GLOBAL MULTI-OBJECTIVE SIMULATION OPTIMIZATION: ERROR BOUNDS AND CONVERGENCE RATES

Burla E. Ondes[1] and Susan R. Hunter[1]

[1] *Edwardson School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, USA*

Abstract

Consider the context of solving a multi-objective simulation optimization problem with one or more continuous objective functions to global optimality on a compact feasible set. For a simple algorithm that consists of selecting a finite set of feasible points using a space-filling design, expending the same number of simulation replications at each point to estimate the objective values, and returning the discretized estimated efficient and Pareto sets, we (i) introduce a mathematically tractable performance indicator for assessing the optimality gap of the discretized estimated efficient set, (ii) derive finite-time probabilistic upper bounds on the optimality gap, and (iii) determine how to trade off the number of feasible points with the number of simulation replications per point to ensure the optimality gap converges to zero in probability at a fast rate. Thus, we identify both an upper bound on the convergence rate for simple algorithms and a lower bound on the convergence rate for other algorithms that exploit structure. In addition, if the optimality gap is measured under the infinity norm, then the required total simulation budget grows slightly faster than logarithmically in the number of objectives. We demonstrate our results through numerical examples having one, two, or three objectives.

1 Introduction

Multi-objective simulation optimization (MOSO) problems are nonlinear optimization problems in which one or more simultaneous objectives can only be observed with stochastic error. Such problems often arise in the modeling and analysis of complex systems when multiple conflicting objective functions are defined implicitly through a Monte Carlo simulation model. Conflicting objectives may arise from the different perspectives of multiple stakeholders or decision makers, such as minimizing the total expected cost of operating a system while simultaneously minimizing its expected environmental or humanitarian impact. While only one decision can be implemented in practice, the goal of solving a MOSO problem is to identify decision variable values that map to the global Pareto set for use as input to the decision-making process; these decision variable values constitute the efficient set. Solving a MOSO problem and presenting decision makers with the results provides a broader perspective on the trade-offs between the objectives when selecting a final decision to implement. Specific applications of MOSO in the literature span a wide variety of settings including agriculture, aviation, energy, healthcare, logistics, manufacturing, and the military; see Hunter et al. [2019] for an overview.

Preprint compiled October 7, 2025 at 6:51pm

We consider the context of solving a MOSO problem with one or more continuous objective functions to global optimality on a compact feasible set, where the objective function values can only be estimated in a pointwise fashion. If computing resources are plentiful, the simulation model runs quickly, or there is no known or exploitable structure in the MOSO problem, one might consider implementing a simple algorithm that consists of selecting m feasible points in the decision space using a space-filling design, expending n simulation replications per point to estimate the objective functions, and returning the discretized estimated efficient and Pareto sets under a total simulation budget b, which equals m times n. If the sampled points are selected on a grid, then the simple algorithm is called *grid search* [Ensor and Glynn, 1997]. Otherwise, the sampled points may be selected as a *low dispersion point set* [Yakowitz et al., 2000]. If a MOSO problem can be solved by a simple algorithm, doing so may have several advantages over implementing a more complex variant: (i) the simple algorithm is intuitive and easily understood by those without particular expertise in MOSO algorithms or their implementation; (ii) the simple algorithm is straightforward to code, and in a parallel computing environment, all simulation replications can be obtained in an embarrassingly parallel fashion; and (iii) common random numbers [Law, 2015] can be fully exploited by obtaining the same number of simulation replications in parallel at each point in the decision space.

Despite its straightforward nature, a simple algorithm must be implemented with care to minimize the error in the returned discretized estimated efficient and Pareto sets. Specifically, the quality of the discretization and the number of selected feasible points m control the discretization error, while the number of simulation replications obtained per point n controls the stochastic error. Assuming discretization points are added according to a space-filling design, at the extreme, holding the number of feasible points m fixed and sending the total budget b to infinity results in an estimator in which discretization error may persist, while holding the number of simulation replications n fixed and sending the total budget b to infinity results in an estimator in which stochastic error may persist. Therefore, determining the correct trade-off of m and n for a given total simulation budget b is key to ensuring the optimality gap of the returned estimator is small with high probability and, ultimately, that the optimality gap converges at a fast rate as the total budget b increases. While Yakowitz et al. [2000] conducted a trade-off analysis for global single-objective simulation optimization, to the best of our knowledge, no such analyses exist for the global MOSO context.

In this article, we provide three major contributions toward the design and analysis of simple algorithms and the identification of error bounds and convergence rates for global MOSO. First, we propose a new, mathematically tractable performance indicator for assessing the convergence of global MOSO algorithms that accounts for the possibility of both false exclusion and false inclusion events; that is, events where truly Pareto points are falsely excluded from the estimated Pareto set, and events where truly dominated points are falsely included in the estimated Pareto set, respectively. While the probability distribution of the Hausdorff distance between the true Pareto set and the true image of the estimated efficient set is a natural performance indicator for assessing convergence of MOSO algorithms [Hunter et al., 2019], to date, the Hausdorff distance is mathematically intractable with regard to our goal of deriving closed-form bounds [Hunter and

Ondes, 2023]. Our proposed performance indicator, described in Section 3 and referred to thereafter as the "optimality gap," exploits a nested set structure to assess convergence in a mathematically tractable way while preserving the desirable characteristics of the Hausdorff distance for convergence analysis (Theorem 1). Second, under the optimality gap defined in Section 3, we use concentration inequalities to derive an upper bound on the probability that the optimality gap exceeds a threshold value which holds for all feasible choices of m and n under mild regularity conditions on the nature of the stochastic error (Theorem 2). In particular, we assume only that the distribution of the stochastic error is sub-exponential. This assumption encompasses a wide variety of distributions including all distributions with bounded support, as well as normal and Laplace error models. Third, we use the probabilistic bound from Theorem 2 to derive implementable values of m and n that ensure convergence in probability of the discretized estimated Pareto set to the true Pareto set at a fast rate for two use cases regarding the nature of the desired guarantee [Hunter and Nelson, 2017]:

(i) *Fixed-precision:* For any user-specified thresholds $\tau > 0$ and $\alpha \in (0, 1)$, Theorem 3 provides finite values of m, n, and b that ensure the probability the optimality gap exceeds τ is less than or equal to α. Theorem 3 implies that for given $\alpha \in (0, 1)$ and any $\delta > 0$, the total budget grows as $b = O\big(\tau^{-(q+2)/(1-\delta)}\big)$ as $\tau \to 0$, where q is the decision space dimension. This rate holds for any fixed number of objectives $d \geq 1$ and, thus, corresponds to the single-objective case; it constitutes an upper bound on the convergence rate for simple algorithms solving global MOSO problems. Further, if the optimality gap is measured under the infinity norm in the objective space, then fixing all other parameters, the total simulation budget grows only slightly faster than logarithmically in the number of objectives d; that is, for any $\epsilon \in (0, 1)$, the budget $b = \Theta((\ln d)^{q/(q-\epsilon)})$ as $d \to \infty$.

(ii) *Fixed-budget:* For a given user-specified total simulation budget b, Theorem 4 provides values of m and n that ensure the probability the optimality gap exceeds $\tau = cd^{1/p}/\lfloor m^{1/q} \rfloor$ is less than or equal to $\alpha = 2d/n$ where c is a constant, $d \geq 1$ is the number of objectives, the optimality gap is measured under the p-norm, and q is the decision space dimension. The allocation rule in Theorem 4 is extremely simple: set $m = \lfloor (\kappa b / \ln b)^{q/(q+2)} \rfloor$ and $n = \lfloor b/m \rfloor$ where κ is a constant whose value depends on the nature of the stochastic error. (For a normal error model, $\kappa = 4/3$, and for a Laplace error model, $\kappa = 1/2$.) This allocation provides a good tradeoff of m and n for a user whose total budget b is fixed in advance but large enough to ensure $m \geq (1 + \tilde{\varepsilon})^q$ for any $\tilde{\varepsilon} > 0$ and $n \geq 1$.

Beyond the design and analysis of simple algorithms, our work also defines a lower bound for the convergence rate of any other global MOSO algorithm that exploits structure. Although the literature on MOSO is relatively small and recent compared to its single-objective counterpart (see, e.g., Hunter et al. [2019] and Fu [2015], respectively), algorithms for solving global MOSO problems with continuous objectives indeed already exist. Examples include Alrefaei and Diabat [2009], Mattila et al. [2013], Huang and Zabinsky [2014, 2025], Zhang et al. [2017], Rojas-Gonzalez et al. [2020], Han and Ouyang [2022], Boffadossi et al. [2025]; also see Andradóttir [2015], Rojas-Gonzalez and Van Nieuwenhuyse [2020]. Our goal is not to argue that our simple algorithm is better than these

existing methods; rather, these existing methods and any future methods should be demonstrably more efficient than a simple algorithm on the same class of global MOSO problems. Thus, from the perspective of a community of researchers creating new and better global MOSO algorithms, e.g. through the *SimOpt* library [Pasupathy and Henderson, 2006, 2011, Eckman et al., 2023a,b], we provide a baseline performance upon which more complex variants should demonstrate improvement. Finally, our work potentially forms a basis for the design of new algorithms that employ aspects of our simple algorithm as a subroutine on compact parts of a feasible space, or that stand on our results for further development of sequential algorithms using low-dispersion point sequences.

The paper is organized as follows. In Section 2, we formally define the MOSO problem and the simple algorithm. In Section 3, we define and discuss the new performance indicator that provides a mathematically tractable way to measure the optimality gap. Section 4 details assumptions required for our analysis, and Section 5 derives closed-form upper bounds on the optimality gap. Section 6 presents results regarding the choice of m, n, and b that ensure convergence at a fast rate. We demonstrate our results on numerical examples having one, two, or three objectives in Section 7. Section 8 contains concluding remarks.

