j-1) \end{bmatrix}. \quad (5.46)$$

Note that $f(k_{1j},k_{1i}|k_{11})$ is found using part 2 of Corollary 1 and the fact that the

marginal distribution over a subset of multivariate normal random variables can be obtained

by dropping the irrelevant variables from the mean vector and the covariance matrix. We

extract the $(i-1)$-th and $(j-1)$-th elements of Σ_1'' because $K(1,i)$ and $K(1,j)$ are the the

$(i-1)$-th and $(j-1)$-th elements in block $\mathbf{K}_{12}^1$. Since (5.45) has a form similar to (5.34), it can be simplified further in the same way. The conditional expectation $\mathrm{E}\big[K(1,j)|K(1,i) = k_{1i}\big]$ is given by (5.47). Note that (5.37) results from (5.47) by setting $i = 2$ and $j = 3$.

Conditional expectation given $K(1,p) = k_{1p}$

To find the conditional expectations given $K(1,p) = k_{1p}$, follow the same procedure used in Subsection 5.2.1 for $K_5 = k_5$.

Conditional expectation given $K(i,j) = k_{ij}$, $2 \le i \le p-1$, $i < j < p$

To evaluate the conditional expectations of other elements given $K(i,j) = k_{ij}$, we find the conditional expectation for $\mathbf{K}_{22}^1$, $K(1,1)$, and $K(l,1), 2 \le l \le p$. We extract a block $\big[K(h,l)\big]_{h,l=i}^j$ and then partition it based on Representation 1 to evaluate

$$f_{K(i,j)}(k_{ij}) = \mathrm{E}_{k_{11}}\big[f_{K(i,j)}(k_{ij}|k_{11})\big] = \mathrm{E}_{k_{11}}\big[N(k_{ij};\alpha,\beta)\big], \qquad (5.48)$$

where $\alpha = k_{11}\mathfrak{C}(\Sigma_1'', j-1)$ and $\beta = k_{11}\Sigma_{22.1}^1(j-i,j-i)$. Note that here the scale matrix Σ used in (5.48) is in fact $\big[\Sigma(h,l)\big]_{h,l=i}^j$, which corresponds to the block $\big[K(h,l)\big]_{h,l=i}^j$. Now we proceed to compute conditional expectations.

For $\mathrm{E}\big[\mathbf{K}_{22}^1|K(i,j) = k_{ij}\big]$ we require the conditional expectation for a Wishart distribution of size $(p-1) \times (p-1)$ and scale matrix Σ_{22}^1 given the value of $K(i-1,j-1)$. We do this recursively.

Representation 2 is used to obtain $\mathrm{E}\big[K(1,1)|K(i,j) = k_{ij}\big]$. Therefore, we must compute the conditional expectation of $K(1,1)$ in a Wishart distribution of size $(p-1) \times (p-1)$ with scale matrix Σ_{11}^2, given $K(i,j) = k_{ij}$.

Now, using Representation 1, for $2 \leq l \leq p$,

$$\mathrm{E}\big[K(l,1)|K(i,j) = k_{ij}\big] = \int k_{l1} f_{K(l,1)}(k_{l1}|k_{ij}) \, dk_{l1}$$

$$= \int k_{l1} \underbrace{\int \cdots \int}_{K(i',j') \in \mathbf{K}_{22}^1} f_{K(l,1)}\big(k_{l1}|k_{i'j'}, k_{ij}\big) f_{K(i',j')}\big(k_{i'j'}|k_{ij}\big) \, dk_{i'j'} \, dk_{l1}$$

$$= \underbrace{\int \cdots \int}_{K(i',j') \in \mathbf{K}_{22}^1} f_{K(i',j')}\big(k_{i'j'}|k_{ij}\big) \int k_{l1} f_{K(l,1)}\big(k_{l1}|k_{i'j'}, k_{ij}\big) \, dk_{l1} \, dk_{i'j'}$$

$$= \underbrace{\int \cdots \int}_{K(i',j') \in \mathbf{K}_{22}^1} f_{K(i',j')}\big(k_{i'j'}|k_{ij}\big) \mathfrak{R}(\mathbf{K}_{22}^1, l-1) \Sigma_1' \, dk_{i'j'}$$

$$= \sum_{j'} \Sigma_1'(j') \mathrm{E}\big[\mathbf{K}_{22}^1(l-1,j')|K(i,j) = k_{ij}\big], \qquad (5.49)$$

where $\mathrm{E}\big[\mathbf{K}_{22}^1(l-1,j')|K(i,j) = k_{ij}\big]$ has been evaluated previously.

5.3 Simulation Results for Wishart Prior

5.3.1 3×3 Covariance Matrix with Wishart Prior

In this section, we analyze the performance of experimental design when the unknown covariance matrix has a Wishart distribution with the scale matrix

$$\Sigma = \begin{bmatrix} 8 & 1.2 & -2.6 \\ 1.2 & 5 & -4.5 \\ -2.6 & -4.5 & 8 \end{bmatrix}, \qquad (5.50)$$

and the aim is to compress 3 random variables to 1 random variable, i.e., $m = 1$. We set the degree of freedom n for the Wishart prior to 4. The experimental design values computed

according to (5.17) for the elements of the covariance matrix are

$$\begin{bmatrix} 55.53 & 55.77 & 56.67 \\ & 54.26 & 55.16 \\ & & 54.59 \end{bmatrix}. \tag{5.51}$$

Based on these values, the best element to be determined first is the element at $(1, 3)$, which has been denoted by K_5 in (5.24), because it has the largest experimental design value. Moreover, determination of unknown elements can be prioritized: for example, the order of determination is $K_5 > K_4 > K_1 > K_6 > K_3 > K_2$, by which we mean that if determining element K_5 is not possible, then the next best option is K_4, and so on.

For performance evaluation, suppose that parameter θ_{j^*} whose true value is μ_{j^*} is chosen first. After conducting the chosen experiment, there is a smaller uncertainty class containing that part of the initial uncertainty class Θ that is compatible with the outcome of the experiment. Denote it by $\Theta|(\theta_{j^*} = \mu_{j^*})$. To evaluate the effectiveness of the chosen experiment, we measure the MSE of the KL expression that is robust for $\Theta|(\theta_{j^*} = \mu_{j^*})$ when it is applied to the underlying process with the true value μ for the unknown parameters of the covariance matrix:

$$\text{MSE} < \mathbf{X}^{\mu}, \mathbf{X}_m^{\Theta|\theta_{j^*}=\mu_{j^*}} >= \sum_{i=1}^{N} \lambda_i^{\mu} - \sum_{i=1}^{m} \left(\mathbf{u}_i^{\Theta|\theta_{j^*}=\mu_{j^*}} \right)^T \mathbf{K}^{\mu} \mathbf{u}_i^{\Theta|\theta_{j^*}=\mu_{j^*}}, \tag{5.52}$$

where λ_k^{μ} is the k-th eigenvalue of the underlying true covariance matrix $\mathbf{K}^{\mu}$ and $\mathbf{u}_k^{\Theta|\theta_{j^*}=\mu_{j^*}}$ is the k-th eigenvector of the expectation of the covariance matrices relative to the uncertainty class $\Theta|\theta_{j^*} = \mu_{j^*}$. We generate a pool of 10^6 covariance matrices using the Wishart distribution and assume that each could be the true covariance matrix. Once the true co-

variance matrix is fixed, we find the MSE corresponding to determining each element of the covariance matrix using (5.52). Then the average MSE corresponding to the determination of each element is obtained by taking the average over different assumed true covariance matrix. To illustrate simulation results, we show the average MSE obtained by determining each element of the covariance matrix over all possible values for a certain element. To enhance evaluation, we also show the empirical marginal distribution of the element. The space of realizations for the element K_i is partitioned into intervals of size $0.2/\left|\mathrm{E}\left[K_i\right]\right|$, where $\mathrm{E}\left[K_i\right]$ is the average of element K_i, and then the empirical marginal distribution is computed as the proportion of the number of sample matrices that fall within the interval. For each interval, the average MSE over those sample matrices which fall within the interval is computed. Figure 5.1 shows the conditional average MSE obtained by determining each element of the covariance matrix for different realizations of each element of the covariance matrix. In this figure, the background gray curves are empirical distributions. Note that, although from the experimental design step the primary parameter is K_5, it is not optimal over the entire space of realizations. K_5 is optimal on average with respect to the Wishart distribution (or the marginal distribution of each element) for the covariance matrix.

The Mahalanobis distance reflects the distance of a realization from the distribution of the random variable. It gets larger as the realization moves away from the mean of the distribution. Because the Mahalanobis distance can be computed for random vectors, if the unknown covariance matrix $\mathbf{K}$ is of size $p \times p$, we first remove its duplicate entries and
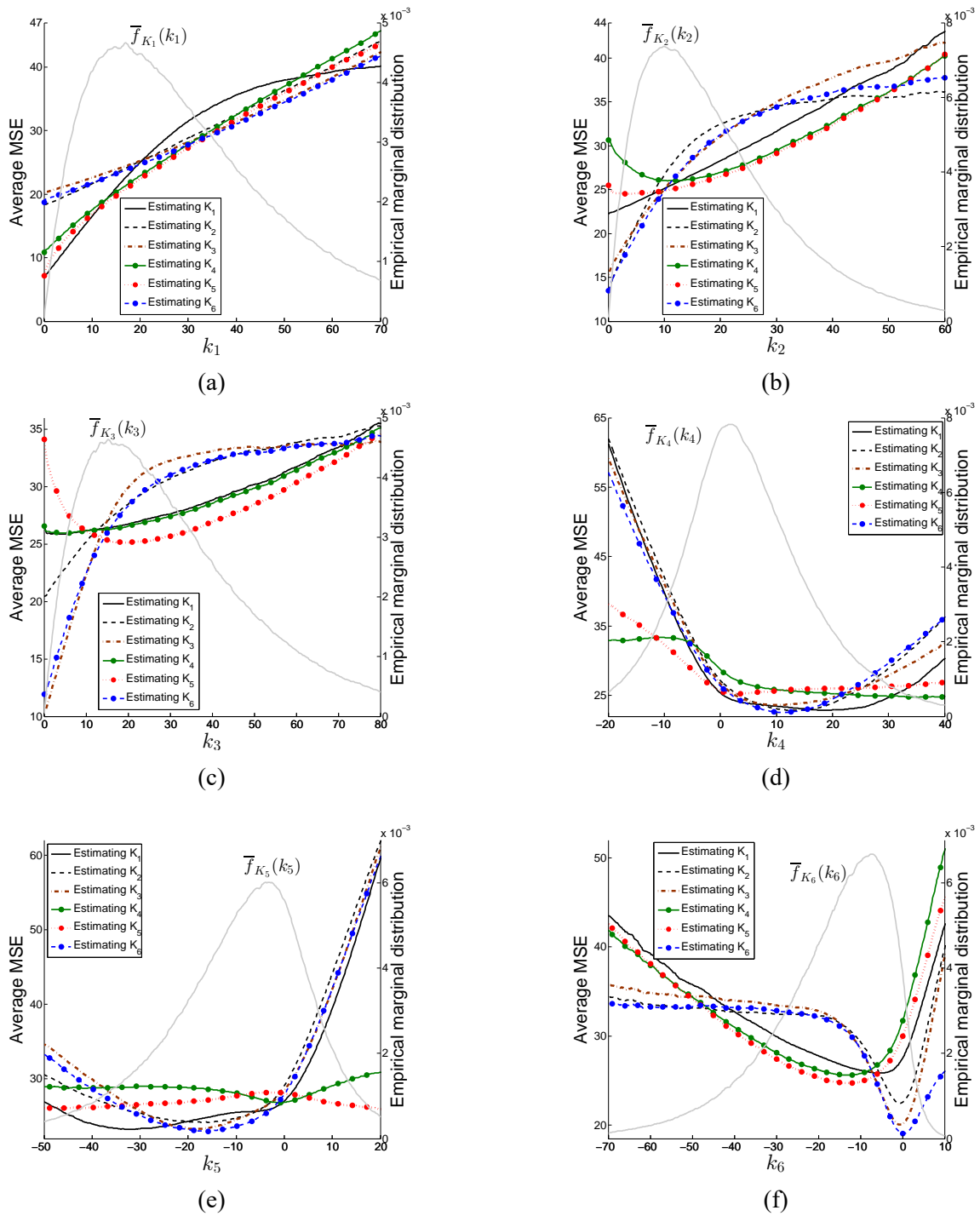
Figure 5.1. The conditional average MSE of determining each element of the covariance matrix over the space of possible realizations of the elements for $n = 4$. The background gray curves are empirical distributions. (a) Conditional average MSE given k_1. (b) Conditional average MSE given k_2. (c) Conditional average MSE given k_3. (d) Conditional average MSE given k_4. (e) Conditional average MSE given k_5. (f) Conditional average MSE given k_6.

