AFRL-OSR-VA-TR-2013-0546

SIMULATION-BASED METHODOLOGIES FOR GLOBAL OPTIMIZATION AND PLANNING

STEVEN MARCUS

UNIVERSITY OF MARYLAND COLLEGE PARK

10/11/2013
Final Report

DISTRIBUTION A: Distribution approved for public release.

AIR FORCE RESEARCH LABORATORY
AF OFFICE OF SCIENTIFIC RESEARCH (AFOSR)/RSL
ARLINGTON, VIRGINIA 22203
AIR FORCE MATERIEL COMMAND

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
09-30-2013	Final	7/1/2010 - 6/30/2013

4. TITLE AND SUBTITLE

Simulation-Based Methodologies For Global Optimization And Planning

5a. CONTRACT NUMBER

5b. GRANT NUMBER

FA9550-10-1-0340

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Steven I. Marcus, Michael C. Fu, and Jiaqiao Hu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Office of Research Administration and Advancement
Rm. 2100 Lee Building
University of Maryland
College Park, MD 20742

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/PKC
875 North Randolph Street, Suite 325, Room 3112, Arlington, VA 22203-1768
VA 22203-1768

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The researchers made significant progress in all of the proposed research areas. The first major task in the proposal involved model-based randomized methods for global optimization. In support of this task, the researchers developed new methods for stochastic derivative estimators for discontinuous payoff functions; the method includes Infinitesimal Perturbation Analysis and the Likelihood Ratio method as special cases and can be applied to functions of more general forms containing indicator functions. The researchers developed a new method of distributed ordinal comparison of selecting the best option, which maximizes the average of local reward function values among available options in a dynamic network. they discovered a new innovative approach to simulation-based global optimization by building a connection between global optimization and evolutionary games, as well as another new approach that exploits particle filtering; they have summarized our model-based results in a comprehensive survey paper. The researchers also made significant progress in other model-based randomized methods, including a stochastic search algorithm for solving general optimization problems with little structure; the algorithm iteratively finds high quality solutions by randomly s

15. SUBJECT TERMS

Simulation-based optimization, global optimization, Markov Decision Processes, risk

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Steven I. Marcus
					19b. TELEPHONE NUMBER *(include area code)* 301-405-7589

Final Performance Report
Grant FA9550-10-1-0340

7/01/2010-6/30/2013

Steven I. Marcus, Michael C. Fu, and Jiaqiao Hu

September 30, 2013

Abstract

The researchers made significant progress in all of the proposed research areas. The first major task in the proposal involved model-based randomized methods for global optimization. In support of this task, the researchers developed new methods for stochastic derivative estimators for discontinuous payoff functions; the method includes Infinitesimal Perturbation Analysis and the Likelihood Ratio method as special cases and can be applied to functions of more general forms containing indicator functions. The researchers developed a new method of distributed ordinal comparison of selecting the best option, which maximizes the average of local reward function values among available options in a dynamic network. they discovered a new innovative approach to simulation-based global optimization by building a connection between global optimization and evolutionary games, as well as another new approach that exploits particle filtering; they have summarized our model-based results in a comprehensive survey paper. The researchers also made significant progress in other model-based randomized methods, including a stochastic search algorithm for solving general optimization problems with little structure; the algorithm iteratively finds high quality solutions by randomly sampling candidate solutions from a parameterized distribution model over the solution space.

In support of the second task, the researchers made progress incorporating simulation-based and sampling methods into Markov Decision Processes (MDPs). They made significant progress in new simulation-based approaches to MDPs, and in applications to problems of supply chains and finance. In particular, the researchers have developed a simulation-based algorithm called Approximate Stochastic Annealing (ASA) for solving finite horizon MDPs.

In addition, the researchers developed a theory and new methods for solving dynamic stochastic optimization problems with non-convex risk-sensitive performance measures, as well as two new methods for simulation optimization.

1 Introduction

In this research project, we proposed to investigate basic questions aimed at challenges in information superiority, logistics, and planning for the Air Force of the future. In particular, we proposed to investigate simulation-based methodologies for global optimization and planning that can be effective tools in an integrated approach to Command and Control (C2), planning, and logistics. The questions we investigated were motivated by future Air Force requirements, which will involve a flexible and world-responsive set of missions. More agile, responsive, and integrated systems will be required. The aim of the research was to facilitate Air Force decision making at many levels of operations, particularly at the mission, campaign, and strategic levels. The proposed simulation-based methodologies were intended to provide approaches to optimization and comparison of Alternative Courses of Action in Air Force simulation-based models.

Such problems and systems are exceedingly complex; in order to solve them, we focused on **using simulation-based methods for global optimization and sequential decision making under uncertainty**.

In particular, we combined three approaches in the study of such problems:

- Developing, studying, and analyzing efficient simulation-based and sampling methodologies for global optimization problems;
- Developing and studying efficient parallel simulation-based and sampling methodologies for problems of dynamic decision making under uncertainty;
- Studying the application of these optimization methodologies to practical problems, such as preventive maintenance, path planning for unmanned aerial vehicles, and financial engineering.

1.1 Model-Based Global Optimization

For a a bounded deterministic measurable function $H : \mathscr{X} \to \mathscr{R}$, where $\mathscr{X}$ is the feasible solution space, the optimization problem is to find

$$x^* \in \arg\max_{x \in \mathscr{X}} H(x). \tag{1}$$

It is common in many situations to introduce a measurable strictly increasing fitness function, $\phi : \mathscr{R} \to \mathscr{R}^+$, and reformulate (1) as

$$x^* \in \arg\max_{x \in \mathscr{X}} \phi\left(H(x)\right), \tag{2}$$

which guarantees the range of the new fitness-objective function will always be non-negative.

Model-based global optimization methods use probability distributions to weight promising areas of the solution space, where the distribution is updated iteratively based on output from the samples drawn according to the current distribution. They are well suited for global optimization problems where there is limited structural information on the optimizing function (e.g., derivatives, convexity).

1.2 Simulation-Based and Sampling Methods for Markov Decision Processes

Simulation optimization problems arising in supply chain management, path planning for unmanned aerial vehicles, financial engineering, and telecommunications are characterized by two critical aspects: changing dynamics and stochastic events. For example, effective supply chain management requires optimal responsive actions in the face of both gradual shifts in demand patterns (e.g., due to technology advances) and sudden unpredictable disruptions in production capacity (e.g., due to an unanticipated manufacturing facility shutdown). Such systems often require computationally expensive simulation models for performance

estimation, such as modeling the operations of an entire semiconductor fabrication facility, where simulation runtime is typically on the order of hours. Markov decision processes (MDPs) provide a powerful paradigm for modeling optimal decision making under uncertainty in these settings, but MDPs suffer from the well-known curse of dimensionality, which can include exponential growth in the size of state spaces and action spaces with the problem size; thus, direct numerical solution of MDPs for large-scale real-world problems presents a formidable computational challenge. In general, heuristics and approximations are employed to simplify the MDP model. Perhaps the most successful example of this approach has been approximate dynamic programming using value function approximation (cf. [4, 16, 29, 37, 40]).

The PIs have developed evolutionary and simulation-based algorithms that provide new advances in the solution of MDPs [9]. These advances include the following approaches: **adaptive multi-stage sampling** — well-suited for large state spaces but relatively smaller action spaces; and **population-based evolutionary randomized policy search** — designed to handle large action spaces. The PIs have developed many such algorithms [7, 8, 9, 10, 12, 23].

2 Research Results

2.1 Model-Based Global Optimization

We distinguish between *instance-based* and *model-based* global optimization solution methods. In instance-based methods, the search for new candidate solutions depends directly on previously generated solutions, e.g., simulated annealing, genetic algorithms (GAs), tabu search, and nested partitions. On the other hand, in model-based algorithms, new candidate solutions are generated via an intermediate *probability model* that is iteratively updated. Our research has focused on the model-based optimization framework, which involves the following ingredients:

(0) specify probability distribution over solution space;
(I) generate candidate solutions by sampling from distribution;
(II) estimate performance of (and possibly improve) candidate solutions;
(III) update distribution based on selected ("elite") set of candidate solutions.

This approach retains the primary strengths of population-based approaches such as genetic algorithms — improving upon simulated annealing, which works with a single iterate at a time, while at the same time providing more flexibility in exploring the entire solution space, introducing more structure in the search procedure, and allowing theoretical properties to be studied regarding both finite-time performance and asymptotic convergence. The theory behind the framework is rigorous, but based on an idealized version of the last three ingredients, specifically the distribution sequence, sampling from the distribution sequence (or from a surrogate or approximation), and estimation of the performance, since it is observed through simulation. Schematically, we seek a sequence of distributions

$$g_0, g_1, g_2, \ldots \longrightarrow g_\infty,$$

where g_∞ concentrates its mass around the optimal solutions.

Examples for the sequence of distributions $\{g_k\}$ include the instantiation of our model reference adaptive search (MRAS) method in [22].

However, the sequence $\{g_k\}$ is unknown explicitly a priori, or else the problem would essentially be solved. Our MRAS approach uses a *projection* onto distributions that are easy to work with, e.g., uses a family of *parameterized* distributions $\{f_\theta\}$, and projects g_k onto the family to obtain a sequence that converges to the (final) target distribution, i.e.,

$$f_{\theta_0}, f_{\theta_1}, f_{\theta_2}, \ldots \longrightarrow g_\infty;$$

a common implementation minimizes the Kullback-Leibler (KL) divergence between f_{θ_k} and g_k at each iteration, because it leads to analytically tractable solutions if the parameterized distributions are from the

exponential family. This leads to a *population* of candidate solutions, from which an elite set is selected and used to update the distribution.

In [42], we propose a new framework for continuous global optimization problems by building a connection between global optimization problems and evolutionary games, and we show that a particular equilibrium set of the evolutionary game is asymptotically stable. Based on this connection, we propose a Model-based Evolutionary Optimization (MEO) algorithm, which uses probabilistic models to generate new candidate solutions and uses dynamics from evolutionary game theory to govern the evolution of the probabilistic models. The MEO algorithm also gives new insight into the mechanism of model updating in model-based global optimization algorithms. Based on the MEO algorithm, a Population Model-based Evolutionary Optimization (PMEO) algorithm is proposed, which captures the multimodal property of global optimization problems. Simulation experiments demonstrate the effectiveness of the proposed algorithm.

We study in [26] a class of random sampling-based algorithms for solving general non-differentiable optimization problems. These are iterative approaches that are based on sampling from and updating an underlying distribution function over the set of feasible solutions. In particular, we propose a novel and systematic framework to investigate the convergence and asymptotic convergence rates of these algorithms by exploiting their connections to the well-known stochastic approximation (SA) method. Such an SA framework unifies our understanding of these randomized algorithms and provides new insight into their design and implementation issues. Our preliminary numerical experiments indicate that new implementations of these algorithms based on the proposed framework may lead to improved performance over existing procedures.