2 Problem context

We formally define the global MOSO problem and the simple algorithm as follows. First, we consider the MOSO problem

$$\text{minimize} \quad [f(x) = (\mathrm{E}[F_1(x,\xi)], \ldots, \mathrm{E}[F_d(x,\xi)])^\intercal] \quad \text{s.t.} \quad x \in \mathcal{X}, \tag{M}$$

where $f\colon \mathcal{X} \to \mathbb{R}^d, \mathcal{X} \subseteq \mathbb{R}^q, q \geq 1$ is an unknown vector-valued function composed of $d \geq 1$ Lipschitz continuous and conflicting objectives $f_k\colon \mathcal{X} \to \mathbb{R}, k = 1, \ldots, d$; the vector-valued random function $F\colon \mathcal{X} \times \Xi \to \mathbb{R}^d$ is composed of individual real-valued random functions $F_k\colon \mathcal{X} \times \Xi \to \mathbb{R}$, $k = 1, \ldots, d$, and for each $x \in \mathcal{X}$, the expected value is $f_k(x) = \mathrm{E}[F_k(x,\xi)] = \int_\Xi F_k(x,u)d\mathrm{P}(u)$. The feasible set $\mathcal{X}$ is nonempty and compact, and the image set is $f(\mathcal{X}) \coloneqq \{f(x) \in \mathbb{R}^d\colon x \in \mathcal{X}\}$. The solution to (M) is the efficient set

$$\mathcal{E} \coloneqq \{x^* \in \mathcal{X}\colon \nexists x \in \mathcal{X} \text{ such that } f(x) \leqslant f(x^*)\} = \{x^* \in \mathcal{X}\colon (f(x^*) - \mathbb{R}^d_{\leqq}) \cap f(\mathcal{X}) = f(x^*)\},$$

where $\mathbb{R}^d_{\leqq} \coloneqq \{y \in \mathbb{R}^d\colon y \leqq 0\}$ and Subsection 2.1 contains notational conventions regarding $\leqslant, \leq$, and $\leqq$. The image of the efficient set is the global Pareto set,

$$\mathcal{P} = f(\mathcal{E}) \coloneqq \{f(x^*) \in \mathbb{R}^d\colon x^* \in \mathcal{E}\} = (f(\mathcal{X}))_{\mathrm{N}},$$

where $\mathcal{S}_{\mathrm{N}}$ is the set of all nondominated points in $\mathcal{S}$ with respect to the cone $\mathbb{R}^d_{\leqq}$.

The formulation in (M) is quite general. In addition to minimizing the expected values of a system's performance measures directly, such as minimizing the expected cost, we can also pose objectives that minimize the expected value of some loss function of the performance measure of interest. Further, our results necessarily assume the version of (M) in which f can only be estimated

in a *pointwise* fashion; that is, if observing an estimator for $f(x)$ at $x \in \mathcal{X}$ requires expending n simulation replications, then observing an estimator for $f(x')$ at a new point $x' \in \mathcal{X}, x \neq x'$ requires expending n additional simulation replications.

We define a *simple algorithm* for solving (M) as a procedure in which we (i) select a finite set $\mathcal{X}_m$ of m points in the feasible set according to a space-filling design and select the total simulation budget $b = m \times n$, (ii) query the simulation oracle to obtain consistent estimators $\bar{F}(\tilde{x}, n) = n^{-1} \sum_{i=1}^{n} F(\tilde{x}, \xi_i)$ of the vector-valued objective at each point $\tilde{x} \in \mathcal{X}_m$ where here, $\xi_1, \ldots, \xi_n$ are n independent and identically distributed (iid) copies of ξ (this assumption does not preclude the use of common random numbers; see Assumption 3 in Section 4), and (iii) return the resulting discretized estimated efficient and Pareto sets. This algorithm is specified in Algorithm 1. Thus, for a given finite discretization of the feasible set $\mathcal{X}_m \subset \mathcal{X}$, $|\mathcal{X}_m| = m$, and simulation budget n, Algorithm 1 solves the *discretized sample-path problem*,

$$\text{minimize} \quad [\bar{F}(x, n) = (\bar{F}_1(x, n), \ldots, \bar{F}_d(x, n))^\intercal] \quad \text{s.t.} \quad x \in \mathcal{X}_m. \tag{$\hat{\text{M}}_m$}$$

We denote the discretized estimated efficient and Pareto sets returned by Algorithm 1 as

$$\hat{\mathcal{E}}_m(b) \coloneqq \{\tilde{x} \in \mathcal{X}_m \colon \nexists \tilde{x}' \in \mathcal{X}_m \text{ such that } \bar{F}(\tilde{x}', n) \leqslant \bar{F}(\tilde{x}, n)\},$$
$$\hat{\mathcal{P}}_m(b) = \bar{F}(\hat{\mathcal{E}}_m(b), n) \coloneqq \{\bar{F}(\tilde{x}, n) \colon \tilde{x} \in \hat{\mathcal{E}}_m(b)\} = (\bar{F}(\mathcal{X}_m))_{\text{N}},$$

where $\hat{\mathcal{E}}_m(b)$ and $\hat{\mathcal{P}}_m(b)$ are discretized estimators of $\mathcal{E}$ and $\mathcal{P}$, respectively.

Algorithm 1: A simple algorithm for global MOSO on a compact set $\mathcal{X}$

Require: a compact feasible set $\mathcal{X}$; an `oracle` capable of producing estimators $\bar{F}(x, n)$ such that $\bar{F}(x, n) \to f(x)$ w.p.1 as $n \to \infty$ for each $x \in \mathcal{X}$ where f is continuous on $\mathcal{X}$

Input: a finite set of feasible points $\mathcal{X}_m \subset \mathcal{X}$ chosen as a low-dispersion point set and having cardinality $m \geq 1$; a finite total simulation budget $b \geq 1$

$n \leftarrow \lfloor b/m \rfloor$ `/determine number of simulation replications` n

foreach $\tilde{x} \in \mathcal{X}_m$ **do**

  $\bar{F}(\tilde{x}, n) \leftarrow$ `oracle`$(\tilde{x}, n)$ `/estimate` $f(\tilde{x})$ `using` n `simulation replications`

$\hat{\mathcal{E}}_m(b) \leftarrow \{\tilde{x} \in \mathcal{X}_m \colon \nexists \tilde{x}' \in \mathcal{X}_m \text{ such that } \bar{F}(\tilde{x}', n) \leqslant \bar{F}(\tilde{x}, n)\}$

$\hat{\mathcal{P}}_m(b) \leftarrow \bar{F}(\hat{\mathcal{E}}_m(b), n) = \{\bar{F}(\tilde{x}, n) \colon \tilde{x} \in \hat{\mathcal{E}}_m(b)\}$

return $\hat{\mathcal{E}}_m(b), \hat{\mathcal{P}}_m(b)$ `/return discretized estimated efficient, Pareto sets`

There is another relevant problem whose notation we require: the discretized version of (M) without estimation. That is, Algorithm 1 implicitly defines the following discretization of (M),

$$\text{minimize} \quad [f(x) = (\text{E}[F_1(x, \xi)], \ldots, \text{E}[F_d(x, \xi)])] \quad \text{s.t.} \quad x \in \mathcal{X}_m. \tag{M_m}$$

The solution to (M_m) is a finite approximation to the solution to (M). We refer to this quantity as

the *discretized efficient set*, $\mathcal{E}_m$, and its image as the *discretized Pareto set*, $\mathcal{P}_m$,

$$\mathcal{E}_m \coloneqq \{\tilde{x}^* \in \mathcal{X}_m \colon \nexists\, \tilde{x} \in \mathcal{X}_m \text{ such that } f(\tilde{x}) \leqslant f(\tilde{x}^*)\}$$
$$\mathcal{P}_m \coloneqq \{f(\tilde{x}^*) \colon \tilde{x}^* \in \mathcal{E}_m\} = (f(\mathcal{X}_m))_{\mathrm{N}}.$$

2.1 Notation and terminology

When comparing two vectors $y, \tilde{y} \in \mathbb{R}^d$, we write $\tilde{y} \leqq y$ if $\tilde{y}_k \leq y_k$ for all $k \in \{1, \ldots, d\}$ and write $\tilde{y} \leqslant y$ if $\tilde{y} \leqq y$ and $\tilde{y} \neq y$; analogous definitions hold for $\geqq, \geqslant$. A d-dimensional vector of ones is 1_d. For $p \geq 1$, the p-norm is $\|y\|_p \coloneqq (|y_1|^p + \ldots + |y_d|^p)^{1/p}$, and the ∞-norm is $\|y\|_\infty \coloneqq \max\{|y_1|, \ldots, |y_d|\}$. If $1 \leq p < r \leq \infty$, then $\|y\|_r \leq \|y\|_p \leq d^{(1/p - 1/r)}\|y\|_r$. The cone $\mathbb{R}^d_{\geqq} \coloneqq \{y \in \mathbb{R}^d \colon y \geqq 0\}$ is the nonnegative orthant, and $\mathbb{R}^d_{>} \coloneqq \{y \in \mathbb{R}^d \colon y > 0\}$. For sets $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^d$, $\mathcal{A} + \mathcal{B} \coloneqq \{a + b \colon a \in \mathcal{A}, b \in \mathcal{B}\}$ is the Minkowski sum. For nonempty $\mathcal{A}, \mathcal{B}$ and $x \in \mathbb{R}^d$, $\mathbb{D}_p(x, \mathcal{B}) \coloneqq \inf_{y \in \mathcal{B}} \|x - y\|_p$ and $\mathbb{D}_p(\mathcal{A}, \mathcal{B}) \coloneqq \sup_{x \in \mathcal{A}} \mathbb{D}_p(x, \mathcal{B})$. The Hausdorff distance is $\mathbb{H}_p(\mathcal{A}, \mathcal{B}) \coloneqq \max\{\mathbb{D}_p(\mathcal{A}, \mathcal{B}), \mathbb{D}_p(\mathcal{B}, \mathcal{A})\}$. A set $\mathcal{S} \subseteq \mathbb{R}^d$ has boundary $\mathrm{bd}(\mathcal{S})$, closure $\mathrm{cl}(\mathcal{S})$, complement $\mathcal{S}^c$, diameter $\mathrm{diam}_p(\mathcal{S}) = \sup_{x,y \in \mathcal{S}} \|x - y\|_p$, interior $\mathrm{int}(\mathcal{S})$, upper image $\mathcal{S}^{\uparrow} \coloneqq \mathcal{S} + \mathbb{R}^d_{\geqq}$, lower image $\mathcal{S}^{\downarrow} \coloneqq \mathcal{S} - \mathbb{R}^d_{\geqq}$, and $\mathcal{S}_{\mathrm{N}}$ and $\mathcal{S}_{\mathrm{wN}}$ are the sets of all nondominated and weakly nondominated (or, all Pareto and weakly Pareto) points in $\mathcal{S}$ with respect to the ordering cone $\mathbb{R}^d_{\geqq}$, respectively: $\mathcal{S}_{\mathrm{N}} \coloneqq \{s \in \mathcal{S} \colon \nexists\, s' \in \mathcal{S} \text{ such that } s' \leqslant s\}$, $\mathcal{S}_{\mathrm{wN}} \coloneqq \{s \in \mathcal{S} \colon \nexists\, s' \in \mathcal{S} \text{ such that } s' < s\}$. In general, $\mathcal{S}_{\mathrm{N}}$ may be empty (e.g. if $\mathcal{S}$ is open). $\mathcal{S}_{\mathrm{N}}$ is *externally stable* if $\mathcal{S} \subseteq \mathcal{S}_{\mathrm{N}} + \mathbb{R}^d_{\geqq}$. The set $\mathcal{S}$ is bounded if there exists a constant $c < \infty$ such that $\forall x, y \in \mathcal{S}$, $\|x - y\| \leq c$. For a function $f \colon \mathcal{D} \to \mathbb{R}^d, \mathcal{D} \subseteq \mathbb{R}^q$ and a set $\mathcal{S} \subseteq \mathcal{D}$, $f(\mathcal{S}) \coloneqq \{f(x) \colon x \in \mathcal{S}\}$. Given a probability space $(\Omega, \mathfrak{F}, \mathrm{P})$, an event $\mathcal{A} \in \mathfrak{F}$ occurs *with probability one (w.p.1)* or *almost surely (a.s.)* if and only if $\mathrm{P}(\mathcal{A}) = \mathrm{P}(\{\omega \in \Omega \colon \omega \in \mathcal{A}\}) = 1$. For sequences of reals $\{a_n\}$ and $\{b_n\}$, $a_n = O(1)$ if there exists $c \in (0, \infty)$ with $|a_n| < c$ for large enough n; $a_n = \Omega(1)$ if there exists $c \in (0, \infty)$ with $|a_n| > c$ for large enough n. For positive-valued sequences of reals $\{a_n\}$ and $\{b_n\}$, we say $a_n = O(b_n)$ if $a_n / b_n = O(1)$ as $n \to \infty$. Further, $a_n = \Theta(1)$ if $0 < \liminf a_n \leq \limsup a_n < \infty$.

