

MULTI-OBJECTIVE SIMULATION OPTIMIZATION ON FINITE SETS: OPTIMAL ALLOCATION VIA SCALARIZATION

Guy Feldman

Department of Statistics
Purdue University
West Lafayette, IN 47907, USA

Susan R. Hunter

School of Industrial Engineering
Purdue University
West Lafayette, IN 47907, USA

Raghu Pasupathy

Department of Statistics
Purdue University
West Lafayette, IN 47907, USA

ABSTRACT

We consider the multi-objective simulation optimization problem on finite sets, where we seek the Pareto set corresponding to systems evaluated on multiple performance measures, using only Monte Carlo simulation observations from each system. We ask how a given simulation budget should be allocated across the systems, and a Pareto surface retrieved, so that the estimated Pareto set minimally deviates from the true Pareto set according to a rigorously defined metric. To answer this question, we suggest scalarization, where the performance measures associated with each system are projected using a carefully considered set of weights, and the Pareto set is estimated as the union of systems that dominate across the weight set. We show that the optimal simulation budget allocation under such scalarization is the solution to a bi-level optimization problem, for which the outer problem is concave, but some inner problems are non-convex. We comment on the development of tractable approximations for use when the number of systems is large.

1 INTRODUCTION

Decision-makers increasingly are using Monte Carlo simulation to model complex stochastic systems in a variety of domains (Powers, Sanchez, and Lucas 2012) — from traffic networks (Osorio and Bierlaire 2013), to airline scheduling (Lee, Lee, and Tan 2007), to plant breeding experiments (Hunter and McClosky 2015). In these settings, the simulation model can be viewed as a function that produces only estimates $f(x, \xi)$ of some underlying function $E[f(x, \xi)]$ at design points x , where ξ represents the randomness in the model. For example, in a Monte Carlo simulation of a traffic network, if x is a decision vector that characterizes the type of road network to be built, the decision-maker can only observe an estimate $f(x, \xi)$ of the expected sojourn time of cars in the traffic network.

Often, decision-makers want to do more than observe estimates of the underlying function at some pre-determined set of design points; they want to *optimize* — that is, they wish to find a point in the design space \mathcal{D} that produces the minimum or maximum value of the objective function $E[f(x, \xi)]$. For example, in the traffic network example, a decision-maker may wish to find the set of stoplight timings that minimize the network-wide expected sojourn time of cars in the network. A popular problem formulation in such contexts is the *simulation optimization* (SO) problem,

$$\text{Find: } \arg \min_{x \in \mathcal{D}} E[f(x, \xi)],$$

where one seeks the location of the minimum of the function $E[f(x, \xi)]$ over points x in a known design space \mathcal{D} , where the objective and constraint functions are observed as output from a Monte Carlo simulation.

To date, most of the algorithms developed to solve SO problems have considered a single performance measure, posed as the objective. For overviews of this literature, see Fu (1994), Kim and Nelson (2006), Henderson and Nelson (2006), Pasupathy and Ghosh (2013). However, decision-makers often consider more than one performance measure, such as the total expected sojourn time of cars in the traffic network *and* the total expected fuel consumption by vehicles in the traffic network. When one performance measure can be posed as the objective and the rest as stochastic constraints, recent work provides methods to find the best feasible system (see, e.g., Andradóttir and Kim 2010, Park and Kim 2011, Luo and Lim 2013, Nagaraj and Pasupathy 2014, Pasupathy et al. 2014, Hu and Andradóttir 2014). When multiple stochastic objectives are considered simultaneously, very little work has been done to efficiently identify the *Pareto set*, the image of which is illustrated in Figure 1(a), which is the set containing all non-dominated points, that is, points for which no other point in the design space \mathcal{D} is preferable on all objectives. (The question of identifying the Pareto set is a problem of recognized importance in the deterministic multi-objective context, with well-developed literature, e.g., see Miettinen 1999, Marler and Arora 2004).

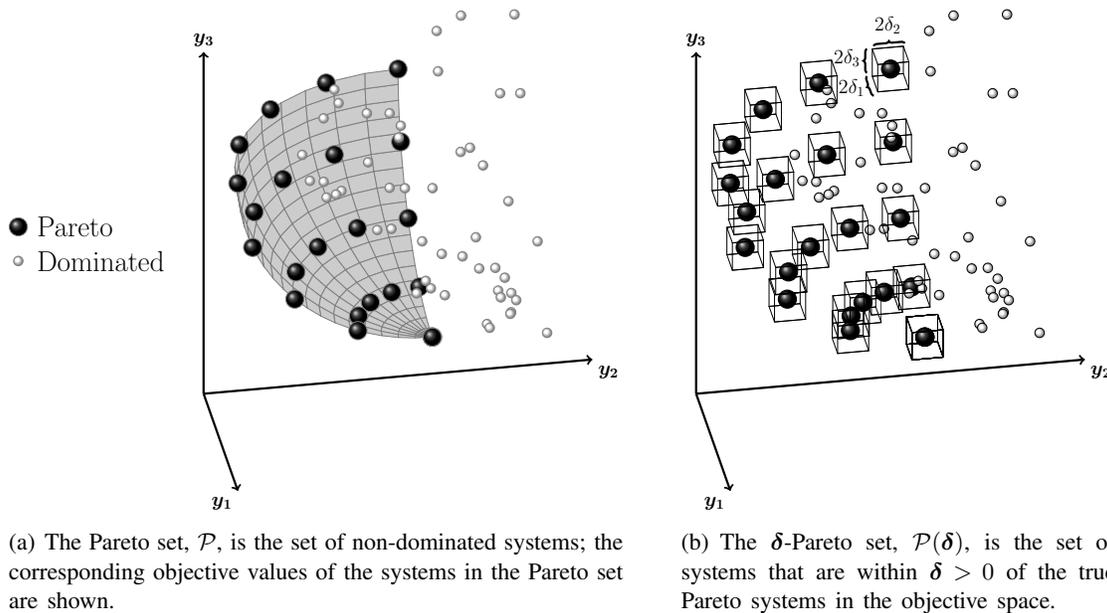


Figure 1: Consider the multi-objective SO problem on finite sets with minimization for all objectives. Also consider the multi-objective SO problem with an indifference zone $\delta > 0$, which allows us to specify the smallest difference worth detecting between systems. In three dimensions, the graph in (a) shows the objective values corresponding to the Pareto set \mathcal{P} , and the graph in (b) shows the objective values corresponding to the δ -Pareto set $\mathcal{P}(\delta)$. The Pareto set \mathcal{P} and δ -Pareto set $\mathcal{P}(\delta)$ are discussed in §1.1.

