Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

CENTRAL LIMIT THEOREMS FOR CONSTRUCTING CONFIDENCE REGIONS IN STRICTLY CONVEX MULTI-OBJECTIVE SIMULATION OPTIMIZATION

Susan R. Hunter

School of Industrial Engineering
Purdue University
West Lafayette, IN 47906, USA

Department of Operations Research
Naval Postgraduate School
Monterey, CA 93943, USA

Raghu Pasupathy

Department of Statistics
Purdue University
West Lafayette, IN 47906, USA

Department of Computer Science & Engineering
Indian Institute of Technology, Madras
Chennai, Tamil Nadu, INDIA

ABSTRACT

We consider the context of multi-objective simulation optimization (MOSO) with strictly convex objectives. We show that under certain types of scalarizations, a $(1-\alpha)$-confidence region on the efficient set can be constructed if the scaled error field (over the scalarization parameter) associated with the estimated efficient set converges weakly to a mean-zero Gaussian process. The main result in this paper proves such a "Central Limit Theorem." A corresponding result on the scaled error field of the image of the efficient set also holds, leading to an analogous confidence region on the Pareto set. The suggested confidence regions are still hypothetical in that they may be infinite-dimensional and therefore not computable, an issue under ongoing investigation.

1 INTRODUCTION

We consider the construction of a Central Limit Theorem (CLT) on the empirical mean of the error when estimating the efficient and Pareto sets in strictly convex multi-objective simulation optimization (MOSO) with continuous decision variables (Pasupathy and Henderson 2006; Hunter et al. 2019). Such a CLT is a first step toward our ultimate goal of creating an appropriate statistic for constructing asymptotically valid, frequentist, $(1-\alpha)$-confidence regions on the solution and optimal value of a MOSO problem.

The general MOSO problem takes the following form:

$$\text{minimize} \quad \{f(x) = (f_1(x),\ldots,f_d(x)) = (\mathrm{E}[F_1(x,Y)],\ldots,\mathrm{E}[F_d(x,Y)])\} \quad \text{s.t.} \quad x\in\mathcal{X}, \qquad (M)$$

where $f\colon \mathcal{D}\to\mathbb{R}^d, \mathcal{D}\subseteq\mathbb{R}^q, d\geq 2$ is a vector-valued function comprised of unknown, real-valued, continuous objective functions $f_k\colon \mathcal{D}\to\mathbb{R}$, $k=1,\ldots,d$, that can only be observed with stochastic error as the output of a Monte Carlo simulation oracle. Each objective $k=1,\ldots,d$ is an expected value

$$f_k(x) = \mathrm{E}[F_k(x,Y)] = \int_{\mathcal{Y}} F_k(x,y)P(dy)$$

where $F_k\colon \mathcal{D}\times\mathcal{Y}\to\mathbb{R}$ and P is the probability measure induced by the random element $Y\colon \Omega\to\mathcal{Y}$. We denote as $F\colon \mathcal{D}\times\mathcal{Y}\to\mathbb{R}^d$ the vector-valued function comprised of real-valued $F_1,\ldots,F_d$. The feasible set $\mathcal{X}\subseteq\mathcal{D}$ is specified by a collection of deterministic functions. (We formalize our assumptions in §3.3.)

The solution to Problem (M) is the efficient set $\mathcal{E}\subseteq\mathcal{X}$, which is the set of all feasible points whose images are non-dominated,

$$\mathcal{E} := \{x^*\in\mathcal{X}\colon \nexists x\in\mathcal{X} \text{ such that } f(x)\leq f(x^*)\}.$$

This set is also called the set of all expected value efficient solutions (Caballero et al. 2001). (We use the notation $f(x) \leq f(x^*)$ to imply that $f_k(x) \leq f_k(x^*)$ for all $k \in \{1,\ldots,d\}$ and $f(x) \neq f(x^*)$; we use $f(x) \leqq f(x^*)$ when equality is allowed and $f(x) < f(x^*)$ when all inequalities are strict.) The Pareto set, also called the Pareto front, is the image of the efficient set, $\mathcal{P} = \{f(x^*) \colon x^* \in \mathcal{E}\}$. When solving Problem ($M$), we characterize the entire efficient and Pareto sets as input to the decision-making process.

1.1 Problem Statement

Suppose we are *given* a significance level $\alpha \in (0,1)$ and a stochastic oracle that, for any point $x \in \mathcal{X}$, sample size n, and random objects $Y_1,\ldots,Y_n$ identically distributed to Y, returns the output random variables $F(x,Y_1)$, $\ldots$, $F(x,Y_n)$ having common mean $f(x)$. Let $\tilde{\mathcal{X}} \subseteq \mathcal{X}$ be some nonempty subset of the feasible set for which we are able to query the simulation oracle and observe its output data. Our problem is to *find*, using only the available output data $\{F(x,Y_1),\ldots,F(x,Y_n) \colon x \in \tilde{\mathcal{X}}\}$,

1. a set $\hat{\mathfrak{E}}_{n,1-\alpha} \subset \mathcal{X}$ such that $\lim_{n\to\infty} \mathbb{P}\{\mathcal{E} \subset \hat{\mathfrak{E}}_{n,1-\alpha}\} = 1-\alpha$;
2. a set $\hat{\mathfrak{P}}_{n,1-\alpha} \subset \{f(x) \colon x \in \mathcal{X}\}$ such that $\lim_{n\to\infty} \mathbb{P}\{\mathcal{P} \subset \hat{\mathfrak{P}}_{n,1-\alpha}\} = 1-\alpha$.

1.2 Motivation

Since the objective functions in Problem (M) are unknown and can only be estimated, providing some form of confidence regions on the efficient and Pareto sets is crucial to quantifying the uncertainty in a real-world estimated solution to (M). Many, if not most, real-world problems naturally involve both uncertainty and multiple competing objectives. Examples include maximizing the mean and minimizing the variance of the return in portfolio allocation (Markowitz 1952; Steinbach 2001), determining power plant designs that maximize the expected efficiency while minimizing the expected carbon dioxide emissions (Subramanyan et al. 2011), and determining irrigation designs that balance expected economical, environmental, and social performance metrics (Crespo et al. 2010); see Hunter et al. (2019) for other applications. While only one solution can be implemented in practice, it is often desirable to provide decision-makers with a characterization of the entire efficient and Pareto sets; such a practice often allows decision-makers to account for factors external to the model in selecting a solution. In such a context, the estimated characterization of the efficient and Pareto sets should also include a characterization of the uncertainty. Further, such theory can later be used as a basis for sequential-sampling algorithms that efficiently estimate the Pareto set in MOSO problems.

1.3 Related Work and Our Contributions

Despite the importance and prominence of the MOSO problem class, the development of theory, methods, and algorithms for MOSO is just beginning. Very little work toward quantifying uncertainty has been done to date. While Zhang et al. (2017) and Rojas-Gonzalez and Van Nieuwenhuyse (2020) quantify uncertainty in MOSO using bootstrapping and stochastic kriging, respectively, neither work's goal is to provide a complete theory for constructing confidence regions as detailed in §1.1.

Most related to our work is Vogel (2017) which, to the best of our knowledge, is the first paper to address confidence regions in the context of multi-objective optimization with objective functions that can only be observed with stochastic error. The confidence regions constructed in Vogel (2017) are valid for each n and do not rely on any assumptions about the distribution of a suitably constructed statistic. Thus, as noted by the author, the resulting confidence regions are conservative.