3 A tractable performance indicator for convergence

In a single-objective simulation optimization context, the optimality gap is straightforwardly defined as $f_1(X^*_m(b)) - f_1(x^*)$, where $X^*_m(b)$ is a random variable representing the solution to the discretized sample-path problem ($\hat{\mathrm{M}}_m$) with $d = 1$ after expending a total of b simulation replications, $f_1(X^*_m(b))$ denotes the true image of this estimated solution, and $f_1(x^*)$ is the true optimal value [Yakowitz et al., 2000, Pasupathy and Henderson, 2006]. In a MOSO context, however, the definition of the optimality gap must account for the fact that the estimated solution is a set rather than a point. To this end, the deterministic multi-objective optimization literature contains over one hundred ways to assess the quality of a returned solution [Zitzler et al., 2003, Faulkenberg and Wiecek, 2010, Li and Yao, 2019, Audet et al., 2021]. However, far fewer performance indicators are specifically designed for a MOSO context; see Hunter et al. [2019], Branke [2024], Thengvall et al. [2025]. Of these performance indicators, we are interested in the ones that assess *convergence*.

For assessing the convergence of Algorithm 1, performance indicators associated with the probability distribution of the Hausdorff distance between the true Pareto set $\mathcal{P}$ and the true image of

the discretized estimated Pareto set $f(\hat{\mathcal{E}}_m(b))$, evaluated under the p-norm,

$$\mathbb{H}_p(\mathcal{P}, f(\hat{\mathcal{E}}_m(b))) = \max\{\mathbb{D}_p(\mathcal{P}, f(\hat{\mathcal{E}}_m(b))), \mathbb{D}_p(f(\hat{\mathcal{E}}_m(b)), \mathcal{P})\}, \tag{1}$$

are a natural choice [Hunter et al., 2019, Hunter and Ondes, 2023]. Assessing algorithms on the expected value of (1), or numerically on the quantiles of its distribution as in Cooper et al. [2020], is consistent with the philosophies of evaluating simulation optimization algorithms in Pasupathy and Henderson [2006]. In particular, as a metric, the Hausdorff distance in (1) has several desirable properties for assessing convergence: (i) the Hausdorff distance between $\mathcal{P}$ and $f(\hat{\mathcal{E}}_m(b))$ equals zero if and only if the two sets have the same closure, which accounts for the fact that the true image of the estimated efficient set may contain points that strictly dominate other points in the same set, and (ii) the triangle inequality holds, which implies we can determine an upper bound on the error by separately analyzing its deterministic and stochastic parts. That is, the overall error is less than or equal to the discretization error plus the selection error for a given discretization. However, as discussed in Hunter and Ondes [2023], Ondes and Hunter [2023], to date, the Hausdorff distance in (1) is mathematically intractable with respect to obtaining the types of closed-form bounds required to perform an analysis of Algorithm 1.

To overcome the intractability of the Hausdorff distance in (1), we propose a new performance indicator for assessing convergence of global MOSO algorithms, or the "optimality gap," defined as follows and illustrated for bi-objective example problems in Figures 1 and 2. First, define the "upper" aspect of the estimated efficient set with respect to the true image,

$$\hat{\mathcal{U}}_m(b) \coloneqq \{x \in \hat{\mathcal{E}}_m(b) \colon \nexists x' \in \hat{\mathcal{E}}_m(b) \text{ such that } f(x') \geqslant f(x)\}.$$

That is, $\hat{\mathcal{U}}_m(b)$ contains all points in $\hat{\mathcal{E}}_m(b)$ whose true images do not dominate the true images of any other points in $\hat{\mathcal{E}}_m(b)$. Using the idea that maximizing f is equivalent to minimizing $-f$,

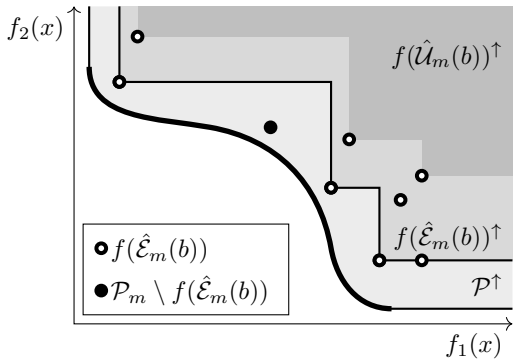
$$f(\hat{\mathcal{U}}_m(b)) = -\big(\big(-f(\hat{\mathcal{E}}_m(b))\big)_{\mathrm{N}}\big) \subseteq f(\hat{\mathcal{E}}_m(b)). \tag{2}$$

While it is possible that $f(\hat{\mathcal{E}}_m(b)) = f(\hat{\mathcal{U}}_m(b))$, it is also possible that $f(\hat{\mathcal{E}}_m(b)) \neq f(\hat{\mathcal{U}}_m(b))$, and to assess convergence, we would prefer a performance indicator that accounts for it. To this end, we propose the (random) performance indicator

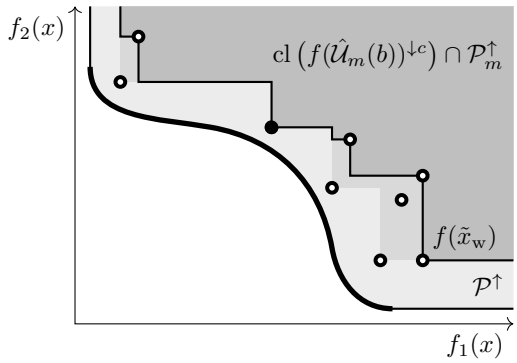
$$G_m(b) \coloneqq \max\left\{\mathbb{D}_p\big(\mathcal{P}^{\uparrow}, f(\hat{\mathcal{E}}_m(b))^{\uparrow}\big), \mathbb{D}_p\big(\mathcal{P}^{\uparrow}, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}\big)\right\}; \tag{3}$$

recall that for a set $\mathcal{S} \subseteq \mathbb{R}^d$, $\mathcal{S}^{\uparrow} \coloneqq \mathcal{S} + \mathbb{R}^d_{\geqq}$ and $\mathcal{S}^{\downarrow} \coloneqq \mathcal{S} - \mathbb{R}^d_{\geqq}$ denote its upper and lower images, respectively. Henceforth, for simplicity, we suppress notational dependence of the distance $G_m(b)$ on the chosen value of p in the p-norm; we default to the infinity norm [Sayin, 2000].

The performance indicator $G_m(b)$ from (3) measures the maximum of the distances from the upper image of the Pareto set, $\mathcal{P}^{\uparrow}$, to the sets $f(\hat{\mathcal{E}}_m(b))^{\uparrow}$ and $\mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}$, respectively. The first distance $\mathbb{D}_p\big(\mathcal{P}^{\uparrow}, f(\hat{\mathcal{E}}_m(b))^{\uparrow}\big)$ in (3) is the *modified coverage error*, which was introduced in Hunter and Ondes [2023] and based on the coverage error [Sayin, 2000]. Loosely speaking, this



(a) The distance $\mathbb{D}_p\big(\mathcal{P}^{\uparrow}, f(\hat{\mathcal{E}}_m(b))^{\uparrow}\big)$ measures the error due to false exclusion.



(b) The distance $\mathbb{D}_p\big(\mathcal{P}^{\uparrow}, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}\big)$ measures the error due to false inclusion; it is not affected by the discretized weakly Pareto point $f(\tilde{x}_{\mathrm{w}}) \in f(\hat{\mathcal{E}}_m(b))$.

Figure 1: For an example with $d = 2$ objectives, the images show the component sets that make up the distances in the performance indicator $G_m(b)$, which is a random variable.

distance measures how well the upper image of $f(\hat{\mathcal{E}}_m(b))$ "covers" the upper image of the true Pareto set. Thus, it measures the distance due to *false exclusion* events where points in the Pareto set are falsely excluded from $f(\hat{\mathcal{E}}_m(b))$ by a discretization that is not fine enough or by estimation error. The second distance $\mathbb{D}_p\big(\mathcal{P}^{\uparrow}, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}\big)$ in (3) measures the error due to *false inclusion* events where points that do not belong to the weakly Pareto set are falsely included in $f(\hat{\mathcal{E}}_m(b))$; this term is unique to the stochastic multi-objective optimization context. That is, even if we magically ensure our algorithm locates all points in the discretized Pareto set, so that $f(\hat{\mathcal{E}}_m(b)) \supseteq \mathcal{P}_m$, if the MOSO algorithm also returns points in $f(\hat{\mathcal{E}}_m(b))$ that dominate other points in $f(\hat{\mathcal{E}}_m(b))$ lying in the interior of the upper image of the discretized Pareto set, $\mathrm{int}(\mathcal{P}_m^{\uparrow})$, then the error term due to false inclusion is the maximum in (3). Thus, like the Hausdorff distance in (1), the performance indicator in (3) accounts for the possibility that a MOSO algorithm returns points in $f(\hat{\mathcal{E}}_m(b))$ that strictly dominate other points in $f(\hat{\mathcal{E}}_m(b))$.

To ensure proper intuition, we emphasize that the optimality gap $G_m(b)$ ignores false inclusion events by weakly Pareto points in both problem (M) and the discretized problem (M_m). That is, $G_m(b)$ ignores false inclusion events by points in $(f(\mathcal{X}))_{\mathrm{wN}}$ and $(f(\mathcal{X}_m))_{\mathrm{wN}}$. Since weakly Pareto points cannot be classified as Pareto or non-Pareto in problem (M) or (M_m) with finite sample size, when proving convergence, it is customary to either (i) assume weakly Pareto points do not exist, or (ii) assume that if weakly Pareto points do exist, then we do not care whether they are classified correctly or incorrectly; see, e.g., Applegate et al. [2020], Cooper et al. [2020], Hunter and Pasupathy [2022] for various combinations of such assumptions and results. Our proposed optimality gap $G_m(b)$ takes the perspective in (ii) when calculating the distance due to false inclusion. Therefore, in Section 4 detailing our assumptions, we require no assumption excluding the existence of weakly Pareto points. That $G_m(b)$ ignores false inclusion by weakly Pareto points in (M_m) is illustrated by the point $f(\tilde{x}_{\mathrm{w}})$ in Figure 1(b) — it belongs to the discretized weakly Pareto set, but the shaded region representing $\mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}$ is the same whether or not this point belongs to $f(\hat{\mathcal{E}}_m(b))$.