137
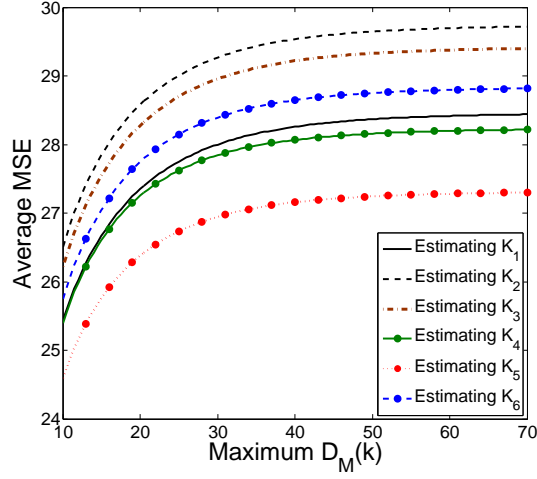
Figure 5.2. The effect of the Mahalanobis distance on the average MSE resulting from the determination of different elements of the covariance matrix with the Wishart prior and degree of freedom $n = 4$.

then re-arrange the distinct $\frac{p(p+1)}{2}$ random variables into a column vector of size $\frac{p(p+1)}{2}$.

We denote this random vector and its realizations by $\widetilde{\mathbf{K}}$ and $\widetilde{\mathbf{k}}$ respectively. We represent the

covariance and mean of $\widetilde{\mathbf{K}}$ by $\text{Cov}[\widetilde{\mathbf{K}}]$ and $\text{E}[\widetilde{\mathbf{k}}]$, respectively. The Mahalanobis distance

for a re-arranged realization $\widetilde{\mathbf{k}}$ is defined as

$$D_M(\widetilde{\mathbf{k}}) = \sqrt{(\widetilde{\mathbf{k}} - \text{E}[\widetilde{\mathbf{K}}])^T \text{Cov}^{-1}[\widetilde{\mathbf{K}}](\widetilde{\mathbf{k}} - \text{E}[\widetilde{\mathbf{K}}])}. \tag{5.53}$$

For $\text{Cov}[\widetilde{\mathbf{K}}]$, we use the sample covariance matrix. Moreover, the mean of a Wishart distribution with parameters n and Σ is $n\Sigma$, which should be re-arranged to be used as $\text{E}[\widetilde{\mathbf{K}}]$

in (5.53). Figure 5.2 presents the average MSE corresponding to the determination of each

element of $\mathbf{K}$ for realizations whose Mahalanobis distance is below a certain value. We

observe that as the maximum value for the Mahalanobis distance gets larger, the average

MSE increases and determining K_5, which has been chosen by experimental design, always

achieves the lowest average MSE.

138

5.3.2   Performance Evaluation for Sequential Experiments

We now evaluate the performance of experimental design for a sequence of experiments. For a covariance matrix of size 3×3 we need to conduct six experiments to fully characterize the covariance matrix. At each step we select the primary parameter and then incorporate its true value in the covariance matrix to find the resulting reduced uncertainty class. We then utilize this new uncertainty class to find the next parameter to be determined. We keep repeating this process until all unknown parameters are determined. With the sequential experiments, after determining the first element, the distribution of the remaining unknown elements of the covariance matrix given the value of the determined element is no longer Wishart and there is no known analytic form for it. Therefore, except the first experiment, where we use the equations in Section 5.2 to find conditional expectations, for the remaining experiments Monte Carlo simulations are employed to approximate the conditional expectations required for the experimental design process. To do so, we initially generate a sample pool $\mathscr{S}$ of covariance matrices for the given Wishart distribution. When the fist element is determined, we obtain its value from the assumed-to-be-true covariance matrix. To find the next experiment to be conducted, we only keep those matrices from the original sample pool $\mathscr{S}$ whose value of the element just determined is very close to that of the assumed-to-be-true covariance matrix and denote the new sample pool by $\mathscr{S}_1$. Now, to compute the conditional expectations required in the experimental design calculations for finding the second parameter to be determined, we employ a Monte Carlo approach: to find the conditional expectation given the value k of a certain unknown element $K(i, j)$ of the
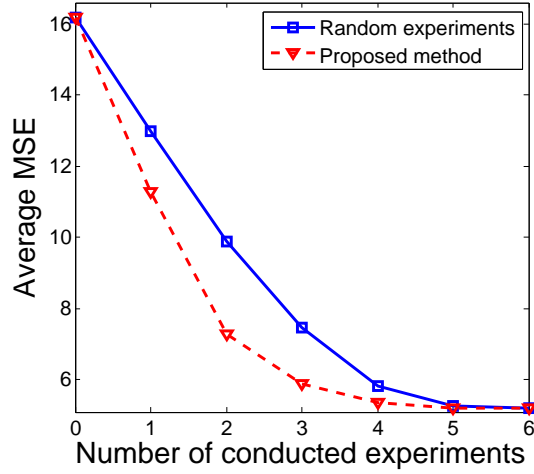
Figure 5.3. The average MSE obtained after conducting each experiment in a sequence of experiments for an unknown covariance matrix with the Wishart prior when experiments are chosen randomly or based on the proposed experimental design method.

covariance matrix, we take the average of those samples in $\mathscr{S}_1$ whose element $K(i,j)$ has a close value to the value k being considered for the conditional expectation. After taking the average of those samples, we use the entries of the average matrix as the conditional expectations given $K(i,j) = k$.

To evaluate the performance of experimental design, each time a parameter is chosen to be determined we get its value from the assumed-to-be-true covariance matrix, find the robust KL expression for the new reduced uncertainty class, and calculate the MSE of the robust KL expression relative to the underlying true covariance matrix using (5.52), where here μ, Θ, and θ_{j*} are the assumed-to-be-true covariance matrix, the uncertainty class prior to conducting the experiment, and the chosen parameter to be determined, respectively.

Consider a Wishart distribution with the scale matrix given in (5.50) and degree of freedom $n = 2$. We initially generate a sample pool of 4×10^7 covariance matrices and assume that the true covariance matrix is fixed during the sequential experiments to be one of

the generated matrices. Figure 5.3 shows the average MSE obtained after conducting each experiment using experimental design and compares it with randomly chosen experiments. The reported average MSE in the figure is averaged over 15,000 different true covariance matrices. Both curves begin from the same point and ultimately reach the same point because initially no experiment has been done and after conducting all experiments the covariance matrix is completely determined. The average MSE obtained via experimental design is always lower than the random selection policy. The decrease in the average MSE is sharper when the first few experiments are conducted based on the proposed method. Note that MSE reduction for the first few experiments is larger even when experiments are chosen randomly because random variables in a Wishart distribution are not independent and therefore determining the value of a random variable can provide information about the true value of the remaining random variables.

Table 5.1 presents the proportion of each element of the covariance matrix chosen to be determined for each step in the sequential experiments for the 15,000 different assumed-to-be-true covariance matrices. Observe that while element K_5 is always chosen for the first experiment, for the remaining experiments different elements are chosen. For example, for the third experiment T_3, elements K_1, K_2, K_3, and K_4 are chosen for 57%, 2%, 32%, and 9% of the assumed-to-be-true covariance matrices, respectively. This is because, keeping in mind that we run simulations over 15,000 different true covariance matrices, different realizations of a chosen parameter result in different posterior distributions, which are then used as the prior distributions for finding the next experiment. As the decision making

Table 5.1. The proportion of choosing each element for each experiment in a sequence of experiments.

	T_1	T_2	T_3	T_4	T_5	T_6
K_1	0	0.010	0.574	0.304	0.081	0.030
K_2	0	0	0.017	0.127	0.533	0.324
K_3	0	0.002	0.319	0.217	0.245	0.217
K_4	0	0	0.086	0.348	0.140	0.426
K_5	1	0	0	0	0	0
K_6	0	0.988	0.004	0.004	0.001	0.003

in our experimental design method depends on the prior distribution, for different true covariance matrices, different parameters, especially for intermediate experiments, might be chosen .

5.4 Blocked Covariance Matrix with Unknown Parameters

A common covariance model is a blocked matrix, in which each block corresponds to a group of correlated random variables, there is no dependency between random variables in different blocks, and within each block the variance and correlation do not change. For example, the following covariance matrix contains 2 blocks of size 2×2 and 3×3, where σ_1^2 and ρ_1 denote the variance and correlation for the first block and σ_2^2 and ρ_2 denote the

variance and correlation for the second block:

$$
\begin{bmatrix}
\sigma_1^2 & \rho_1\sigma_1^2 & 0 & 0 & 0 \\
\rho_1\sigma_1^2 & \sigma_1^2 & 0 & 0 & 0 \\
0 & 0 & \sigma_2^2 & \rho_2\sigma_2^2 & \rho_2\sigma_2^2 \\
0 & 0 & \rho_2\sigma_2^2 & \sigma_2^2 & \rho_2\sigma_2^2 \\
0 & 0 & \rho_2\sigma_2^2 & \rho_2\sigma_2^2 & \sigma_2^2
\end{bmatrix}.
$$

Among other applications, the model has been used to study feature selection in systems where blocks correspond to uncorrelated subsystems [115, 116], for instance, in genomics, where each block represents the correlation between genes within a pathway and genes from different pathways are assumed to be uncorrelated.

Assume that the parameters of the covariance matrix, which are the variance and correlation coefficient for each block, are not known. We apply experimental design to decide which parameter should be determined first to improve the performance of KL compression. Unknown parameters are uniformly distributed over a certain interval and different parameters are statistically independent from each other.