Formally, we define the multi-objective SO problem as

$$\text{Find: } \arg \min_{x \in \mathcal{D}} (E[f_1(x, \xi)], E[f_2(x, \xi)], \dots, E[f_d(x, \xi)]),$$

where the solution to this problem is the Pareto set. Table 1 categorizes the existing work on the multi-objective SO problem based on the nature of the feasibility set \mathcal{D} , and whether the method is specific to two objectives or can handle greater than two objectives. We note here that all papers cited in Table 1 provide some form of minimum guarantee on the returned solution or on the efficiency of the method; see, e.g., Lee et al. (2009), Andradóttir (2015) and the references therein for work on the multi-objective SO problem

that does not provide such guarantees. The work in this paper lies in the top-right corner of Table 1; that is, we consider unordered, finite sets with a small number of objectives, and derive the simulation sampling allocation that maximizes the rate of decay of the probability of misclassifying systems in \mathcal{D} as Pareto or non-Pareto. (See Hunter and Feldman 2015 for methods analogous to those in this paper but that are specific to the bi-objective case.)

Table 1: The table contains references of example papers containing algorithms to solve multi-objective SO problems that provide a guarantee on the returned solutions. The papers are characterized by the nature of the feasibility set \mathcal{D} and the number of objectives considered.

Feasibility Set \mathcal{D} , solution Global or Local	Bi-Objective Only ($d = 2$)	Multi-Objective ($d \geq 2$)
Unordered, Finite (Global)	Hunter and McClosky (2015), Hunter and Feldman (2015)	Butler et al. (2001), MOCBA (Lee et al. 2010, Teng et al. 2010), this work
Integer-ordered, Local		Li et al. (2015)
Continuous, Local	Ryu et al. (2009), Kim and Ryu (2011b), Kim and Ryu (2011a)	
Integer-ordered, Global		Huang and Zabinsky (2014)
Continuous, Global		Huang and Zabinsky (2014)

The methods we develop can be viewed as a direct competitor to Multi-objective Optimal Computing Budget Allocation (MOCBA) (Lee et al. 2010, Teng et al. 2010), a heuristic sampling framework for solving multi-objective SO problems on finite sets that was developed along the lines of the popular OCBA method (Chen et al. 2000, Chen and Lee 2010). While our methods differ from MOCBA in several ways, one of the most notable differences is that our optimal allocation framework explicitly accounts for correlation among the objectives, while MOCBA does not. However, we acknowledge that the methods we outline in this paper are not nearly as fast or easy to implement as MOCBA; therefore we emphasize that we view this work as the first step in deriving Sampling Criteria for Optimization using Rate Estimators (SCORE) allocations for the multi-objective SO case (Pasupathy et al. 2014). SCORE allocations are easily-computed sampling allocations that are optimal in a certain rigorous sense, as the sampling budget and the number of systems tend to infinity, under mild conditions. While we consider the multi-objective SCORE allocations to be future research, this paper provides the crucial first steps by characterizing the asymptotically optimal sampling allocation as the sampling budget tends to infinity.

1.1 Problem Formulation

Consider an optimization problem in the presence of multiple performance measures, in which each performance measure is posed as an objective. When the known feasible set \mathcal{D} of the problem is finite, we refer to the points in the set \mathcal{D} as “systems,” consistent with the ranking and selection (R&S) literature (Kim and Nelson 2006). Let $1, \dots, r$ denote the elements of the set \mathcal{D} . Let the unknown performance of system ℓ on objective k be $y_{\ell k}$; without loss of generality, let $y_{\ell k} \in \mathbb{R}_+$ for all systems $\ell = 1, \dots, r$ and objectives $k = 1, \dots, d$, where \mathbb{R}_+ is the set of positive real numbers. Thus each system has unknown objective values $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_r$, where we let $\mathbf{y}_\ell = (y_{\ell 1}, \dots, y_{\ell d})$. Consider the vector optimization problem

$$\text{Problem P: Find } \arg \min_{\ell \in \{1, \dots, r\}} (y_{\ell 1}, \dots, y_{\ell d}),$$

where the \mathbf{y}_ℓ 's are expectations, and an estimate of each \mathbf{y}_ℓ is observed through Monte Carlo simulation.

Due to conflicting objectives, often there is no single best system. Thus the solution to Problem P is the set of non-dominated systems, that is, the set of systems for which no system is preferable on *all* objectives. System i *dominates* system j , written as $i \preceq j$, if $y_{ik} \leq y_{jk}$ for all objectives $k \in \{1, \dots, d\}$ and $y_{ik^*} < y_{jk^*}$ for at least one objective k^* . System i *strictly dominates* system j , written as $i \prec j$, if

$y_{ik} < y_{jk}$ for all objectives $k \in \{1, \dots, d\}$. The set of non-dominated (Pareto) systems

$$\mathcal{P} := \{\text{systems } i \in \{1, \dots, r\} : \nexists \text{ system } j \in \{1, \dots, r\} \text{ such that } j \preceq i\}$$

is the solution to Problem P.

Our broad objective is to estimate the set \mathcal{P} efficiently, by expending a given simulation budget across the various systems in a manner that would in some rigorous and reasonable sense be optimal. Suppose $\alpha = (\alpha_1, \dots, \alpha_r)$ is a vector denoting the proportion of the total sampling budget n given to each system, so that $\sum_{\ell=1}^r \alpha_\ell = 1$ and $\alpha_\ell \geq 0$ for all $\ell \in \{1, \dots, r\}$. We say that a *misclassification* event occurs if, after the sampling budget n has been expended and the Pareto set estimator constructed, a system that is truly Pareto is falsely estimated as non-Pareto, or a system that is truly non-Pareto is falsely estimated as Pareto. Then we ask, what vector of proportions α^* maximizes the rate of decay of the probability that a misclassification event occurs?

Furthermore, since the optimal allocation vector α^* resulting from the previous formulation may allocate a large portion of the sampling budget to distinguish between two systems that are very close on one or more objectives, we also define the notion of a δ -Pareto set through the use of an *indifference zone* — the minimum difference worth detecting between systems on each objective. Specifically, for $\delta \in \mathbb{R}_+^d$, we define the δ -Pareto set $\mathcal{P}(\delta) := \cup_{i \in \mathcal{P}} \mathcal{N}_i(\delta)$, where the system neighborhood $\mathcal{N}_i(\delta) := \{\ell : |y_{\ell k} - y_{ik}| \leq \delta_k \text{ for all } k \in \{1, \dots, d\}\}$ effectively places a d -dimensional box around each Pareto system i (see Figure 1(b)). In the indifference zone setting, our goal of solving Problem P remains the same, but we are indifferent to the misclassification of systems that are within the indifference zone. We thus update our optimal allocation question to: what vector of proportions $\alpha^*(\delta)$ maximizes the rate of decay of the probability that a system not in $\mathcal{P}(\delta)$ is falsely estimated as Pareto, or that a Pareto system $i \in \mathcal{P}$ is falsely excluded by some system outside its neighborhood, $\mathcal{N}_i(\delta)$?

We note here that one could consider objective functions of α formulated using quantities other than the probability of a misclassification event, such as the expected number of misclassifications, or some loss function that is linear in the estimated objective values. Hunter and McClosky (2015) prove that maximizing the rate of decay of the probability of a misclassification event is equivalent to maximizing the rate of decay of the expected number of misclassifications in two objectives. A similar equivalence result holds for loss functions in the stochastically constrained SO context; see Pujowidianto, Lee, and Chen (2013). We expect that these results also hold for multi-objective SO with $d > 2$ objectives.