Given the nested structure of the sets in Figure 1, the following Theorem 1 holds for the proposed

performance indicator $G_m(b)$. Specifically, Theorem 1 demonstrates that $G_m(b)$ retains the desirable properties of the Hausdorff distance for assessing convergence, including that it obeys a triangle inequality. This triangle inequality enables us to obtain an upper bound on $G_m(b)$ as the sum of two terms corresponding to the discretization error and the selection error, where the selection error is the maximum of the false inclusion and false exclusion errors. When there is only one objective such that $d = 1$, the proposed performance indicator $G_m(b)$ simplifies to the standard optimality gap definition, $f_1(X_m^*(b)) - f_1(x^*)$, where x^* is a solution to (M), $\tilde{x}_m^*$ is a solution to (M_m), and $X_m^*(b)$ is a solution to ($\hat{\mathrm{M}}_m$). Figure 2 illustrates Theorem 1 Parts 2 and 3 for an example with $d = 2$ objectives.

Theorem 1. *Suppose $\mathcal{P} \neq \emptyset$ is externally stable, $m \geq 1$, $n = \lfloor b/m \rfloor \geq 1$, and $f(\hat{\mathcal{E}}_m(b)) \neq \emptyset$.*

1. *$G_m(b) = 0$ if and only if $\mathrm{cl}(\mathcal{P}^\uparrow) = \mathrm{cl}(f(\hat{\mathcal{E}}_m(b))^\uparrow) = \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^\uparrow$.*
2. *If $f(\hat{\mathcal{E}}_m(b)) = f(\hat{\mathcal{U}}_m(b))$, then $G_m(b) = \mathbb{D}_p\big(\mathcal{P}^\uparrow, f(\hat{\mathcal{E}}_m(b))^\uparrow\big)$.*
3. *If $f(\hat{\mathcal{E}}_m(b)) \supseteq \mathcal{P}_m$, then $G_m(b) = \mathbb{D}_p\big(\mathcal{P}^\uparrow, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^\uparrow\big)$.*
4. *Define the false exclusion and false inclusion aspects of the selection error as*

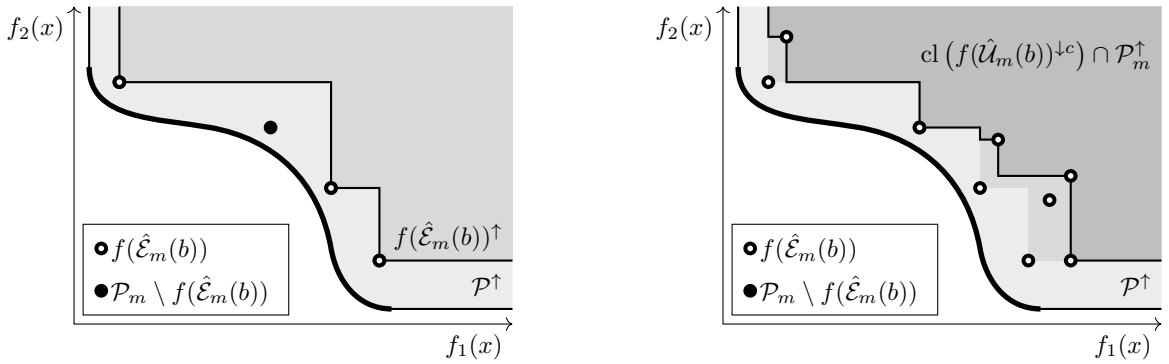
$$G_m^{\mathrm{FE}}(b) \coloneqq \underbrace{\mathbb{D}_p\big(\mathcal{P}_m^\uparrow, f(\hat{\mathcal{E}}_m(b))^\uparrow\big)}_{\text{false exclusion error}}, \quad G_m^{\mathrm{FI}}(b) \coloneqq \underbrace{\mathbb{D}_p\big(\mathcal{P}_m^\uparrow, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^\uparrow\big)}_{\text{false inclusion error}}$$

Then the following inequality holds w.p.1:

$$G_m(b) \leq \underbrace{\mathbb{D}_p\big(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow\big)}_{\text{discretization error}} + \underbrace{\max\left\{G_m^{\mathrm{FE}}(b), G_m^{\mathrm{FI}}(b)\right\}}_{\text{selection error}}. \tag{4}$$

5. *If $d = 1$, then (i) false inclusion cannot occur without false exclusion, (ii) $f(\hat{\mathcal{E}}_m(b)) = f(\hat{\mathcal{U}}_m(b))$, (iii) the discretization error is $\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow) = f_1(\tilde{x}_m^*) - f_1(x^*)$, and (iv) the selection error and the optimality gap are, respectively,*

$$G_m^{\mathrm{FE}}(b) = G_m^{\mathrm{FI}}(b) = f_1(X_m^*(b)) - f_1(\tilde{x}_m^*), \quad G_m(b) = f_1(X_m^*(b)) - f_1(x^*). \tag{5}$$



(a) In Theorem 1 Part 2, $f(\hat{\mathcal{E}}_m(b)) = f(\hat{\mathcal{U}}_m(b)$ implies no false inclusion, and $G_m(b) = \mathbb{D}_p\big(\mathcal{P}^\uparrow, f(\hat{\mathcal{E}}_m(b))^\uparrow\big)$.

(b) In Theorem 1 Part 3, $f(\hat{\mathcal{E}}_m(b)) \supseteq \mathcal{P}_m$ implies no false exclusion, and $G_m(b) = \mathbb{D}_p\big(\mathcal{P}^\uparrow, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^\uparrow\big)$.

Figure 2: For an example with $d = 2$ objectives, the images illustrate Theorem 1 Parts 2 and 3.

Proof. For Part 1, since $\mathcal{P} = \big(f(\mathcal{X})\big)_{\mathrm{N}}$ is externally stable, $f(\mathcal{X}) \subset \mathcal{P} + \mathbb{R}^d_{\geqq} = \mathcal{P}^{\uparrow}$. Further, $\emptyset \neq f(\hat{\mathcal{U}}_m(b)) \subseteq f(\hat{\mathcal{E}}_m(b)) \subseteq f(\mathcal{X}_m) \subseteq f(\mathcal{X})$. Then external stability also implies

$$\emptyset \neq f(\hat{\mathcal{U}}_m(b))^{\uparrow} \subseteq f(\hat{\mathcal{E}}_m(b))^{\uparrow} \subseteq \mathcal{P}_m^{\uparrow} \subseteq \mathcal{P}^{\uparrow}. \tag{6}$$

Therefore $\mathbb{D}_p\big(f(\hat{\mathcal{E}}_m(b))^{\uparrow}, \mathcal{P}^{\uparrow}\big) = 0$, and

$$\mathbb{D}_p\big(\mathcal{P}^{\uparrow}, f(\hat{\mathcal{E}}_m(b))^{\uparrow}\big) = \mathbb{H}_p\big(\mathcal{P}^{\uparrow}, f(\hat{\mathcal{E}}_m(b))^{\uparrow}\big). \tag{7}$$

Likewise, (6) implies $\mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow} \subseteq \mathcal{P}^{\uparrow}$, $\mathbb{D}_p\big(\mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}, \mathcal{P}^{\uparrow}\big) = 0$, and

$$\mathbb{D}_p\big(\mathcal{P}^{\uparrow}, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}\big) = \mathbb{H}_p\big(\mathcal{P}^{\uparrow}, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}\big). \tag{8}$$

By its definition in (3), $G_m(b)$ equals the maximum of the two Hausdorff distances in (7) and (8). Thus, the result follows by the properties of the Hausdorff distance.

For Part 2, suppose $f(\hat{\mathcal{E}}_m(b)) = f(\hat{\mathcal{U}}_m(b))$, and notice that $f(\hat{\mathcal{E}}_m(b))^{\uparrow} = f(\hat{\mathcal{E}}_m(b)) + \mathbb{R}^p_{\geqq} \subseteq (f(\hat{\mathcal{E}}_m(b)) - \mathbb{R}^p_{\geqq})^c = f(\hat{\mathcal{E}}_m(b))^{\downarrow c}$. Using (6), $f(\hat{\mathcal{E}}_m(b))^{\uparrow} = f(\hat{\mathcal{E}}_m(b))^{\uparrow} \cap \mathcal{P}_m^{\uparrow} \subseteq \mathrm{cl}\left(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}$, which implies the result.

For Part 3, if $f(\hat{\mathcal{E}}_m(b)) \supseteq \mathcal{P}_m$, then $f(\hat{\mathcal{E}}_m(b))^{\uparrow} = \mathcal{P}_m^{\uparrow}$. Substituting into (3),

$$G_m(b) = \max\left\{\mathbb{D}_p\big(\mathcal{P}^{\uparrow}, \mathcal{P}_m^{\uparrow}\big), \mathbb{D}_p\big(\mathcal{P}^{\uparrow}, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}\big)\right\} = \mathbb{D}_p\big(\mathcal{P}^{\uparrow}, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}\big).$$

For Part 4, using (7), the triangle inequality, and the nesting in (6), we have

$$\begin{aligned}\mathbb{D}_p(\mathcal{P}^{\uparrow}, f(\hat{\mathcal{E}}_m(b))^{\uparrow}) = \mathbb{H}_p(\mathcal{P}^{\uparrow}, f(\hat{\mathcal{E}}_m(b))^{\uparrow}) &\leq \mathbb{H}_p(\mathcal{P}^{\uparrow}, \mathcal{P}_m^{\uparrow}) + \mathbb{H}_p(\mathcal{P}_m^{\uparrow}, f(\hat{\mathcal{E}}_m(b))^{\uparrow}) \\ &= \mathbb{D}_p(\mathcal{P}^{\uparrow}, \mathcal{P}_m^{\uparrow}) + \mathbb{D}_p(\mathcal{P}_m^{\uparrow}, f(\hat{\mathcal{E}}_m(b))^{\uparrow}).\end{aligned} \tag{9}$$

Similarly, using (8), the triangle inequality, and the nesting in (6) yields

$$\mathbb{D}_p\big(\mathcal{P}^{\uparrow}, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}\big) \leq \mathbb{D}_p\big(\mathcal{P}^{\uparrow}, \mathcal{P}_m^{\uparrow}\big) + \mathbb{D}_p\big(\mathcal{P}_m^{\uparrow}, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}\big). \tag{10}$$

Combining (9), (10), and the definition of $G_m(b)$ from (3) yields the result.