For a numerical illustration, assume a covariance matrix with two blocks of size 2×2 whose unknown parameters are uniformly distributed as follows (nominal intervals): $\sigma_1^2 \in [0.1, 4]$, $\rho_1 \in [-0.3, 0.3]$, $\sigma_2^2 \in [0.1, 3]$, and $\rho_2 \in [-0.1, 0.1]$. After compression, one random variable remains. In Figure 5.4, the interval of one uncertain parameter changes while other intervals are fixed to the nominal intervals. We then find which uncertain parameter is better to be determined first by looking at the experimental design values. For example, according to Figure 5.4 (a), if we change the interval of σ_1^2 such that $\sigma_1^2 \in$

$[0.1,\ \sigma^2_{1\,max}]$ and $\sigma^2_{1\,max}$ varies from 0.5 to 7, then the primary parameter is given by

$$\begin{cases} \rho_2 & \sigma^2_{1\,max} \leq 1.4 \\[2mm] \sigma^2_2 & 1.4 \leq \sigma^2_{1\,max} \leq 2.9 \\[2mm] \sigma^2_1 & 2.9 \leq \sigma^2_{1\,max} \leq 3.7 \\[2mm] \rho_1 & 3.7 \leq \sigma^2_{1\,max} \end{cases}. \tag{5.54}$$

In Figure 5.4 (e), the average MSE resulting from obtaining the chosen parameter is always lower than the MSE of obtaining other parameters. Suppose we change the interval of σ^2_2 such that $[0.1,\ \sigma^2_{2\,max}]$, where $\sigma^2_{2\,max}$ varies from 0.5 to 7. As seen from Figure 5.4 (c), when the interval of σ^2_2 is very small, the primary parameter is ρ_1, as the interval becomes larger; the primary parameter changes to σ^2_1, and for very large intervals, the primary parameter is σ^2_2.

Figure 5.5 presents the same type of simulations when the covariance matrix contains a block of size 2×2 (block 1) and a block of size 3×3 (block 2) and the unknown parameters have the following nominal uncertainty intervals: $\sigma^2_1 \in [0.1,\ 4]$, $\rho_1 \in [-0.3,\ 0.3]$, $\sigma^2_2 \in [0.1,\ 3]$, and $\rho_2 \in [0.01,\ 0.4]$. Now if $\sigma^2_1 \in [0.1,\ \sigma^2_{1\,max}]$ and change $\sigma^2_{1\,max}$ from 0.5 to 7, then the primary parameter is given by

$$\begin{cases} \sigma^2_1 \ \text{ or } \ \rho_1 & \sigma^2_{1\,max} \leq 1 \\[2mm] \sigma^2_2 & 1 \leq \sigma^2_{1\,max} \leq 4.2 \\[2mm] \sigma^2_1 & 4.2 \leq \sigma^2_{1\,max} \leq 5.3 \\[2mm] \rho_1 & 5.3 \leq \sigma^2_{1\,max} \end{cases}. \tag{5.55}$$
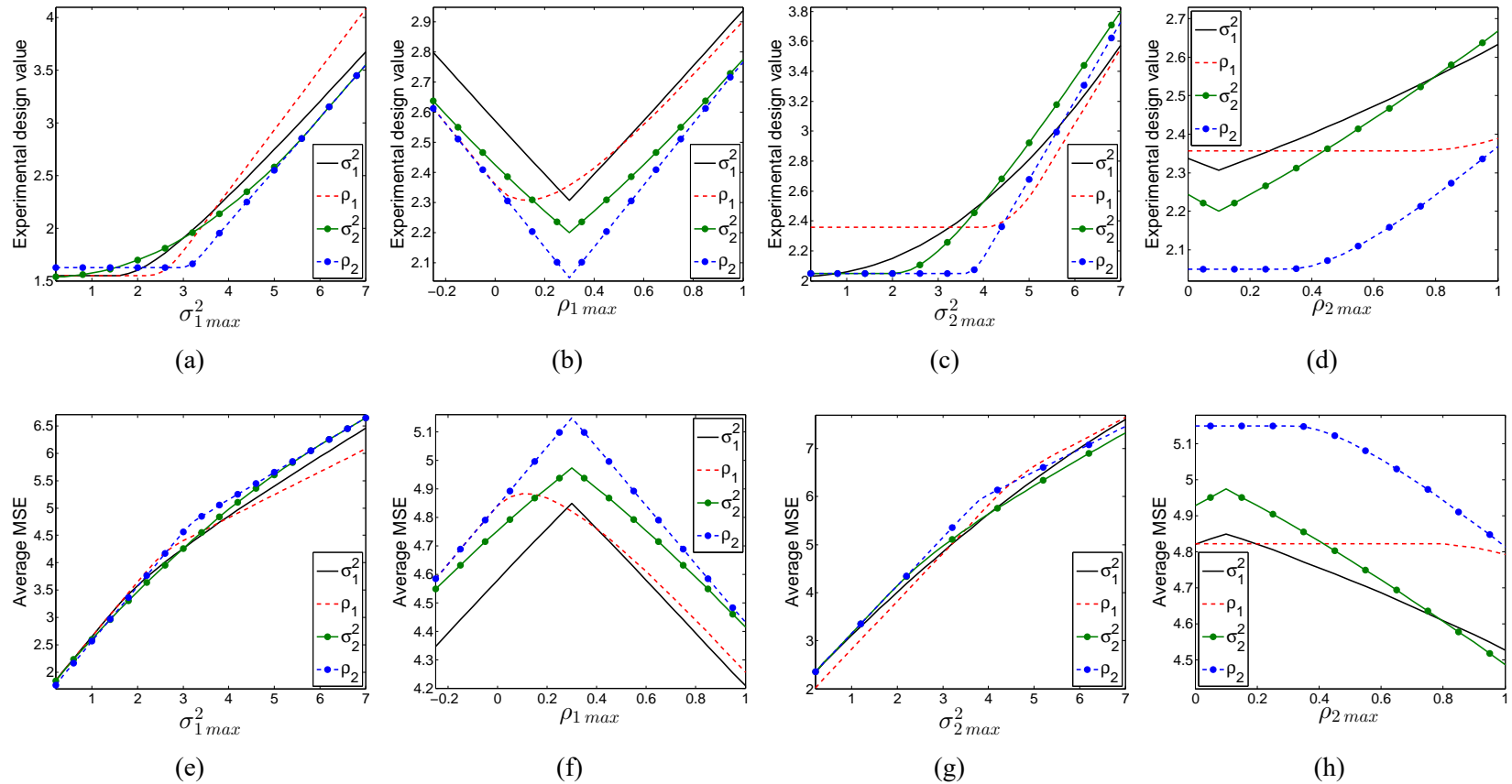
Figure 5.4. The effect of increasing the uncertainty interval of a particular uncertain parameter on the experimental design performance for an unknown covariance matrix with two blocks of size 2×2. (a) Experimental design values as $\sigma^2_{1\,max}$ increases. (b) Experimental design values as $\rho_{1\,max}$ increases. (c) Experimental design values as $\sigma^2_{2\,max}$ increases. (d) Experimental design values as $\rho_{2\,max}$ increases. (e) The average MSE as $\sigma^2_{1\,max}$ increases. (f) The average MSE as $\rho_{1\,max}$ increases. (g) The average MSE as $\sigma^2_{2\,max}$ increases. (h) The average MSE as $\rho_{2\,max}$ increases.
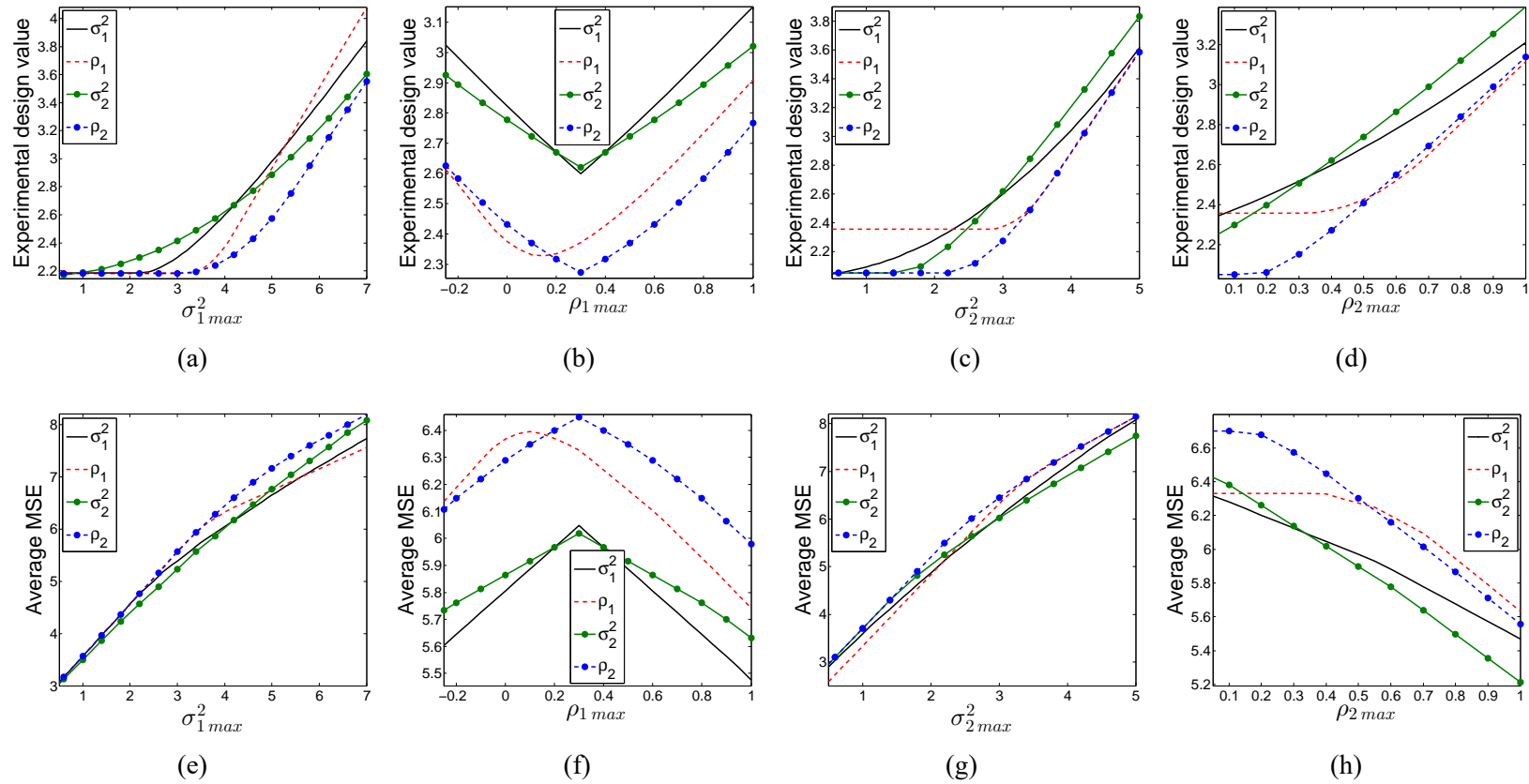
Figure 5.5. The effect of increasing the uncertainty interval of a particular uncertain parameter on the experimental design performance for an unknown covariance matrix with two blocks of size 2×2 and 3×3. (a) Experimental design values as $\sigma^2_{1\,max}$ increases. (b) Experimental design values as $\rho_{1\,max}$ increases. (c) Experimental design values as $\sigma^2_{2\,max}$ increases. (d) Experimental design values as $\rho_{2\,max}$ increases. (e) The average MSE as $\sigma^2_{1\,max}$ increases. (f) The average MSE as $\rho_{1\,max}$ increases. (g) The average MSE as $\sigma^2_{2\,max}$ increases. (h) The average MSE as $\rho_{2\,max}$ increases.
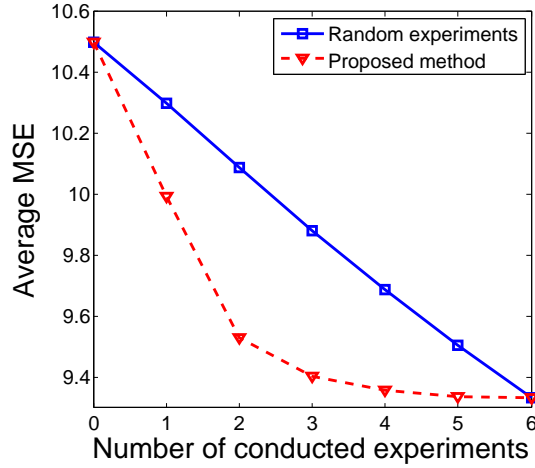
Figure 5.6. The average MSE obtained after conducting each experiment in a sequence of experiments for an unknown covariance matrix with disjoint blocks model when experiments are chosen randomly or based on the proposed experimental design method.