The Annealing Adaptive Search (AAS) algorithm for global optimization searches the solution space by sampling from a sequence of Boltzmann distributions. For a class of optimization problems, it has been shown that the complexity of AAS increases at most linearly in the problem dimension. However, despite its desirable property, sampling from a Boltzmann distribution at each iteration of the algorithm remains a practical challenge. Prior work to address this issue has focused on embedding Markov chain-based sampling techniques within the AAS framework. In [24, 25], based on ideas from the recent Cross-Entropy method and our Model Reference Adaptive Search method, we propose an algorithm, called Model-based Annealing Random Search (MARS), that complements prior work by sampling solutions from a sequence of surrogate distributions that iteratively approximate the target Boltzmann distributions. We establish a novel connection between MARS and the well-known Stochastic Approximation method. By exploiting this connection, we prove the global convergence of MARS and characterize its asymptotic convergence rate behavior. Our empirical results indicate promising performance of the algorithm in comparison with some of the existing methods.

The paper [45] presents a novel interpretation to transform an optimization problem into a filtering problem, where the goal is to compute the conditional distribution of the unobserved state given the observation history. We prove that in our formulation the conditional distribution converges asymptotically to a degenerate distribution concentrated on the global optimum. Hence, the goal of searching for the global optimum can be achieved by computing the conditional distribution sequentially. That is done through the application of particle filtering, a class of sequential Monte Carlo methods for filtering, which has proven convergence in tracking the conditional distribution. The resultant algorithmic framework unifies some recent randomized optimization algorithms as well as providing new insights into their connection. More importantly, the framework opens up the possibility of new improved algorithms. In particular, we develop a new improved cross-entropy method under this framework, and the numerical results show that our method is very effective in preventing premature convergence of the Cross-Entropy method.

The paper [21] aims to improve the sampling efficiency of model-based methods for global optimization by considering a generalization where a population of distribution models is maintained and subsequently propagated from generation to generation. A key issue in the proposed approach is how to efficiently allocate the sampling budget among the population of models to maximize the algorithm performance. We formulate

this problem as a generalized max k-armed bandit problem, and derive an efficient dynamic sample allocation scheme based on Markov decision theory to adaptively allocate computational resources. The proposed allocation scheme is then further used to update the current population to produce an improving population of models. Our preliminary numerical results indicate that the proposed procedure may considerably reduce the number of function evaluations needed to obtain high quality solutions, and thus further enhance the value of model-based methods for optimization problems that require expensive function evaluations for performance evaluation.

We propose in [15] to improve the efficiency of simulation optimization by integrating the notion of optimal computing budget allocation into the Cross-Entropy (CE) method. This paper focuses on continuous optimization problems. In the stochastic simulation setting where replications are expensive but noise in the objective function estimate could mislead the search process, the allocation of simulation replications can make a significant difference in the performance of such global optimization search algorithms. A new allocation scheme is developed based on the notion of optimal computing budget allocation. The proposed approach improves the updating of the sampling distribution by carrying out this computing budget allocation in an efficient manner, by minimizing the expected mean-squared error of the CE weight function. Numerical experiments indicate that the computational efficiency of the CE method can be substantially improved if the ideas of computing budget allocation are applied.

We have also made considerable advances in applying our sampling and model-based framework and related techniques to various applications. For example, the assessment of dose-response is an integral component of the drug development process. Parallel dose-response studies are conducted, customarily, in preclinical and phase 1, 2 clinical trials for this purpose. Practical constraints on dose range, dose levels and dose proportions are intrinsic issues in the design of dose response studies because of drug toxicity, efficacy, FDA regulations, protocol requirements, clinical trial logistics, and marketing issues. We have developed (see [27]) a free on-line software package called Controlled Optimal Design 2.0 for generating controlled optimal designs that can incorporate prior information and multiple objectives, and meet multiple practical constraints at the same time. Researchers can either run the web-based design program or download its stand-alone version to construct the desired multiple-objective controlled Bayesian optimal designs. Because researchers often adopt ad-hoc design schemes such as the equal allocation rules without knowing how efficient such designs would be for the design problem, the program also evaluates the efficiency of user-supplied designs.

In another application of simulation optimization, we have studied a finance problem in [5]. Assuming the underlying assets follow a Variance-Gamma (VG) process, we consider the problem of estimating sensitivities such as the Greeks on a basket of stocks when Monte Carlo simulation is employed. We focus on a class of derivatives called mountain range options, comparing indirect methods (finite difference techniques such as forward differences) and two direct methods: infinitesimal perturbation analysis (IPA) and the likelihood ratio (LR) method, where the latter is also implemented via a recently proposed numerical technique developed by Glasserman and Liu using the characteristic function. We carry out numerical simulation experiments to evaluate the efficiency of the different estimators and discuss the strengths and weakness of each method.

Motivated by IPA and the LR method, we derive in [44, 43] a new unbiased stochastic derivative estimator for a class of discontinuous payoff functions that arise in many options pricing settings from finance. Our method includes IPA and the LR method as special cases and can be applied to functions of more general forms containing indicator functions. This new estimator can be computed from a single sample path or simulation, whereas existing estimators in the literature require additional simulations. We apply this method to sensitivity analysis for European call options and American style call options. For pricing American-style derivatives using a gradient-based stochastic approximation algorithm, numerical experiments indicate that the estimator is computationally more efficient than other estimators in the literature.

In [9], we analyze a call center with multiple customer types and dynamic priority service discipline, in which a low-priority customer becomes high priority when its waiting time exceeds a given deterministic service level threshold. Within each priority queue, the service discipline is first come, first served. Based on a fluid approximation of the system, we apply infinitesimal perturbation analysis (IPA) to derive estimators for the derivatives of the queue lengths with respect to the threshold parameter. Numerical examples illustrate the validity of the fluid model approximation and the accuracy of the IPA estimators.

In [11], we consider distributed ordinal comparison of selecting the best option, which maximizes the average of local reward function values among available options in a dynamic network. Each node in the network knows only its reward function, and edge-connectivity across the nodes changes over time according to Calafiores model. To estimate each options global reward function value, local samples for each option are generated at each node, and those are iteratively mixed over the network by a weighted average of local estimates of instantaneous neighbors. Each node selects an option that achieves the maximum of the current global estimates as an estimate of the best option. We establish a lower bound on the probability of correct local-selection at any node, which uniformly converges over the nodes to a lower bound on the probability of correct global-selection by a virtual centralized node with the total available samples.

In [47, 46], we propose a stochastic search algorithm for solving general optimization problems with little structure. The algorithm iteratively finds high quality solutions by randomly sampling candidate solutions from a parameterized distribution model over the solution space. The basic idea is to convert the original (possibly non-continuous, non-differentiable) problem into a differentiable optimization problem on the parameter space of the parameterized sampling distribution, and then use a direct gradient search method to find improved sampling distributions. Thus, the algorithm combines the robustness feature of stochastic search from considering a population of candidate solutions with the relative fast convergence speed of classical gradient methods by exploiting local differentiable structures. We analyze the convergence and convergence rate properties of the proposed algorithm, and carry out a numerical study to illustrate its performance.

2.2 Simulation Optimization

2.2.1 Direct Gradient Augmented Regression (DiGAR)

The classical linear regression model takes the form:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_d x_{id} + \varepsilon_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i,$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{id})^T \in \mathbb{R}^d$ and $\boldsymbol{\beta}$ is to be estimated from the data. The Direct Gradient Augmented Regression (DiGAR) model in [18] adds the gradient in an intuitive manner:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \tag{3}$$
$$\mathbf{g}_i = \boldsymbol{\beta} + \boldsymbol{\delta}_i, \tag{4}$$

where y_i and $\mathbf{g}_i$, $i = 1, 2, \ldots, k$ are the performance measures and gradient estimates with residuals $\{\varepsilon_i\}$ and $\{\boldsymbol{\delta}_i\}$, respectively. Since (4) is also a linear relationship, it could in principle be combined with (3) to obtain the traditional model of the same form but with higher dimension (see Fu and Qu [18, 32]), but at a loss of the intuition that will become evident by keeping them separate. For illustrative purposes, consider the one-dimensional problem, i.e., the given data points are (x_1, y_1, g_1),..., (x_k, y_k, g_k) and $\mathbf{X}\boldsymbol{\beta} = \beta_0 + x\beta_1$. Using ordinary least squares, the function to be minimized is the sum of the squared deviations in both y_i and g_i:

$$L = \sum_{i=1}^{k} (y_i - \beta_0 - \beta_1 x_i)^2 + \sum_{i=1}^{k} (g_i - \beta_1)^2. \tag{5}$$

Here, for simplicity, the two components are equally weighted; [18] includes the general convex combination case. Denoting $\hat{\beta}_i^D$ and $\hat{\beta}_i^L$, $i = 0,1$, as the respective DiGAR and classical linear regression estimators, the

resulting estimators that minimize (5) are

$$\hat{\beta}_0^D = \bar{y} - \hat{\beta}_1^D \bar{x}, \tag{6}$$

$$\hat{\beta}_1^D = \frac{\sum_{i=1}^{k}(x_i - \bar{x})(y_i - \bar{y}) + k\bar{g}}{\sum_{i=1}^{k}(x_i - \bar{x})^2 + k}, \tag{7}$$

whereas the estimators in classical linear regression are

$$\hat{\beta}_0^L = \bar{y} - \hat{\beta}_1^L \bar{x}, \qquad \hat{\beta}_1^L = \frac{\sum_{i=1}^{k}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{k}(x_i - \bar{x})^2}, \tag{8}$$

where $\bar{x}$, $\bar{y}$ and $\bar{g}$ are the corresponding sample means of x_i, y_i and g_i. Note that in the DiGAR model, the form of the intercept estimator given by (6) remains unchanged, whereas the slope estimator given by (7) has the additional terms $n\bar{g}$ and n in the numerator and denominator, respectively, reflecting the additional direct gradient information.

Assumption 1.

i) The estimators for the responses and gradients are unbiased, i.e., $E(\delta_i) = E(\varepsilon_i) = 0$.

ii) All the residuals are uncorrelated, i.e., $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$, $\mathrm{Cov}(\delta_i, \delta_j) = 0$, $\mathrm{Cov}(\varepsilon_i, \delta_j) = 0$, $\forall\, i, j$.

The quality of the slope estimator is critical for use in sequential RSM, as it provides the basis for the search direction. Denoting the respective response and gradient residual variances by $\mathrm{Var}(\varepsilon_i) = \sigma^2$ and $\mathrm{Var}(\delta_i) = \sigma_g^2$, the following result provides one simple sufficient condition under which the variance of the slope estimator is smaller for the DiGAR model.