1.2 Assumptions

We require the following assumptions for our results to hold. First, we make a standard assumption in optimal allocation literature — that no two Pareto systems have identical objective values.

Assumption 1 No two Pareto systems have identical objective values on any objective; that is, we assume that there exists $\epsilon > 0$ such that $\min \{|y_{ik} - y_{\ell k}| : i, \ell \in \mathcal{P} \text{ and } k \in \{1, \dots, d\}\} > \epsilon$.

Further, without loss of generality, we assume that all true objective values are positive, that is, $y_{\ell k} > 0$ for all $\ell = 1, \dots, r, k = 1, \dots, d$.

To estimate the unknown objective vectors \mathbf{y}_ℓ , we assume we obtain replicates of the random variables $\mathbf{Y}_\ell = (Y_{\ell 1}, \dots, Y_{\ell d})$ from each system, where each system is simulated independently of the others.

Assumption 2 The systems are simulated independently of each other, that is, the random vectors \mathbf{Y}_ℓ are mutually independent for all $\ell \in \{1, \dots, r\}$.

We note here that, unlike MOCBA, we *do* explicitly account for dependence between the objectives in our methods; that is, we account for possible dependence between the random variables $Y_{\ell 1}, \dots, Y_{\ell d}$.

We also require the following two assumptions, which are standard in literature using large deviation theory and are included here only for completeness. Similar versions of these assumptions are required in Glynn and Juneja (2004), Glynn and Juneja (2006), Hunter and Pasupathy (2013), Pasupathy et al. (2014). We refer the reader to Dembo and Zeitouni (1998) for further explanation. We first define notation.

For every system ℓ , with n copies of \mathbf{Y}_ℓ , we construct an estimator of \mathbf{y}_ℓ as $\bar{\mathbf{Y}}_\ell = (n)^{-1} \sum_{i=1}^n \mathbf{Y}_{\ell i}$. With $n\alpha_\ell$ copies of \mathbf{Y}_ℓ , we likewise construct an estimator of \mathbf{y}_ℓ as $\hat{\mathbf{Y}}_\ell = (n\alpha_\ell)^{-1} \sum_{i=1}^{n\alpha_\ell} \mathbf{Y}_{\ell i}$, where we assume $\alpha_\ell > 0$ and ignore that $n\alpha_\ell$ is not necessarily an integer. Let the k th component of these vectors be $\bar{Y}_{\ell k} = n^{-1} \sum_{i=1}^n Y_{\ell ki}$ and $\hat{Y}_{\ell k} = (n\alpha_\ell)^{-1} \sum_{i=1}^{n\alpha_\ell} Y_{\ell ki}$, respectively. Let the cumulant generating function of $\bar{\mathbf{Y}}_\ell$ be $\Lambda_\ell^{(n)}(\boldsymbol{\theta}) := \log \mathbb{E}[\exp(\langle \boldsymbol{\theta}, \bar{\mathbf{Y}}_\ell \rangle)]$, where $\boldsymbol{\theta} \in \mathbb{R}^d$ and $\langle \cdot, \cdot \rangle$ denotes the dot product. Let the effective domain of a function $f(\cdot)$ be denoted by $\mathcal{D}_f = \{\mathbf{x} : f(\mathbf{x}) < \infty\}$ and its interior by \mathcal{D}_f° . Let $\nabla f(\mathbf{x})$ be the gradient of f with respect to \mathbf{x} .

Assumption 3 The following hold for each $\ell \in \{1, \dots, r\}$ and $k \in \{1, \dots, d\}$:

- (1) the limit $\Lambda_\ell(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_\ell^{(n)}(n\boldsymbol{\theta})$ exists as an extended real number for all $\boldsymbol{\theta} \in \mathbb{R}^d$;
- (2) the origin belongs to the interior of $\mathcal{D}_{\Lambda_\ell}$, that is, $\mathbf{0} \in \mathcal{D}_{\Lambda_\ell}^\circ$;
- (3) $\Lambda_\ell(\boldsymbol{\theta})$ is strictly convex and C^∞ on $\mathcal{D}_{\Lambda_\ell}^\circ$;
- (4) $\Lambda_\ell(\boldsymbol{\theta})$ is steep, that is for any sequence $\{\boldsymbol{\theta}(n)\} \in \mathcal{D}_{\Lambda_\ell}$ converging to a boundary point of $\mathcal{D}_{\Lambda_\ell}$, then $\lim_{n \rightarrow \infty} |\nabla \Lambda_\ell(\boldsymbol{\theta}(n))| = \infty$.

Under Assumption 3, the large deviations principle (LDP) holds for the estimator $\bar{\mathbf{Y}}_\ell$ and with strictly convex rate function $I_\ell(\mathbf{x}_\ell) = \sup_{\boldsymbol{\theta} \in \mathbb{R}^d} \{\langle \boldsymbol{\theta}, \mathbf{x}_\ell \rangle - \Lambda_\ell(\boldsymbol{\theta})\}$, where $\mathbf{x}_\ell = (x_{\ell 1}, \dots, x_{\ell d}) \in \mathbb{R}^d$ (Dembo and Zeitouni 1998, p. 44). Let $\mathcal{F}_\ell^\circ := \text{int}\{\nabla \Lambda_\ell(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathcal{D}_{\Lambda_\ell}^\circ\}$. Then under Assumption 3, we also have that $I_\ell(\mathbf{x}_\ell)$ is strictly convex and C^∞ for $\mathbf{x}_\ell \in \mathcal{F}_\ell^\circ$ for all $\ell \in \{1, \dots, r\}$.

The following assumption ensures that each system can be estimated as dominating any other system.

Assumption 4 The closure of the convex hull of all points $\mathbf{y}_\ell \in \mathbb{R}^d$ is a subset of the intersection of the interiors of the effective domains of the rate functions $I_\ell(\mathbf{x}_\ell)$ for all $\ell \in \{1, \dots, r\}$, that is, $\mathcal{F}_d^c \subset \cap_{\ell=1}^r \mathcal{F}_\ell^\circ$.

2 SCALARIZATION IN MULTI-OBJECTIVE OPTIMIZATION

A common approach to solving deterministic multi-objective optimization problems is to use a scalarization function to convert the multi-objective problem into a parameterized series of single-objective problems. In this section, we introduce ideas that we use to determine the optimal allocation in the stochastic context.

We consider the single-objective scalarized optimization problem, which is

$$\text{Problem } SP_f(\mathbf{w}) : \arg \min_{\ell \in \{1, \dots, r\}} f(\mathbf{w}, \mathbf{y}_\ell),$$

where f is some scalarizing function, and \mathbf{w} is a weight in some weight set $\mathcal{W} \subset \mathbb{R}_+^d$. (Note that for the scalarizations considered in this paper, if system i is the solution to Problem $SP_f(\mathbf{w})$, then it is also the solution to $SP_f(c\mathbf{w})$ for $c > 0$. This implies that the weights can be standardized. For the rest of the paper, we assume that the weights are not standardized.)