We also evaluate the performance for sequential experiments. To do so, we assume a covariance matrix that has one 2×2 block (block 1) and two 3×3 blocks (blocks 2 and 3). The parameters are distributed as follows: $\sigma_1^2 \in [0.1, \ 4]$, $\rho_1 \in [-0.3, \ 0.3]$, $\sigma_2^2 \in [0.1, \ 3]$, $\rho_2 \in [0.01, \ 0.4]$, $\sigma_3^2 \in [0.5, \ 2]$, and $\rho_3 \in [0.5, \ 0.9]$. One random variable remains after compression. The framework for running simulations is the same as done for the Wishart prior case except that here there is no need to perform Monte Carlo simulations to find conditional expectations after conducting the first experiment because parameters are uniformly distributed and statistically independent. Therefore, the posterior distribution after determining each parameter can be found analytically via multiplying the marginal distributions of the remaining unknown parameters. Figure 5.6 shows the average MSE obtained after conducting each experiment chosen either via experimental design or randomly. Note the large difference between experimental design and random experiments for the first few

147

Table 5.2. The proportion of choosing each element for each experiment in a sequence of experiments for a blocked covariance matrix.

	T_1	T_2	T_3	T_4	T_5	T_6
σ_1	1	0	0	0	0	0
ρ_1	0	0.43	0.18	0.27	0.07	0.05
σ_2	0	0.43	0.31	0.21	0.06	0
ρ_2	0	0	0	0.25	0.55	0.20
σ_3	0	0.14	0.51	0.22	0.03	0.10
ρ_3	0	0	0	0.05	0.29	0.65

experiments. Note also that here, unlike the simulations for the Wishart prior case, the average MSE reduction after conducting each experiment in the random selection policy is almost the same because parameters are independent from each other.

Finally, Table 5.2 presents the proportion of times that each unknown parameter is chosen based on all experiments. The primary parameter for the first experiment is always σ_1^2 but for the remaining experiments the primary parameter depends on the outcomes of preceding experiments.

5.5 Discussion

In this chapter, we studied optimal covariance uncertainty reduction for KL compression. Having found the optimal parameter to be determined first in (5.17), we then turn to the case of an unknown covariance matrix with a Wishart prior. A recursive method is developed to solve the problem for an arbitrary size matrix. Simulations are then provided to demonstrate the experimental design method, including its advantage over randomly

selected experiments, for the Wishart model and for a uniform model with a blocked co-variance matrix. This comparison is critical because the pragmatic goal of experimental design is to obtain a large gain in relevant knowledge with as few experiments as possible. The sequential graphs in Figures 5.3 and 5.6 demonstrate this advantage, with the latter looking very similar to a corresponding graph in [41] for optimal experimental design in the case of gene regulatory networks, where the objective is drug intervention to reduce the long-term risk of disease. Owing to the many applications of canonical expansions in science and engineering, canonical-expansion-based experimental design can be helpful in many problems where accessibility to data is limited or costly. Admittedly, there are computational issues and these need to be addressed within specific applications, as they have been in the case of gene regulatory networks [42].

6   CONCLUSION

In this dissertation, we introduced an objective-based experimental design framework for uncertainty reduction in complex dynamical systems. The standard engineering approach to operator design is to construct a mathematical model and then find an operator in a class of feasible operators that minimizes a cost function relative to an objective. In the presence of uncertainty, the aim becomes to find a robust operator in the sense that its performance is acceptable across the uncertainty class. Focusing on the Bayesian setting, model uncertainty negatively impacts operator performance relative to the true model; however, model uncertainty in itself, such as entropy, is not of prime importance. What matters is the extent to which the uncertainty results in the loss of operator performance. In that regard, we developed an objective-based experimental design framework that takes into account the performance degradation of the operator due to the presence of uncertainty.

The proposed experimental design method utilizes the concept of the mean objective cost of uncertainty which captures the pertinent uncertainty in the model. In brief, to find out which uncertainty is better to be estimated first, we calculate the expected remaining MOCU given each uncertain parameter is estimated. The parameter whose expected remaining MOCU is minimum is selected for estimation.

In Chapter 2, we explained how experimental design can be used for uncertainty reduction in gene regulatory networks. Through simulations on both synthetic and real gene regulatory networks, we showed that the proposed strategy significantly outperforms both

the random selection policy and the selection policy based on pure statistical information, such as entropy.

In the following chapter, we presented a computationally efficient method to mitigate the computational burden of the optimal experimental design for gene regulatory networks. We used a network reduction scheme to approximately estimate the MOCU at a reduced computational cost without disrupting the ranking of potential experiments. Our network reduction scheme is analogous to the reduction of gene regulatory networks to facilitate design of optimal controllers, except that reduction must be accomplished in such a way as to preserve (to the extent possible) the MOCU calculations. Simulation results verified that the proposed approximate method clearly outperforms random selection and is comparable to the optimal experimental design method.

To further explore the utility of the proposed framework in systems biology, in an on-going project, we are currently working on an experimental design methodology to reduce dynamics uncertainty in a class of dynamical gene network models at the process level [117, 118]. In these models, network dynamics can be updated in different ways, thereby giving multiple dynamic trajectories, that is, dynamics uncertainty [119]. The goal is to find the experiment yielding the largest reduction of the pertinent dynamics uncertainty – equivalently, the number of dynamic trajectories in the network dynamical model – that affects operational cost.

In the second half of this dissertation, we focused on canonical expansions. Canonical expansions are convenient expressions that can facilitate the study of random functions.

Canonical expansions play a major role in many engineering applications. We developed a rigorous mathematical methodology for the experimental design in the context of canonical expansions and solved it for three important signal processing applications: linear filtering, signal detection, and signal compression. A crucial step in the proposed experimental design for canonical expansions is to derive closed-form solutions for the intrinsically Bayesian robust operators. In the case of Wiener filtering, this can be done via the concept of effective characteristics, or for the KL compression, the intrinsically Bayesian robust KL compression is shown to be obtained using the expected covariance matrix.

As a final comment, we should emphasize the far-reaching utility of the proposed experimental design framework for various disciplines, including signal processing, engineering of new materials, communication systems, and drug design, which may require mathematical models involving covariance matrices, regression models, graphical models, systems of differential equations, or other parameterized mathematical structures. In this dissertation, we solved it for gene regulatory networks and canonical expansions, but we believe that it can be further applied to other engineering problems. In this regard, we are currently investigating the application of the proposed framework for the problem of materials design and discovery where the outstanding challenge is to reduce the number of experiments required to find new materials with desired properties.

REFERENCES

[1] A. C. Atkinson and A. N. Donev, *Optimum Experimental Designs*. Oxford: Oxford University Press, 1992.

[2] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Adminitration, Harvard University, 1961.

[3] V. V. Fedorov, *Theory of Optimal Experiments*. Elsevier Science, 1972.

[4] K. Smith, "On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations," *Biometrika*, vol. 12, no. 1/2, pp. 1–85, 1918.

[5] A. Wald, "On the efficient design of statistical investigations," *The Annals of Mathematical Statistics*, vol. 14, no. 2, pp. 134–140, 1943.

[6] G. Elfving, "Optimum allocation in linear regression theory," *The Annals of Mathematical Statistics*, vol. 23, no. 2, pp. 255–262, 1952.

[7] J. Kiefer and J. Wolfowitz, "Optimum designs in regression problems," *The Annals of Mathematical Statistics*, pp. 271–294, 1959.

[8] H. Chernoff, "Locally optimal designs for estimating parameters," *The Annals of Mathematical Statistics*, pp. 586–602, 1953.

[9] D. V. Lindley, "On a measure of the information provided by an experiment," *The Annals of Mathematical Statistics*, pp. 986–1005, 1956.

[10] M. Stone, "Application of a measure of information to the design and comparison of regression experiments," *The Annals of Mathematical Statistics*, pp. 55–70, 1959.

[11] M. H. DeGroot, "Uncertainty, information, and sequential experiments," *The Annals of Mathematical Statistics*, pp. 404–419, 1962.

[12] M. H. DeGroot, "Concepts of information based on utility," in *Recent Developments in the Foundations of Utility and Risk Theory*, pp. 265–275, Springer, 1986.

[13] J. M. Bernardo, "Expected information as expected utility," *The Annals of Statistics*, pp. 686–690, 1979.

[14] P. Sebastiani and H. P. Wynn, "Maximum entropy sampling and optimal Bayesian experimental design," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 62, no. 1, pp. 145–157, 2000.

[15] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[16] P. J. Huber, "A robust version of the probability ratio test," *The Annals of Mathematical Statistics*, vol. 36, no. 6, pp. 1753–1758, 1965.

[17] W. L. Root, "Communications through unspecified additive noise," *Information and Control*, vol. 4, no. 1, pp. 15–29, 1961.

[18] L.-H. Zetterberg, "Signal detection under noise interference in a game situation," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 47–52, 1962.

[19] S. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proceedings of the IEEE*, vol. 73, no. 3, pp. 433–481, 1985.

[20] V. Kuznetsov, "Stable detection when the signal and spectrum of normal noise are inaccurately known," *Telecommunications and Radio Engineering*, vol. 30, p. 31, 1976.

[21] S. A. Kassam and T. L. Lim, "Robust Wiener filters," *Journal of the Franklin Institute*, vol. 304, no. 4, pp. 171–185, 1977.

[22] H. V. Poor, "On robust Wiener filtering," *IEEE Transactions on Automatic Control*, vol. 25, no. 3, pp. 531–536, 1980.

[23] K. S. Vastola and H. V. Poor, "Robust Wiener-Kolmogorov theory," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 316–327, 1984.

[24] V. Poor and D. P. Looze, "Minimax state estimation for linear stochastic systems with noise uncertainty," *IEEE Transactions on Automatic Control*, vol. 26, no. 4, pp. 902–906, 1981.

[25] S. Verdu and H. V. Poor, "Minimax linear observers and regulators for stochastic systems with uncertain second-order statistics," *IEEE Transactions on Automatic Control*, vol. 29, no. 6, pp. 499–511, 1984.