Proposition 1. Under Assumption 1, the variance of DiGAR estimator $\mathrm{Var}(\hat{\beta}_1^D) \leq \mathrm{Var}(\hat{\beta}_1^L)$ if $\sigma_g^2 \leq C \cdot \sigma^2$, where

$$C = \frac{k + \sum_{i=1}^{k} x_i^2 - k\bar{x}^2}{\sum_{i=1}^{k} x_i^2 - k\bar{x}^2} > 1.$$

Since $C > 1$, as long as the gradient estimate is not too much noisier than the variance of the performance estimate, the DiGAR slope estimator is guaranteed to provide statistical improvement.

Assumption 2. *The residuals are normally distributed, i.e.,* $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, $\delta_i \sim \mathcal{N}(0, \sigma_g^2)$.

Assumptions 1 and 2 imply that the performance and gradient estimates are independent due to the residuals being uncorrelated, and the likelihood function is given by

$$L(\beta_0, \beta_1, \sigma^2, \sigma_g^2) = (2\pi)^{-k}(\sigma\sigma_g)^{-k} \exp\left\{ -\frac{1}{2\sigma^2}\sum_{i=1}^{k}(y_i - \beta_0 - \beta_1 x_i)^2 - \frac{1}{2\sigma_g^2}\sum_{i=1}^{k}(g_i - \beta_1)^2 \right\},$$

which leads to the following respective maximum likelihood estimators (MLEs) for β_0 and β_1:

$$\hat{\beta}_0^D = \bar{y} - \hat{\beta}_1 \bar{x}, \qquad \hat{\beta}_1^D = \frac{\frac{1}{\sigma^2}\sum_{i=1}^{k} x_i y_i + \frac{k}{\sigma_g^2}\bar{g} - \frac{k}{\sigma^2}\bar{x}\bar{y}}{\frac{1}{\sigma^2}\sum_{i=1}^{k} x_i^2 + \frac{k}{\sigma_g^2} - \frac{k}{\sigma^2}\bar{x}^2}, \tag{9}$$

for which theoretical superiority of the slope estimator can also be established (proof in [18]).

Proposition 2. Under Assumptions 1 and 2, the MLE $\hat{\beta}_1^D$ in (9) has smaller variance than $\hat{\beta}_1^L$.

These results provide theoretical support for the intuition that the availability of direct gradient estimates is beneficial. However, in most practical applications, the most critical assumptions are generally not satisfied, e.g., direct gradient estimates are generally correlated with and less precise than the performance estimates. Furthermore, the implicit assumption that the variance is constant over the input parameter space is also violated. Even so, preliminary simulation experiments with queueing systems indicate that the conclusions are robust to violations in the assumptions, i.e., substantial gains are observed using the DiGAR models. The following example illustrates typical results that have been observed.

Queueing Example

A single-server, first-come, first-served queue with Poisson arrivals and i.i.d. exponentially distributed service times (i.e., an M/M/1 queue) is considered, where the arrival rate is fixed (at 0.2). The performance measure of interest is the expected system time. For simplicity, the queue is assumed to start empty, and the output performance is the expected service time of the ith customer, denoted by $y^{(i)}$, which can be calculated analytically as a function of the mean service time, denoted by x, for the purposes of evaluating the quality of the fits of the different regression models. For this and many queueing examples, direct gradient estimates are available using IPA, LR/SF, and WD; for the numerical results reported here, the IPA estimator is used, as it has the lowest variance in this setting. To compare with traditional regression, three DiGAR models are considered: uncorrelated, independent Gaussian, and a correlated Gaussian model to be described shortly. The simulated data, true model and fitted models (DiGAR = uncorrelated, DiGARn = independent normal, DiGAR* = correlated normal) are plotted in Figure 1, along with the 10 data points (at $x = 3.6, 3.7, ..., 4.5$), which themselves are sample means based on 10 replications. All methods fit the model reasonably well for the 2nd and 3rd customers, but there are dramatic differences in $y^{(4)}$ and $y^{(5)}$, where the slope of the traditional model has the *incorrect sign*, because the small number of replications leads to very noisy estimates of the system time that indicate a negative trend. The direct gradient estimates (also based on the small number of replications), however, provide critical additional information that leads to each of the DiGAR models capturing the correct orientation of the curve. Similar results can be observed for higher-order functional forms, e.g., quadratic (see Fu and Qu [18]).

Preliminary theoretical and experimental results such as these and others in Fu and Qu [18] are highly encouraging. In our proposed research, we will use generalized least squares (GLS) to handle correlations and heteroscedasticity. We assume the residuals have zero mean, i.e., $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, so $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, and $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \mathbf{V}$, where the covariance matrix $\mathbf{V}$ is non-diagonal due to the correlations between y_i and g_i. The generalized least squares estimator is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X^T V^{-1} X})^{-1} \mathbf{X^T V^{-1} y},$$

and the covariance matrix for $\hat{\boldsymbol{\beta}}$ is

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X^T V^{-1} X})^{-1}.$$

If the residuals are assumed to be normally distributed, the MLE of $\boldsymbol{\beta}$ is the same as the GLS estimator. To make the analysis of the slope estimator tractable, we assume that y_i is correlated with g_j only when $i = j$ and $\mathrm{corr}(y_i, g_i) = \rho$, $i = 1, 2, \cdots, k$. Under these assumptions, the variance of $\hat{\beta}_1^D$ is given by

$$\mathrm{Var}(\hat{\beta}_1^D) = \frac{\sigma^2}{\frac{1}{1-\rho^2}\left(\sum_{i=1}^{k} x_i^2 - k\bar{x}^2\right) + k\frac{\sigma^2}{\sigma_g^2}}, \tag{10}$$

If $0 < \sigma^2 < \infty$, $0 < \sigma_g^2 < \infty$ and $-1 < \rho < 1$, then we can show that $\mathrm{Var}(\hat{\beta}_1^D)$ in (10) is smaller than $\mathrm{Var}(\hat{\beta}_1^L)$. Generally, ρ, and more generally $\mathbf{V}$, is unknown and must be estimated based on data. Although the theoretical analysis indicates potential for variance reduction from using the correlated DiGAR model,

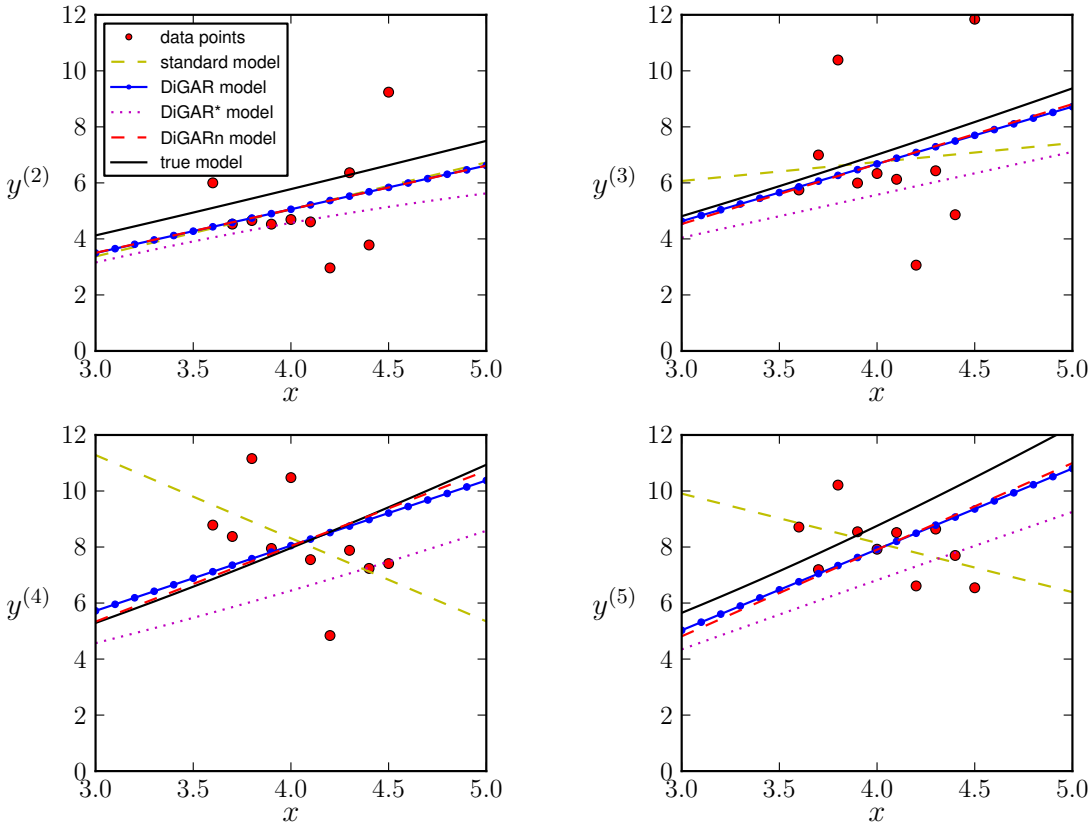

Figure 1: Expected time in system for each customer. For $y^{(4)}$ and $y^{(5)}$, the slope of the standard regression model has the incorrect sign (negative rather than positive).

the extra computational budget spent on accurately estimating correlations must be traded off with any potential performance gains.

2.2.2 Gradient-Enhanced Stochastic Kriging (GESK)

Stochastic kriging (SK) was introduced by Ankenman, Nelson, and Staum [1] to handle the stochastic simulation setting, where the noise in the fitted curve could be viewed as coming from both uncertainty in the fit and stochastic nature of the underlying system. Thus, unlike in regular (deterministic) kriging, the fitted curve in stochastic kriging need not go through every data point, making it closer to regression rather than, e.g., spline interpolation. In the same spirit as DiGAR and the second direction we propose here, Chen, Ankenman, and Nelson [13] introduced stochastic kriging with gradient estimators (SKG), showing that SKG provides better prediction than ordinary SK, in the sense of smaller mean squared error (MSE). SKG is similar to cokriging used in deterministic simulations and directly differentiates correlation functions. The approach introduced in Qu and Fu [32, 31] and summarized here is fundamentally different, because it generates *a set of completely new data points* rather than improving the estimated fit at the originally provided points.

Given an experiment design $(\mathbf{x}_i, n_i)$, $i = 1, 2, \cdots, k$, stochastic kriging models the simulation output $y_j(\mathbf{x}_i)$ from jth replication at design point $\mathbf{x}_i$ as:

$$y_j(\mathbf{x}_i) = \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta} + \mathsf{M}(\mathbf{x}_i) + \varepsilon_j(\mathbf{x}_i),$$

where $\mathbf{f}(\mathbf{x}_i) \in \mathbb{R}^p$ with known functions of $\mathbf{x}_i$, $\boldsymbol{\beta} \in \mathbb{R}^p$ with unknown parameters to be estimated, M is a realization of a zero-mean random field. The trend term $\mathbf{f}(\mathbf{x}_i)^T\boldsymbol{\beta}$ represents the overall surface mean and the measurement error is denoted as $\varepsilon_j(\mathbf{x}_i)$. The uncertainties in M and ε_j are referred as extrinsic and intrinsic uncertainties, respectively. Denote the sample mean of response output and the average simulation noise at $\mathbf{x}_i$ as

$$\bar{y}(\mathbf{x}_i) = \frac{1}{n_i}\sum_{j=1}^{n_i} y_j(\mathbf{x}_i), \quad \bar{\varepsilon}(\mathbf{x}_i) = \frac{1}{n_i}\sum_{j=1}^{n_i} \varepsilon_j(\mathbf{x}_i),$$

with $\bar{\mathbf{y}} = (\bar{y}(\mathbf{x}_1), \bar{y}(\mathbf{x}_2), \cdots, \bar{y}(\mathbf{x}_k))^T$.