In this paper, we lay the groundwork for developing a theory for constructing asymptotically valid frequentist confidence regions on the efficient and Pareto sets for Problem (M). The key step in our proposed methodology involves constructing certain scalarized sample average estimators of the efficient and Pareto sets, and then demonstrating that under common regularity conditions, a centered and scaled version of these estimators converges to a mean-zero Gaussian process. The implication of such a result is that a $(1-\alpha)$-confidence region on the efficient and Pareto sets can then be constructed in a manner analogous to what is done in classical settings, provided the second moment associated with the limiting Gaussian is appropriately handled; see §2 for more detail. Finally, we remark that such a sample average framework is among the most popular approaches for solving MOSO problems; see, e.g., Fliege and Xu (2011), Kim and Ryu (2011a), Kim and Ryu (2011b), Bonnel and Collonge (2014), Wang (2017), Cooper et al. (2020).

We organize the remainder of the paper as follows. §2, contains an overview of confidence regions for MOSO and motivates the weak convergence results. In §3, we provide some mathematical preliminaries, characterize the error when solving Problem (M) with and without parameterization, and detail the assumptions required for our main results. We provide the main results in §4 and concluding remarks in §5.

2 CONSTRUCTING CONFIDENCE REGIONS ON THE EFFICIENT AND PARETO SETS

In this section, we lay the foundations for constructing confidence regions on the efficient and Pareto sets associated with MOSO. In the process, we motivate the main results of the paper, which appear in §4.

2.1 Classical Confidence Regions

Recall the three heuristic steps for constructing a confidence region on an unknown parameter using observed data: (a) choose an appropriate random object called the *root* or the *statistic* that is some convenient function of the parameter, (b) identify (or approximate) the root's distribution, and (c) "invert" the expression for the root to obtain the $(1-\alpha)$-confidence region (Politis et al. 1999).

For example, in the familiar setting of estimating the mean θ of a real-valued random variable Y using data $Y_1,\ldots,Y_n$, the frequently chosen root is the *Studentized* statistic $(\sqrt{n}/\hat{\sigma}_n)(\bar{Y}-\theta)$ for $\bar{Y}=n^{-1}\sum_{i=1}^{n}Y_i$ and $\hat{\sigma}_n=\sqrt{n^{-1}\sum_{j=1}^{n}(Y_j-\bar{Y})^2}$. If a CLT on $(\sqrt{n}/\hat{\sigma}_n)(\bar{Y}-\theta)$ holds, then we can approximate the distribution of the root using the standard normal cdf $\Phi(\cdot)$, yielding $P(\theta\in(\bar{Y}\pm z_{\alpha/2}\hat{\sigma}_n/\sqrt{n}))\to 1-\alpha$ as $n\to\infty$, where $z_{\alpha/2}:=\Phi^{-1}(1-\alpha/2)$. This justifies the classical $(1-\alpha)$-confidence interval $(\bar{Y}\pm z_{\alpha/2}\hat{\sigma}_n/\sqrt{n})$.

In a more general context, suppose the desired parameter is $\theta\in\mathbb{R}^d$, and that the chosen root is the L_p norm of the *scaled error*

$$\tau_n\|\hat{\theta}_n-\theta\|_p, \tag{1}$$

where $\hat{\theta}_n$ estimates θ, $\tau_n>0$ is a scaling parameter and $p\geq 1$ is fixed. If $\tau_n\|\hat{\theta}_n-\theta\|_p$ exhibits weak convergence to a real-valued random variable with distribution F_T, then an asymptotically valid $(1-\alpha)$-confidence region on θ is $\{y\in\mathbb{R}^d\colon \tau_n\|\hat{\theta}_n-y\|_p\leq F_T^{-1}(1-\alpha)\}$.

2.2 Confidence Regions in MOSO

Now consider our MOSO context, in which the desired parameters are the efficient set $\mathcal{E}$ and the Pareto set $\mathcal{P}$. First, we construct estimators for $\mathcal{E}$ and $\mathcal{P}$ by replacing the objective f in Problem (M) with its estimated counterpart $\bar{F}_n(x):=\left(n^{-1}\sum_{i=1}^{n}F_1(x,Y_i),\ldots,n^{-1}\sum_{i=1}^{n}F_d(x,Y_i)\right)$, yielding the *sample-path problem*,

$$\text{minimize}\quad \bar{F}_n(x)\quad \text{s.t.}\quad x\in\mathcal{X}. \tag{$\overline{M}(n)$}$$

Implicit in ($\overline{M}(n)$) is the fact that we use the same set of iid random variables $Y_1,\ldots,Y_n$ at each x. Then, natural estimators for $\mathcal{E},\mathcal{P}$ are the solution to ($\overline{M}(n)$) and its image, constructed as

$$\hat{\mathcal{E}}_n:=\{X^*\in\mathcal{X}\colon \nexists x\in\mathcal{X}\text{ such that }\bar{F}_n(x)\leq\bar{F}_n(X^*)\},\quad \hat{\mathcal{P}}_n=\{\bar{F}_n(X^*)\colon X^*\in\hat{\mathcal{E}}_n\}.$$

Now if we attempt to construct a root as in (1), e.g. as $\tau_n\|\hat{\mathcal{E}}_n-\mathcal{E}\|$ for appropriate scaling τ_n and norm $\|\cdot\|$, the subtraction is not well-defined: $\hat{\mathcal{E}}_n$ and $\mathcal{E}$ are sets that have no clear mapping between them. While we could try to subtract them anyway using, e.g., the Minkowski difference (Molchanov 2017), doing so would result in an undesirable increase in the "size" of the set under consideration.

To create a clear mapping between the sets $\mathcal{E},\mathcal{P}$ and their estimators, we *parameterize* (or scalarize) the objective functions, which effectively transforms Problem (M) into a collection of single-objective problems indexed by the parameter s in the parameter set $\mathcal{S}$. (The linear weighted sum is arguably the most common scalarization approach: each objective $k \in \{1,\dots,d\}$ is multiplied by a weight $s_k \in \mathbb{R}^+$, yielding $s_k f_k(x)$, and the weighted objectives are summed to form a new single-objective problem $\sum_k s_k f_k(x)$. Often, $s = (s_1,\dots,s_d) \in \mathbb{R}^d$ and $\mathcal{S} \subset \mathbb{R}^d$; if we require that the weights sum to one, $\mathcal{S} \subset \mathbb{R}^{d-1}$. We discuss parameterization/scalarization and the linear weighted sum in more detail in §3.2.2.) Let $(M(s))$, $s \in \mathcal{S}$ denote the true scalarized problems, and suppose both f and the scalarization are such that the solution to each scalarized problem x_s^* is unique and the sets $\mathcal{E},\mathcal{P}$ can be written as the sets

$$\mathcal{E} = \{x_s^*\colon s \in \mathcal{S}\}, \quad \mathcal{P} = \{f(x_s^*)\colon s \in \mathcal{S}\}. \tag{2}$$

(Whether or not we can write (2) depends on both the nature of the function f and the nature of the scalarization; see, e.g., Miettinen (1999). For instance, sufficient conditions for these requirements include that f is strictly convex and the scalarization is the linear weighted sum. We formalize our assumptions in §3.3.) Likewise, let $(\overline{M}(s,n))$, $s \in \mathcal{S}$ denote the scalarized sample-path problems, where both $\bar{F}$ and the scalarization are such that the solution to each scalarized problem $X_s^*(n)$ is unique almost surely and the estimators for $\mathcal{E},\mathcal{P}$ can be written as

$$\hat{\mathcal{E}}_n = \{X_s^*(n)\colon s \in \mathcal{S}\}, \quad \hat{\mathcal{P}}_n = \{f(X_s^*(n))\colon s \in \mathcal{S}\}. \tag{3}$$

Now there is a clear correspondence between $\mathcal{E},\mathcal{P}$ and their respective estimators $\hat{\mathcal{E}}_n,\hat{\mathcal{P}}_n$: For each $s \in \mathcal{S}$, there is a point $x_s^* \in \mathcal{E}$ and a corresponding estimator $X_s^*(n) \in \hat{\mathcal{E}}_n$. Thus, we construct the scaled *error fields*

$$\{\sqrt{n}\,(X_s^*(n) - x_s^*)\,, s \in \mathcal{S}\} \text{ and } \{\sqrt{n}\,(f(X_s^*(n)) - f(x_s^*))\,, s \in \mathcal{S}\}. \tag{4}$$

These scaled error fields represent the error in the estimated efficient and Pareto sets, respectively, and form the basis for constructing a root in our MOSO context. (As a word on terminology, by a *field*, we mean a collection of random variables indexed by some variable. In this sense, $\{\sqrt{n}\,(X_s^*(n) - x_s^*)\,, s \in \mathcal{S}\}$ is a field of random variables indexed by $s \in \mathcal{S}$; $\{\sqrt{n}\,(X_s^*(n) - x_s^*)\,, s \in \mathcal{S}\}$ is also a *stochastic process*.)