[26] H. V. Poor, "Robust matched filters," *IEEE Transactions on Information Theory*, vol. 29, no. 5, pp. 677–687, 1983.

[27] C.-T. Chen and S. A. Kassam, "Robust multiple-input matched filtering: frequency and time-domain results," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 812–821, 1985.

[28] S. Verdu and H. V. Poor, "On minimax robustness: A general approach and applications," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 328–340, 1984.

[29] A. M. Grigoryan and E. R. Dougherty, "Bayesian robust optimal linear filters," *Signal Processing*, vol. 81, no. 12, pp. 2503–2521, 2001.

[30] L. Dalton and E. Dougherty, "Intrinsically optimal Bayesian robust filtering," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 657–670, 2014.

[31] R. Dehghannasiri, M. S. Esfahani, and E. R. Dougherty, "A Bayesian framework for robust Kalman filtering under uncertain noise statistics," in *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, 2016.

[32] L. A. Dalton and E. R. Dougherty, "Optimal classifiers with minimum expected error within a Bayesian framework — Part I: Discrete and Gaussian models," *Pattern Recognition*, vol. 46, no. 5, pp. 1288–1300, 2013.

[33] E. R. Dougherty and Y. Chen, "Robust optimal granulometric bandpass filters," *Signal Processing*, vol. 81, no. 7, pp. 1357–1372, 2001.

[34] A. M. Grigoryan and E. R. Dougherty, "Design and analysis of robust binary filters in the context of a prior distribution for the states of nature," *Journal of Mathematical Imaging and Vision*, vol. 11, no. 3, pp. 239–254, 1999.

[35] L. A. Dalton and E. R. Dougherty, "Bayesian minimum mean-square error estimation for classification error — Part I: Definition and the Bayesian MMSE error estimator for discrete classification," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 115–129, 2011.

[36] A. C. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2222–2233, 2004.

[37] H. Chen and P. K. Varshney, "Theory of the stochastic resonance effect in signal detection — Part II: Variable detectors," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5031–5041, 2008.

[38] J. J. Martin, *Bayesian Decision Problems and Markov Chains*. Wiley, 1967.

[39] R. Pal, A. Datta, and E. R. Dougherty, "Bayesian robustness in the control of gene regulatory networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3667–3678, 2009.

[40] B. J. Yoon, X. Qian, and E. R. Dougherty, "Quantifying the objective cost of uncertainty in complex dynamical systems," *IEEE Transactions on Signal Processing*, vol. 61, no. 9, pp. 2256–2266, 2013.

[41] R. Dehghannasiri, B.-J. Yoon, and E. Dougherty, "Optimal experimental design for gene regulatory networks in the presence of uncertainty," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 12, pp. 938–950, July 2015.

[42] R. Dehghannasiri, B.-J. Yoon, and E. R. Dougherty, "Efficient experimental design for uncertainty reduction in gene regulatory networks," *BMC Bioinformatics*, vol. 16, no. Suppl 13, p. S2, 2015.

[43] R. Dehghannasiri, X. Qian, and E. R. Dougherty, "Intrinsically Bayesian robust Karhunen-Loève compression," *Submitted*.

[44] C. Christensen, J. Thakar, and R. Albert, "Systems-level insights into cellular regulation: inferring, analysing, and modelling intracellular networks," *IET Systems Biology*, vol. 1, no. 2, pp. 61–77, 2007.

[45] I. Shmulevich, E. R. Dougherty, and W. Zhang, "Gene perturbation and intervention in probabilistic Boolean networks," *Bioinformatics*, vol. 18, no. 10, pp. 1319–1331, 2002.

[46] I. Shmulevich, E. R. Dougherty, and W. Zhang, "Control of stationary behavior in probabilistic Boolean networks by means of structural intervention," *Biological Systems*, vol. 10, no. 4, pp. 431–446, 2002.

[47] I. Shmulevich, E. R. Dougherty, and W. Zhang, "From Boolean to probabilistic Boolean networks as models of genetic regulatory networks," *Proceedings of the IEEE*, vol. 90, no. 11, pp. 1778–1792, 2002.

[48] Y. Xiao and E. R. Dougherty, "The impact of function perturbations in Boolean networks," *Bioinformatics*, vol. 23, no. 10, pp. 1265–1273, 2007.

[49] X. Qian and E. R. Dougherty, "Effect of function perturbation on the steady-state distribution of genetic regulatory networks: Optimal structural intervention," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4966–4976, 2008.

[50] R. Pal, A. Datta, and E. R. Dougherty, "Robust intervention in probabilistic Boolean networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1280–1294, 2008.

[51] G. Chesi, "On the steady states of uncertain genetic regulatory networks," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 42, pp. 1020–1024, July 2012.

[52] Z. Wang, H. Wu, J. Liang, J. Cao, and X. Liu, "On modeling and state estimation for genetic regulatory networks with polytopic uncertainties," *IEEE Transactions on NanoBioscience*, vol. 12, pp. 13–20, March 2013.

[53] F.-X. Wu, "Global and robust stability analysis of genetic regulatory networks with time-varying delays and parameter uncertainties," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 5, pp. 391–398, Aug 2011.

[54] S. Z. Denic, B. Vasic, C. D. Charalambous, and R. Palanivelu, "Robust control of uncertain context-sensitive probabilistic Boolean networks," *IET Systems Biology*, vol. 3, pp. 279–295, July 2009.

[55] C.-H. Yeang, H. C. Mak, S. McCuine, C. Workman, T. Jaakkola, and T. Ideker, "Validation and refinement of gene-regulatory pathways on a network of physical interactions," *Genome Biology*, vol. 6, no. 7, p. R62, 2005.

[56] T. E. Ideker, V. Thorsson, and R. M. Karp, "Discovery of regulatory interactions through perturbation: Inference and experimental design," in *Proceedings of the Pacific Symposium on Biocomputing*, 2000.

[57] R. D. King, K. E. Whelan, F. M. Jones, P. G. K. Reiser, C. H. Bryant, S. H. Muggleton, D. B. Kell, and S. G. Oliver, "Functional genomic hypothesis generation and experimentation by a robot scientist," *Nature*, no. 6971, p. 247–252, 2004.

[58] A. Almudevar and P. Salzman, "Using a Bayesian posterior density in the design of perturbation experiments for network reconstruction," in *Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pp. 1–7, Nov. 2005.

[59] S. A. Kauffman, *The Origins of Order*. Oxford: Oxford University Press, 1993.

[60] P. Trairatphisan, A. Mizera, J. Pang, A. A. Tantar, J. Schneider, and T. Sauter, "Recent development and biomedical applications of probabilistic Boolean networks," *Cell Communication and Signaling*, vol. 4, no. 6, pp. 1–25, 2013.

[61] A. Wuensche, "Genomic regulation modeled as a network with basins of attraction," in *Pacific Symposium on Biocomputing*, vol. 3, p. 44, 1998.

[62] S. Huang, "Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery," *Journal of Molecular Medicine*, vol. 77, no. 6, pp. 469–480, 1999.

[63] Y. Zhang, M. Qian, Q. Ouyang, M. Deng, F. Li, and C. Tang, "Stochastic model of yeast cell-cycle network," *Physica D: Nonlinear Phenomena*, vol. 219, no. 1, pp. 35–39, 2006.

[64] M. Flöttmann, T. Scharp, and E. Klipp, "A stochastic model of epigenetic dynamics in somatic cell reprogramming," *Frontiers in Physiology*, vol. 3, 2012.

[65] M. I. Davidich and S. Bornholdt, "Boolean network model predicts cell cycle sequence of fission yeast," *PloS One*, vol. 3, no. 2, p. e1672, 2008.

[66] A. Faure, A. Naldi, C. Chaouiya, and D. Thieffry, "Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle," *Bioinformatics*, vol. 22, no. 14, pp. 124–131, 2006.

[67] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, vol. 356. van Nostrand Princeton, NJ, 1960.

[68] P. J. Schweitzer, "Perturbation theory and finite Markov chains," *Journal of Applied Probability*, pp. 401–413, 1968.

[69] A. Datta, R. Pal, A. Choudhary, and E. R. Dougherty, "Control approaches for probabilistic gene regulatory networks-what approaches have been developed for addreassinig the issue of intervention?," *IEEE Signal Processing Magazine*, vol. 24, no. 1, pp. 54–63, 2007.

[70] N. Bouaynaya, R. Shterenberg, and D. Schonfeld, "Inverse perturbation for optimal intervention in gene regulatory networks," *Bioinformatics*, vol. 27, no. 1, pp. 103–110, 2011.

[71] G. Vahedi, B. Faryabi, J. Chamberland, A. Datta, and E. R. Dougherty, "Intervention in gene regulatory networks via a stationary mean-first-passage-time control policy," *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 2319–2331, Oct 2008.

[72] W.-K. Ching, S.-Q. Zhang, Y. Jiao, T. Akutsu, N.-K. Tsing, and A. Wong, "Optimal control policy for probabilistic Boolean networks with hard constraints," *IET Systems Biology*, vol. 3, no. 2, pp. 90–99, 2009.

[73] C. Yang, C. Wai-Ki, T. Nam-Kiu, and L. Ho-Yin, "On finite-horizon control of genetic regulatory networks with multiple hard-constraints," *BMC Systems Biology*, vol. 4, no. Suppl 2, p. S14, 2010.

[74] K. Lau, S. Ganguli, and C. Tang, "Function constrains network architecture and dynamics: A case study on the yeast cell cycle Boolean network," *Physical Review E*, vol. 75, p. 051907, 2007.

[75] E. Batchelor, A. Loewer, and G. Lahav, "The ups and downs of p53: understanding protein dynamics in single cells," *Nature Reviews Cancer*, vol. 9, pp. 371–377, Apr. 2009.

[76] R. A. Weinberg, *The Biology of Cancer*. New York: Garland Science, 2007.

[77] R. K. Layek, A. Datta, and E. R. Dougherty, "From biological pathways to regulatory networks," *Molecular BioSystems*, vol. 7, pp. 843–851, 2011.

[78] X. Qian, N. Ghaffari, I. Ivanov, and E. R. Dougherty, "State reduction for network intervention in probabilistic Boolean networks," *Bioinformatics*, vol. 26, pp. 3098–3104, Dec 2010.

[79] I. Ivanov, P. Simeonov, N. Ghaffari, X. Qian, and E. R. Dougherty, "Selection policy-induced reduction mappings for Boolean networks," *IEEE Transactions on Signal Processing*, vol. 58, pp. 4871–4882, Sept 2010.

[80] I. Ivanov, R. Pal, and E. R. Dougherty, "Dynamics preserving size reduction mappings for probabilistic Boolean networks," *IEEE Transactions on Signal Processing*, vol. 55, pp. 2310–2322, May 2007.

[81] I. O. Ryzhov, W. B. Powell, and P. I. Frazier, "The knowledge gradient algorithm for a general class of online learning problems," *Operations Research*, vol. 60, no. 1, pp. 180–195, 2012.

[82] N. Ghaffari, I. Ivanov, X. Qian, and E. R. Dougherty, "A CoD-based reduction algorithm for designing stationary control policies on Boolean networks," *Bioinformatics*, vol. 26, no. 12, pp. 1556–1563, 2010.

[83] E. R. Dougherty, S. Kim, and Y. Chen, "Coefficient of determination in nonlinear signal processing," *Signal Processing*, vol. 80, no. 10, pp. 2219–2235, 2000.