Suppose we want to predict the response $y(\mathbf{x}_0)$ at $\mathbf{x}_0$. Let $\boldsymbol{\Sigma}_{\mathsf{M}}$ be the $k \times k$ covariance matrix implied by the random field M and $\boldsymbol{\Sigma}_{\varepsilon}$ be the $k \times k$ covariance matrix implied by the simulation noise across all design point $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$. Let $\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot) = (\mathrm{Cov}(y(\mathbf{x}_0), y(\mathbf{x}_1), \cdots, \mathrm{Cov}(y(\mathbf{x}_0), y(\mathbf{x}_k))^T$ denote the covariances between $y(\mathbf{x}_0)$ and the responses from all design points. Also, let $\mathbf{F} = (\mathbf{f}(\mathbf{x}_1), \mathbf{f}(\mathbf{x}_2), \cdots, \mathbf{f}(\mathbf{x}_k))$ be the design matrix. The MSE-optimal predictor is of the form

$$\hat{y}(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0)^T\hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot)^T(\boldsymbol{\Sigma}_{\mathsf{M}} + \boldsymbol{\Sigma}_{\varepsilon})^{-1}(\bar{\mathbf{y}} - \mathbf{F}^T\hat{\boldsymbol{\beta}}), \tag{11}$$

and the optimal MSE is

$$\mathrm{MSE}(\hat{y}(\mathbf{x}_0)) = \boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \mathbf{x}_0) - \boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot)^T[\boldsymbol{\Sigma}_{\mathsf{M}} + \boldsymbol{\Sigma}_{\varepsilon}]^{-1}\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot). \tag{12}$$

In an enhanced data setting, we observe the responses $y_j(\mathbf{x}_i)$ and the gradient estimates $\mathbf{g}_j(\mathbf{x}_i)$ for the jth simulation replication at design points $\mathbf{x}_i$. Instead of modeling the gradient estimates with the partial derivative of the random field M as in [13], we model $\mathbf{g}_j(\mathbf{x}_i)$ as a noise measurement of the true gradient $\mathbf{g}(\mathbf{x}_i)$, i.e., $\mathbf{g}_j(\mathbf{x}_i) = \mathbf{g}(\mathbf{x}_i) + \boldsymbol{\delta}_j(\mathbf{x}_i)$. Denote the sample mean of gradient estimates and the average simulation noise at $\mathbf{x}_i$ as

$$\bar{\mathbf{g}}(\mathbf{x}_i) = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{g}_j(\mathbf{x}_i), \quad \bar{\boldsymbol{\delta}}(\mathbf{x}_i) = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{\delta}_j(\mathbf{x}_i).$$

Notice that the response and the gradient estimates are noisy and usually correlated, and we assume that $\bar{\boldsymbol{\delta}}(\mathbf{x}_i)$ is independent of the random field M.

To incorporate gradient estimates into stochastic kriging, we extrapolate in the neighborhood of the original design points $\{\mathbf{x}_i\}$, $i = 1, 2, \cdots, k$, i.e., additional response data is generated via linear extrapolations using the gradient estimates as follows:

$$\mathbf{x}_i^+ = \mathbf{x}_i + \Delta\mathbf{x}_i, \qquad \bar{y}^+(\mathbf{x}_i^+) = \bar{y}(\mathbf{x}_i) + \bar{\mathbf{g}}(\mathbf{x}_i)\cdot\Delta\mathbf{x}_i. \tag{13}$$

Different extrapolation techniques can be applied in (13), and we can also add multiple points to the neighborhood of $\mathbf{x}_i$. In this preliminary study we assume that the same step size is used for all design points, i.e., $\Delta\mathbf{x}_i = \Delta x$, $i = 1, 2, \cdots, k$. We also assume that only one additional point is added in the neighborhood of $\mathbf{x}_i$. Let $\bar{y}_i = \bar{y}(\mathbf{x}_i)$ and $\bar{y}_i^+ = \bar{y}(\mathbf{x}_i^+)$ for simplicity and $\bar{\mathbf{y}}^*$ be the $2k \times 1$ vector containing all of the original response outputs and the additional response outputs in (13):

$$\bar{\mathbf{y}}^* = (\bar{y}_1, \bar{y}_2, \cdots, \bar{y}_k; \bar{y}_1^+, \bar{y}_2^+, \cdots, \bar{y}_k^+).$$

Similarly, $\mathbf{x}^+$ is defined as

$$\mathbf{x}^+ = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k; \mathbf{x}_1^+, \mathbf{x}_2^+, \cdots, \mathbf{x}_k^+).$$

To fit this augmented dataset into the stochastic kriging approach, we model the additional points similar to the original response output, i.e.,

$$\bar{y}^+(\mathbf{x}_i^+) = \mathbf{f}(\mathbf{x}_i^+)^T\boldsymbol{\beta} + \mathsf{M}(\mathbf{x}_i^+) + \varepsilon^+(\mathbf{x}_i^+),$$

and the variance of the noise $\varepsilon^+(\mathbf{x}_i^+)$ and the covariance between $\varepsilon^+(\mathbf{x}_i^+)$ and $\varepsilon(\mathbf{x}_i)$ are approximated by

$$\begin{aligned}\mathrm{Var}\left(\varepsilon^+(\mathbf{x}_i^+)\right) &= \mathrm{Var}\left(\varepsilon(\mathbf{x}_i)\right) + (\Delta x)^2 \,\mathrm{tr}\left[\mathrm{Cov}(\bar{\boldsymbol{\delta}}(\mathbf{x}_i))\right] + 2(\Delta x)\mathbf{1}^T \,\mathrm{Cov}\left(\bar{\varepsilon}(\mathbf{x}_i), \bar{\boldsymbol{\delta}}(\mathbf{x}_i)\right),\\ \mathrm{Cov}\left(\varepsilon^+(\mathbf{x}_i^+), \varepsilon(\mathbf{x}_i)\right) &= \mathrm{Var}\left(\varepsilon(\mathbf{x}_i)\right) + \Delta x \mathbf{1}^T \,\mathrm{Cov}\left(\varepsilon(\mathbf{x}_i), \bar{\boldsymbol{\delta}}(\mathbf{x}_i)\right).\end{aligned}$$

Let $\boldsymbol{\Sigma}_{\mathsf{M}}^{\dagger} = \mathrm{Cov}[\mathsf{M}(\mathbf{x}_i), \mathsf{M}(\mathbf{x}_j^+)]$, $\boldsymbol{\Sigma}_{\mathsf{M}}^{+} = \mathrm{Cov}[\mathsf{M}(\mathbf{x}_i^+), \mathsf{M}(\mathbf{x}_j^+)]$, $i, j = 1, 2, \cdots, k$, and $\boldsymbol{\Sigma}_{\mathsf{M}}^{*}$ be a $2k \times 2k$ covariance matrix across all the original design points and additional design points, which takes the form

$$\boldsymbol{\Sigma}_{\mathsf{M}}^{*} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathsf{M}} & \boldsymbol{\Sigma}_{\mathsf{M}}^{\dagger} \\ \boldsymbol{\Sigma}_{\mathsf{M}}^{\dagger} & \boldsymbol{\Sigma}_{\mathsf{M}}^{+} \end{bmatrix}.$$

Similarly, let

$$\boldsymbol{\Sigma}_{\varepsilon}^{*} = \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon} & \boldsymbol{\Sigma}_{\varepsilon}^{\dagger} \\ \boldsymbol{\Sigma}_{\varepsilon}^{\dagger} & \boldsymbol{\Sigma}_{\varepsilon}^{+} \end{bmatrix},$$

where

$$\begin{aligned}\boldsymbol{\Sigma}_{\varepsilon}^{\dagger} &= \mathrm{diag}\left\{\mathrm{Cov}\left(\varepsilon^+(\mathbf{x}_1^+), \varepsilon(\mathbf{x}_1)\right), \cdots, \mathrm{Cov}\left(\varepsilon^+(\mathbf{x}_k^+), \varepsilon(\mathbf{x}_k)\right)\right\},\\ \boldsymbol{\Sigma}_{\varepsilon}^{+} &= \mathrm{diag}\left\{\mathrm{Var}\left(\varepsilon^+(\mathbf{x}_1^+)\right), \cdots, \mathrm{Var}\left(\varepsilon^+(\mathbf{x}_k^+)\right)\right\}.\end{aligned}$$

Let $\boldsymbol{\Sigma}_{\mathsf{M}}^{*}(\mathbf{x}_0, \cdot)$ be the covariance between $y(\mathbf{x}_0)$ and all $2k$ design points. Also, let $\mathbf{F}^* = (\mathbf{f}(\mathbf{x}_1), \mathbf{f}(\mathbf{x}_2), \cdots, \mathbf{f}(\mathbf{x}_k), \mathbf{f}(\mathbf{x}_1^+), \mathbf{f}(\mathbf{x}_2^+), \cdots, \mathbf{f}(\mathbf{x}_k^+))^T$ be the design matrix. Under the enhanced data setting, we can easily find the MSE-optimal predictor and the corresponding MSE by substituting $\bar{\mathbf{y}}^*$, $\mathbf{F}^*$, $\boldsymbol{\Sigma}_{\mathsf{M}}^{*}(\mathbf{x}_0, \cdot)$, $\boldsymbol{\Sigma}_{\mathsf{M}}^{*}$ and $\boldsymbol{\Sigma}_{\varepsilon}^{*}$ for $\bar{\mathbf{y}}$, $\mathbf{F}$, $\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot)$, $\boldsymbol{\Sigma}_{\mathsf{M}}$ and $\boldsymbol{\Sigma}_{\varepsilon}$ in (11) and (12), respectively.

The random field M is assumed to be second-order stationary, i.e.,

$$\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_i, \mathbf{x}_j) = \tau^2 R(\mathbf{x}_i - \mathbf{x}_j; \boldsymbol{\theta}),$$

where $\tau^2 = \mathrm{Var}[\mathsf{M}(\mathbf{x})]$ and $R(\mathbf{x}_i - \mathbf{x}_j; \boldsymbol{\theta})$ is a correlation function with parameter $\boldsymbol{\theta}$ depending on the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$. The extended covariance matrix $\boldsymbol{\Sigma}_{\mathsf{M}}^{*}$ follows the same correlation structure, and the parameters $(\tau^2, \boldsymbol{\theta})$ and $\boldsymbol{\beta}$ can be estimated from maximum likelihood estimators (MLEs) provided that $\boldsymbol{\Sigma}_{\varepsilon}^{*}$ is known.