For Part 5, use $\mathcal{P} = \{f_1(x^*)\}$, $\mathcal{P}_m = \{f_1(\tilde{x}_m^*)\}$ and $\hat{\mathcal{E}}_m(b) = \{X_m^*(b)\}$ in (3). □

Henceforth, when we refer to the optimality gap at the returned discretized estimated efficient set $\hat{\mathcal{E}}_m(b)$ in Algorithm 1, the optimality gap is assessed under our proposed performance indicator $G_m(b)$ in (3). We demonstrate the tractability of this performance indicator by deriving upper bounds for the optimality gap in Section 5. These upper bounds hold under mild regularity conditions that we formalize next, in Section 4.

4 Assumptions for analysis

In this section, we formalize the required assumptions for our analysis of the optimality gap corresponding to the estimated discretized efficient set returned by Algorithm 1. We require assumptions on the feasible set and objectives, the discretization quality, and the type of stochastic error.

4.1 Characteristics of the feasible set and objective functions

To begin, we formalize the assumption that the feasible set $\mathcal{X}$ is compact and the true function f is ℓ-Lipschitz continuous on $\mathcal{X}$. We also require mild conditions on the continuity of the objective function estimators $\bar{F}(\cdot, n)$ for all $n \geq 1$ that ensure the estimated efficient set exists, where $\bar{F}(\cdot, n) = (\bar{F}_1(\cdot, n), \ldots, \bar{F}_d(\cdot, n))^\intercal$.

Assumption 1. *We assume the following:*

1. *The feasible set $\mathcal{X} \subset \mathbb{R}^q$, $q \geq 1$, is nonempty and compact, and its discretization $\mathcal{X}_m \subset \mathcal{X}$ is nonempty and finite.*
2. *For given $r, p \geq 1$, the vector-valued random function f is ℓ-Lipschitz continuous on $\mathcal{X}$, that is, for all $x_1, x_2 \in \mathcal{X}$, $\|f(x_1) - f(x_2)\|_p \leq \ell \|x_1 - x_2\|_r$, where $\ell > 0$.*
3. *For all values of $n \geq 1$, $\bar{F}(\cdot, n)$ is $\mathbb{R}^d_{\geqq}$-semicontinuous w.p.1; that is, for all $y \in \mathbb{R}^d$, the set $\{x \in \mathcal{X} \colon y - \bar{F}(x, n) \in \mathbb{R}^d_{\geqq}\}$ is almost surely closed.*

Assumption 1 Part 3 ensures the non-discretized estimated efficient set is nonempty w.p.1 [Ehrgott, 2005, Theorem 2.19, p. 33]. It holds if and only if all of the component functions of $\bar{F}(\cdot, n)$ are almost surely lower semicontinuous on $\mathcal{X}$ for all $n \geq 1$; see Ehrgott [2005, Lemma 2.17, p. 32].

Assumption 1 implies several important properties of the relevant sets under consideration, which are stated in the following Lemma 1. In particular, it implies that the efficient and Pareto sets are nonempty and $\mathcal{P}$, $\mathcal{P}_m$, and $f(\hat{\mathcal{E}}_m(b))_{\mathrm{N}}$ are externally stable. We present the results without proof since they follow directly from Assumption 1 and the results in Ehrgott [2005, p. 32-33].

Lemma 1. *Under Assumption 1, for all finite sets $\mathcal{X}_m \subset \mathcal{X}$, $m \geq 1$ and all $n \geq 1$:*

1. *$f(\mathcal{X})$, $f(\mathcal{X}_m)$, are nonempty and compact;*
2. *$\mathcal{E}$, $\mathcal{P}$, $\mathcal{E}_m$, $\mathcal{P}_m$ are nonempty, and $\mathcal{P}$ and $\mathcal{P}_m$ are externally stable;*
3. *$\hat{\mathcal{E}}_m(b)$, $\hat{\mathcal{P}}_m(b)$, $\hat{\mathcal{U}}_m(b)$, $f(\hat{\mathcal{U}}_m(b))$, and $f(\hat{\mathcal{E}}_m(b))$ are nonempty and finite w.p.1;*
4. *$f(\hat{\mathcal{E}}_m(b))_{\mathrm{N}}$ is externally stable w.p.1.*

4.2 Discretization quality

How well the discretized efficient and Pareto sets $\mathcal{E}_m$ and $\mathcal{P}_m$ approximate the true sets $\mathcal{E}$ and $\mathcal{P}$ has to do with how well the discretized feasible set $\mathcal{X}_m$ "covers" the feasible set $\mathcal{X}$. Thus, we measure the quality of the discretization $\mathcal{X}_m$ by its *dispersion* under the r-norm as

$$t \coloneqq \mathbb{D}_r(\mathcal{X}, \mathcal{X}_m) = \sup_{x \in \mathcal{X}} \inf_{\tilde{x} \in \mathcal{X}_m} \|x - \tilde{x}\|_r. \tag{11}$$

Letting $B_r(x, t) \subset \mathbb{R}^q$ be a closed q-dimensional ball under the r-norm having radius t centered at $x \in \mathcal{X}$ and denoting the discretized feasible set as $\mathcal{X}_m = \{\tilde{x}_1, \ldots, \tilde{x}_m\}$, (11) is also the smallest value of t such that the q-balls $B_r(\tilde{x}_1, t), \ldots, B_r(\tilde{x}_m, t)$ cover the feasible set $\mathcal{X}$ [Niederreiter, 1992, Yakowitz et al., 2000].

As in Yakowitz et al. [2000], we make the following Assumption 2 on the dispersion of the points in the discretized feasible set.

Assumption 2 (see Yakowitz et al. [2000, p. 943]). *The chosen points in the discretization* $\mathcal{X}_m \subset \mathcal{X} \subset \mathbb{R}^q$ *have* $m \geq (1+\tilde{\varepsilon})^q$ *for some* $\tilde{\varepsilon} > 0$ *and dispersion* $t = \mathbb{D}_r(\mathcal{X}, \mathcal{X}_m) \leq \tilde{c}_q / \lfloor m^{1/q} \rfloor$ *where* $0 < \tilde{c}_q < \infty$ *is a constant that may depend on the dimension* q *but not on* m.

Choosing points on a grid gives a lower bound on the problem of choosing $\mathcal{X}_m$ to minimize $\mathbb{D}_2(\mathcal{X}, \mathcal{X}_m)$ in $q \geq 2$ dimensions [Yakowitz et al., 2000, p. 942]. The following Example 1 shows how to calculate $\tilde{c}_q$ for grid search on the q-dimensional unit hypercube.

Example 1 (Yakowitz et al. [2000, p. 942]). Let $\mathcal{X} = [0,1]^q$ be the q-dimensional unit hypercube and set $\nu = \lfloor m^{1/q} \rfloor$. Choose $\mathcal{X}_m$ as $\left\{(x_1, \ldots, x_q) \colon x_j \in \left\{\frac{1}{2\nu}, \frac{3}{2\nu}, \ldots, \frac{2\nu-1}{2\nu}\right\} \text{ for each } j\right\} \cup \Psi$, where Ψ is an arbitrary set of $m - \nu^q$ points in $\mathcal{X}$. Then under the $r \in \{1, 2, \ldots\}$ and infinity norms, the respective dispersions of the grid $\mathcal{X}_m$ are $\mathbb{D}_r(\mathcal{X}, \mathcal{X}_m) = q^{1/r}/(2\lfloor m^{1/q} \rfloor)$ and $\mathbb{D}_\infty(\mathcal{X}, \mathcal{X}_m) = 1/(2\lfloor m^{1/q} \rfloor)$. Therefore, a grid design on the q-dimensional unit hypercube has $\tilde{c}_q = (1/2)q^{1/r}$ under the r-norm and $\tilde{c}_q = 1/2$ under the infinity norm in the decision space.

Any quantity that depends on the dispersion t also depends on m, in the sense that $t \to 0$ implies $m \to \infty$. For appropriately space-filling choices of $\{\mathcal{X}_m, m \geq (1+\tilde{\varepsilon})^q\}$ satisfying Assumption 2, then $m \to \infty$ implies $t \to 0$.

4.3 Types of stochastic error

Since F is defined implicitly through a Monte Carlo simulation oracle in Section 2, given iid copies of ξ, denoted $\xi_j, j = 1, \ldots, n$, the stochastic error is governed by the properties of the vector-valued error random fields $Z_j(\cdot) := F(\cdot, \xi_j) - f(\cdot)$ where $Z_j(\cdot) = (Z_{j1}(\cdot), \ldots, Z_{jd}(\cdot))^\intercal$. The vector-valued objective function estimator is $\bar{F}(\cdot, n) := f(\cdot) + n^{-1} \sum_{j=1}^n Z_j(\cdot)$. To ensure the estimators are unbiased, as in Ragavan et al. [2022], we impose the following Assumption 3 on the stochastic error fields.

Assumption 3. *The sequence* $\xi_j, j = 1, \ldots, n$ *is iid with* $\mathrm{E}[Z_{jk}(x)] = 0$ *for each objective* $k \in \{1, \ldots, d\}$ *and feasible point* $x \in \mathcal{X}$.

Assumption 3 implies that the error fields $Z_j(\cdot), j = 1, \ldots, n$ are iid, and hence, that $\bar{F}(\cdot, n)$ is an unbiased estimator for $f(\cdot)$. Note that this assumption allows the use of common random numbers, which is the practice of using the same $\xi_j, j = 1, \ldots, n$ when estimating f as $\bar{F}(\tilde{x}_i, n)$ for all points $\tilde{x}_i \in \mathcal{X}_m$. Further, since $Z_j(\cdot)$ is a vector, Assumption 3 allows the error to be correlated across the objectives.

Next, we assume the unbiased estimator is of sufficient quality by ensuring the error is sub-exponential and that the sub-exponential norms are uniformly bounded away from infinity on the feasible set. As in Vershynin [2018, p. 31], for a sub-exponential random variable X, let the sub-exponential norm of X be

$$\|X\|_{\psi_1} := \inf\{s > 0 \colon \mathrm{E}[\exp(|X|/s)] \leq 2\}. \tag{12}$$

Assumption 4. *The error* $Z_{jk}(x)$ *is sub-exponential for each objective* $k \in \{1, \ldots, d\}$, *feasible point* $x \in \mathcal{X}$, *and simulation replication* $j = 1, 2, \ldots$. *Further, the sub-exponential norms of the* $Z_{jk}(x)$*'s*

are uniformly bounded such that there exists a constant

$$\eta_1^* \coloneqq \sup\{\|Z_{jk}(x)\|_{\psi_1} \colon k \in \{1,\dots,d\},\ x \in \mathcal{X},\ j \in \{1,2,\dots\}\} < \infty. \tag{13}$$

Assumption 4 ensures that for each random variable $Z_{jk}(x)$, the moment generating function of $|Z_{jk}(x)|$ exists and is finite in an open interval around zero. Further, it ensures that we do not encounter any difficulties due to considering the supremum of the norm over the objectives or feasible space. We use Assumption 4 when we employ concentration inequalities to bound the error in Section 6. Specifically, under Assumptions 3 and 4, Bernstein's inequality implies that given $k \in \{1,\dots,d\}$ and $x \in \mathcal{X}$, for every $\delta > 0$,

$$\mathrm{P}\big\{\big|\tfrac{1}{n}\textstyle\sum_{j=1}^n Z_{jk}(x)\big| \geq \delta\big\} \leq 2\exp\big(-n\kappa_1 \min\big\{\delta^2/\|Z_{1k}(x)\|_{\psi_1}^2, \delta/\|Z_{1k}(x)\|_{\psi_1}\big\}\big) \tag{14}$$

where $\kappa_1 > 0$ is an absolute constant [Vershynin, 2018, p. 34]. Assumption 4 is mild in the sense that the class of sub-exponential distributions is quite large. It encompasses all sub-Gaussian distributions, which includes normal, Bernoulli, and all bounded distributions. It also includes exponential, Laplace, and Poisson, and the distributions of the square of sub-Gaussian random variables. In particular, a random variable X is sub-Gaussian if and only if X^2 is sub-exponential [Vershynin, 2018, p. 31]. Finally, we remark that the inequality in (14) and our application of it in Section 5 does not require the stochastic errors to have the same distributional family across the objectives k or across the values of $x \in \mathcal{X}$.