[84] D. C. Martins, U. M. Braga-Neto, R. F. Hashimoto, M. L. Bittner, and E. R. Dougherty, "Intrinsically multivariate predictive genes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 3, pp. 424–439, 2008.

[85] S. Bornholdt, "Boolean network models of cellular regulation: prospects and limitations," *Journal of the Royal Society Interface*, vol. 5, no. Suppl 1, pp. S85–S94, 2008.

[86] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang, "The yeast cell-cycle network is robustly designed," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 14, pp. 4781–4786, 2004.

[87] Y. Wu, X. Zhang, J. Yu, and Q. Ouyang, "Identification of a topological characteristic responsible for the biological robustness of regulatory networks," *PLoS Computational Biology*, vol. 5, no. 7, p. e1000442, 2009.

[88] J. Hua, C. Sima, M. Cypert, G. C. Gooden, S. Shack, L. Alla, E. A. Smith, J. M. Trent, E. R. Dougherty, and M. L. Bittner, "Tracking transcriptional activities with

high-content epifluorescent imaging," *Journal of Biomedical Optics*, vol. 17, no. 4, pp. 0460081–04600815, 2012.

[89] M. S. Esfahani and E. R. Dougherty, "Incorporation of biological pathway knowledge in the construction of priors for optimal Bayesian classification," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 11, no. 1, pp. 202–218, 2014.

[90] D. E. Levy and J. E. Darnell, "Stats: transcriptional control and biological impact," *Nature Reviews Molecular Cell Biology*, vol. 3, no. 9, pp. 651–662, 2002.

[91] T. Kusaba, T. Nakayama, K. Yamazumi, Y. Yakata, A. Yoshizaki, K. Inoue, T. Nagayasu, and I. Sekine, "Activation of stat3 is a marker of poor prognosis in human colorectal cancer," *Oncology Reports*, vol. 15, no. 6, pp. 1445–1451, 2006.

[92] X. T. Ma, S. Wang, Y. J. Ye, R. Y. Du, Z. R. C, and M. Somsouk, "Constitutive activation of stat3 signaling pathway in human colorectal carcinoma," *World Journal of Gastroenterology*, vol. 10, no. 11, pp. 1569–1573, 2004.

[93] L. Klampfer, "Signal transducers and activators of transcription (stats): Novel targets of chemopreventive and chemotherapeutic drugs," *Current Cancer Drug Targets*, vol. 6, no. 2, pp. 107–121, 2006.

[94] M. S. Esfahani and E. R. Dougherty, "An optimization-based framework for the transformation of incomplete biological knowledge into a probabilistic structure

and its application to the utilization of gene/protein signaling pathways in discrete phenotype classification," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 12, no. 6, pp. 1304–1321, 2015.

[95] V. Pugachev, *Theory of Random Functions and Its Applications to Control Problems*. Reading, MA, USA: Pergamon Press; [U.S.A. ed. distributed by] Addison-Wesley Pub. Co., 1965.

[96] K. Karhunen, *Ueber Lineare Methoden in der Wahrscheinlichkeitsrechnung*. Annales Academiae scientiarum Fennicae. Series A. 1, Mathematica-physica, 1947.

[97] E. R. Dougherty, *Random Processes for Image and Signal Processing*. SPIE Optical Engineering Press, 1999.

[98] V. Tuzlukov, *Signal Detection Theory*. Springer Science & Business Media, 2001.

[99] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. Springer Science & Business Media, 2008.

[100] D. C. Youla, "The use of the method of maximum likelihood in estimating continuous-modulated intelligence which has been corrupted by noise," *Transactions of the IRE Professional Group on Information Theory*, vol. 3, pp. 90–105, March 1954.

[101] J. Thomas and E. Wong, "On the statistical theory of optimum demodulation," *IRE Transactions on Information Theory*, vol. 6, pp. 420–425, September 1960.

[102] S. Watanabe, "Karhunen-loeve expansion and factor analysis, theoretical remarks and applications," in *Transactions of the 4th Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, 1965.

[103] Y.-S. Chien and K.-S. Fu, "On the generalized karhunen-loève expansion," *IEEE Transactions on Information Theory*, vol. 13, no. 3, pp. 518–520, 1967.

[104] J. Brown, Jr, "Mean square truncation error in series expansions of random functions," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 1, pp. 28–32, 1960.

[105] A. K. Jain, "Image data compression: A review," *Proceedings of the IEEE*, vol. 69, no. 3, pp. 349–389, 1981.

[106] L. Torres-Urgell and R. L. Kirlin, "Adaptive image compression using karhunen-loeve transform," *Signal Processing*, vol. 21, no. 4, pp. 303–313, 1990.

[107] M. Effros and P. Chou, "Weighted universal transform coding: Universal image compression with the karhunen-loeve transform," in *Proceedings of International Conference on Image Processing*, vol. 2, pp. 61–64, IEEE, 1995.

[108] H.-H. Kang, D.-H. Shin, and E.-S. Kim, "Compression scheme of sub-images using karhunen-loeve transform in three-dimensional integral imaging," *Optics Communications*, vol. 281, no. 14, pp. 3640–3647, 2008.

[109] R. Temam, *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, vol. 68. Springer Science & Business Media, 2012.

[110] H. Banks, R. C. Del Rosario, and R. C. Smith, "Reduced-order model feedback control design: numerical implementation in a thin shell model," *IEEE Transactions on Automatic Control*, vol. 45, no. 7, pp. 1312–1324, 2000.

[111] G. Berkooz, P. Holmes, and J. L. Lumley, "The proper orthogonal decomposition in the analysis of turbulent flows," *Annual Review of Fluid Mechanics*, vol. 25, no. 1, pp. 539–575, 1993.

[112] M. Kirby, J. Boris, and L. Sirovich, "A proper orthogonal decomposition of a simulated supersonic shear layer," *International Journal for Numerical Methods in Fluids*, vol. 10, no. 4, pp. 411–428, 1990.

[113] O. P. Le Maître, M. T. Reagan, H. N. Najm, R. G. Ghanem, and O. M. Knio, "A stochastic projection method for fluid flow: II. Random process," *Journal of Computational Physics*, vol. 181, no. 1, pp. 9–44, 2002.

[114] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, vol. 197. John Wiley & Sons, 2009.

[115] J. Hua, Z. Xiong, J. Lowey, E. Suh, and E. R. Dougherty, "Optimal number of features as a function of sample size for various classification rules," *Bioinformatics*, vol. 21, no. 8, pp. 1509–1515, 2005.

[116] J. Hua, W. D. Tembe, and E. R. Dougherty, "Performance of feature-selection methods in the classification of high-dimension data," *Pattern Recognition*, vol. 42, no. 3, pp. 409–424, 2009.

[117] D. N. Mohsenizadeh, R. Dehghannasiri, and E. R. Dougherty, "Optimal dynamics uncertainty reduction in gene networks in the presence of experimental error," in *Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 581–582, IEEE, 2016.

[118] D. N. Mohsenizadeh, R. Dehghannasiri, and E. R. Dougherty, "Optimal objective-based experimental design for uncertain dynamical gene networks with experimental error," *Submitted*.

[119] D. N. Mohsenizadeh, J. Hua, M. Bittner, and E. R. Dougherty, "Dynamical modeling of uncertain interaction-based genomic networks," *BMC Bioinformatics*, vol. 16, no. 13, p. S3, 2015.

APPENDIX A

AN ILLUSTRATIVE EXAMPLE TO DEMONSTRATE THE

APPROXIMATE METHOD IN CHAPTER 3

In this supplemental document we provide an illustrative example to demonstrate the algorithm used for the approximate experimental design proposed.

We consider a 3-gene toy network as shown in Figure A.1. This network consists of three genes $\{X_1, X_2, X_3\}$ and three regulations. We assume that the activating regulation from gene X_2 to gene X_3 and the suppressive regulation from gene X_1 to gene X_3 are unknown and denote them by θ_1 and θ_2 respectively. Each uncertain parameter can take two values: 1 for being activating and 2 for being suppressive. Uncertainty class Θ as shown in Figure A.2 contains 4 different networks: $\{\theta^1, \theta^2, \theta^3, \theta^4\}$ such that:

$$\begin{cases} \theta^1 : (\theta_1 = 1, \theta_2 = 1) \\ \theta^2 : (\theta_1 = 1, \theta_2 = 2) \\ \theta^3 : (\theta_1 = 2, \theta_2 = 1) \\ \theta^4 : (\theta_1 = 2, \theta_2 = 2) \end{cases}$$

Let us assume that the probability density function governing the uncertainty class is uniform and two uncertain parameters are independent from each other. Therefore, all networks within Θ are equally likely having probability $1/4$, i.e., $P(\theta^i) = 1/4$.

The first step in the proposed experimental design method is to decide which gene is better to be removed. In this example, we assume that the expression state of gene X_3
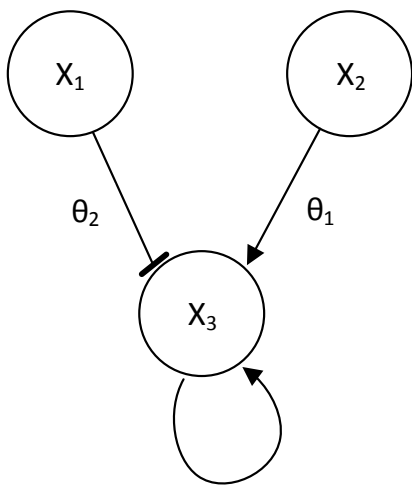
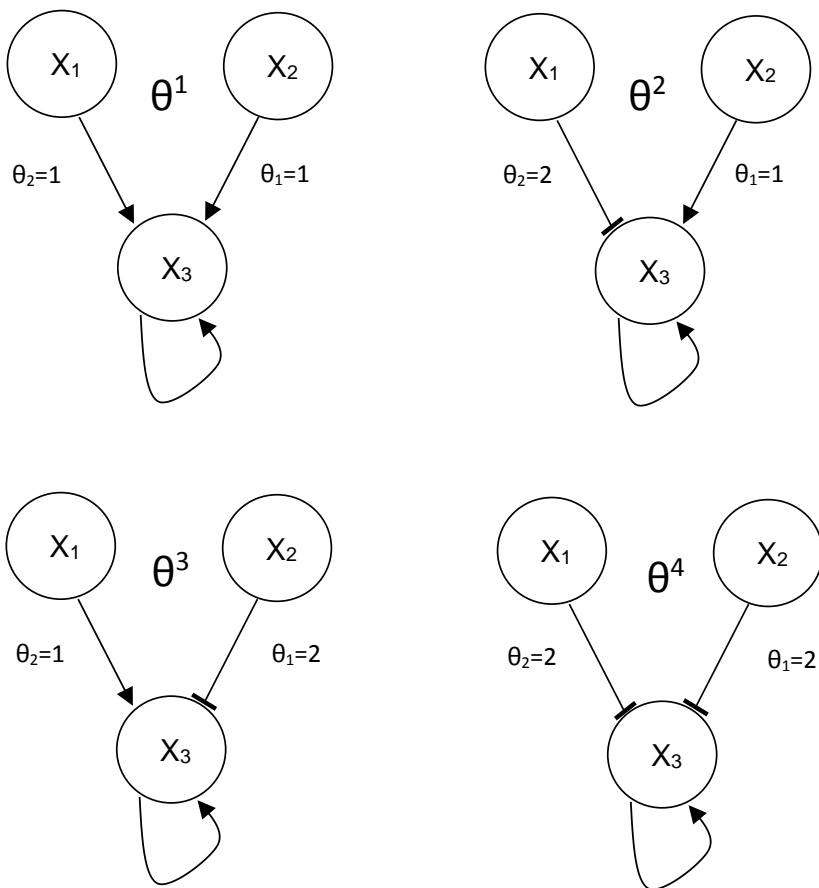Figure A.1. The 3-gene toy network used for illustrative example.