The motivation behind scalarization methods is that, for appropriate choices of f , if system i is optimal for Problem $SP_f(\mathbf{w}_i)$ for some weight \mathbf{w}_i , then system i is a Pareto system. Thus different systems in the Pareto set can be retrieved by varying the weights \mathbf{w} in some weight set \mathcal{W} and solving each resulting Problem $SP_f(\mathbf{w})$. The converse, that if system i is a Pareto system then there exists a scalarized Problem $SP_f(\mathbf{w}_i)$ for which it is optimal, depends on the choice of scalarization function and the shape of the Pareto front. For more details, see Eichfelder (2008). Linear scalarization, which is a weighted sum, and Chebyshev scalarization, which is a weighted max function, are among the commonly used scalarization functions. In the next two sections, we discuss the details of linear and Chebyshev scalarizations.

2.1 Linear Scalarization

Given a weight \mathbf{w} and a vector \mathbf{y}_ℓ , linear scalarization forms a single objective by taking a weighted sum across the d individual objective values $y_{\ell k}$ for $k = 1, \dots, d$. That is, $f_{\text{Linear}}(\mathbf{w}, \mathbf{y}_\ell) = \sum_{k=1}^d w_k y_{\ell k}$. For a Pareto system i with objective vector \mathbf{y}_i , consider the question of which weight \mathbf{w}_i will retrieve \mathbf{y}_i as the solution to Problem $SP_{f_{\text{Linear}}}(\mathbf{w}_i)$. To find a weight \mathbf{w}_i , if possible, construct a ‘‘supporting’’ hyperplane to

the Pareto front that passes through \mathbf{y}_i . Then, assuming a supporting hyperplane exists, set \mathbf{w}_i equal to the normal vector of the supporting hyperplane that passes through \mathbf{y}_i to retrieve system i as the solution to Problem $SP_{f_{\text{Linear}}}(\mathbf{w}_i)$. (See Bertsekas et al. (2003), Proposition 2.4.1). If no such supporting hyperplane exists, then no weight \mathbf{w}_i will retrieve system i . Therefore it only makes sense to use linear scalarization when all Pareto points have supporting hyperplanes; and in this sense, the Pareto front is “convex.” Weight vectors that retrieve the Pareto front are shown in Figure 2(a). To find a function f and weights $\mathbf{w} \in \mathcal{W}$ that will retrieve the entire Pareto front, we turn to Chebyshev scalarization, which we discuss in the next section. Many of the results we pursue under Chebyshev scalarization hold under linear scalarization and require less computational effort, should one know in advance that the Pareto front is “convex.”

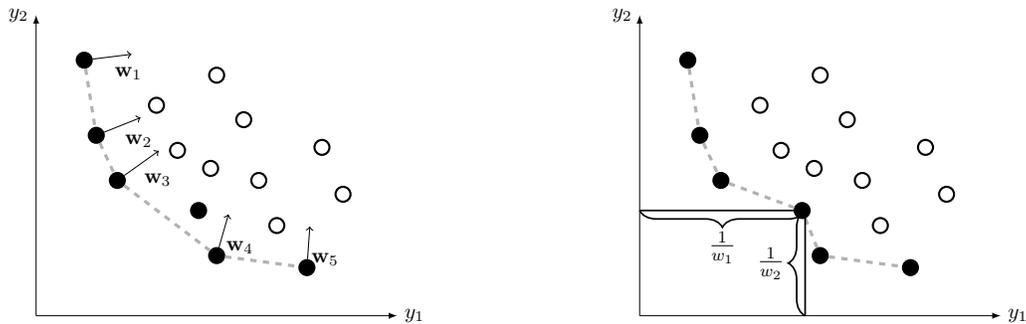
2.2 Chebyshev Scalarization

To retrieve the entire Pareto set, including the Pareto systems that do not lie on the “convex hull” of the Pareto front, we use Chebyshev scalarization. When all objective vectors are positive, which we assume, the Chebyshev scalarization is a weighted max function, that is, $f_{\text{Chebyshev}}(\mathbf{w}, \mathbf{y}_\ell) = \max_{k \in \{1, \dots, d\}} w_k y_{\ell k}$.

Lightner and Director (1981) show that, under the Chebyshev scalarization, the weight vector $\mathbf{w}_i = (y_{i1}^{-1}, \dots, y_{id}^{-1})$ will retrieve system i as the solution to Problem $SP_{f_{\text{Chebyshev}}}(\mathbf{w}_i)$ for all $i \in \mathcal{P}$ (see Figure 2(b)). Thus, if the values \mathbf{y}_ℓ were known for all $\ell \in \{1, \dots, r\}$, we could retrieve the Pareto set by solving Problem $SP_{f_{\text{Chebyshev}}}(\mathbf{w}_i)$ for $\mathbf{w}_i = (y_{i1}^{-1}, \dots, y_{id}^{-1})$ for all $i \in \mathcal{P}$. That is,

$$\mathcal{P} = \bigcup_{i \in \mathcal{P}} \left\{ \arg \min_{\ell \in \{1, \dots, r\}} \max_{k \in \{1, \dots, d\}} w_{ik} y_{\ell k} : \mathbf{w}_i = (y_{i1}^{-1}, \dots, y_{id}^{-1}) \right\}. \tag{1}$$

Since in our context, we do not know the values of \mathbf{y}_ℓ , for all $\ell \in \{1, \dots, r\}$, we cannot use these weights to retrieve the Pareto set. We discuss retrieving the Pareto set with unknown objectives in the next section.



(a) Linear scalarization: there exists a Pareto system that cannot be retrieved by any weight.

(b) Chebyshev scalarization: there exist weights that can retrieve every Pareto system.

Figure 2: The weights $\mathbf{w} \in \mathcal{W}$ that retrieve each Pareto system are shown for linear scalarization in (a) and for Chebyshev scalarization in (b).

2.3 Chebyshev Scalarization with Unknown Objectives

Since we do not know the objective vector values \mathbf{y}_ℓ for all $\ell \in \{1, \dots, r\}$ in our context, and our original goal in Problem P is accurate estimation of \mathbf{y}_ℓ for all $\ell \in \{1, \dots, r\}$, it is intuitive that we might also estimate the weights $\mathbf{w}_i = (y_{i1}^{-1}, \dots, y_{id}^{-1})$ for all $i \in \mathcal{P}$ in equation (1). However, doing so would not only complicate our probabilistic analysis, but may also introduce undesirable correlation between the chosen weights and the systems that, by chance, have already been estimated as Pareto. Therefore we instead use a fixed weight set.