Suppose we are able to demonstrate the weak convergence (as $n \to \infty$) of the scaled error field $\{\sqrt{n}\,(X_s^*(n) - x_s^*)\,, s \in \mathcal{S}\}$ to a characterizable stable random object such as a zero-mean Gaussian process $\{W(s), s \in \mathcal{S}\}$ (see §3); that is, suppose $C(\mathcal{S})$ is the space of $\mathbb{R}^q$-valued continuous functions on $\mathcal{S}$ equipped with an appropriate norm $\|\cdot\|$ and we can show that

$$\{\sqrt{n}\,(X_s^*(n) - x_s^*)\,, s \in \mathcal{S}\} \xrightarrow{\text{d}} \{W(s), s \in \mathcal{S}\} \text{ in } C(\mathcal{S}), \tag{5}$$

then a $(1-\alpha)$-confidence region on $\mathcal{E}$ is

$$\hat{\mathfrak{E}}_{n,1-\alpha} := \left\{\tilde{\mathcal{X}} = \{\tilde{x}_s\colon s \in \mathcal{S}\} \subset \mathbb{R}^q\colon \sqrt{n}\big\|\hat{\mathcal{E}}_n - \tilde{\mathcal{X}}\big\| \le w_{1-\alpha}\right\}, \tag{6}$$

where $w_{1-\alpha}$ is the radius of the ball to which the process $\{W(s), s \in \mathcal{S}\}$ assigns probability $(1-\alpha)$; that is,

$$w_{1-\alpha} := \inf\{r > 0\colon\ \mathbb{P}(\|W\| \le r)\} \ge 1-\alpha.$$

In writing (6) we assume, again for ease of exposition, that the positive definite kernel associated with $\{W(s), s \in \mathcal{S}\}$ is known when computing $w_{1-\alpha}$. The reader should recognize the weak convergence result in (5) and the confidence region in (6) as the infinite-dimensional analogues of the CLT and the confidence interval, respectively, in the classical case presented in §2.1. In the rest of the paper, we undertake the rigorous demonstration of (5).

3 PRELIMINARIES

In this section, we discuss preliminaries for our main results. First, in §3.1, we discuss mathematical preliminaries. In §3.2, we characterize the error in solving Problem (M) using the sample-path problem $(\overline{M}(n))$ with and without parameterization. Finally, in §3.3, we discuss our assumptions and their implications.

3.1 Mathematical Preliminaries

Definition 1 (ε-net, Billingsley 1999, p. 239) Let (M,ρ) be a metric space. An *ε-net* for a set $\mathcal{A} \subseteq M$ is a set of points $\{x_j\}$ with the property that for each $x \in \mathcal{A}$, there is an x_j such that $\rho(x,x_j) < \varepsilon$.

Definition 2 (Total boundedness, Billingsley 1999, p. 239) Let (M,ρ) be a metric space. A set $\mathcal{A} \subseteq M$ is *totally bounded* if for every $\varepsilon > 0$, it has a finite ε-net.

Definition 3 (Weak convergence, Billingsley 1999, p. 26) For random variables $X_n, n \geq 1$ and X in a σ-field $\mathcal{M}$, we say that X_n converges weakly to X, denoted $X_n \xrightarrow{\text{d}} X$, if $P(X_n \in A) \to P(X \in A)$ for X-continuity sets A, that is, sets $A \in \mathcal{M}$ such that $P(X \in \partial A) = 0$, where ∂A is the boundary of A.

Note that several other equivalent notions of weak convergence exist.

Let $(C_d(\mathcal{D}), \|\cdot\|)$ be a Banach space of all $\mathbb{R}^d$-valued continuous functions on $\mathcal{D} \subset \mathbb{R}^q$, and let $C_d(\mathcal{D})^*$ denote its *dual space*, that is, the space of bounded linear functionals on $C_d(\mathcal{D})$. Recall that a linear functional $J \in C_d(\mathcal{D})^*$ is bounded if and only if it is continuous with respect to the operator norm $\|J\|_* := \sup\{|Jx| \colon x \in C_d(\mathcal{D}),\, \|x\| = 1\}$ (Kreyszig 1978).

Definition 4 (Gaussian element) A $C_d(\mathcal{D})$-valued random element W is a *Gaussian element* if for any $J \in C_d(\mathcal{D})^*$, the random variable $JW \in \mathbb{R}$ is normally distributed.

Notice that any random element W in $C_d(\mathcal{D})$ can be seen both as a random function $W \colon \mathcal{D} \to \mathbb{R}^d$ and as a stochastic process $\{W(x) \colon x \in \mathcal{D}\}$. Therefore, a *Gaussian process* in $C_d(\mathcal{D})$ is a Gaussian random element in $C_d(\mathcal{D})$.

Further, W is said to be *centered* or *mean-zero* if $\mathbb{E}[JW] = 0$ for all $J \in C_d(\mathcal{D})^*$. For a fixed $x = (x_1,\dots,x_q) \in \mathcal{D}$ and $\alpha = (\alpha_1,\dots,\alpha_d) \in \mathbb{R}^d$, the functional $J_{\alpha,x} \colon h = (h_1,\dots,h_d) \mapsto \sum_{i=1}^d \alpha_i h_i(x) \in C_d(\mathcal{D})^*$ and $J_{\alpha,x}W = \sum_{i=1}^d \alpha_i W_i(x)$ is univariate Gaussian. However, α is arbitrary and thus $W(x)$ is d-variate Gaussian. Similarly, for fixed $x = (x_1,\dots,x_n) \in \mathbb{R}^d \times \mathbb{R}^n$, and any $d \times n$ matrix (of reals) α, since the functional $J_{\alpha,x} : h = (h_1,\dots,h_d) \mapsto \sum_{i=1}^d \sum_{j=1}^n \alpha_i h_i(x_j) \in C_d(\mathcal{D})^*$, we see that $J_{\alpha,x}W = \sum_{i=1}^d \sum_{j=1}^n \alpha_i W_i(x_j)$ is normally distributed and thus $(W(x_1),\dots,W(x_n))$ is vector-valued multivariate normal.

3.2 Error Characterizations

In this section, we characterize the error in solving Problem (M) using the sample-path problem ($\overline{M}(n)$). We also discuss parameterization and characterize the error under parameterization.