If it is known that the errors belong to the smaller class of sub-Gaussian random variables — for example, the errors are assumed to be iid mean-zero random variables with bounded support, or iid mean-zero normals with variance $\sigma_k^2(x)$ for each $x \in \mathcal{X}$ — then we can apply Hoeffding's inequality instead of (14) to obtain a tighter bound, as follows. First, for a sub-Gaussian random variable Y, let the sub-Gaussian norm of Y be

$$\|Y\|_{\psi_2} \coloneqq \inf\{s > 0 \colon \mathrm{E}[\exp(Y^2/s^2)] \leq 2\} \tag{15}$$

where we note that Y^2 is sub-exponential and the two norms are related by $\|Y^2\|_{\psi_1} = \|Y\|_{\psi_2}^2$ [Vershynin, 2018, p. 24, 31]. Thus, for sub-Gaussian errors under Assumption 4,

$$\eta_2^* \coloneqq \sup\{\|Z_{jk}(x)\|_{\psi_2} \colon k \in \{1,\dots,d\},\ x \in \mathcal{X},\ j \in \{1,2,\dots\}\} < \infty. \tag{16}$$

Given $k \in \{1,\dots,d\}$, $x \in \mathcal{X}$, if the errors $Z_{jk}(x), j = 1,\dots,n$ are iid from a sub-Gaussian distribution and Assumption 3 holds, then for every $\delta > 0$,

$$\mathrm{P}\big\{|\tfrac{1}{n}\textstyle\sum_{j=1}^n Z_{jk}(x)| \geq \delta\big\} \leq 2\exp\big(-n\kappa_2\big(\delta^2/\|Z_{1k}(x)\|_{\psi_2}^2\big)\big) \tag{17}$$

where $\kappa_2 > 0$ is an absolute constant [Vershynin, 2018, p. 27].

In what follows, we first make the more general assumption of sub-exponential errors via Assumption 4. Then, we provide explicit modifications to the results in case all objectives have sub-Gaussian error. If at least one objective has sub-exponential error and another has sub-Gaussian error, then

the largest error dominates the bound, and we should employ (14). Appendix A contains examples of how to compute the relevant constants κ_1, η_1^* and κ_2, η_2^*. For example, if all objectives have Laplace error, then $\kappa_1 = 1/2$. If all objectives have normal error, then $\kappa_2 = 4/3$.

5 Upper bounds on the optimality gap

Recall that our performance indicator for the optimality gap, $G_m(b)$ defined in (3), is the maximum of the modified coverage error and a distance term that measures the error due to false inclusion. We separately analyze the discretization error, false exclusion error, and false inclusion error terms from (4) to determine how they change as a function of the chosen discretization $\mathcal{X}_m$ and the total sampling budget b. (Under Assumption 1, Lemma 1 implies that the requirements of Theorem 1 are satisfied, which enables us to analyze these errors separately.) Then, we combine the results to obtain probabilistic upper bounds on the optimality gap.

5.1 Upper bounds on the error terms

First, we analyze the discretization error term in (4), ultimately yielding the result that $\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow) = O(t)$, where t is the dispersion of the discretized feasible set $\mathcal{X}_m$ from equation (11). We formally state the upper bound in Lemma 3; this result relies on Lemma 2, which relates the distance between two sets to the distance between their upper images.

Lemma 2. *Let $\mathcal{Y}, \mathcal{A} \subset \mathbb{R}^d$ be nonempty sets, and let $\mathcal{Y}$ be compact. If $\mathbb{D}_p(\mathcal{Y}_{\mathrm{N}}, \mathcal{A}) \leq \alpha$, then $\mathbb{D}_p(\mathcal{Y}_{\mathrm{N}}^\uparrow, \mathcal{A}^\uparrow) \leq \alpha d^{1/p}$.*

Proof. Since $\mathcal{Y}$ is nonempty and compact, $\mathcal{Y}_{\mathrm{N}}$ is nonempty. Then since $\emptyset \neq \mathcal{A} \subseteq \mathcal{A}^\uparrow$, we have $\mathbb{D}_p(\mathcal{Y}_{\mathrm{N}}, \mathcal{A}^\uparrow) \leq \mathbb{D}_p(\mathcal{Y}_{\mathrm{N}}, \mathcal{A}) \leq \alpha$, which implies that for any $y^* \in \mathcal{Y}_{\mathrm{N}} \subseteq \mathrm{bd}(\mathcal{Y})$, there exists a point $x^* \in \mathcal{A}^\uparrow$ such that $\|y^* - x^*\|_p \leq \alpha$. Using inequalities for the p-norm (see Subsection 2.1),

$$\begin{aligned}
\mathbb{D}_p(y^* + \mathbb{R}^d, \mathcal{A}^\uparrow) &\leq \mathbb{D}_p(y^* + \mathbb{R}^d, x^* + \mathbb{R}^d) \\
&\leq d^{1/p}\mathbb{D}_\infty(y^* + \mathbb{R}^d, x^* + \mathbb{R}^d) = d^{1/p}\mathbb{D}_\infty(y^*, x^*) \\
&\leq d^{1/p}\mathbb{D}_p(y^*, x^*) \leq \alpha d^{1/p}.
\end{aligned}$$

Since this inequality holds for any $y^* \in \mathcal{Y}_{\mathrm{N}}$ and $\mathcal{Y}_{\mathrm{N}}^\uparrow = \cup_{y^* \in \mathcal{Y}_{\mathrm{N}}}(y^* + \mathbb{R}^d_{\geqq})$, it follows that

$$\mathbb{D}_p(\mathcal{Y}_{\mathrm{N}}^\uparrow, \mathcal{A}^\uparrow) = \sup_{y \in \mathcal{Y}_{\mathrm{N}}^\uparrow} \inf_{x \in \mathcal{A}^\uparrow} \|y - x\|_p = \sup_{y^* \in \mathcal{Y}_{\mathrm{N}}} \left(\sup_{y \in (y^* + \mathbb{R}^d_{\geqq})} \inf_{x \in \mathcal{A}^\uparrow} \|x - y\| \right) \leq \alpha d^{1/p}. \quad \square$$

Lemma 3. *Under Assumption 1, for each $\mathcal{X}_m \subset \mathcal{X}$, $\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow) \leq \ell t d^{1/p}$.*

Proof. Since $\mathcal{P}_m \subseteq f(\mathcal{X}_m)$ is externally stable by Lemma 1, that is, $f(\mathcal{X}_m) \subset \mathcal{P}_m^\uparrow$, it follows that $\mathcal{P}_m^\uparrow = f(\mathcal{X}_m)^\uparrow$. Thus, we consider $\mathbb{D}_p(\mathcal{P}^\uparrow, f(\mathcal{X}_m)^\uparrow) = \mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$. We first show that $\mathbb{D}_p(\mathcal{P}, f(\mathcal{X}_m)) \leq \ell t$, then we invoke Lemma 2 to yield the result for $\mathbb{D}_p(\mathcal{P}^\uparrow, f(\mathcal{X}_m)^\uparrow)$.

First, under the ℓ-Lipschitz continuity of f in Assumption 1 and the dispersion in (11), for each $y^* \in \mathcal{P}$, there exists another point $\tilde{y} \in f(\mathcal{X}_m)$ such that $\|y^* - \tilde{y}\|_p \leq \ell t$ [Niederreiter, 1992, p. 149]. Thus, $\mathbb{D}_p(\mathcal{P}, f(\mathcal{X}_m)) \leq \ell t$. Then we can apply Lemma 2 to the nonempty sets $f(\mathcal{X}), f(\mathcal{X}_m)$

where $f(\mathcal{X})$ is compact by Lemma 1 and $\mathcal{P} = f(\mathcal{X})_{\mathrm{N}}$ by definition. Thus, $\mathbb{D}_p(f(\mathcal{X})_{\mathrm{N}}, f(\mathcal{X}_m)) = \mathbb{D}_p(\mathcal{P}, f(\mathcal{X}_m)) \leq \ell t$ implies the result under Lemma 2. □

Next, we turn our attention to determining an upper bound on the selection error term from (4) for a given discretization $\mathcal{X}_m$. Unlike Hunter and Ondes [2023], who construct an upper bound on the modified coverage error by analyzing the selection error in a pointwise fashion and then applying Lemma 2, we instead construct a tighter upper bound in a "surface-wise" fashion in Lemma 4, which is key to all following results. Specifically, Lemma 4 Part 1a traps the points in the set $f(\hat{\mathcal{E}}_m(b))$ within a piecewise linear "delta tube" of the discretized Pareto set. This result forms the linchpin of our analysis and is illustrated in Figure 3. Before we state the result, we require the following notation. Let

$$f_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b)) \coloneqq \{f(\tilde{x}') \in f(\hat{\mathcal{E}}_m(b)) \colon \exists\, f(\tilde{x}^*) \in \mathcal{P}_m \text{ such that } \bar{F}(\tilde{x}', n) \leqslant \bar{F}(\tilde{x}^*, n)\}$$

be the subset of points in $f(\hat{\mathcal{E}}_m(b))$ that falsely exclude a discretized Pareto point, and let

$$f^*_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b)) \coloneqq \big(f(\hat{\mathcal{E}}_m(b)) \cap \mathcal{P}_m\big) \cup f_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b)) \subseteq f(\hat{\mathcal{E}}_m(b))$$

be the set that also includes the correctly classified discretized Pareto points.