Figure A.2. The uncertainty class Θ which contains all possible networks.

determines whether a given state is desirable or undesirable. Therefore, we need to find the best gene for deletion among X_1 and X_2. We go through lines 3 to 12 of Algorithm 1 to calculate the cost of deleting gene X_1 (the sign ✓ is used for comments):

1. Line 3: $g \leftarrow X_1$

2. Line 4: $cost(X_1) \leftarrow 0$

3. Line 5: $i \leftarrow 1$

   ✓ we compute the cost of deleting gene X_1 related to uncertain parameter θ_1.

4. Line 6: $\theta_1 \leftarrow 1$

5. Line 7: $\Theta|\theta_1 = 1 \leftarrow \{\theta^1, \theta^2\}$, $\Theta^{X_1}|\theta_1 = 1 \leftarrow \{\theta^{1,X_1}, \theta^{2,X_1}\}$

   ✓ remaining uncertainty class when $\theta_1 = 1$

   ✓ θ^{1,X_1} and θ^{2,X_1} are obtained from θ^1 and θ^2 respectively by deleting gene X_1.

   ✓ we find the reduced networks in $\Theta^{X_1}|\theta_1 = 1$ using the procedure given in section "Reduction mappings and induced interventions".

6. Line 8: $P(\theta^1) \leftarrow 1/4$, $P(\theta^2) \leftarrow 1/4$

   ✓ probabilities of two networks inside $\Theta|\theta_1 = 1$

7. Line 9: $\psi_{\text{IBR}}\left(\Theta^{X_1}|\theta_1 = 1\right) \leftarrow \arg\min_{\psi \in \Psi} \left\{ P(\theta^1)\eta_{\theta^{1,X_1}}(\psi) + P(\theta^2)\eta_{\theta^{2,X_1}}(\psi) \right\}$

   ✓ we found the robust intervention for the uncertainty class $\Theta^{X_1}|\theta_1 = 1$ of reduced networks.

   ✓ we can store all costs such as $\eta_{\theta^{1,X_1}}(\psi)$ and $\eta_{\theta^{2,X_1}}(\psi)$ calculated in this step for future computations.

8. Line 9: find $\psi_{\text{IBR}}^{\Theta|\theta_1=1}(\text{ind}; X_1)$ using Algorithm 3

9. Line 10: $h_{X_1}(\theta_1 = 1) \leftarrow P(\theta^1)\eta_{\theta^1}\left(\psi_{\text{IBR}}^{\Theta|\theta_1=1}(\text{ind}; X_1)\right) + P(\theta^2)\eta_{\theta^2}\left(\psi_{\text{IBR}}^{\Theta|\theta_1=1}(\text{ind}; X_1)\right)$

   ✓ the average performance of induced intervention across $\Theta|\theta_1 = 1$

10. Line 6: $\theta_1 \leftarrow 2$

11. Line 7: $\Theta|\theta_1 = 2 \leftarrow \{\theta^3, \theta^4\}$, $\Theta^{X_1}|\theta_1 = 2 \leftarrow \{\theta^{3,X_1}, \theta^{4,X_1}\}$

   ✓ remaining uncertainty class when $\theta_1 = 2$

   ✓ we find the reduced networks in $\Theta^{X_1}|\theta_1 = 2$ using the procedure given in section

   "Reduction mappings and induced interventions".

12. Line 8: $P(\theta^3) \leftarrow 1/4$, $P(\theta^4) \leftarrow 1/4$

   ✓ probabilities of two networks inside $\Theta|\theta_1 = 2$

13. Line 9: $\psi_{\text{IBR}}^{\Theta^{X_1}|\theta_1=2} \leftarrow \underset{\psi \in \Psi}{\arg\min}\left\{P(\theta^3)\eta_{\theta^{3,X_1}}(\psi) + P(\theta^4)\eta_{\theta^{4,X_1}}(\psi)\right\}$

   ✓ we found the robust intervention for the uncertainty class $\Theta^{X_1}|\theta_1 = 2$ of reduced

   networks.

   ✓ we can store all costs such as $\eta_{\theta^{3,X_1}}(\psi)$ and $\eta_{\theta^{3,X_1}}(\psi)$ calculated in this step for

   future computations.

14. Line 9: find $\psi_{\text{IBR}}^{\Theta|\theta_1=2}(\text{ind}; X_1)$ using Algorithm 3

15. Line 10: $h_{X_1}(\theta_1 = 2) \leftarrow P(\theta^3)\eta_{\theta^3}\left(\psi_{\text{IBR}}^{\Theta|\theta_1=2}(\text{ind}; X_1)\right) + P(\theta^4)\eta_{\theta^4}\left(\psi_{\text{IBR}}^{\Theta|\theta_1=2}(\text{ind}; X_1)\right)$

16. Line 11: $P(\theta_1 = 1) \leftarrow 1/2$, $P(\theta_1 = 2) \leftarrow 1/2$

17. Line 11: $cost(X_1) \leftarrow cost(X_1) + P(\theta_1 = 1)h_{X_1}(\theta_1 = 1) + P(\theta_1 = 2)h_{X_1}(\theta_1 = 2)$

   ✓ we obtained cost for gene X_1 caused by θ_1.

18. Line 5: $i \leftarrow 2$

    ✓ we now consider uncertain parameter θ_2.

    ✓ steps 20-31 (for θ_2) are similar to steps 5-16 (for θ_1).

19. Line 6: $\theta_2 \leftarrow 1$

20. Line 7: $\Theta|\theta_2 = 1 \leftarrow \{\theta^1, \theta^3\}$, $\Theta^{X_1}|\theta_2 = 1 \leftarrow \{\theta^{1,X_1}, \theta^{3,X_1}\}$

21. Line 8: $P(\theta^1) \leftarrow 1/4$, $P(\theta^3) \leftarrow 1/4$

22. Line 9: $\psi_{\text{IBR}}^{\Theta^{X_1}|\theta_2=1} \leftarrow \underset{\psi \in \Psi}{\arg\min} \left\{ P(\theta^1)\eta_{\theta^{1,X_1}}(\psi) + P(\theta^3)\eta_{\theta^{3,X_1}}(\psi) \right\}$

23. Line 9: find $\psi_{\text{IBR}}^{\Theta|\theta_2=1}(\text{ind}; X_1)$ using Algorithm 3

24. Line   10: $h_{X_1}(\theta_2 = 1) \leftarrow P(\theta^1)\eta_{\theta^1}\left(\psi_{\text{IBR}}^{\Theta|\theta_2=1}(\text{ind}; X_1)\right) +$

    $P(\theta^3)\eta_{\theta^3}\left(\psi_{\text{IBR}}^{\Theta|\theta_2=1}(\text{ind}; X_1)\right)$

25. Line 6: $\theta_2 \leftarrow 2$

26. Line 7: $\Theta|\theta_2 = 2 \leftarrow \{\theta^2, \theta^4\}$, $\Theta^{X_1}|\theta_2 = 2 \leftarrow \{\theta^{2,X_1}, \theta^{4,X_1}\}$

27. Line 8: $P(\theta^2) \leftarrow 1/4$, $P(\theta^4) \leftarrow 1/4$

28. Line 9: $\psi_{\text{IBR}}^{\Theta^{X_1}|\theta_2=2} \leftarrow \underset{\psi \in \Psi}{\arg\min} \left\{ P(\theta^2)\eta_{\theta^{2,X_1}}(\psi) + P(\theta^4)\eta_{\theta^{4,X_1}}(\psi) \right\}$

29. Line 9: find $\psi_{\text{IBR}}^{\Theta|\theta_2=2}(\text{ind}; X_1)$ using Algorithm 3

30. Line   10: $h_{X_1}(\theta_2 = 2) \leftarrow P(\theta^2)\eta_{\theta^2}\left(\psi_{\text{IBR}}^{\Theta|\theta_2=2}(\text{ind}; X_1)\right) +$

    $P(\theta^4)\eta_{\theta^4}\left(\psi_{\text{IBR}}^{\Theta|\theta_2=2}(\text{ind}; X_1)\right)$

31. Line 11: $P(\theta_2 = 1) \leftarrow 1/2$, $P(\theta_2 = 2) \leftarrow 1/2$

32. Line 11: $cost(X_1) \leftarrow cost(X_1) + P(\theta_2 = 1)h_{X_1}(\theta_2 = 1) + P(\theta_2 = 2)h_{X_1}(\theta_2 = 2)$

    ✓ we found the cost of deleting gene X_1 by adding the cost related to θ_2 to the

    cost related to θ_1.

At this point we have calculated the cost of deleting gene X_1, $cost(X_1)$. We need to calculate the cost of deleting gene X_2, $cost(X_2)$, as well. The steps for calculating the cost of gene X_2 are similar to those for X_1. Therefore, we skip illustrating these steps and proceed to the next stage to use the induced optimal and robust interventions found via deleting the optimal gene for the experimental design. Suppose that the optimal gene for deletion is gene X_1 meaning that $cost(X_1) < cost(X_2)$. Therefore, we go through the rest of Algorithm 1 (lines 13 to 19) to estimate the optimal experiment E_{i^*} to be conducted first:

- Line 13: $i \leftarrow 1$

- Line 14: $\theta_1 \leftarrow 1$

- Line 15: $\Theta|\theta_1 = 1 \leftarrow \{\theta^1, \theta^2\}$

- Line 16: $P(\theta^1) \leftarrow 1/4$, $P(\theta^2) \leftarrow 1/4$

- Line 17: $M_\Psi^{X_1}(\Theta|\theta_1 = 1) \leftarrow P(\theta^1)\left\{ \eta_{\theta^1}\left(\psi_{\mathrm{IBR}}^{\Theta|\theta_1=1}(\mathrm{ind}; X_1)\right) - \eta_{\theta^1}\left(\psi^{\mathrm{ind}}(\theta^1; X_1)\right) \right\}$
  $+ P(\theta^2)\left\{ \eta_{\theta^2}\left(\psi_{\mathrm{IBR}}^{\Theta|\theta_1=1}(\mathrm{ind}; X_1)\right) - \eta_{\theta^2}\left(\psi^{\mathrm{ind}}(\theta^2; X_1)\right) \right\}$

  ✓ we estimated the remaining MOCU when $\theta_1 = 1$ via deleting gene X_1.