3.2.1 The Function Estimator's Error

When solving the sample-path problem ($\overline{M}(n)$), let $G \colon \mathcal{D} \times \mathcal{Y} \to \mathbb{R}^d$ be a vector-valued function that specifies the error. Thus, for each $(x,Y) \in \mathcal{D} \times \mathcal{Y}$, we have $G(x,Y) := F(x,Y) - f(x)$. For notational convenience, for each $x \in \mathcal{D}$, further define $G_x \colon \mathcal{Y} \to \mathbb{R}^d$ so that for each $Y \in \mathcal{Y}$,

$$G_x(Y) := G(x,Y) = F(x,Y) - f(x).$$

Then, we can think of G_x as a random vector that specifies the error for a given $x \in \mathcal{D}$. Let

$$\mathcal{G} := \{G_x, x \in \mathcal{X}\}$$

be a random field where x is the index labeling the collection of random vectors.

For $Y_1,\dots,Y_n$ and for each $x \in \mathcal{D}$, define the empirical mean of the function estimator's error as $\bar{G}(x) := n^{-1}\sum_{i=1}^n G(x,Y_i)$, so that

$$\bar{G}_x := \bar{G}(x) = n^{-1}\textstyle\sum_{i=1}^n G(x,Y_i) = n^{-1}\sum_{i=1}^n G_x(Y_i).$$

Thus, the oracle allows us to observe the estimator $\bar{F}_n(x) = f(x) + \bar{G}(x)$. (For notational simplicity, we omit the implied subscript of n on $\bar{G}(x)$.) Since we have the same $Y_1,\dots,Y_n$ at each x in ($\overline{M}(n)$), when we perform any operation across the index set, such as writing $\bar{G}_{x_1} - \bar{G}_{x_2}$ for $x_1, x_2 \in \mathcal{D}$, we have the same $Y_1,\dots,Y_n$ at each point in the index set. For example, $\bar{G}_{x_1} - \bar{G}_{x_2} = n^{-1}\sum_{i=1}^n \big(G_{x_1}(Y_i) - G_{x_2}(Y_i)\big)$.

3.2.2 The Parameterized Function Estimator's Error

As discussed in §1.3, we require a mapping between the efficient set $\mathcal{E}$ and its estimator $\hat{\mathcal{E}}_n$. We achieve this mapping through parameterization, which we describe first. Then, we define the function estimator's error under parameterization.

We parameterize the objective functions for the true multi-objective Problem (M) and the sample-path multi-objective problem ($\overline{M}(n)$) as follows. For some parameter set $\mathcal{S}$, we assume the existence of a parameterization ϕ taking the form of an operator $\phi\colon \mathcal{S}\times C_d(\mathcal{D})\to C_1(\mathcal{D})$. Such a parameterization may arise from special structure in Problem (M) or from scalarization (Miettinen 1999; Marler and Arora 2004; Audet et al. 2008; Eichfelder 2008), which is used to convert the multi-objective problem in (M) into a single-objective stochastic optimization problem (Romanko et al. 2012).

In our context, given $s\in\mathcal{S}$, the parameterized true objective function f is the function $\phi(s,f)\in C_1(\mathcal{D})$, and the parameterized version of $F(\cdot,Y)$ is the function $\phi(s,F(\cdot,Y))$, where $\phi(s,F(\cdot,Y))\in C_1(\mathcal{D})$ for a.e. $Y\in\mathcal{Y}$. Further, given a parameter value $s\in\mathcal{S}$, the scalarized version of Problem (M) is

$$\text{minimize}\quad \phi(s,f)(x)\quad \text{s.t.}\quad x\in\mathcal{X}, \tag{$M(s)$}$$

and, given a parameter value $s\in\mathcal{S}$ and sample size n, the scalarized version of ($\overline{M}(n)$) is

$$\text{minimize}\quad \phi(s,\bar{F}_n)(x)\quad \text{s.t.}\quad x\in\mathcal{X}. \tag{$\overline{M}(s,n)$}$$

(This notation is not ideal, but it is well-considered; see Kreyszig (1978) for examples of similar notation.)

Example 1 (Linear weighted sum) Given a parameter value s in the simplex $\mathcal{S}=\{s\in\mathbb{R}^d\colon \sum_{i=1}^d s_i=1, 0\le s\}$ and a vector-valued function $f\in C_d(\mathcal{D})$, the linear weighted sum parameterization implies that for $x\in\mathcal{D}$, we have: $\phi(s,f)(x)=\langle s,f(x)\rangle$, $\phi(s,F(\cdot,Y))(x)=\langle s,F(x,Y)\rangle$, and $\phi(s,\bar{F}_n)(x)=\langle s,\bar{F}_n(x)\rangle$. Given these parameterizations, scalarized versions of Problems (M) and ($\overline{M}(n)$) are formulated as above.

Now to characterize the parameterized function estimator's error, let $V\colon \mathcal{S}\times\mathcal{D}\times\mathcal{Y}\to\mathbb{R}$ be a real-valued random variable that specifies the error; for each $(s,x,Y)\in\mathcal{S}\times\mathcal{D}\times\mathcal{Y}$, let $V(s,x,Y):=\phi(s,F(\cdot,Y))(x)-\phi(s,f)(x)$. For notational convenience, for each $(s,x)\in\mathcal{S}\times\mathcal{D}$, further define $V_{s,x}\colon \mathcal{Y}\to\mathbb{R}$ so that for each $Y\in\mathcal{Y}$,

$$V_{s,x}(Y):=V(s,x,Y)=\phi(s,F(\cdot,Y))(x)-\phi(s,f)(x).$$

Then, we can think of $V_{s,x}$ as a random variable that specifies the error for a given $(s,x)\in\mathcal{S}\times\mathcal{D}$. For a given parameterization ϕ, let

$$\mathcal{V}:=\{V_{s,x},(s,x)\in\mathcal{S}\times\mathcal{X}\}$$

be a random field where (s,x) is the index labeling the collection of random variables.

For $Y_1,\ldots,Y_n$ and for each $(s,x)\in\mathcal{S}\times\mathcal{D}$, likewise define the empirical mean of the parameterized estimator's error as $\overline{V}(s,x):=n^{-1}\sum_{i=1}^n V(s,x,Y_i)$, so that

$$\overline{V}_{s,x}:=\overline{V}(s,x)=n^{-1}\textstyle\sum_{i=1}^n V(s,x,Y_i)=n^{-1}\sum_{i=1}^n V_{s,x}(Y_i).$$

Example 2 (Linear weighted sum, continued) Given $(s,x)\in\mathcal{S}\times\mathcal{X}$, the linear weighted sum parameterization implies that the error random variable is $V_{s,x}(Y)=\langle s,F(x,Y)\rangle-\langle s,f(x)\rangle=\langle s,F(x,Y)-f(x)\rangle=\langle s,G(x,Y)\rangle$, and the empirical mean of the parameterized estimator's error is $\overline{V}_{s,x}=\langle s,\bar{F}_n(x)\rangle-\langle s,f(x)\rangle=\langle s,\bar{F}_n(x)-f(x)\rangle=\langle s,\bar{G}(x)\rangle=\phi(s,\bar{G})(x)$.

Recall that we have *the same* $Y_1,\ldots,Y_n$ at each point x in ($\overline{M}(n)$). Likewise, when we perform any operation across the index set, such as writing $\overline{V}_{s_1,x_1}-\overline{V}_{s_2,x_2}$ for $(s_1,x_1),(s_2,x_2)\in\mathcal{S}\times\mathcal{D}$, we have the same $Y_1,\ldots,Y_n$ at each point in the index set. For example, $\overline{V}_{s_1,x_1}-\overline{V}_{s_2,x_2}=n^{-1}\sum_{i=1}^n\big(V_{s_1,x_1}(Y_i)-V_{s_2,x_2}(Y_i)\big)$.