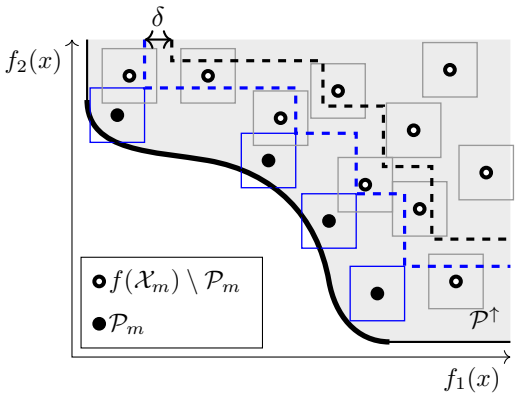
Lemma 4. *Let Assumption 1 hold, and let $\delta > 0$. If we have $|\bar{F}_k(\tilde{x}, n) - f_k(\tilde{x})| < \delta$ for all objectives $k = 1, \ldots, d$ and all points $\tilde{x} \in \mathcal{X}_m$, then w.p.1:*

1. *For false exclusion error,*
   (a) $f(\hat{\mathcal{U}}_m(b)) \subseteq f(\hat{\mathcal{E}}_m(b)) \subseteq \mathcal{P}_m^{\uparrow} \cap \big((\mathcal{P}_m + 2\delta 1_d)^{\uparrow}\big)^c$,
   (b) $f^*_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b)) \subseteq \mathcal{P}_m^{\uparrow} \cap (\mathcal{P}_m + 2\delta 1_d)^{\downarrow}$,
   (c) $D_m^{\mathrm{FE}}(b) = \mathbb{D}_p(\mathcal{P}_m^{\uparrow}, f(\hat{\mathcal{E}}_m(b))^{\uparrow}) \leq 2\delta d^{1/p}$;
2. *For false inclusion error,*
   (a) $(\mathcal{P}_m + 2\delta 1_d)^{\uparrow} \subseteq \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}$,
   (b) $D_m^{\mathrm{FI}}(b) = \mathbb{D}_p\big(\mathcal{P}_m^{\uparrow}, \mathrm{cl}\left(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\right) \cap \mathcal{P}_m^{\uparrow}\big) \leq 2\delta d^{1/p}$.
3. *The selection error term has upper bound* $\max\left\{D_m^{\mathrm{FE}}(b),\, D_m^{\mathrm{FI}}(b)\right\} \leq 2\delta d^{1/p}$.

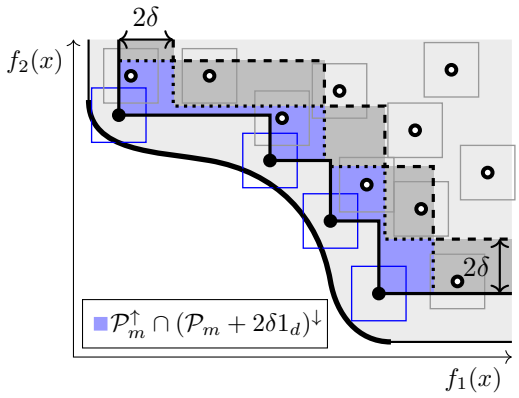
Proof. For Part 1a, we show that if $f(\tilde{x}') \in (\mathcal{P}_m + 2\delta 1_d)^{\uparrow}$ for some $\tilde{x}' \in \mathcal{X}_m$, then $f(\tilde{x}') \notin f(\hat{\mathcal{E}}_m(b))$, which implies $f(\hat{\mathcal{E}}_m(b)) \subseteq \big((\mathcal{P}_m + 2\delta 1_d)^{\uparrow}\big)^c$; that $f(\hat{\mathcal{U}}_m(b)) \subseteq f(\hat{\mathcal{E}}_m(b)) \subseteq \mathcal{P}_m^{\uparrow}$ follows from (2) and Lemma 1, respectively. For a contradiction, suppose there exists $\tilde{x}' \in \hat{\mathcal{E}}_m(b)$ which is estimated as efficient, but such that its true image is "far away" from $\mathcal{P}_m$ in the sense that $f(\tilde{x}') \in (\mathcal{P}_m + 2\delta 1_d)^{\uparrow}$, as shown in Figure 3(a). Then, there must exist $f(\tilde{x}^*) \in \mathcal{P}_m$ such that $f(\tilde{x}^*) + 2\delta 1_d \leqq f(\tilde{x}')$, which implies $f(\tilde{x}^*) + \delta 1_d \leqq f(\tilde{x}') - \delta 1_d$. Since every point is estimated within a 2δ-box of its true value, we have $\bar{F}(\tilde{x}^*, n) \leqslant f(\tilde{x}^*) + \delta 1_d \leqq f(\tilde{x}') - \delta 1_d \leqslant \bar{F}(\tilde{x}', n)$. Since $\bar{F}(\tilde{x}^*, n) \leqslant \bar{F}(\tilde{x}', n)$, then $\tilde{x}' \notin \hat{\mathcal{E}}_m(b)$, which is a contradiction. The region $\mathcal{P}_m^{\uparrow} \cap \big((\mathcal{P}_m + 2\delta 1_d)^{\uparrow}\big)^c$ is shown in Figure 3(b).

For Part 1b, for every $f(\tilde{x}') \in f_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))$, let $f(\tilde{x}^*) \in \mathcal{P}_m$ be the point it falsely excludes. Then $f(\tilde{x}') \leqq \bar{F}(\tilde{x}', n) + \delta 1_d \leqslant \bar{F}(\tilde{x}^*, n) + \delta 1_d \leqq f(\tilde{x}^*) + 2\delta 1_d$, which implies

$$f(\tilde{x}') \leqslant f(\tilde{x}^*) + 2\delta 1_d. \tag{18}$$



(a) Points in $\hat{\mathcal{P}}_m(b)$ are below $\mathrm{bd}((\mathcal{P}_m + \delta 1_d)^{\uparrow})$ (lower dashed line). Points above $\mathrm{bd}((\mathcal{P}_m + 2\delta 1_d)^{\uparrow})$ (upper dashed line) are estimated above $\mathrm{bd}((\mathcal{P}_m + \delta 1_d)^{\uparrow})$.



(b) Points in $f^*_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))$ belong to the false exclusion region, $\mathcal{P}_m^{\uparrow} \cap (\mathcal{P}_m + 2\delta 1_d)^{\downarrow}$ (shaded blue, below dotted line). The false inclusion only region is its complement within $\mathcal{P}_m^{\uparrow} \cap \left((\mathcal{P}_m + 2\delta 1_d)^{\uparrow}\right)^c$ (the false inclusion only region is shaded dark gray, between dotted and dashed lines).

Figure 3: The figures illustrate the results in Lemma 4 for an example problem in $d = 2$ objectives where for $\delta > 0$, $|\bar{F}_k(\tilde{x}, n) - f_k(\tilde{x})| < \delta$ for all objectives $k = 1, 2$ and all points $\tilde{x} \in \mathcal{X}_m$. The 2δ-boxes are indicated by squares around each point in the figure.

Recalling that $\mathcal{S}^{\downarrow} := \mathcal{S} - \mathbb{R}^d_{\geqq}$ is the lower image of a set $\mathcal{S} \subseteq \mathbb{R}^d$, (18) implies

$$f(\tilde{x}') \in (f(\tilde{x}^*) + 2\delta 1_d) - \mathbb{R}^d_{\geqq} = (f(\tilde{x}^*) + 2\delta 1_d)^{\downarrow} \subseteq (\mathcal{P}_m + 2\delta 1_d)^{\downarrow}. \tag{19}$$

Using Part 1a and (19), $f(\tilde{x}') \in f(\hat{\mathcal{E}}_m(b))$ implies $f(\tilde{x}') \in \mathcal{P}_m^{\uparrow} \cap \left((\mathcal{P}_m + 2\delta 1_d)^{\uparrow}\right)^c \cap (\mathcal{P}_m + \delta 1_d)^{\downarrow}$. Since $(\mathcal{P}_m + 2\delta 1_d)^{\downarrow} \subseteq \left((\mathcal{P}_m + 2\delta 1_d)^{\uparrow}\right)^c$, we have $f(\tilde{x}') \in \mathcal{P}_m^{\uparrow} \cap (\mathcal{P}_m + 2\delta 1_d)^{\downarrow}$. Since this is true for every $f(\tilde{x}') \in f_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))$, then $f_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b)) \subseteq \mathcal{P}_m^{\uparrow} \cap (\mathcal{P}_m + 2\delta 1_d)^{\downarrow}$. Since $\left(f(\hat{\mathcal{E}}_m(b)) \cap \mathcal{P}_m\right) \subseteq \mathcal{P}_m^{\uparrow} \cap (\mathcal{P}_m + 2\delta 1_d)^{\downarrow}$ by definition, it follows that $f^*_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b)) \subseteq \mathcal{P}_m^{\uparrow} \cap (\mathcal{P}_m + 2\delta 1_d)^{\downarrow}$. This region is the false exclusion (FE) region, shown as the lower blue shaded region in Figure 3(b).

For Part 1c, if $\mathbb{D}_p\big(\mathcal{P}_m^{\uparrow}, f(\hat{\mathcal{E}}_m(b))^{\uparrow}\big) > 0$, then $\mathcal{P}_m \setminus f(\hat{\mathcal{E}}_m(b)) \neq \emptyset$ and a false exclusion event occurred. For every falsely excluded point $f(\tilde{x}^*) \in \mathcal{P}_m \setminus f(\hat{\mathcal{E}}_m(b))$, $\tilde{x}^* \in \mathcal{E}_m$, there must exist another point $f(\tilde{x}') \in f(\hat{\mathcal{E}}_m(b))$ such that $\tilde{x}' \in \hat{\mathcal{E}}_m(b)$ and $\bar{F}(\tilde{x}', n) \leqslant \bar{F}(\tilde{x}^*, n)$, resulting in the false exclusion event. Therefore, $f^*_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b)) \neq \emptyset$. Now, the result holds if we show that $(\mathcal{P}_m + 2\delta 1_d)^{\uparrow} \subseteq f^*_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))^{\uparrow}$, because this result and $f^*_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))^{\uparrow} \subseteq f(\hat{\mathcal{E}}_m(b))^{\uparrow}$ imply

$$\mathbb{D}_p(\mathcal{P}_m^{\uparrow}, f(\hat{\mathcal{E}}_m(b))^{\uparrow}) \leq \mathbb{D}_p(\mathcal{P}_m^{\uparrow}, f^*_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))^{\uparrow}) \leq \mathbb{D}(\mathcal{P}_m^{\uparrow}, (\mathcal{P}_m + 2\delta 1_d)^{\uparrow}) \leq 2\delta d^{1/p}.$$

To see that $(\mathcal{P}_m + 2\delta 1_d)^{\uparrow} \subseteq f^*_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))^{\uparrow}$, first, for every $f(\tilde{x}') \in f_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))$, there exists $f(\tilde{x}^*) \in \mathcal{P}_m$ such that (18) implies $f(\tilde{x}^*) + 2\delta 1_d \subseteq f(\tilde{x}') + \mathbb{R}^d_{\geqq}$ and

$$(f(\tilde{x}^*) + 2\delta 1_d) + \mathbb{R}^d_{\geqq} \subseteq f(\tilde{x}') + \mathbb{R}^d_{\geqq} \subseteq f_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))^{\uparrow} \subseteq f^*_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))^{\uparrow}. \tag{20}$$

Since every point in $\mathcal{P}_m$ is either correctly included or falsely excluded, then it either belongs to $f^*_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))$ due to correct inclusion, or (20) applies for some $f(\tilde{x}') \in f_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))$. Then, it must be the case that $(\mathcal{P}_m + 2\delta 1_d)^{\uparrow} \subseteq f^*_{\mathrm{FE}}(\hat{\mathcal{E}}_m(b))^{\uparrow}$.