Let $\mathcal{W}(i)$ be the set of weights for which system i is the solution to Problem $SP_{f_{\text{Chebyshev}}}(\mathbf{w})$, that is,

$$\mathcal{W}(i) = \{\mathbf{w} : i = \arg \min_{\ell \in \{1, \dots, r\}} \max_{k \in \{1, \dots, d\}} w_k y_{\ell k} \text{ and } \mathbf{w} \in \mathcal{W}\},$$

and recall that under Assumption 1, there is a separation of magnitude at least $\epsilon > 0$ between the Pareto systems on all objectives. The following Theorem 1 essentially states that, because $\mathcal{W}(i)$ has a nonempty interior if and only if $i \in \mathcal{P}$, then we can find a finite weight set \mathcal{W} that retrieves \mathcal{P} when we solve Problem $SP_{f_{\text{Chebyshev}}}(\mathbf{w})$ for every $\mathbf{w} \in \mathcal{W}$.

Theorem 1 Under Assumption 1, the following hold:

- (i) the set of weights for which system i is optimal for Problem $SP_{f_{\text{Chebyshev}}}(\mathbf{w})$, $\mathcal{W}(i)$, has non-empty interior if and only if $i \in \mathcal{P}$;
- (ii) there exists a fixed and finite weight set \mathcal{W} such that solving Problem $SP_{f_{\text{Chebyshev}}}(\mathbf{w})$ for every $\mathbf{w} \in \mathcal{W}$ retrieves the Pareto set \mathcal{P} .

We do not go into the details of selecting the weight set \mathcal{W} in our context, except to note that one can generate a set of weights using the standard technique of discretizing the surface of the d -dimensional unit sphere \mathbb{S}_+^d , see, e.g. Ghosh and Chakraborty (2014). As a result of Theorem 1, for fine-enough discretization, the weight set \mathcal{W} will retrieve the Pareto set \mathcal{P} .

3 MISCLASSIFICATION EVENT FORMULATION

Recall that our problem context is Problem P, and our broad solution context involves two steps: first, sample from each of the systems to obtain objective vector estimates; and second, construct the estimated Pareto set as the set of systems that are estimated as non-dominated. Thus we consider $\hat{\mathcal{P}}$ as

$$\hat{\mathcal{P}} := \{\text{systems } i \in \{1, \dots, r\} : \nexists \text{ system } j \in \{1, \dots, r\} \text{ such that } j \succeq i\},$$

where system j is estimated as dominating system i , denoted $j \succeq i$, if the estimated objective values are such that $\hat{Y}_{jk} \leq \hat{Y}_{ik}$ for all $k = 1, \dots, d$, and $\hat{Y}_{jk} < \hat{Y}_{ik}$ for at least one objective $k \in \{1, \dots, d\}$. A misclassification (MC) error occurs when the estimated Pareto set is not equal to the true Pareto set, that is, when $\hat{\mathcal{P}} \neq \mathcal{P}$. In this section, we formulate the MC event naively, which is difficult to analyze. We then reformulate the MC event using scalarization, which yields a tractable form for deriving the large deviations rate function for the rate of decay of the probability of an MC event in §4.

3.1 Naïve Formulation of the Misclassification Event

A misclassification (MC) event can occur in one of two ways. First, a truly Pareto system may be falsely excluded from the estimated Pareto set by being estimated as dominated by another system, be it Pareto or non-Pareto. We call this event *misclassification by exclusion*, or MCE, since a truly Pareto system was *excluded* from $\hat{\mathcal{P}}$. The second type of misclassification event occurs when a non-Pareto system is falsely included in the estimated Pareto set by being estimated as non-dominated. We call this event *misclassification by inclusion*, or MCI, since a truly non-Pareto system was *included* in $\hat{\mathcal{P}}$. That is,

$$\text{MCE} := \bigcup_{i \in \mathcal{P}} \bigcup_{\ell \in \{1, \dots, r\}} \bigcap_{k \in \{1, \dots, d\}} \hat{Y}_{\ell k} \leq \hat{Y}_{ik} \quad \text{and} \quad \text{MCI} := \bigcup_{j \notin \mathcal{P}} \bigcap_{\ell \in \{1, \dots, r\}} \bigcup_{k \in \{1, \dots, d\}} \hat{Y}_{jk} \leq \hat{Y}_{\ell k}.$$

Then an intuitive way of writing the MC event is $\text{MC} := \text{MCE} \cup \text{MCI}$, such that the probability of an MC event is $P\{\text{MC}\} = P\{\text{MCE} \cup \text{MCI}\}$.

Now consider formulating the MC event under the IZ formulation. Whereas before we counted all false inclusions and false exclusions as MC events, under the IZ we only count an exclusion if the falsely excluded Pareto system is excluded by a system outside its IZ neighborhood, $\mathcal{N}_i(\delta)$ for $i \in \mathcal{P}$, and we only count an inclusion if the falsely included system lives outside the IZ Pareto front $\mathcal{P}(\delta)$.

Under this formulation, we may also write the MCE and MCI events as a function of the IZ δ . Updating the statements above to incorporate the IZ, we have

$$\text{MCE}(\delta) := \bigcup_{i \in \mathcal{P}} \underbrace{\bigcup_{\ell \notin \mathcal{N}_i(\delta)} \bigcap_{k \in \{1, \dots, d\}} \hat{Y}_{\ell k} \leq \hat{Y}_{ik}}_{i \in \mathcal{P} \text{ estimated as dominated by a system outside its IZ}} \quad \text{and} \quad \text{MCI}(\delta) := \underbrace{\bigcup_{j \notin \mathcal{P}(\delta)} \bigcap_{\ell \in \{1, \dots, r\}} \bigcup_{k \in \{1, \dots, d\}} \hat{Y}_{jk} \leq \hat{Y}_{\ell k}}_{\text{at least one } j \text{ outside all of the IZ's estimated as non-dominated}},$$

such that we can define a new MC event that depends on the IZ δ , $\text{MC}(\delta) := \text{MCE}(\delta) \cup \text{MCI}(\delta)$. Since the MC and $\text{MC}(\delta)$ events are equivalent when $\delta = (0, \dots, 0)$, henceforth, we only consider the more general formulation of $\text{MC}(\delta)$.

Since the rate of decay of the $\text{MC}(\delta)$ event will correspond to the minimum of the corresponding rates of decay for $\text{MCE}(\delta)$ and $\text{MCI}(\delta)$, each of these terms can be analyzed separately. However, while the $\text{MCE}(\delta)$ event is easy to analyze, the $\text{MCI}(\delta)$ event is difficult to analyze due to dependence. To obtain an equivalent event that is easier to analyze, we reformulate the $\text{MCI}(\delta)$ event using scalarization.

3.2 Reformulation of the MC Event

Recall that we have a two-step solution context for Problem P: first, sample from each of the systems to obtain objective vector estimates; and second, construct the estimated Pareto set $\hat{\mathcal{P}}$ as the set of systems that are estimated as non-dominated. Using the Chebyshev scalarization, we consider a different procedure to solve Problem P: first, sample from each of the systems to obtain objective vector estimates, and second, construct the estimated Pareto set $\hat{\mathcal{P}}_{\text{Chebyshev}}$ by solving Problem $SP_{f_{\text{Chebyshev}}}(\mathbf{w})$ for all weight vectors $\mathbf{w} \in \mathcal{W}$, where \mathcal{W} is a fixed, finite weight set that we assume is “dense” enough to return the true Pareto set as the sampling budget n tends to infinity. Then, we return the set $\hat{\mathcal{P}}_{\text{Chebyshev}}$, where

$$\hat{\mathcal{P}}_{\text{Chebyshev}} := \{i \in \{1, \dots, r\} : \exists \mathbf{w} \in \mathcal{W} \text{ such that } i = \arg \min_{\ell \in \{1, \dots, r\}} \max_{k \in \{1, \dots, d\}} w_k \hat{Y}_{\ell k}\}$$

is the set of systems that are estimated as the minima of each of the scalarized Problems $SP_{f_{\text{Chebyshev}}}(\mathbf{w})$.