3.3 Assumptions and Their Implications

In this section, we formalize our standing assumptions about the functions that comprise the true problem (M), the sample-path problem ($\overline{M}(n)$), and the parameterization resulting in problems ($M(s)$) and ($\overline{M}(s,n)$). In brief, we require compactness, Lipschitz continuity, and uniqueness for the solutions to the parameterized problems and the estimated parameterized problems. We also discuss some implications of these assumptions.

3.3.1 Compactness and Lipschitz Continuity

To begin, Assumption 1 details sufficient conditions for a CLT to hold on the random field that specifies the error $\mathcal{V} = \{V_{s,x}, (s,x) \in \mathcal{S} \times \mathcal{X}\}$. In brief, we require the feasible set and the parameter set to be compact, and we also require Lipschitz continuity.

Assumption 1 We assume the random elements $Y_1, \ldots, Y_n$ are iid copies of Y having probability measure P and the following:

1. The feasible set $\mathcal{X} \subset \mathbb{R}^q$ and the real-valued parameter set $\mathcal{S}$ are compact. (Typically, for d objectives, the parameter set $\mathcal{S}$ is such that $\mathcal{S} \subset \mathbb{R}^d$ or $\mathcal{S} \subset \mathbb{R}^{d-1}$.)
2. The vector-valued function $F(\cdot, Y)$ is $L(Y)$-Lipschitz continuous for a.e. $Y \in \mathcal{Y}$ under the Euclidean norm; that is, for all $x_1, x_2 \in \mathcal{D}$ and a.e. $Y \in \mathcal{Y}$, $\|F(x_1,Y) - F(x_2,Y)\| \leq L(Y)\|x_1 - x_2\|$ where $0 < \mathrm{E}[L(Y)^2] < \infty$ and $\ell := \mathrm{E}[L(Y)] > 0$.
3. The parameterization ϕ *preserves Lipschitz continuity on* $\mathcal{S} \times \mathcal{D}$ (to be defined in Definition 5) where the resulting Lipschitz constant for $\phi(\cdot, F(\cdot, Y))$, called $\Lambda(Y)$, is such that $0 < \mathrm{E}[\Lambda(Y)^2] < \infty$ and $\lambda := \mathrm{E}[\Lambda(Y)] > 0$.

Under Assumption 1, Problem (M) is well-posed in the sense that it implies the efficient and Pareto sets exist and are bounded. In particular, Assumption 1 Part 2 implies the following Lemma 1 which provides several implications of Lipschitz continuity. The proof for Lemma 1 appears in §A.1.

Lemma 1 Suppose Assumption 1 Part 2 holds, so that $F(\cdot, Y)$ is $L(Y)$-Lipschitz continuous for a.e. $Y \in \mathcal{Y}$. Let $\bar{L} := n^{-1} \sum_{i=1}^{n} L(Y_i)$. Then

1. f is ℓ-Lipschitz continuous,
2. $G(\cdot, Y)$ is $(L(Y) + \ell)$-Lipschitz continuous for a.e. $Y \in \mathcal{Y}$, and
3. $\bar{G}(\cdot)$ is $(\bar{L} + \ell)$-Lipschitz continuous a.s.

Further, in Assumption 1 Part 3, we require the parameterization ϕ to preserve Lipschitz continuity, a term we define in Definition 5.

Definition 5 We say that the parameterization ϕ *preserves Lipschitz continuity on* $\mathcal{S} \times \mathcal{D}$ if for each ℓ-Lipschitz continuous function $h \in C_d(\mathcal{D})$, there exists $\lambda(h) \in (0, \infty)$ such that the function $\phi(\cdot, h)\colon \mathcal{S} \times \mathcal{D} \to \mathbb{R}$ is $\lambda(h)$-Lipschitz continuous; that is, for all $(s_1, x_1), (s_2, x_2) \in \mathcal{S} \times \mathcal{D}$,

$$|\phi(s_1,h)(x_1) - \phi(s_2,h)(x_2)| \leq \lambda(h)\|(s_1,x_1) - (s_2,x_2)\|.$$

Lemma 2 states that the linear weighted sum parameterization preserves Lipschitz continuity under Assumption 1 Part 1; a proof of this lemma appears in §A.2.

Lemma 2 If Assumption 1 Part 1 holds, so that $\mathcal{S}$ and $\mathcal{D}$ are compact, then the linear weighted sum parameterization preserves Lipschitz continuity on $\mathcal{S} \times \mathcal{D}$.

Now, the following Lemma 3 regarding the Lipschitz continuity of the error also holds, in a similar fashion to Lemma 1. The proof for Lemma 3 appears in §A.3.

Lemma 3 Suppose Assumption 1 Parts 2 and 3 hold, so that $F(\cdot, Y) \in C_d(\mathcal{D})$ is $L(Y)$-Lipschitz continuous for a.e. $Y \in \mathcal{Y}$ and ϕ preserves Lipschitz continuity on $\mathcal{S} \times \mathcal{D}$. Let $\bar{\Lambda} := n^{-1} \sum_{i=1}^{n} \Lambda(Y_i)$. Then

1. $\phi(\cdot, f)$ is λ-Lipschitz continuous,
2. $V(\cdot, \cdot, Y)$ is $(\Lambda(Y) + \lambda)$-Lipschitz continuous for a.e. $Y \in \mathcal{Y}$, and
3. $\overline{V}(\cdot, \cdot)$ is $(\bar{\Lambda} + \lambda)$-Lipschitz continuous a.s.

3.3.2 Uniqueness of the Solution Under Parameterization

Next, Assumption 2 details conditions that, together with Assumption 1, are sufficient for a CLT to hold on the infinite-dimensional root $\{\sqrt{n}\,(X_s^*(n)-x_s^*)\,, s\in\mathcal{S}\}$.

Assumption 2 We assume the following:

1. Every objective function $f_1,\ldots,f_d$ is strictly convex.
2. Every sample-path objective function $\bar{F}_{1,n},\ldots,\bar{F}_{d,n}$ is strictly convex almost surely.
3. The parameterization ϕ is the linear weighted sum.

Under Assumptions 1 and 2, for each $s\in\mathcal{S}$, $\phi(s,f)(\cdot)$ and $\phi(s,\bar{F}_n)(\cdot)$ are strictly convex problems. Thus, the solution to the parameterized true problem $x_s^* := \arginf\{\phi(s,f)(x)\colon x\in\mathcal{X}\}$ exists and is unique, and the solution to the parameterized sample-path problem $X_s^*(n) := \arginf\{\phi(s,\bar{F}_n)(x)\colon x\in\mathcal{X}\}$ exists and is unique almost surely; see, e.g., Miettinen (1999), Fliege and Xu (2011).

4 MAIN RESULTS

We are now ready to present the main results of the paper. First, as background information, §4.1 provides a CLT on the empirical mean of the function estimator's error; the proofs in §4.2 follow similar steps to those presented here. Thus, this section builds intuition and preliminary results for the remainder of the paper in a slightly simpler setting than the one we ultimately consider. Then, the main CLT on the empirical mean of the parameterized function estimator's error appears in §4.2. Finally, we present the main weak convergence result in §4.3.

4.1 CLT on the Empirical Mean of the Function Estimator's Error

For $G_{x_1}, G_{x_2}\in\mathcal{G}$, define the $L_2(P)$-distance on $\mathcal{G}$ as

$$\rho(G_{x_1},G_{x_2}) := \sqrt{\mathrm{E}[\|G_{x_1}-G_{x_2}\|^2]} := \left(\int_{\mathcal{Y}}\|G_{x_1}(y)-G_{x_2}(y)\|^2 P(dy)\right)^{1/2},$$

so that $(\mathcal{G},\rho)$ is a metric space. Also, we define the terms ε-*net* and *totally bounded* in Definitions 1 and 2.