- Line 14: $\theta_1 \leftarrow 2$

- Line 15: $\Theta|\theta_1 = 2 \leftarrow \{\theta^3, \theta^4\}$

- Line 16: $P(\theta^3) \leftarrow 1/4$, $P(\theta^4) \leftarrow 1/4$

- Line 17: $M_\Psi^{X_1}(\Theta|\theta_1 = 2) \leftarrow P(\theta^3)\left\{ \eta_{\theta^3}\left(\psi_{\mathrm{IBR}}^{\Theta|\theta_1=2}(\mathrm{ind}; X_1)\right) - \eta_{\theta^3}\left(\psi^{\mathrm{ind}}(\theta^3; X_1)\right) \right\}$
  $+ P(\theta^4)\left\{ \eta_{\theta^4}\left(\psi_{\mathrm{IBR}}^{\Theta|\theta_1=2}(\mathrm{ind}; X_1)\right) - \eta_{\theta^4}\left(\psi^{\mathrm{ind}}(\theta^4; X_1)\right) \right\}$

  ✓ we estimated the remaining MOCU when $\theta_1 = 2$ via deleting gene X_1.

- Line 18: $M_\Psi^{X_1}(\Theta, 1) \leftarrow P(\theta_1 = 1)M_\Psi^{X_1}(\Theta|\theta_1 = 1) + P(\theta_1 = 2)M_\Psi^{X_1}(\Theta|\theta_1 = 2)$

175

- ✓ we estimated the expected remaining MOCU when θ_1 is assumed to be known via deleting gene X_1.

- Line 13: $i \leftarrow 2$

- Line 14: $\theta_2 \leftarrow 1$

- Line 15: $\Theta | \theta_2 = 1 \leftarrow \{\theta^1, \theta^3\}$

- Line 16: $P(\theta^1) \leftarrow 1/4, P(\theta^3) \leftarrow 1/4$

- Line 17: $M_{\Psi}^{X_1}(\Theta | \theta_2 = 1) \leftarrow P(\theta^1)\left\{ \eta_{\theta^1}\left(\psi_{IBR}^{\Theta | \theta_2 = 1}(\text{ind}; X_1) \right) - \eta_{\theta^1}(\psi^{\text{ind}}(\theta^1; X_1)) \right\}$
$+ P(\theta^3)\left\{ \eta_{\theta^3}\left(\psi_{IBR}^{\Theta | \theta_2 = 1}(\text{ind}; X_1) \right) - \eta_{\theta^3}\left(\psi^{\text{ind}}(\theta^3; X_1) \right) \right\}$

  ✓ we estimated the remaining MOCU when $\theta_2 = 1$ via deleting gene X_1.

- Line 14: $\theta_2 \leftarrow 2$

- Line 15: $\Theta | \theta_2 = 2 \leftarrow \{\theta^2, \theta^4\}$

- Line 16: $P(\theta^2) \leftarrow 1/4, P(\theta^4) \leftarrow 1/4$

- Line 17: $M_{\Psi}^{X_1}(\Theta | \theta_2 = 2) \leftarrow P(\theta^2)\left\{ \eta_{\theta^2}\left(\psi_{IBR}^{\Theta | \theta_2 = 2}(\text{ind}; X_1) \right) - \eta_{\theta^2}\left(\psi^{\text{ind}}(\theta^2; X_1) \right) \right\}$
$+ P(\theta^4)\left\{ \eta_{\theta^4}\left(\psi_{IBR}^{\Theta | \theta_2 = 2}(\text{ind}; X_1) \right) - \eta_{\theta^4}\left(\psi^{\text{ind}}(\theta^4; X_1) \right) \right\}$

  ✓ we estimated the remaining MOCU when $\theta_2 = 2$ via deleting gene X_1.

- Line 18: $M_{\Psi}^{X_1}(\Theta, 2) \leftarrow P(\theta_2 = 1)M_{\Psi}^{X_1}(\Theta | \theta_2 = 1) + P(\theta_2 = 2)M_{\Psi}^{X_1}(\Theta | \theta_2 = 2)$

  ✓ we estimated the expected remaining MOCU when θ_2 is assumed to be known via deleting gene X_1.

- Line 19: $i^* \leftarrow \arg\min_{i \in 1,2} M_{\Psi}^{X_1}(\Theta; \theta_i)$

APPENDIX B

PROOF OF THEOREM 1 IN CHAPTER 5

In this appendix, we prove (5.9), which shows how to find the MSE between an original random process $\mathbf{X}$ and a compressed random process $\mathbf{X}'_m$ obtained using an arbitrary covariance matrix $\mathbf{K}'$ that may not be necessarily equal to that of the original random process, $\mathbf{K}$.

First we find the MSE for each n:

$$
\begin{aligned}
\mathrm{MSE} < X(n), X'_m(n) > \ &= \mathrm{E}\left[\left|X(n) - X'_m(n)\right|^2\right] \\
&= \mathrm{E}\left[\left(\sum_{i=1}^{N} Z_i u_i(n) - \sum_{i=1}^{m} Z'_i u'_i(n)\right)\overline{\left(\sum_{l=1}^{N} Z_l u_l(n) - \sum_{l=1}^{m} Z'_l u'_l(n)\right)}\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{N}\sum_{l=1}^{N} Z_i \overline{Z_l} u_i(n)\overline{u_l(n)} + \sum_{i=1}^{m}\sum_{l=1}^{m} Z'_i \overline{Z'_l} u'_i(n)\overline{u'_l(n)}\right.\\
&\qquad\left. - \sum_{i=1}^{m}\sum_{l=1}^{N} Z'_i \overline{Z_l} u'_i(n)\overline{u_l(n)} - \sum_{i=1}^{N}\sum_{l=1}^{m} Z_i \overline{Z'_l} u_i(n)\overline{u'_l(n)}\right] \\
&= \sum_{i=1}^{N} \lambda_i \left|u_i(n)\right|^2 + \sum_{i=1}^{m} \mathrm{E}\left[\left|Z'_i\right|^2\right]\left|u'_i(n)\right|^2 \\
&\qquad - \sum_{i=1}^{m}\sum_{l=1}^{N} \mathrm{E}\left[Z'_i \overline{Z_l}\right] u'_i(n)\overline{u_l(n)} - \sum_{i=1}^{N}\sum_{l=1}^{m} \mathrm{E}\left[Z_i \overline{Z'_l}\right] u_i(n)\overline{u'_l(n)}.
\end{aligned}
$$

$$(\mathrm{B.1})$$

Now the MSE over all n is

$$\text{MSE} < \mathbf{X}, \mathbf{X}'_m > = \sum_{n=1}^{N} \text{E}\left[\left|X(n) - X'_m(n)\right|^2\right]$$

$$= \sum_{i=1}^{N} \lambda_i \sum_{n=1}^{N} |u_i(n)|^2 + \sum_{i=1}^{m} \text{E}\left[|Z'_i|^2\right] \sum_{n=1}^{N} |u'_i(n)|^2$$

$$- \sum_{i=1}^{m} \sum_{l=1}^{N} \text{E}\left[Z'_i \overline{Z_l}\right] \sum_{n=1}^{N} u'_i(n)\overline{u_l(n)}$$

$$- \sum_{i=1}^{N} \sum_{l=1}^{m} \text{E}\left[Z_i \overline{Z'_l}\right] \sum_{n=1}^{N} u_i(n)\overline{u'_i(n)}. \qquad \text{(B.2)}$$

$\text{E}\left[|Z'_i|^2\right]$ can be calculated as follows:

$$\text{E}\left[|Z'_i|^2\right] = \text{E}\left[\sum_{n=1}^{N}\sum_{n_1=1}^{N} X(n)\overline{X(n_1)}\,\overline{u'_i(n)}u'_i(n_1)\right]$$

$$= \sum_{n=1}^{N}\sum_{n_1=1}^{N} \text{E}\left[X(n)\overline{X(n_1)}\right]\overline{u'_i(n)}u'_i(n_1)$$

$$= (\overline{\mathbf{u}'_i})^T \mathbf{K} \mathbf{u}'_i. \qquad \text{(B.3)}$$

To compute the third term in (B.2), first we find $\text{E}\left[\overline{Z_i}Z'_l\right]$:

$$\text{E}\left[Z'_i\overline{Z_l}\right] = \text{E}\left[\sum_{n=1}^{N}\sum_{n_1=1}^{N} X(n)\overline{X(n_1)}\,\overline{u'_i(n)}u_l(n_1)\right]$$

$$= \sum_{n=1}^{N}\left(\sum_{n_1=1}^{N} K(n,n_1)u_l(n_1)\right)\overline{u'_i(n)}$$

$$= \sum_{n=1}^{N} \lambda_l u_l(n)\overline{u'_i(n)}, \qquad \text{(B.4)}$$

where the third equality follows from the fact that $\mathbf{u}_i$ is the eigenvector of $\mathbf{K}$ such that $\sum_{n_1=1}^{N} K(n,n_1)u_l(n_1) = \lambda_l u_l(n)$.

Compute the third term in (B.2) by plugging in (B.4):

$$\sum_{i=1}^{m}\sum_{l=1}^{N}\mathrm{E}\big[Z_i'\overline{Z_l}\big]\Big(\sum_{n_1=1}^{N}u_i'(n_1)\overline{u_l(n_1)}\Big) = \sum_{i=1}^{m}\sum_{l=1}^{N}\Big(\sum_{n=1}^{N}\lambda_l u_l(n)\overline{u_i'(n)}\Big)\Big(\sum_{n_1=1}^{N}u_i'(n_1)\overline{u_l(n_1)}\Big)$$

$$= \sum_{i=1}^{m}\sum_{n=1}^{N}\sum_{n_1=1}^{N}\Big(\sum_{l=1}^{N}\lambda_l u_l(n)\overline{u_l(n_1)}\Big)u_i'(n_1)\overline{u_i'(n)}$$

$$= \sum_{i=1}^{m}\sum_{n=1}^{N}\sum_{n_1=1}^{N}K(n,n_1)u_i'(n_1)\overline{u_i'(n)}$$

$$= \sum_{i=1}^{m}(\overline{\mathbf{u}_i'})^T\mathbf{K}\mathbf{u}_i', \tag{B.5}$$

where the third equality follows from

$$\sum_{i=l}^{N}\lambda_l u_l(n)\overline{u_l(n_1)} = K(n,n_1). \tag{B.6}$$

Note that in (B.2), the fourth term is the conjugate of the third term; therefore, it equals the

conjugate of (B.5). Using (B.3) and (B.5), we simplify (B.2):

$$\mathrm{MSE} < \mathbf{X},\mathbf{X}_m' > \; = \sum_{i=1}^{N}\lambda_i + \sum_{i=1}^{m}\mathrm{E}\Big[\big|Z_i'\big|^2\Big] - \sum_{i=1}^{m}\sum_{l=1}^{N}\mathrm{E}\big[Z_i'\overline{Z_l}\big]\sum_{n=1}^{N}u_i'(n)\overline{u_l(n)}$$

$$- \sum_{i=1}^{N}\sum_{l=1}^{m}\mathrm{E}\big[Z_i\overline{Z_l'}\big]\sum_{n=1}^{N}u_i(n)\overline{u_l'(n)}$$

$$= \sum_{i=1}^{N}\lambda_i + \sum_{i=1}^{m}(\overline{\mathbf{u}_i'})^T\mathbf{K}\mathbf{u}_i' - \sum_{i=1}^{m}(\overline{\mathbf{u}_i'})^T\mathbf{K}\mathbf{u}_i' - \sum_{i=1}^{m}(\mathbf{u}_i')^T\overline{\mathbf{K}}\overline{\mathbf{u}_i'}$$

$$= \sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{m}(\mathbf{u}_i')^T\overline{\mathbf{K}}\overline{\mathbf{u}_i'}. \tag{B.7}$$