Now, under this new procedure, we can reformulate $\text{MCI}(\delta)$ as

$$\widetilde{\text{MCI}}(\delta) = \bigcup_{j \notin \mathcal{P}(\delta)} \underbrace{\bigcup_{\mathbf{w} \in \mathcal{W}} \max_{k \in \{1, \dots, d\}} w_k \hat{Y}_{jk} \leq \min_{i \in \mathcal{P}(\delta)} \max_{k \in \{1, \dots, d\}} w_k \hat{Y}_{ik}}_{j \notin \mathcal{P}(\delta) \text{ estimated as dominating all } i \in \mathcal{P}(\delta) \text{ for some Chebyshev scalarization weight } \mathbf{w}}.$$

Unfortunately, formulating the $\text{MCE}(\delta)$ event under the Chebyshev scalarization procedure results in a term that contains dependence and is hence difficult to analyze. For a false exclusion event to occur under Chebyshev scalarization, the Pareto system i must be beaten by some system ℓ that is outside its IZ neighborhood, $\mathcal{N}_i(\delta)$, on *all* Chebyshev weights \mathbf{w} , that is,

$$\widetilde{\text{MCE}}(\delta) = \bigcup_{i \in \mathcal{P}} \underbrace{\bigcup_{\ell \notin \mathcal{N}_i(\delta)} \bigcap_{\mathbf{w} \in \mathcal{W}} \max_{k \in \{1, \dots, d\}} w_k \hat{Y}_{\ell k} \leq \max_{k \in \{1, \dots, d\}} w_k \hat{Y}_{ik}}_{i \in \mathcal{P} \text{ estimated as dominated by } \ell \notin \mathcal{N}_i(\delta) \text{ on all Chebyshev scalarization weights } \mathbf{w}}.$$

Since $\widetilde{\text{MCE}}(\delta)$ under Chebyshev scalarization is difficult to analyze, we analyze the $\text{MCE}(\delta)$ event using its original naïve formulation. Therefore, we reformulate the $\text{MC}(\delta)$ event as $\text{MC}(\delta) = \text{MCE}(\delta) \cup \widetilde{\text{MCI}}(\delta)$.

Formulating the $\text{MC}(\delta)$ event this way presents a conundrum: which procedure should we use for estimating the Pareto set? We propose using the naïve procedure and returning $\hat{\mathcal{P}}$, while using Chebyshev scalarization only for allocating samples according to the asymptotically optimal allocation that results from analyzing rate of decay of the probability of $\widetilde{\text{MCI}}(\delta)$. If the weight set \mathcal{W} is large enough to retrieve the Pareto set as n tends to infinity, then for large enough n , $\hat{\mathcal{P}} = \hat{\mathcal{P}}_{\text{Chebyshev}}$.

4 OPTIMAL ALLOCATION WITH CHEBYSHEV SCALARIZATION

We are now ready to analyze the rate of decay of the probability of an $\text{MC}(\delta)$ event as a function of $\alpha = (\alpha_1, \dots, \alpha_r)$, which is the vector representing proportion of the total sampling budget n that is allocated to each system. We first derive the rate function, then we characterize the optimal allocation as the solution to a concave problem in the allocation vector α . Due to space constraints, we omit all proofs.

First, the rate of decay of $P\{\text{MC}(\delta)\}$ is the minimum of the rates of decay of the probabilities of $\text{MCE}(\delta)$ and $\widetilde{\text{MCI}}(\delta)$. To analyze the rate of decay of $P\{\text{MCE}(\delta)\}$, we decompose it into the minimum of rates involving pairs of systems (i, ℓ) where $i \in \mathcal{P}$ and $\ell \notin \mathcal{N}_i(\delta)$. Define this pairwise rate as

$$R_{\text{MCE}(i,\ell)}(\alpha_i, \alpha_\ell) := \inf_{\mathbf{x}_\ell \leq \mathbf{x}_i} \alpha_i I_i(\mathbf{x}_i) + \alpha_\ell I_\ell(\mathbf{x}_\ell).$$

To analyze the rate of decay of $P\{\widetilde{\text{MCI}}(\delta)\}$, we decompose it into the minimum of rates involving a system and a Chebychev weight (j, \mathbf{w}) , where $j \notin \mathcal{P}(\delta)$ and $\mathbf{w} \in \mathcal{W}$. Define the pairwise rate as

$$R_{\text{MCI}(j,\mathbf{w})}(\alpha_j, \alpha_1, \dots, \alpha_{|\mathcal{P}(\delta)|}) := \max_{k \in \{1, \dots, d\}} w_k x_{jk} \inf_{i \in \mathcal{P}(\delta)} \min_{k \in \{1, \dots, d\}} \max_{k \in \{1, \dots, d\}} w_k x_{ik} \alpha_j I_j(\mathbf{x}_j) + \sum_{i \in \mathcal{P}(\delta)} \alpha_i I_i(\mathbf{x}_i).$$

Theorem 2 states that the overall rate of decay of the probability of $\text{MC}(\delta)$ is the minimum of the rates corresponding to $\text{MCE}(\delta)$ and $\widetilde{\text{MCI}}(\delta)$.

Theorem 2 The rate of decay of $P\{\text{MC}(\delta)\}$ is

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MC}(\delta)\} = \min\left\{ \min_{i \in \mathcal{P}} \min_{\ell \notin \mathcal{N}_i(\delta)} R_{\text{MCE}(i,\ell)}(\alpha_i, \alpha_\ell), \min_{j \notin \mathcal{P}(\delta)} \min_{\mathbf{w} \in \mathcal{W}} R_{\text{MCI}(j,\mathbf{w})}(\alpha_j, \alpha_1, \dots, \alpha_{|\mathcal{P}(\delta)|}) \right\}.$$