The following theorem is an analogue of Theorem 6.1 on p. 89 of van de Geer (2000); see also van der Vaart and Wellner (1996), van der Vaart (1998).

Theorem 1 Let $Y_1,\ldots,Y_n$ be iid, and let $\mathcal{B}(\delta) := \{(x_1,x_2)\colon x_1,x_2\in\mathcal{X},\ \rho(G_{x_1},G_{x_2})\le\delta\}$. Suppose that

1. $\mathcal{G}=\{G_x, x\in\mathcal{X}\}$ is totally bounded and
2. for all $\eta>0$, there exists a $\delta>0$ such that

$$\limsup_{n\to\infty} P\left(\sup_{(x_1,x_2)\in\mathcal{B}(\delta)}\left\{\sqrt{n}\left\|\bar{G}_{x_1}-\bar{G}_{x_2}\right\|>\eta\right\}\right)<\eta.$$

Then $\mathcal{G}$ is P-Donsker; that is, as $n\to\infty$, the random field $\{\sqrt{n}\,\bar{G}_x, x\in\mathcal{X}\}$ converges in distribution to a zero-mean Gaussian process $\{W_x\in\mathbb{R}^d, x\in\mathcal{X}\}$.

We now demonstrate conditions under which the postulates of Theorem 1 are satisfied.

Proposition 1 The postulates of Theorem 1 hold under Assumption 1; that is, if $Y_1,\ldots,Y_n$ are iid, the feasible set $\mathcal{X}$ is compact, and $F(\cdot,Y)$ is $L(Y)$-Lipschitz continuous for a.e. $Y\in\mathcal{Y}$ where $0<\mathrm{E}[L(Y)^2]<\infty$.

Proof. We demonstrate the second postulate of Theorem 1, then the first.

Let $\eta>0$. Then for each $x_1,x_2\in\mathcal{D}$, since $G(\cdot,Y)$ is $(L(Y)+\ell)$-Lipschitz continuous for a.e. $Y\in\mathcal{Y}$ by Lemma 1, we have

$$\rho(G_{x_1},G_{x_2}) = \sqrt{\mathrm{E}[\|G_{x_1}-G_{x_2}\|^2]} \le \sqrt{\mathrm{E}[(L(Y)+\ell)^2]}\,\|x_1-x_2\|.$$

Let $\delta > \|x_1 - x_2\| \sqrt{\mathrm{E}[(L(Y)+\ell)^2]}$. For a.e. $Y \in \mathcal{Y}$, Lemma 1 also implies that for every $x_1, x_2 \in \mathcal{X}$,

$$\sqrt{n}\left\|\bar{G}_{x_1} - \bar{G}_{x_2}\right\| \leq \sqrt{n}(\bar{L}+\ell)\|x_1 - x_2\| < \sqrt{n}\left(\frac{\bar{L}+\ell}{\sqrt{\mathrm{E}[(L(Y)+\ell)^2]}}\right)\delta.$$

Then

$$P\left(\sup_{(x_1,x_2)\in\mathcal{B}(\delta)} \left\{\sqrt{n}\left\|\bar{G}_{x_1} - \bar{G}_{x_2}\right\| > \eta\right\}\right) \leq P\left(\sqrt{n}\left(\frac{\bar{L}+\ell}{\sqrt{\mathrm{E}[(L(Y)+\ell)^2]}}\right) > \frac{\eta}{\delta}\right).$$

Under the postulates of the proposition, $\bar{L}$ obeys a CLT. Thus, for small enough δ, the second postulate of Theorem 1 holds.

Next, we show that $\mathcal{G}$ is totally bounded by constructing a finite ε-net. Let $\varepsilon > 0$ and $G_x \in \mathcal{G}$. Then if, for some $j \in \{1,2,\ldots\}$, we have an x_j such that $\|x - x_j\| < \varepsilon / \sqrt{\mathrm{E}[(L(Y)+\ell)^2]}$, it follows that

$$\rho(G_x, G_{x_j}) = \sqrt{\mathrm{E}[\|G_x - G_{x_j}\|^2]} \leq \sqrt{\mathrm{E}[(L(Y)+\ell)^2]}\|x - x_j\| < \varepsilon.$$

Then we construct a finite ε-net $\{G_{x_j}, j = 1,\ldots,r\}$ by selecting $\{x_j\}$ so that

$$\mathrm{dist}(\mathcal{X}, \{x_j\}) = \sup_{x\in\mathcal{X}} \inf_j \|x - x_j\| < t := \varepsilon / \sqrt{\mathrm{E}[(L(Y)+\ell)^2]};$$

then, $\cup_j \mathrm{int}(B(x_j,t)) \supseteq \mathcal{X}$ is an open cover of $\mathcal{X}$. Since $\mathcal{X}$ is compact, every open cover has a finite subcover. Thus, $\exists r < \infty$ large enough that $\cup_{j\leq r} \mathrm{int}(B(x_j,t)) \supseteq \mathcal{X}$, implying $\{x_j, j = 1,\ldots,r\}$ is a finite ε-net. $\square$

4.2 CLT on the Empirical Mean of the Parameterized Function Estimator's Error

In this section, we derive a CLT on the empirical mean of the parameterized function estimator's error. This section mirrors §4.1 in intuition, but with an extra index due to the parameterization.

For $V_{s_1,x_1}, V_{s_2,x_2} \in \mathcal{V}$, recall that the $L_2(P)$-distance on $\mathcal{V}$ is $\rho(V_{s_1,x_1}, V_{s_2,x_2}) = \sqrt{\mathrm{E}[(V_{s_1,x_1} - V_{s_2,x_2})^2]}$, so that $(\mathcal{V}, \rho)$ is a metric space. As with Theorem 1, in the context of parameterization, we also have the following Theorem 2, which is an analogue of Theorem 6.1 on p. 89 of van de Geer (2000).

Theorem 2 Let $Y_1,\ldots,Y_n$ be iid and let

$$\tilde{\mathcal{B}}(\delta) := \{((s_1,x_1),(s_2,x_2)) \colon (s_1,x_1),(s_2,x_2) \in \mathcal{S}\times\mathcal{X}, \rho(V_{s_1,x_1}, V_{s_2,x_2}) < \delta\}.$$

Suppose that

1. $\mathcal{V} = \{V_{s,x}, (s,x) \in \mathcal{S}\times\mathcal{X}\}$ is totally bounded and
2. for all $\eta > 0$, there exists a $\delta > 0$ such that

$$\limsup_{n\to\infty} P\left(\sup_{((s_1,x_1),(s_2,x_2))\in\tilde{\mathcal{B}}(\delta)} \left\{\sqrt{n}|\overline{V}_{s_1,x_1} - \overline{V}_{s_2,x_2}| > \eta\right\}\right) < \eta.$$

Then $\mathcal{V}$ is P-Donsker; that is, the random field $\{\sqrt{n}\,\overline{V}_{s,x}, (s,x) \in \mathcal{S}\times\mathcal{X}\}$ converges in distribution to a zero-mean Gaussian process $\{W_{s,x} \in \mathbb{R}, (s,x) \in \mathcal{S}\times\mathcal{X}\}$.

The following Proposition 2 states conditions under which the postulates of Theorem 2 are satisfied.

Proposition 2 The postulates of Theorem 2 hold under Assumption 1; that is, if $Y_1,\ldots,Y_n$ are iid, the feasible set $\mathcal{X}$ and the parameter set $\mathcal{S}$ are compact, $F(\cdot,Y)$ is $L(Y)$-Lipschitz continuous for a.e. $Y \in \mathcal{Y}$ where $0 < \mathrm{E}[L(Y)^2] < \infty$, and ϕ preserves Lipschitz continuity on $\mathcal{S}\times\mathcal{D}$ where the resulting Lipschitz constant for $\phi(\cdot, F(\cdot,Y))$, called $\Lambda(Y)$, is such that $0 < \mathrm{E}[\Lambda(Y)^2] < \infty$ and $\lambda := \mathrm{E}[\Lambda(Y)] > 0$.