Illustrative Numerical Example

Here, we consider a stylized example adopted from [39] where direct gradient estimates are just assumed to be available in the canonical form:

$$y(x) = \exp(-1.4x)\cos(7\pi x/2) + \varepsilon, \;\; -2 \le x \le 0, \;\; \varepsilon \sim \mathcal{N}(0,1), \quad g(x) = y'(x) + \delta, \;\; \delta \sim \mathcal{N}(0,25).$$

This is just an exponentially damped sinusoidal function with added noise. A Gaussian correlation function $R(x,x') = \exp\{\theta(x-x')^2\}$ is used for the kriging. Four experiments will be used to illustrate some characteristics of SK, SKG, and GESK, with respect to the choice of design points and number of replications.

In the upper left graph in Figure 2 we plot predictions using six equally spaced design points based on 20 samples at each point. Because GESK is able to explore the design space more via extrapolation, it does a better job of capturing the fluctuations of the response surface. When the number of points is increased to eight (again equally spaced, with 20 samples per point), which in addition to being a larger number of points also happens to place the points in more critical parts of the design space, all three of the approaches are able to do a decent job of capturing the true shape, as indicated in the upper right graph.

In the next experiment we increase the number of replications by an order of magnitude (from 20 to 200 per each of the eight design points), which surprisingly leads to a dramatically worse fit for SK, as shown in the lower left graph. The last experiment then increases the number of points again, this time to 20 (with 200 replications per design point), at which point both SK and SKG perform poorly, whereas GESK continues to provide an improved fit. Again, because the location of the 20 points is not ideally suited to the shape of

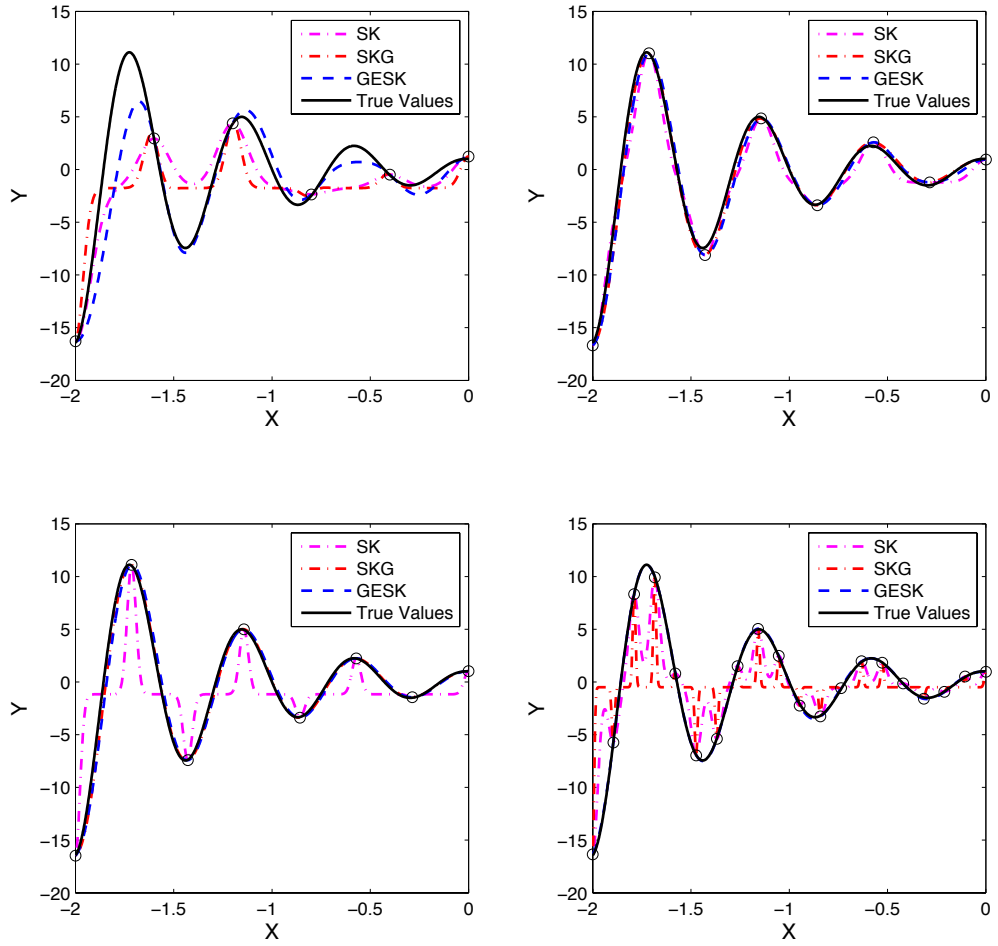

Figure 2: Test function: $y(x) = \exp(-1.4x)\cos(7\pi x/2) + \varepsilon$, using equally spaced design points; top row 20 samples per design point; bottom row 200 samples per design point; SK and SKG have poor fits when design points not well placed (cf. top left and bottom right), whereas GESK robust to placement.

the curve, SKG is lead astray, whereas GESK is still able to capture the shape due to the additional points generated that can serve to compensate.

This simple numerical example indicates how both GESK and SKG can dramatically improve the meta-model fit over ordinary stochastic kriging.

2.3 Simulation-Based and Sampling Methods for Markov Decision Processes

We define an MDP $\{X_i, i = 0, 1, ..., T\}$ on state space $\mathscr{S}$ and action space $\mathscr{A}$ (cf. e.g., [3, 9]). In period (stage) i, the MDP in state $X_i \in \mathscr{S}$ takes action $a_i \in \mathscr{A}$, incurs cost $C_i(X_i, a_i, \omega_i)$, where ω_i denotes the stochastic element (e.g., random number), and then transitions according to

$$X_{i+1} = f_{i+1}(X_i, a_i, \omega_i),$$

where $f_i(x, a, \cdot)$ denotes the (stochastic) transition function in period i for action a taken in state x. For notational simplicity, we have not made state and action spaces period dependent.

The objective is to find a feedback control **policy** $\pi \equiv \{\pi_i(x)\}_{i=0}^{T-1}$, which is a sequence of decision rules

specifying the action a_i taken when in state x in period i, that minimizes an expected cost function, usually either finite horizon total cost, finite horizon discounted total cost, infinite horizon average cost, or infinite horizon discounted total cost. While we will consider all of these types of cost criteria in our research, in this proposal we focus on the discounted total cost setting, both finite and infinite horizon; for the risk-neutral case, we define the **value function** associated with a policy and initial state:

$$V^{\pi}(x) = E\left[\sum_{i=0}^{T-1} \alpha^i C_i(X_i, a_i, \omega_i) \middle| X_0 = x\right], \tag{14}$$

where α is the (one-period) discount factor and T could be infinite, under the assumption that the limit is then well defined. As stated earlier, the chief context is the setting in which simulation is required to generate the system dynamics (state transitions) and/or period costs.

We begin by defining some familiar quantities:

$$\begin{aligned} Q_i(x,a) &= \text{(expected) cost-to-go (Q-function) in period } i \text{ for action } a \text{ taken in state } x \\ &\quad \text{and optimal actions taken henceforth;} \\ V_i(x) &= \text{optimal value function in period } i \text{ for state } x. \end{aligned}$$

Then we have the usual Bellman optimality equation [3, 30]:

$$V_i(x) = \inf_a \{E\left[C_i(x,a,\omega_i) + \alpha V_{i+1}(f_{i+1}(x,a,\omega_i))\right]\}, \tag{15}$$

written here in two-part form:

$$Q_i(x,a) = E\left[C_i(s,a,\omega_i) + \alpha V_{i+1}(f_{i+1}(x,a,\omega_i))\right], \tag{16}$$

$$V_i(x) = \inf_a Q_i(x,a). \tag{17}$$

An optimal policy in period i will be denoted by

$$\pi_i^*(x) \in \arg\inf_a Q_i(x,a), \;\; i = 0,...,T-1, \;\; x \in \mathscr{S}. \tag{18}$$

When the policy is stationary, the subscript/argument i will be dropped. In the infinite horizon stationary case, (15) takes the following form:

$$V(x) = \inf_a \{E\left[C(x,a,\omega) + \alpha V(f(x,a,\omega))\right]\}, \tag{19}$$

and we will assume there exists an optimal stationary policy such that

$$\pi^*(x) \in \arg\inf_a Q(x,a), \; x \in \mathscr{S}.$$

Traditional methods of policy iteration, value iteration, and variants based on linear programming all suffer from the curse of dimensionality. Furthermore, the transition function f_i is generally not known in closed form (note that in traditional MDP formulations, it is expressed in terms of explicit *transition probabilities* assumed given), but may be generated by a complicated stochastic simulation model, so in such a setting, the traditional methods are not directly applicable.

2.3.1 Non-Convex Dynamic Measures

In recent years, with particular motivation from the field of mathematical finance, many new approaches to the incorporation of risk into decision making have been developed. One well-studied approach involves coherent or convex risk measures (cf. [2, 17]); here, we briefly describe some of the work of Ruszczyński and colleagues in applying this approach to multi-period problems [38, 6]. In this work, expected value operators are replaced by more general risk measures. It is shown that time-consistent dynamic risk measures can be written recursively in terms of one-step conditional risk measures; furthermore, these one-step conditional risk measures are assumed to satisfy properties of coherent measures of risk. Examples of one-step conditional risk measures satisfying these conditions are mean-semideviation models and conditional Value-at-Risk (cf. Rockafellar and Uryasev [35, 36]). These risk measures are applied to controlled Markov processes by defining *Markov risk measures*.

To motivate the form of these risk measures, consider the cost function (14) with $\alpha = 1$ and where C_i is only a function of X_i and a_i. Using the tower property of conditional expectation, $V^\pi(X_0)$ can be written as

$$V^\pi(X_0) = C_0(X_0, a_0) + \rho_0(C_1(X_1, a_1) + \rho_1(C_2(X_2, a_2) + \cdots + \rho_{T-2}(C_{T-1}(X_{T-1}, a_{T-1}))\cdots)), \tag{20}$$

where, for any function h,

$$\rho_t(h(X_{t+1}, a_{t+1})) = E[h(X_{t+1}, a_{t+1})|X_t].$$

The work of Ruszczyński and colleagues generalizes the cost function (20) to more general Markov risk measures satisfying a number of properties. Under appropriate hypotheses, value functions are defined and dynamic programming equations are shown to lead to optimal Markov policies. Discounted infinite-horizon problems and undiscounted transient problems are also studied, along with their corresponding analogs of value iteration and policy.