Since $R_{\text{MCE}(i,\ell)}(\alpha_i, \alpha_\ell)$ and $R_{\text{MCI}(j,\mathbf{w})}(\alpha_j, \alpha_1, \dots, \alpha_{|\mathcal{P}(\delta)|})$ are concave functions of (α_i, α_ℓ) and $(\alpha_j, \alpha_1, \dots, \alpha_{|\mathcal{P}(\delta)|})$ respectively for all $i \in \mathcal{P}$, $\ell \notin \mathcal{N}_i(\delta)$, $j \notin \mathcal{P}(\delta)$, and $\mathbf{w} \in \mathcal{W}$, the asymptotically optimal sample allocation can be expressed as the solution to the concave maximization problem

$$\begin{aligned} \text{Problem Q:} \quad & \text{maximize } z \text{ s.t.} \\ & R_{\text{MCE}(i,\ell)}(\alpha_i, \alpha_\ell) \geq z \text{ for all } i \in \mathcal{P} \text{ and } \ell \notin \mathcal{N}_i(\delta) \\ & R_{\text{MCI}(j,\mathbf{w})}(\alpha_j, \alpha_1, \dots, \alpha_{|\mathcal{P}(\delta)|}) \geq z \text{ for all } j \notin \mathcal{P}(\delta) \text{ and } \mathbf{w} \in \mathcal{W} \\ & \sum_{\ell=1}^r \alpha_\ell = 1, \alpha_\ell \geq 0 \text{ for all } \ell \in \{1, \dots, r\}. \end{aligned}$$

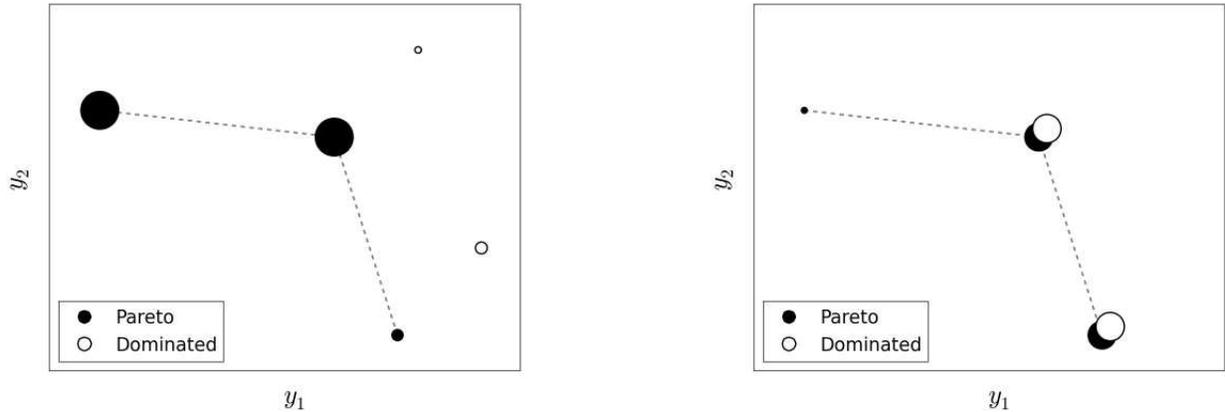
The computational burden in Problem Q arises from computing the values of the constraint functions. For a given value of α , computing the value of $R_{\text{MCE}(i,\ell)}(\alpha_i, \alpha_\ell)$ involves solving a convex optimization problem. However computing the value of $R_{\text{MCI}(j,\mathbf{w})}(\alpha_j, \alpha_1, \dots, \alpha_{|\mathcal{P}(\delta)|})$ requires solving a non-convex optimization problem, where the non-convexity arises in the feasible set. We term these problems Problem $R_{\text{MCE}(i,\ell)}$ and Problem $R_{\text{MCI}(j,\mathbf{w})}$ respectively. We focus on Problem $R_{\text{MCI}(j,\mathbf{w})}$ since it is computationally demanding.

Note that we can write Problem $R_{\text{MCI}(j,\mathbf{w})}$ as

$$\text{minimize}_{\mathbf{x}_j, \mathbf{x}_1, \dots, \mathbf{x}_{|\mathcal{P}(\delta)|}} \alpha_j I_j(\mathbf{x}_j) + \sum_{i \in \mathcal{P}(\delta)} \alpha_i I_i(\mathbf{x}_i) \quad \text{s.t.} \quad \max_{k \in \{1, \dots, d\}} w_k x_{jk} - \max_{k \in \{1, \dots, d\}} w_k x_{ik} \leq 0, \forall i \in \mathcal{P}(\delta).$$

While the objective function in Problem $R_{\text{MCI}(j,\mathbf{w})}$ is strictly convex, the constraint functions are not convex. We are investigating techniques by which we can decompose Problem $R_{\text{MCI}(j,\mathbf{w})}$ into a set of convex problems. Figure 3 depicts an optimal allocation, that is, a solution to Problem Q, based on such a decomposition where the objective vectors are independent Gaussian random variables with unit variance.

Any algorithm that solves Problem Q has to compute the constraints related to the MCE and the scalarized MCI events at each step. Evaluation of the constraints related to the scalarized MCI event require solving $|\mathcal{P}(\delta)|^c \times |\mathcal{W}|$ non-convex optimization Problems $R_{\text{MCI}(j,w)}$. To evaluate the constraints related to the MCE event, we would need to solve $|\mathcal{P}| \times |\mathcal{N}_i(\delta)|^c$ convex optimization Problems $R_{\text{MCE}(i,\ell)}$. Although these problems can be solved in parallel, efficient solutions to these problems are crucial to solving Problem Q efficiently.



(a) The MCE event is the driving misclassification event. The dominated systems are allocated approximately 6% of the sample.

(b) The MCI event is the driving misclassification event. Approximately 50% of the sample is allocated to dominated systems.

Figure 3: The figures show the asymptotically optimal budget allocation obtained by solving Problem Q in the case where all objectives are independent Gaussian random variables with unit variance. In (a), dominated systems are located approximately one standard deviation away from their dominating system, while in (b), Dominated systems are located approximately 0.1 standard deviations away from their dominating system.

5 CONCLUDING REMARKS

Multi-objective optimization in the simulation context is an important problem with application in a wide variety of important real-world contexts. Despite such importance, currently, little is known on how to solve such problems efficiently. We believe, following literature in the deterministic context, that scalarization is a powerful mechanism for increased tractability in identifying optimal budget allocations.

The main challenge in our approach is the computational complexity associated with identifying good simulation budget allocations, that is, those that ensure that the retrieved Pareto set deviates from the true Pareto set minimally. Further computational gains through SCORE-type approximations will likely prove valuable for implementation.

REFERENCES

Andradóttir, S. 2015. “A Review of Random Search Methods”. In *Handbook of Simulation Optimization*, edited by M. C. Fu, Volume 216 of *International Series in Operations Research & Management Science*, 277–292. Springer New York.

Andradóttir, S., and S.-H. Kim. 2010. “Fully Sequential Procedures for Comparing Constrained Systems via Simulation”. *Naval Research Logistics* 57 (5): 403–421.

Bertsekas, D., A. Nedic, and A. Ozdaglar. 2003. “Convex Analysis and Optimization”. *Athena Scientific*.

Butler, J. C., D. J. Morrice, and P. Mullarkey. 2001. “A Multiple Attribute Utility Theory Approach to Ranking and Selection”. *Management Science* 47 (6): 800–816.