4.3 CLTs on the Solutions to the Estimated Scalarized Problems and Their Images

Equipped with the general result that the random field $\mathcal{V} = \{V_{s,x}, (s,x) \in \mathcal{S} \times \mathcal{X}\}$ is P-Donsker from Theorem 2, next, we employ the continuous mapping theorem (see, e.g., Billingsley 1999, p. 21) to obtain a CLT-type result on the solutions to the scalarized problems ($(M(s))$), indexed by the parameters $s \in \mathcal{S}$. We present this result as the following Theorem 3, without proof.

Theorem 3 Let Assumptions 1 and 2 hold. Then

$$\{\sqrt{n}(X_s^*(n) - x_s^*), s \in \mathcal{S}\} \xrightarrow{\text{d}} \{W_s, s \in \mathcal{S}\},$$

where $\{W_s, s \in \mathcal{S}\}$ is a mean-zero Gaussian process; recall $W_s \in \mathbb{R}^q$ for each $s \in \mathcal{S}$. Furthermore, if f is Fréchet differentiable at each $x \in \mathcal{D}$, that is, there exists a $d \times q$ matrix $\nabla f(x)$ such that for $u \in \mathbb{R}^q$, $\|f(x+u) - (f(x) + \nabla f(x)u)\| = o(\|u\|)$, then

$$\{\sqrt{n}(f(X_s^*(n)) - f(x_s^*)), s \in \mathcal{S}\} \xrightarrow{\text{d}} \{\nabla f(x_s^*)W_s, s \in \mathcal{S}\}.$$

A proof of the first part of Theorem 3 follows from the application of the continuous mapping theorem to Theorem 2. The proof of the second part follows from the first part and from Taylor's theorem; heurisically, due to Fréchet differentiability, we see $f(X_s^*(n)) - f(x_s^*) \approx \nabla f(x_s^*)(X_s^*(n) - x_s^*)$; then apply the first part.

5 CONCLUDING REMARKS

We provide weak convergence results that lay the foundation for confidence regions on the efficient and Pareto sets in MOSO. Future research includes translating these objects into something computable, e.g., through a sequence of finite-dimensional approximations and ensuring tightness.

ACKNOWLEDGMENTS

The authors thank David Eckman for comments that improved the paper and Susan Sanchez, Dashi Singham, Roberto Szechtman, and the Department of Operations Research at the Naval Postgraduate School in Monterey, CA for their hospitality during AY21-22. The first author thanks the National Science Foundation for support under grant CMMI-1554144, and the second author gratefully acknowledges the Office of Naval Research for support under grants N000141712295 and 13000991.

A APPENDICES

We present the proofs of Lemmas 1–3 below.

A.1 Proof of Lemma 1

Proof. Let $x_1, x_2 \in \mathcal{D}$, and suppose the postulates hold. *Proof of Part 1.* Using Jensen's inequality,

$$\begin{aligned}\|f(x_1) - f(x_2)\| &= \|\mathrm{E}[F(x_1,Y) - F(x_2,Y)]\| \le \mathrm{E}[\|F(x_1,Y) - F(x_2,Y)\|] \\ &\le \mathrm{E}[L(Y)\|x_1 - x_2\|] = \mathrm{E}[L(Y)]\|x_1 - x_2\|.\end{aligned} \tag{7}$$

Proof of Part 2. Now consider $G(x,Y) = G_x(Y)$. Then for a.e. $Y \in \mathcal{Y}$, we have

$$\begin{aligned}\|G_{x_1}(Y) - G_{x_2}(Y)\| &= \|F(x_1,Y) - F(x_2,Y) + (f(x_2) - f(x_1))\| \le \|F(x_1,Y) - F(x_2,Y)\| + \|f(x_2) - f(x_1)\| \\ &\le (L(Y) + \mathrm{E}[L(Y)])\|x_1 - x_2\|.\end{aligned}$$

Proof of Part 3. Finally, consider $\bar{G}(x) = \bar{G}_x$. Then almost surely,

$$\begin{aligned}\|\bar{G}_{x_1} - \bar{G}_{x_2}\| &= \left\|n^{-1}\textstyle\sum_{i=1}^n G_{x_1}(Y_i) - G_{x_2}(Y_i)\right\| = \left\|f(x_2) - f(x_1) + n^{-1}\textstyle\sum_{i=1}^n F(x_1,Y_i) - F(x_2,Y_i)\right\| \\ &\le \|f(x_2) - f(x_1)\| + n^{-1}\textstyle\sum_{i=1}^n \|F(x_1,Y_i) - F(x_2,Y_i)\| \\ &\le (\mathrm{E}[L(Y)] + n^{-1}\textstyle\sum_{i=1}^n L(Y_i))\|x_1 - x_2\| = (\bar{L} + \mathrm{E}[L(Y)])\|x_1 - x_2\|.\end{aligned}$$

□

A.2 Proof of Lemma 2

Proof. Without loss of generality, let $\mathcal{S} = \{s \in \mathbb{R}^d : \sum_{i=1}^d s_i = 1, 0 \leq s\}$. Let $\mathcal{D}$ be compact, let $h \in C_d(\mathcal{D})$ be an ℓ-Lipschitz continuous function, and set $u := \sup_{x \in \mathcal{D}} \|h(x)\| < \infty$. Set $\lambda = \sqrt{2}\max\{\ell\sqrt{d}, u\}$, which depends on h. Then for (s_1, x_1), $(s_2, x_2) \in \mathcal{S} \times \mathcal{D}$,

$$\begin{aligned}
&|\phi(s_1,h)(x_1) - \phi(s_2,h)(x_2)| = |\langle s_1, h(x_1)\rangle - \langle s_2, h(x_2)\rangle| \\
&\quad \leq |\langle s_1, h(x_1)\rangle - \langle s_1, h(x_2)\rangle| + |\langle s_1, h(x_2)\rangle - \langle s_2, h(x_2)\rangle| = |\langle s_1, h(x_1) - h(x_2)\rangle| + |\langle s_1 - s_2, h(x_2)\rangle| \\
&\quad \leq \|s_1\|\|h(x_1) - h(x_2)\| + \|s_1 - s_2\|\|h(x_2)\| \text{ by Cauchy-Schwartz} \\
&\quad \leq \ell\sqrt{d}\|x_1 - x_2\| + u\|s_1 - s_2\| \leq \max\{\ell\sqrt{d}, u\}(\|x_1 - x_2\| + \|s_1 - s_2\|) \\
&\quad \leq \sqrt{2}\max\{\ell\sqrt{d}, u\}\sqrt{\|x_1 - x_2\|^2 + \|s_1 - s_2\|^2} = \lambda\|(s_1, x_1) - (s_2, x_2)\|. \qquad \square
\end{aligned}$$