Cumulative Prospect Theory (CPT)

Prospect theory was proposed in the 1970s by Kahneman and Tversky to model observed behavior in human decision making that could not be adequately explained by existing utility theory. Although prospect theory did a better job of explaining the experimental data, and served as the basis on which the Nobel Prize was awarded to Kahneman in 2002 (Tversky had already passed away), there were still some theoretical deficiencies that led to various proposed alternatives to and extensions of prospect theory. Cumulative prospect theory (CPT), also developed by Kahneman and Tversky [41], posits a utility function that has a reference point against which gains and losses are measured, and is concave on gains and convex on losses (i.e., horizontal S-shape), along with a probability weighting function that transforms the probability measure such that a small probability is inflated and a large probability is deflated.

Based on the special utility and weighting functions, a CPT performance measure can be characterized.

A *CPT performance measure has the following form:*

$$\rho(R) = \int_0^\infty w^+\left(\mathbb{P}\left(u_+\left((R-B)_+\right) > t\right)\right) dt - \int_0^\infty w^-\left(\mathbb{P}\left(u_-\left((R-B)_-\right) > t\right)\right) dt, \tag{21}$$

where $w^+ : [0,1] \to [0,1]$ *and* $w^- : [0,1] \to [0,1]$ *are two continuous non-decreasing functions.* $u_+ : \mathbb{R}^+ \to \mathbb{R}^+$ *and* $u_- : \mathbb{R}^+ \to \mathbb{R}^+$ *are two utility functions, and the random variable B represents the reference point against which the performance is measured.*

The key to understanding Equation (21) is

$$\int_0^\infty w\left(\mathbb{P}\left(u(x) > t\right)\right) dt = \int_{u(x)\geq 0} u(x) d\left[w\left(F_X\left(x\right)\right)\right], \tag{22}$$

so that the CPT performance measure can be interpreted as the sum of the distorted expected gains and losses. As mentioned earlier, CPT and other related approaches better explain empirically observed human behavior in many cases. One way to understand these approaches is the following: recall that an expected utility assigns a utility to each outcome, and then weights each outcome according to its probability of occurrence. In CPT, not only are the outcomes transformed using the utility function, but the probabilities by which they are weighted are also transformed using the weighting function.

Probabilistic Sensitivity

Combining a number of other approaches, including both expected utility maximization and CPT, He and Zhou [19] proposed a method of risk-averse optimization via quantiles. At the heart of the unification is the incorporation of the idea of probabilistic sensitivity into risk measures, which is a source of technical challenges. Overcoming these challenges, He and Zhou restrict their investigation to the drift-diffusion case, which can be tackled using the martingale approach. Probabilistic sensitivity replaces the (probabilistic) linearity of expected utility theory with more expressive but also more challenging nonlinear weighting functions. Because of this modification, the resulting stochastic optimal control problems do not enjoy the

desirable properties of time-consistency and convexity, which lead to fresh challenges in applying dynamic programming. Our proposed research builds upon ideas from this work in the MDP setting.

Research Results

The class of coherent risk measures satisfies several properties, one of which is convexity. In many cases, the convexity requirement is too strong for practical applications. Specifically, the CPT risk measures, which have nonlinear probabilistic sensitivity, are not convex, but empirical evidence provides support of their use for predicting human decision-making processes. The class of CPT-inspired measures bestows upon practitioners the flexibility of choosing gain/loss measure distortion functions, which is important in expressing risk. This alternative approach of representing risk-sensitivity using probabilistic sensitivity sets CPT-inspired performance measures apart from the existing risk-sensitive measures. A CPT-inspired performance measure can be used as the one-step conditional risk measure $\rho_t(\cdot)$ (i.e., condition on the knowledge at time t) in Equation (20).

As noted before, we are interested in a broader class of risk-sensitive measures than expected utility; however, we still need to impose some conditions on the performance measure such that dynamic programming is still applicable. We focus on a class of risk-sensitive measures that contains a subset of CPT-inspired risk measures. The members of this class of risk-sensitive measures satisfy the following conditions.

Assumption 3. *The one-step conditional risk measure, $\rho_t(\cdot)$, satisfies the following conditions:*

1. *If $Z \leq W$ then $\rho_t(Z) \leq \rho_t(W)$, $\forall Z, W \in \mathscr{L}_{t+1}$;*
2. *$\rho_t(\beta Z) = \beta \rho_t(Z)$, $\forall Z \in \mathscr{L}_{t+1}, \beta \geq 0$,*

where $\mathscr{L}_{t+1}$ is the space of $t+1$ measurable and integrable functions.

A desirable property of the class of risk measures satisfying Assumption 3 is that it include expected utility and coherent risk measures. As we will see, this class also contains a large subset of CPT-inspired risk measures. We denote the space of probability measures over A by $\mathscr{P}(A)$ and the state transition probability measure by $Q_t(\cdot|x,a)$. The first result presents the optimality criteria for the class of risk measures satisfying Assumption 3. (The proofs of all theorems in this section can be found in [28].)

Theorem 1. *Assume Assumption 3 and the following conditions hold:*

1) $\forall x \in \mathscr{S}$, the stochastic kernels $Q_{t,x} : a \to Q_t(\cdot|x,a)$ are continuous.

2) The one-step dynamic risk measure $\{\rho_t\}_{t=0}^{T-1}$ is Markov, and $\exists$ a sequence of corresponding risk transition mappings $\sigma_t : m \to \sigma(\psi, x, m)$, $t = 0, \ldots, T-1$ that are lower semi-continuous.

3) The functions $\{C_t(\cdot,\cdot,\cdot)\}_{t=0}^{T-1}$ are bounded, measurable, and $a \to C_t(\cdot, a, \cdot)$ is lower semi-continuous.

4) $\forall x \in \mathscr{S}$ and $t \in [0, \ldots, T-1]$ the set $A_t(x)$ is compact.

5) The function C_T is bounded and measurable.

Then a minimizer for the dynamic programing equations,

$$v_t(x) = \min_{\delta \in \mathscr{P}(A(x))} \sigma_t\left(C_t(x,\cdot,\cdot) + v_{t+1}(\cdot), x, \delta \circ Q_{t,x}\right)$$
$$v_T(x) = C_T(x) \quad x \in \mathscr{S},\ t = 1, \ldots, T-1, \tag{23}$$

exists. Furthermore, an optimal policy, $\pi^* = \left\{\pi_0^*, \ldots, \pi_{T-1}^*\right\}$, exists and each $\pi_{t,x}^*$ is a minimizer for the right-hand side of Equation (23). In addition, every measurable solution of Equation (23) at time 0, v_0, is an optimal solution for Equation (20).

In Theorem 1, the assumptions are standard with the exception of the second. In the second assumption, $\rho_t(\cdot)$ is required to be Markov, which means it can be expressed as a function of the risk, the current state, and the transition probability (i.e., $\rho_t(z) = \sigma_t(z, x, m)$, where x is the current state, and m is the transition

probability). Other more technical requirements (i.e., boundedness) for a conditional risk measure to be Markov have been omitted here.

Theorem 1 provides the optimality criteria for a class of dynamic risk measures. The next goal is to identify the subset of CPT-inspired risk measures that belongs to this class, i.e., satisfy Assumption 3 and the second assumption of Theorem 1. The following result provides conditions on the gain/loss probability weighting functions such that the corresponding CPT risk measure belongs to the class of risk measures satisfying Assumption 3; hence, it is suitable for dynamic programming.

Theorem 2. *If w^+ and w^- are continuous and monotonically non-decreasing functions in Equation (21), then the assumptions in Theorem 1 are satisfied; hence, dynamic programming is applicable.*

Since the requirement on these probability weighting functions is minimal (i.e., continuity and monotonicity), the class of CPT-inspired measures is large. Examples of popular probability weighting functions and their empirical support can be found in existing literature from the decision theory community. The finite-horizon case can be extended to the infinite-horizon transient case via assumptions such as k-step contraction and the Markov model being uniformly transient, as shown in [28].

2.3.2 Additional Results on Simulation-Based and Sampling Methods for Markov Decision Processes

In a simulation-based approach to MDPs, we have developed in [20] a simulation-based algorithm called Approximate Stochastic Annealing (ASA) for solving finite horizon MDPs. The algorithm iteratively estimates the optimal policy by sampling from a sequence of probability distribution functions over the policy space. By exploiting a novel connection of ASA to the stochastic approximation method, we show that the sequence of distribution functions generated by the algorithm converges to a degenerate distribution that concentrates only on the optimal policy. Numerical examples are also provided to illustrate the algorithm.

In a supply chain application, we consider in [34] a make-to-order business that serves customers in multiple priority classes. Orders from customers in higher classes bring greater revenue, but they expect shorter lead times than customers in lower classes. In making lead time promises, the firm must recognize preexisting order commitments, uncertainty over future demand from each class, and the possibility of supply chain disruptions. We model this scenario as an MDP and use reinforcement learning to determine the firms lead-time policy. In order to achieve tractability on large problems, we utilize a sequential decision making approach that effectively allows us to eliminate one dimension from the state space of the system. Initial numerical results from the sequential dynamic approach suggest that the resulting policies more closely approximate optimal policies than static optimization approaches.

Portfolio Selection as introduced by Harry Markowitz laid the foundation for Modern Portfolio Theory. However, the assumption that underlying asset returns follow a normal distribution and that investors are indifferent to skew and kurtosis is not practically suited for the hedge fund environment. Additionally, the lockup and notice provisions built into hedge fund contracts make portfolio rebalancing difficult and justify the need for dynamic allocation strategies. Market conditions are dynamic; therefore, rebalancing constraints in the face of changing market environments can have a severe impact on return generation. There is a need for sophisticated yet tractable solutions to the multi-period problem of hedge fund portfolio construction and rebalancing. In [14], we generalize the hedge fund asset return distribution to a multivariate K-mean Gaussian mixture distribution; model the multi-period hedge fund allocation problem as a Partially Observable Markov Decision Process (POMDP); and propose practical rebalancing strategies that represent a convergence of literature on hedge fund investing, regime switching, and dynamic portfolio optimization.

In an application to semiconductor manufacturing, we present in [33] the architecture and implementation of a preventive maintenance optimization software tool (PMOST), based on algorithms for the optimal scheduling of preventive maintenance (PM) tasks in semiconductor manufacturing operations. We also present results from four complex simulation case studies, based on real industrial data and employing full fab models, to illustrate the use, data needs and outcomes produced by PMOST. These results demonstrate

significant improvements in tool production and consolidation of PM tasks. We give a description of the different software modules that compose PMOST, to provide guidelines as well as a template for other implementations of the PM optimization algorithms utilized by PMOST.

References

[1] Bruce Ankenman, Barry L. Nelson, and Jeremy Staum. Stochastic Kriging for Simulation Metamodeling. *Operations research*, 58(2):371–382, March 2010.

[2] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[3] D. P. Bertsekas. *Dynamic Programming and Optimal Control, Vols. 1 & 2*. Athena Scientific, 1995.