- Chen, C.-H., and L. H. Lee. 2010. *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*. World Scientific.
- Chen, C.-H., J. Lin, E. Yücesan, and S. E. Chick. 2000. "Simulation Budget Allocation for Further Enhancing the Efficiency of Ordinal Optimization". *Discrete Event Dynamic Systems* 10 (3): 251–270.
- Dembo, A., and O. Zeitouni. 1998. *Large Deviations Techniques and Applications*. 2nd ed. New York: Springer.
- Eichfelder, G. 2008. *Adaptive Scalarization Methods in Multi-Objective Optimization*. Springer.
- Fu, M. 1994. "Optimization via Simulation: A Review". *Annals of Operations Research* 53:199–247.
- Ghosh, D., and D. Chakraborty. 2014. "Ideal Cone: A New Method to Generate Complete Pareto Set of Multi-criteria Optimization Problems". In *Mathematics and Computing 2013*, 171–190. Springer.
- Glynn, P. W., and S. Juneja. 2004. "A Large Deviations Perspective on Ordinal Optimization". In *Proc. of the 2004 Winter Simulation Conference*, edited by R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 577–585. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.
- Glynn, P. W., and S. Juneja. 2006. "Ordinal Optimization: A Large Deviations Perspective". Technical report, Indian School of Business. Working Paper Series.
- Henderson, S. G., and B. L. Nelson. (Eds.) 2006. *Simulation*. Handbooks in Operations Research and Management Science, Volume 13. Elsevier.
- Hu, L., and S. Andradóttir. 2014. "A Penalty Function Approach for Simulation Optimization with Stochastic Constraints". In *Proc. of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 3730–3736. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.
- Huang, H., and Z. B. Zabinsky. 2014. "Multiple Objective Probabilistic Branch and Bound for Pareto Optimal Approximation". In *Proc. of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 3916–3927. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.
- Hunter, S. R., and G. Feldman. 2015. "Optimal Sampling Laws for Bi-Objective Simulation Optimization on Finite Sets". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.
- Hunter, S. R., and B. McClosky. 2015. "Maximizing Quantitative Traits in the Mating Design Problem via Simulation-Based Pareto Estimation". *Under 3rd review for IIE Transactions*.
- Hunter, S. R., and R. Pasupathy. 2013. "Optimal Sampling Laws for Stochastically Constrained Simulation Optimization on Finite Sets". *INFORMS Journal on Computing* 25 (3): 527–542.
- Kim, S., and J. Ryu. 2011a. "The Sample Average Approximation Method for Multi-Objective Stochastic Optimization". In *Proc. of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 4026–4037. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.
- Kim, S., and J. Ryu. 2011b. "A Trust-Region Algorithm for Bi-Objective Stochastic Optimization". *Procedia Computer Science* 4:1422–1430.
- Kim, S.-H., and B. L. Nelson. 2006. "Selecting the Best System". In *Simulation*, edited by S. G. Henderson and B. L. Nelson, Handbooks in Operations Research and Management Science, Volume 13, 501–534. Elsevier.
- Lee, L. H., E. P. Chew, S. Teng, and D. Goldsman. 2010. "Finding the Non-Dominated Pareto Set for Multi-Objective Simulation Models". *IIE Transactions* 42:656–674.
- Lee, L. H., E. P. Chew, S. Teng, and J. Li. 2009. "Application of Evolutionary Algorithms for Solving Multi-Objective Simulation Optimization Problems". In *Multi-objective memetic algorithms*, edited by C.-K. Goh, Y.-S. Ong, and K.-C. Tan, Studies in Computational Intelligence, 91–110. Springer.
- Lee, L. H., C. U. Lee, and Y. P. Tan. 2007. "A Multi-Objective Genetic Algorithm for Robust Scheduling using Simulation". *European Journal of Operational Research* 177:1948–1968.

- Li, H., L. H. Lee, E. P. Chew, and P. Lendermann. 2015. "MO-COMPASS: A Fast Convergent Search Algorithm for Multi-Objective Discrete Optimization via Simulation". *IIE Transactions*.
- Lightner, M. R., and S. Director. 1981. "Multiple Criterion Optimization for the Design of Electronic Circuits". *Circuits and Systems, IEEE Transactions on* 28 (3): 169–179.
- Luo, Y., and E. Lim. 2013. "Simulation-Based Optimization Over Discrete Sets with Noisy Constraints". *IIE Transactions* 45 (7): 699–715.
- Marler, R. T., and J. S. Arora. 2004. "Survey of Multi-Objective Optimization Methods for Engineering". *Structural and Multidisciplinary Optimization* 26:369–395.
- Miettinen, K. 1999. *Nonlinear Multiobjective Optimization*. Boston: Kluwer Academic Publishers.
- Nagaraj, K., and R. Pasupathy. 2014. "Stochastically Constrained Simulation Optimization on Integer-Ordered Spaces: The cgR-SPLINE Algorithm". *Submitted*. http://www.stat.purdue.edu/~nagaraj5/papers/OPRE_cgR-SPLINE.pdf.
- Osorio, C., and M. Bierlaire. 2013. "A Simulation-Based Optimization Framework for Urban Transportation Problems". *Operations Research* 61 (6): 1333–1345.
- Park, C., and S.-H. Kim. 2011. "Handling Stochastic Constraints in Discrete Optimization via Simulation". In *Proc. of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 4217–4226. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.
- Pasupathy, R., and S. Ghosh. 2013. "Simulation Optimization: A Concise Overview and Implementation Guide". In *TutORials in Operations Research*, edited by H. Topaloglu, Chapter 7, 122–150. INFORMS.
- Pasupathy, R., S. R. Hunter, N. A. Pujowidianto, L. H. Lee, and C. Chen. 2014. "Stochastically Constrained Ranking and Selection via SCORE". *ACM Transactions on Modeling and Computer Simulation* 25 (1): 1–26.
- Powers, M. J., S. M. Sanchez, and T. W. Lucas. 2012. "The Exponential Expansion of Simulation in Research". In *Proc. of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 1552–1563. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.
- Pujowidianto, N. A., L. H. Lee, and C.-H. Chen. 2013. "Minimizing Opportunity Cost in Selecting the Best Feasible Design". In *Proc. of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 898–907. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.
- Ryu, J., S. Kim, and H. Wan. 2009. "Pareto Front Approximation with Adaptive Weighted Sum Method in Multi-Objective Simulation Optimization". In *Proc. of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 623–633: Institute of Electrical and Electronics Engineers, Inc.
- Teng, S., L. H. Lee, and E. P. Chew. 2010. "Integration of Indifference-Zone with Multi-Objective Computing Budget Allocation". *European Journal of Operational Research* 203 (2): 419–429.

AUTHOR BIOGRAPHIES

GUY FELDMAN is a Ph.D. student in the Department of Statistics at Purdue University. His website is <http://www.stat.purdue.edu/~gfeldman/>.

SUSAN R. HUNTER is an assistant professor in the School of Industrial Engineering at Purdue University. Her e-mail address is susanhunter@purdue.edu, and her website is <http://web.ics.purdue.edu/~hunter63/>.

RAGHU PASUPATHY is an associate professor in the Department of Statistics at Purdue University. His email address is pasupath@purdue.edu. His website is <http://web.ics.purdue.edu/~pasupath/>.