A.3 Proof of Lemma 3

Proof. Suppose the postulates hold. *Proof of Part 1.* Since $F(\cdot, Y)$ is $L(Y)$-Lipschitz continuous for a.e. $Y \in \mathcal{Y}$, then f is ℓ-Lipschitz continuous by Lemma 1. Since ϕ preserves Lipschitz continuity on $\mathcal{S} \times \mathcal{D}$, by Definition 5, then there exists $\lambda \in (0, \infty)$ such that $\phi(\cdot, f)$ is λ-Lipschitz continuous; that is, for all (s_1, x_1), $(s_2, x_2) \in \mathcal{S} \times \mathcal{D}$, $|\phi(s_1, f)(x_1) - \phi(s_2, f)(x_2)| \leq \lambda\|(s_1, x_1) - (s_2, x_2)\|$, where λ depends on ℓ. *Proof of Part 2.* Let $(s_1, x_1), (s_2, x_2) \in \mathcal{S} \times \mathcal{D}$ and consider $V_{s,x}(Y) = V(s, x, Y)$. Then for a.e. $Y \in \mathcal{Y}$,

$$\begin{aligned}
|V_{s_1,x_1}(Y) - V_{s_2,x_2}(Y)| &= |\phi(s_1, F(\cdot, Y))(x_1) - \phi(s_1, f)(x_1) - (\phi(s_2, F(\cdot, Y))(x_2) - \phi(s_2, f)(x_2))| \\
&\leq |\phi(s_1, F(\cdot, Y))(x_1) - \phi(s_2, F(\cdot, Y))(x_2)| + |\phi(s_2, f)(x_2) - \phi(s_1, f)(x_1)| \\
&\leq (\Lambda(Y) + \lambda)\|(s_1, x_1) - (s_2, x_2)\|.
\end{aligned}$$

Proof of Part 3. Let $(s_1, x_1), (s_2, x_2) \in \mathcal{W} \times \mathcal{D}$ and consider $\overline{V}_{s,x} = \overline{V}(s, x)$. Then almost surely,

$$\begin{aligned}
|\overline{V}_{s_1,x_1} - \overline{V}_{s_2,x_2}| &= |n^{-1}\textstyle\sum_{i=1}^n (V_{s_1,x_1}(Y_i) - V_{s_2,x_2}(Y_i))| \\
&\leq |\phi(s_2, f)(x_2) - \phi(s_1, f)(x_1)| + n^{-1}\textstyle\sum_{i=1}^n |\phi(s_1, F(\cdot, Y_i))(x_1) - \phi(s_2, F(\cdot, Y_i))(x_2)| \\
&\leq (\bar{\Lambda} + \lambda)\|(s_1, x_1) - (s_2, x_2)\|. \qquad \square
\end{aligned}$$

References

Audet, C., G. Savard, and W. Zghal. 2008. "Multiobjective optimization through a series of single-objective formulations". *SIAM J. Optim.* 19(1):188–210.

Billingsley, P. 1999. *Convergence of Probability Measures*. Second ed. Wiley Series in Probability and Statistics. New York: Wiley-Interscience.

Bonnel, H., and C. Collonge. 2014. "Stochastic optimization over a Pareto set associated with a stochastic multi-objective optimization problem". *Journal of Optimization Theory and Applications* 162:405–427.

Caballero, R., E. Cerdá, M. M. M. noz, L. Rey, and I. M. Stancu-Minasian. 2001, July. "Efficient solution concepts and their relations in stochastic multiobjective programming". *Journal of Optimization Theory and Applications* 110(1):53–74.

Cooper, K., S. R. Hunter, and K. Nagaraj. 2020. "Bi-objective simulation optimization on integer lattices using the epsilon-constraint method in a retrospective approximation framework". *INFORMS Journal on Computing* 32(4):1080–1100.

Crespo, O., J. E. Bergez, and F. Garcia. 2010. "Multiobjective optimization subject to uncertainty: Application to irrigation strategy management". *Computers and Electronics in Agriculture* 74:145–154.

Eichfelder, G. 2008. *Adaptive Scalarization Methods in Multiobjective Optimization*. Berlin Heidelberg: Springer.

Fliege, J., and H. Xu. 2011. "Stochastic multiobjective optimization: sample average approximation and applications". *Journal of Optimization Theory and Applications* 151:135–162.

Hunter, S. R., E. A. Applegate, V. Arora, B. Chong, K. Cooper, O. Rincón-Guevara, and C. Vivas-Valencia. 2019, January. "An introduction to multi-objective simulation optimization". *ACM Transactions on Modeling and Computer Simulation* 29(1):7:1–7:36.

Kim, S., and J. Ryu. 2011a. "The sample average approximation method for multi-objective stochastic optimization". In *Proceedings 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: IEEE.

Kim, S., and J. Ryu. 2011b, May. "A trust-region algorithm for bi-objective stochastic optimization". *Procedia Computer Science* 4:1422–1430.

Kreyszig, E. 1978. *Introductory Functional Analysis with Applications*. Wiley Classics Library. London: John Wiley & Sons.

Markowitz, H. 1952. "Portfolio selection". *The Journal of Finance* 7(1):77–91.

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.

Molchanov, I. 2017. *Theory of random sets*. second ed, Volume 87 of *Probability Theory and Stochastic Modelling*. London: Springer-Verlag.

Pasupathy, R., and S. G. Henderson. 2006. "A testbed of simulation-optimization problems". In *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 255–263. Piscataway, NJ: IEEE.

Politis, D. N., J. P. Romano, and M. Wolf. 1999. *Subsampling*. Springer Series in Statistics. New York: Springer.

Rojas-Gonzalez, S., and I. Van Nieuwenhuyse. 2020. "A survey on kriging-based infill algorithms for multiobjective simulation optimization". *Computers and Operations Research* 116.

Romanko, O., A. Ghaffari-Hadigheh, and T. Terlaky. 2012. "Multiobjective optimization via parametric optimization: Models, algorithms, and applications". In *Modeling and Optimization: Theory and Applications*, edited by T. Terlaky and F. Curtis, Volume 21 of *Springer Proceedings in Mathematics & Statistics*, 77–119. New York, NY: Springer.

Steinbach, M. C. 2001. "Markowitz revisited: Mean-variance models in financial portfolio analysis". *SIAM Review* 43(1):31–85.

Subramanyan, K., U. Diwekar, and S. E. Zitney. 2011. "Stochastic modeling and multi-objective optimization for the APECS system". *Computers and Chemical Engineering* 35:2667–2679.

van de Geer, S. 2000. *Empirical Processes in M-Estimation*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK: Cambridge University Press.

van der Vaart, A. W. 1998. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.

van der Vaart, A. W., and J. A. Wellner. 1996. *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer Series in Statistics. New York, NY: Springer.

Vogel, S. 2017. "Random approximations in multiobjective optimization". *Mathematical Programming* 164:29–53.

Wang, H. 2017, December. "Multi-objective retrospective optimization using stochastic zigzag search". *European Journal of Operational Research* 263(3):946–960.

Zhang, J., Y. Ma, T. Yang, and L. Liu. 2017. "Estimation of the Pareto front in stochastic simulation through stochastic Kriging". *Simulation Modelling Practice and Theory* 79:69–86.

AUTHOR BIOGRAPHIES

SUSAN R. HUNTER is an associate professor in the School of Industrial Engineering at Purdue University. Her research interests include Monte Carlo methods and stochastic optimization with a specific focus on multi-objective simulation optimization, a topic for which she received an NSF CAREER award in 2016. Her published works have been recognized by the INFORMS Computing Society in 2011, by *IISE Transactions* in 2017, and by The Operational Research Society in 2021. She currently serves as co-chair of local arrangements for the 2022 INFORMS Annual Meeting, Program Chair for the 2024 Winter Simulation Conference, and as an associate editor for *Operations Research*. Her email address is susanhunter@purdue.edu, and her website is https://web.ics.purdue.edu/∼hunter63/.

RAGHU PASUPATHY is Professor of Statistics at Purdue University. His current research interests lie broadly in general simulation methodology, statistical computation (especially optimization) and statistical inference. He has been actively involved with the Winter Simulation Conference for the past 15 years. Raghu Pasupathy's email address is pasupath@purdue.edu, and his web page https://web.ics.purdue.edu/~pasupath contains links to papers, software codes, and other material.