[4] D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, 1996.

[5] L. Cao and Michael C. Fu. Estimating greeks for variance-gamma. In *Proceedings of the 2010 Winter Simulation Conference*, pages 2620–2628, Piscataway, New Jersey, December 2010. Institute of Electrical and Electronics Engineers, Inc.

[6] Ö. Çavuş and A. A. Ruszczyński. Risk-averse control of undiscounted transient Markov models. Arxiv preprint arXiv:1203.5437, 2012.

[7] H. S. Chang, M. C. Fu, J. Hu, and S. I. Marcus. An adaptive sampling algorithm for solving Markov decision processes. *Operations Research*, 53:126–139, 2005.

[8] H. S. Chang, M. C. Fu, J. Hu, and S. I. Marcus. Recursive learning automata approach to Markov decision processes. *IEEE Transactions on Automatic Control*, 52(7), 2007.

[9] H. S. Chang, M. C. Fu, J. Hu, and S. I. Marcus. *Simulation-based Algorithms for Markov Decision Processes*. Springer, 2007.

[10] H. S. Chang, R. Givan, and E. K. P. Chong. Parallel rollout for on-line solution of partially observable Markov decision processes. *Discrete Event Dynamic Systems: Theory and Application*, 15(3):309–341, 2004.

[11] H. S. Chang and J. Hu. On the probability of correct selection in ordinal comparison over dynamic networks. *Journal of Optimization Theory and Applications*, 2013.

[12] H. S. Chang, H.-G. Lee, M. C. Fu, and S. I. Marcus. Evolutionary policy iteration for solving Markov decision processes. *IEEE Transactions on Automatic Control*, 50(11):1804–1808, 2005.

[13] X. Chen, B. Ankenman, and B. L. Nelson. Enhancing stochastic kriging metamodels with gradient estimators. Working Paper, 2011.

[14] D. Cru and J. Hu. Dynamic hedge fund asset allocation under multiple regimes. In *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing*, pages 1376–1383, 2010.

[15] C.H. Chen M.C. Fu D. He, L.H. Lee and S. Wasserkrug. Simulation optimization using the cross-entropy method with optimal computing budget allocation. *ACM Transactions on Modeling and Computer Simulation*, 20(1), 2010.

[16] T. K. Das, A. Gosavi, S. Mahadevan, and N. Marchalleck. Solving semi-Markov decision problems using average reward reinforcement learning. *Management Science*, 45:560–574, 1999.

[17] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. de Gruyter, 2004.

[18] M.C. Fu and Huashuai Qu. Augmented regression with direct gradient estimates. Working Paper, 2012.

[19] Xue Dong He and Xun Yu Zhou. Portfolio choice via quantiles. *Mathematical Finance*, 21(2):203–231, 2011.

[20] J. Hu and H.S. Chang. An approximate stochastic annealing algorithm for finite horizon markov decision processes. In *Proceedings of the 49th IEEE Conference on Decision and Control*, pages 5338–5343, Piscataway, New Jersey, December 2010. Institute of Electrical and Electronics Engineers, Inc.

[21] J. Hu, H.S. Chang, M.C. Fu, , and S.I. Marcus. Dynamic sample budget allocation in model-based optimization. *Journal of Global Optimization*, 2010.

[22] J. Hu, M. C. Fu, and S. I. Marcus. A model reference adaptive search method for global optimization. *Operations Research*, 55(3):549–568, 2007.

[23] J. Hu, M. C. Fu, V. Ramezani, and S. I. Marcus. An evolutionary random search algorithm for solving Markov decision processes. *INFORMS Journal on Computing*, 19(2):161–174, 2007.

[24] J. Hu and P. Hu. An approximate annealing search algorithm to global optimization and its connections to stochastic approximation. In *Proceedings of the 2010 Winter Simulation Conference*, pages 1223–1234, Piscataway, New Jersey, December 2010. Institute of Electrical and Electronics Engineers, Inc.

[25] J. Hu and P. Hu. Annealing adaptive search, cross-entropy, and stochastic approximation in global optimization. *Naval Research Logistics*, 58:457–477, 2011.

[26] J. Hu, P. Hu, and H.S. Chang. A stochastic approximation framework for a class of randomized optimization algorithms. *IEEE Trans. Automatic Control*, 57:165–178, 2012.

[27] Y. Su J. Hu, W. Zhu and W. K. Wong. Controlled optimal design program for the logit dose response model. *Journal of Statistical Software*, 35:1–17, 2010.

[28] Kun Lin. *Stochastic Systems with Cumulative Prospect Theory*. PhD thesis, University of Maryland, 2013.

[29] W. B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley, New York, NY, 2007.

[30] M. L. Puterman. *Markov Decision Processes*. John Wiley & Sons, 1994.

[31] Huashuai Qu and M.C. Fu. Augmented stochastic kriging with direct gradient estimates. Working Paper, 2012.

[32] Huashuai Qu and M.C. Fu. On direct gradient enhanced simulation metamodels. In *Proceedings of the 2012 Winter Simulation Conference*, pages forthcoming, to appear, Piscataway, New Jersey, 2012.

[33] J.A. Ramirez-Hernández, J. Crabtree, X. Yao, E. Fernández, M.C. Fu, M. Janakiram, S. I. Marcus, and M. O'Connor. Optimal preventive maintenance scheduling in semiconductor manufacturing systems: Software tool and simulation case studies. *IEEE Transactions on Semiconductor Manufacturing*, 23:477–489, 2010.

[34] M. Reindorp and M.C. Fu. Dynamic lead time promising. In *Proceedings of 2011 IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning*, 2011.

[35] R. T. Rockafellar and S. P. Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2:21–41, 2000.

[36] R. T. Rockafellar and S. P. Uryasev. Conditional value-at-risk for general loss distribution. *The Journal of Banking and Finance*, 26:1443–1471, 2002.

[37] B. Van Roy and J. N. Tsitsiklis. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 14:694–703, 2001.

[38] A. Ruszczyński. Risk-averse dynamic programming for Markov decision processes. *Math. Programming, Ser. B*, 125:235–261, 2010.

[39] T. J. Santner, B. Williams, and W. Notz. *The Design and Analysis of Computer Experiments*. Springer-Verlag, 2003.

[40] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, 1998.

[41] A. Tversky and D. Kahneman. Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5:297–323, 1992.

[42] Yongqiang Wang, Michael C. Fu, and Steven I. Marcus. Model based evolutionary optimization. In *Proc. Winter Simulation Conference*, 2010.

[43] Yongqiang Wang, Michael C. Fu, and Steven I. Marcus. A new derivative estimator for discontinuous payoff functions with financial applications. In *Proc. 2010 INFORMS Annual Meeting*, 2010.

[44] Yongqiang Wang, Michael C. Fu, and Steven I. Marcus. A new stochastic derivative estimator for discontinuous payoff functions with application to financial derivatives. *Operations Research*, 60:447–460, 2012.

[45] E. Zhou, M.C. Fu, and S.I. Marcus. A particle filtering framework for randomized optimization algorithms. *IEEE Transactions on Automatic Control*, 2013.

[46] E. Zhou and J. Hu. Combining gradient-based optimization with stochastic search. In *Proceedings of the 2012 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc, December 2012.

[47] E. Zhou and J. Hu. Gradient guided adaptive stochastic search for non-differentiable optimization. *IEEE Transactions on Automatic Control*, 2013.

ii

3 Research Output from Prior AFOSR Support:

3.1 Publications

1. L. Cao and M.C. Fu, Estimating Greeks for Variance-Gamma, *Proceedings of the Winter Simulation Conference*, 2620–2628, 2010.

2. H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus, *Simulation-Based Algorithms for Markov Decision Processes*, Springer-Verlag, 2007 (research monograph).

3. H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus, *Simulation-Based Algorithms for Markov Decision Processes, Second Edition*, Springer-Verlag, 2013 (research monograph).

4. H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus, An Asymptotically Efficient Simulation-based Algorithm for Finite Horizon Stochastic Dynamic Programming. *IEEE Transactions on Automatic Control*, Vol. 52, 89-94, 2007.

5. H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus, Recursive Learning Automata Approach to Markov Decision Processes. *IEEE Transactions on Automatic Control*, Vol. 52, 1249-1355, 2007.

6. H.S. Chang, M.C. Fu, and S.I. Marcus, Adversarial Multi-Armed Bandit Approach to Two-Person Zero-Sum Markov Games. *Proceedings of the 46th IEEE Conference on Decision and Control*, 2007.

   H.S. Chang and J. Hu, On the Probability of Correct Selection in Ordinal Comparison over Dynamic Networks. *Journal of Optimization Theory and Applications*, to appear, 2013.

7. H.S. Chang, J. Hu, M.C. Fu, and S.I. Marcus, Adaptive Adversarial Multi-Armed Approach to Two-Person Zero-Sum Markov Games. *IEEE Transactions on Automatic Control*, Vol. 55, 463-468, February 2010.

8. C.H. Chen, M.C. Fu, and L. Shi, Simulation and Optimization, *Tutorials in Operations Research,* Z.L. Chen and S. Raghavan, editors, INFORMS, 247-260, 2008.

9. C.H. Chen, D. He, and M.C. Fu, Efficient Simulation Budget Allocation for Selecting an Optimal Subset. *INFORMS Journal on Computing*, Vol. 20, 579-595, 2008.

10. M. Chen, J.Q. Hu, and M.C. Fu, Perturbation Analysis of a Dynamic Priority Call Center, *IEEE Transactions on Automatic Control*, Vol. 55, No. 5, 1191-1196, 2010.

11. D. Cru and J. Hu, Dynamic Hedge Fund Asset Allocation Under Multiple Regimes, *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing*, 1376-1383, 2010.

12. Q. Fan and J. Hu, Adaptive Simulation Budget Allocation for Determining the Best Design, *Proceedings of the 2013 Winter Simulation Conference*, forthcoming, 2013.

13. E. Feinberg, J. Hu, and E. Yuan, Voltage and Reactive Power Control Using Approximate Stochastic Annealing, *Proceedings of the 52nd IEEE Conference on Decision and Control*, forthcoming, 2013.

14. M.C. Fu, Variance-Gamma and Monte Carlo. *Advances in Mathematical Finance*, M.C. Fu, R.A. Jarrow, J.-Y. Yen, and R.J. Elliott, editors, Birkhäuser, 21–35, 2007.

15. M.C. Fu, Are We There Yet? The Marriage Between Simulation & Optimization. *OR/MS Today*, 16–17, June 2007.

16. M.C. Fu and W.C. Howell, Traffic Light Signal Optimization via Simulation. *Proceedings of the International Modeling & Simulation Multiconference*, 241–246, 2007.

17. M.C. Fu, J.Q. Hu, C.H. Chen, and X. Xiong, Simulation Allocation for Determining the Best Design in the Presence of Correlated Sampling. *INFORMS Journal on Computing*, Vol. 19, No. 1, 101-111, 2007.

18. M.C. Fu and H. Qu, Augmented regression with direct gradient estimates, WorkingPaper, 2012.

19. A. Gosavi, S. Murray, J. Hu, and S. Ghosh, Model-Building Adaptive Critics for Semi-Markov Decision Processes, *International Journal of Artificial Intelligence and Soft Computing Research*, forthcoming, 2012.

20. D. He, L.H. Lee, C.H. Chen, M.C. Fu, and S. Wasserkrug, Simulation Optimization Using Optimal Computing Budget Allocation with the Cross-Entropy Method, *ACM Transactions on Modeling and Computer Simulation*, Vol. 20, No. 1, 2010.

21. J.W. Heath, M.C. Fu, and W. Jank, New Global Optimization Algorithms for Model-Based Clustering. *Computational Statistics and Data Analysis*, Vol. 53, 3999-4017, October 2009.

22. J. Hu, Simulation Optimization via Model-based Stochastic Search, in *Handbook of Simulation Optimization*, (Ed.) M. C. Fu, forthcoming, 2013.

23. J. Hu and H.S. Chang, An Approximate Stochastic Annealing Algorithm for Finite Horizon Markov Decision Processes, *Proceedings of the 49th IEEE Conference on Decision and Control*, 5338-5343, 2010.

24. J. Hu and H. S. Chang, Approximate Stochastic Annealing for Online Control of Infinite Horizon Markov Decision Processes, *Automatica*, forthcoming, 2013.

25. J. Hu, M.C. Fu, and S.I. Marcus, A Model Reference Adaptive Search Method for Global Optimization. *Operations Research*, Vol.55, 549-568, 2007.

26. J. Hu, M.C. Fu, and S.I. Marcus, A Model Reference Adaptive Search Method for Stochastic Optimization with Applications to Markov Decision Processes, *Proceedings of the 46th IEEE Conference on Decision and Control*, 2007.

27. J. Hu, M.C. Fu, V. Ramezani, and S.I. Marcus, An Evolutionary Random Search Algorithm for Solving Markov Decision Processes. *INFORMS Journal on Computing*, Vol.19, 161-174, 2007.

28. J. Hu, M.C. Fu, and S.I. Marcus, A Model Reference Adaptive Search Method for Stochastic Global Optimization. *Communications in Information and Systems* Vol.8, 245-276, 2008.

29. J. Hu and P. Hu, An Approximate Annealing Search Algorithm to Global Optimization and Its Connections to Stochastic Approximation, *Proceedings of the 2010 Winter Simulation Conference*, 1223-1234, 2010.

30. J. Hu and P. Hu, Annealing Adaptive Search, Cross-Entropy, and Stochastic Approximation in Global Optimization, *Naval Research Logistics*, Vol. 58, 457-477, 2011.

31. J. Hu, P. Hu, and H. S. Chang. A Stochastic Approximation Framework for a Class of Randomized Optimization Algorithms, *IEEE Trans. Automatic Control*, Vol. 57, 165-178, 2012.

32. J. Hu and H. S. Chang, A Population-Based Cross-Entropy Method with Dynamic Sample Allocation, *Proceedings of the 47th IEEE Conference on Decision and Control*, 2007.

33. J. Hu and Z. Su, Adaptive Resampling Algorithms for Estimating Bootstrap Distributions, *Journal of Statistical Planning and Inference*, Vol. 138, 1763-1777, 2008.

34. J. Hu and Z. Su, Bootstrap Quantile Estimation via Importance Resampling, *Computational Statistics and Data Analysis*, Vol. 52, 5136-5142, 2008.

35. J. Hu and Z. Su, Efficient Error Determination in Sequential Clinical Trial Design, *Journal of Computational and Graphical Statistics*, Vol. 17, 925-948, 2008.

36. J. Hu, H. S. Chang, M.C. Fu, and S.I. Marcus, Dynamic Sample Budget Allocation in Model-Based Optimization, *Journal of Global Optimization*, DOI: 10.1007/s10898-009-9490-3, 2010.

37. J. Hu, W. Zhu, Y. Su, and W. K. Wong, Controlled Optimal Design Program for the Logit Dose Response Model, *Journal of Statistical Software*, Vol. 35, 1-17, 2010.

38. U. Kuter and J. Hu, Computing and Using Lower and Upper Bounds for Action Elimination in MDP Planning, *Proceedings of the 7th Symposium on Abstraction, Reformulation and Approximation (SARA-07)*, Springer Lecture Notes in Computer Science (4612), 243-257, 2007.

39. K. Lin, Stochastic Systems with Cumulative Prospect Theory, Ph.D. Thesis, University of Maryland, 2013.

40. K. Lin and S.I. Marcus, Cumulative Weighting Optimization: The Discrete Case, *Proc. Winter Simulation Conference*, Washington, DC, Dec. 8-11, 2013.

41. K. Lin and S.I. Marcus, Dynamic Programming with Non-Convex Risk-Sensitive Measures, *Proc. 2013 American Control Conference*, Washington, DC, June 17-19, 2013.

42. S.I. Marcus, J. Hu, Y. Wang, E. Zhou, and M. Fu, A Survey of Some Model-Based Methods in Global Optimization, in *Optimization, Control, and Applications of Stochastic Systems*, D. Hernández-Hernández and J. A. Minjárez-Sosa, eds. Berlin: Springer-Verlag, 2012, 157-180.

43. H. Qu and M. C. Fu, Augmented stochastic kriging with direct gradient estimates, Working Paper, 2012.

44. J.A. Ramirez-Hernández, J. Crabtree, X. Yao, E. Fernández, M.C. Fu, M. Janakiram, S.I. Marcus, M. O'Connor, and N. Patel, Optimal Preventive Maintenance Scheduling in Semiconductor Manufacturing Systems: Software Tool and Simulation Case Studies, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 23, 477-489, August 2010.

45. M. Reindorp and M.C. Fu, Dynamic Lead Time Promising, *Proceedings of 2011 IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning*, 2011.

46. Z. Su, J. Hu, and W. Zhu, Multi-Step Variance Minimization in Sequential Tests, *Statistics and Computing*, Vol. 18, 101-108, 2008.

47. Y. Wang, M.C. Fu, and S.I. Marcus, A New Stochastic Gradient Estimator for American Option Pricing, *Proc. of the European Control Conference*, Budapest, Hungary, August 23-26, 2009.

48. Y. Wang, M.C. Fu, and S.I. Marcus, Sensitivity Analysis for Barrier Options, *Proc. Winter Simulation Conference*, Austin, TX, Dec. 13-16, 2009.

49. Y. Wang, M.C. Fu, and S.I. Marcus, A New Derivative Estimator for Discontinuous Payoff Functions with Financial Applications, *Proc. 2010 INFORMS Annual Meeting*, Austin, TX, November 7-10, 2010.

50. Y. Wang, M.C. Fu, and S.I. Marcus, Model Based Evolutionary Optimization, *Proc. Winter Simulation Conference*, Baltimore, MD, Dec. 5-8, 2010.

51. Y. Wang, M.C. Fu, and S.I. Marcus, A New Stochastic Derivative Estimator for Discontinuous Payoff Functions with Application to Financial Derivatives, *Operations Research*, Vol. 60, 447-460, 2012.

52. E. Zhou, M.C. Fu, and S.I. Marcus, Solving Continuous-State POMDPs via Projection Particle Filtering , *Proc. Eighth International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, Montreal, Canada, July 6-11, 2008.

53. E. Zhou, M.C. Fu, and S.I. Marcus, A Particle Filtering Framework for Simulation-Based Optimization, *Proc. 2008 INFORMS Annual Meeting*, Washington, DC, Oct. 12-15, 2008.

54. E. Zhou, M.C. Fu, and S.I. Marcus, A Particle Filtering Framework for Randomized Optimization Algorithms, *Proc. 2008 Winter Simulation Conference*, Miami, FL, Dec. 7-10, 2008.

55. E. Zhou, M.C. Fu, and S.I. Marcus, A Density Projection Approach to Dimension Reduction for Continuous-State POMDPs, *Proc. 47th IEEE Conference on Decision and Control*, Cancun, Mexico, Dec. 9-11, 2008.

56. E. Zhou, M.C. Fu, and S.I. Marcus, A Numerical Method for Financial Decision Problems Under Stochastic Volatility, *Proc. Winter Simulation Conference*, Austin, TX, Dec. 13-16, 2009.

57. E. Zhou, M.C. Fu, and S.I. Marcus Solving Continuous-State POMDPs via Density Projection, *IEEE Transactions on Automatic Control*, Vol. 55, 1101-1116, May 2010.

58. E. Zhou, M.C. Fu, and S.I. Marcus, A Particle Filtering Framework for Randomized Optimization Algorithms, *IEEE Transactions on Automatic Control*, to appear, 2013.

59. E. Zhou and J. Hu, Gradient Guided Adaptive Stochastic Search for Non-Differentiable Optimization, *IEEE Transactions on Automatic Control*, under revision, 2013.

3.2 Awards

- Michael Fu: Elected Fellow of the Institute of Electrical and Electronics Engineers (IEEE).
- Michael Fu: Elected Fellow of the Institute for Operations Research and the Management Sciences (INFORMS).
- Steven Marcus: Elected Fellow of the Society for Industrial and Applied Mathematics (SIAM).
- Yongqiang Wang (University of Maryland Ph.D. student): received the INFORMS Computing Society Student Paper Award 2010 for the paper "A New Stochastic Derivative Estimator for Discontinuous Payoff Functions with Application to Financial Derivatives."
- Yongqiang Wang (University of Maryland Ph.D. student): won the Best Student Paper Award for "Best OR/MS-focused Paper" at the 2010 Winter Simulation Conference, for the paper "Model-based Evolutionary Optimization."

3.3 Ph.D. Students Graduated

- Andrew Hall, Ph.D., 2009, Univ. of Maryland, supervised by M. Fu, "Simulating and Optimizing: Military Manpower Modeling and Mountain Range Options" (currently: US Military Academy, West Point)
- Matthew Reindorp, Ph.D., 2009, Univ. of Maryland, supervised by M. Fu, "Industrial Flexibility in Theory and Practice" (currently: Technical University of Eindhoven)
- Abraham Thomas, Ph.D., 2009, Univ. of Maryland, supervised by S. Marcus, "Learning Algorithms for Markov Decision Processes."
- Enlu Zhou, Ph.D., 2009, Univ. of Maryland, supervised by S. Marcus and M. Fu, "Particle Filtering for Stochastic Control and Global Optimization" (currently: Georgia Tech)

- Ping Hu, Ph.D., 2011, Stony Brook, supervised by J. Hu
- Yongqiang Wang, Ph.D, 2011, Univ. of Maryland, supervised by M. Fu and S. Marcus
- Kun Lin, Ph.D, 2013, Univ. of Maryland, supervised by S. Marcus