For Part 2a, we first show that $(\mathcal{P}_m + 2\delta 1_d)^{\uparrow} \subseteq \big(f(\hat{\mathcal{U}}_m(b))^{\downarrow}\big)^c$, which is equivalent to $f(\hat{\mathcal{U}}_m(b))^{\downarrow} \subseteq \big((\mathcal{P}_m + 2\delta 1_d)^{\uparrow}\big)^c$. By Part 1a, we have that $f(\hat{\mathcal{U}}_m(b)) \subseteq \big((\mathcal{P}_m + 2\delta 1_d)^{\uparrow}\big)^c$, which implies $f(\hat{\mathcal{U}}_m(b))^{\downarrow} \subseteq \big((\mathcal{P}_m + 2\delta 1_d)^{\uparrow}\big)^c$; see Figure 3. Now since $(\mathcal{P}_m + 2\delta 1_d)^{\uparrow} \subseteq \big(f(\hat{\mathcal{U}}_m(b))^{\downarrow}\big)^c$, we have $(\mathcal{P}_m + 2\delta 1_d)^{\uparrow} = (\mathcal{P}_m + 2\delta 1_d)^{\uparrow} \cap \mathcal{P}_m^{\uparrow} \subseteq \big(f(\hat{\mathcal{U}}_m(b))^{\downarrow}\big)^c \cap \mathcal{P}_m^{\uparrow}$.

For Parts 2b and 3, first, $\mathbb{D}_p\big(\mathcal{P}_m^{\uparrow}, \mathrm{cl}\,\big(f(\hat{\mathcal{U}}_m(b))^{\downarrow c}\big) \cap \mathcal{P}_m^{\uparrow}\big) \leq \mathbb{D}_p(\mathcal{P}_m^{\uparrow}, (\mathcal{P}_m + 2\delta 1_d)^{\uparrow}) \leq 2\delta d^{1/p}$ follows from Part 2a. Then, Part 3 follows from Parts 1c and 2b. □

Having bounded the discretization and selection error terms in Lemmas 3 and 4, we now use the inequality in (4) to combine the upper bounds into an upper bound on the performance indicator $G_m(b)$ from (3) in the following Proposition 1.

Proposition 1. *Let Assumption 1 hold, and let $\delta > 0$. If $|\bar{F}_k(\tilde{x}, n) - f_k(\tilde{x})| \leq \delta$ for all $k = 1, \ldots, d$ and all discretization points $\tilde{x} \in \mathcal{X}_m$, then $G_m(b) \leq (\ell t + 2\delta)d^{1/p}$ w.p.1.*

5.2 Probabilistic upper bound on the optimality gap

Since the selection error term in Lemma 4 is random, in the following Lemma 5, we consider upper bounds on the probability that this quantity exceeds a threshold value under Assumptions 3 and 4, which govern the nature of the stochastic error.

Lemma 5. *Let Assumption 1 hold, and let the stochastic error be governed by Assumptions 3 and 4, where the positive constant $\eta_1^* < \infty$ is defined in (13). Then there exists an absolute constant $\kappa_1 > 0$ such that for all $n \geq 1$ and all $\delta > 0$,*

$$\mathrm{P}\big\{\max\big\{G_m^{\mathrm{FE}}(b),\, G_m^{\mathrm{FI}}(b)\big\} > 2\delta d^{1/p}\big\} \leq 2dm \exp\big(-n\kappa_1 \min\big\{(\delta/\eta_1^*)^2, \delta/\eta_1^*\big\}\big). \tag{21}$$

If all errors in Assumption 4 are sub-Gaussian, then the right side of (21) can be modified to $2dm \exp\big(-n\kappa_2(\delta/\eta_2^*)^2\big)$ where $\kappa_2 > 0$ is an absolute constant and η_2^* is defined in (16).

Proof. Let $n \geq 1$ and $\delta > 0$. The contrapositive of Lemma 4 Part 3 states that if the selection error term is larger than $2\delta d^{1/p}$, then there exists $\tilde{x}_i \in \mathcal{X}_m$ and an objective $k \in \{1, \ldots, d\}$ such that $|\bar{F}_k(\tilde{x}_i, n) - f_k(\tilde{x}_i)| \geq \delta$. Then

$$\begin{aligned}
&\mathrm{P}\big\{\max\big\{G_m^{\mathrm{FE}}(b),\, G_m^{\mathrm{FI}}(b)\big\} > 2\delta d^{1/p}\big\} \leq \mathrm{P}\big\{|\bar{F}_k(\tilde{x}_i, n) - f_k(\tilde{x}_i)| \geq \delta \text{ for some } i, k\big\} \\
&\leq \textstyle\sum_{k=1}^{d}\sum_{i=1}^{m} \mathrm{P}\{|\bar{F}_k(\tilde{x}_i, n) - f_k(\tilde{x}_i)| \geq \delta\} = \sum_{k=1}^{d}\sum_{i=1}^{m} \mathrm{P}\{|\tfrac{1}{n}\textstyle\sum_{j=1}^{n} Z_{jk}(\tilde{x}_i)| \geq \delta\} \\
&\leq \textstyle\sum_{k=1}^{d}\sum_{i=1}^{m} 2\exp(-n\kappa_1 \min\{\delta^2/\|Z_{1k}(\tilde{x}_i)\|_{\psi_1}^2, \delta/\|Z_{1k}(\tilde{x}_i)\|_{\psi_1}\}) \qquad (22) \\
&\leq 2dm \exp\big(-n\kappa_1 \min\big\{(\delta/\eta_1^*)^2, \delta/\eta_1^*\big\}\big), \qquad (23)
\end{aligned}$$

where (22) applies Bernstein's inequality in (14) and (23) uses the definition of η_1^* in (13).

If all errors belong to the class of sub-Gaussian random variables, then Hoeffding's inequality, (17), can be applied to (22) instead of Bernstein's inequality, yielding the modified right side of (21). □

Next, we use Lemma 5 and Assumption 2, which governs the discretization quality, to prove the main probabilistic bound on the performance indicator stated in Theorem 2.

Theorem 2. *Let $\tilde{\varepsilon} > 0$. For all $m \geq (1+\tilde{\varepsilon})^q$, let Assumptions 1 and 2 hold, and let the stochastic error be governed by Assumptions 3 and 4. Then for all $m \geq (1+\tilde{\varepsilon})^q$, all $n \geq 1$, and all $\delta > 0$,*

$$\mathrm{P}\left\{G_m(b) > \left(\frac{\tilde{c}_q \ell}{\lfloor m^{1/q} \rfloor} + 2\delta\right) d^{1/p}\right\} \leq 2dm \exp\left(-n\kappa_1 \min\left\{\frac{\delta^2}{\eta_1^{*2}}, \frac{\delta}{\eta_1^*}\right\}\right) \tag{24}$$

where constant $\tilde{c}_q$ is from Assumption 2, constants ℓ, d, p are defined as in Proposition 1, and η_1^*, κ_1 are from (13), (14), and Lemma 5. If all errors in Assumption 4 are sub-Gaussian, then the right side of (24) becomes $2dm \exp\left(-n\kappa_2(\delta/\eta_2^*)^2\right)$ where η_2^*, κ_2 are from (16), (17), and Lemma 5.

Proof. Let $m \geq (1+\tilde{\varepsilon})^q$, $n \geq 1$, and $\delta > 0$. Then $\lfloor m^{1/q} \rfloor \geq 1$ and by Assumption 2, $t\ell \leq \tilde{c}_q \ell (\lfloor m^{1/q} \rfloor)^{-1}$. Then using Theorem 1, Proposition 1, and Lemma 5,

$$\begin{aligned}
\mathrm{P}\left\{G_m(b) > \left(\tilde{c}_q \ell (\lfloor m^{1/q} \rfloor)^{-1} + 2\delta\right) d^{1/p}\right\} &\leq \mathrm{P}\left\{G_m(b) > (\ell t + 2\delta) d^{1/p}\right\} \\
&\leq \mathrm{P}\left\{\ell t d^{1/p} + \max\{G_m^{\mathrm{FE}}(b),\, G_m^{\mathrm{FI}}(b)\} > (\ell t + 2\delta) d^{1/p}\right\} \\
&\leq 2dm \exp\left(-n\kappa_1 \min\left\{(\delta/\eta_1^*)^2, \delta/\eta_1^*\right\}\right).
\end{aligned}$$

Updating the right side of (24) for the sub-Gaussian case follows from Lemma 5. □

6 Balancing sampled points with simulation replications

We now use Theorem 2 to conduct a tradeoff analysis on the number of points to sample in the decision space m versus the number of replications to obtain per point n. Specifically, in Theorem 3, for user-provided values of τ and α, we derive values of m, n, and b such that the optimality gap converges to zero at a fast rate and $\mathrm{P}\{G_m(b) > \tau\} \leq \alpha$ holds. We also derive a good tradeoff of m and n for a given b in Theorem 4.

To begin, recall that b denotes the entire simulation budget across all points, where the sample size at each point is $n = \lfloor b/m \rfloor$. To balance the discretization and selection errors in (24) by ensuring they go to zero at the same rate, as in Yakowitz et al. [2000], we choose $\delta = \Theta(1/\lfloor m^{1/q} \rfloor)$ in the following Lemma 6.

Lemma 6. *Let $\tilde{\varepsilon} > 0$, and let $g\colon \mathbb{Z}_> \to \mathbb{R}_>$ be a positive-valued monotone increasing function of the total simulation budget $b \geq 2$ such that*

$$\lim_{b\to\infty} \left(\frac{\ln g(b)}{b}\right)^{1/(q+2)} = 0 \quad \text{and} \quad \lim_{b\to\infty} \frac{1}{g(b)} \left(\frac{b}{\ln g(b)}\right)^{q/(q+2)} = 0, \tag{25}$$

where $\mathbb{Z}_> := \{1, 2, \ldots\}$ is the set of all positive integers and, in the simplest case, $g(b) = b$ satisfies (25). If Assumptions 1 and 2 hold, the stochastic error is governed by Assumptions 3 and 4, and b is large enough that we can choose

$$m = \left\lfloor \left(\frac{\kappa_1 b}{\ln g(b)}\right)^{q/(q+2)} \right\rfloor \geq (1+\tilde{\varepsilon})^q, \quad n = \left\lfloor \frac{b}{m} \right\rfloor \geq 1, \tag{26}$$

then for constants $\tilde{c}_q, \ell, d, \eta_1^*, \kappa_1$ *as specified in Theorem 2,*

$$\mathrm{P}\left\{G_m(b) > \frac{(\tilde{c}_q\ell + 2\eta_1^*)d^{1/p}}{\left\lfloor \lfloor(\kappa_1 b/\ln g(b))^{q/(q+2)}\rfloor^{1/q}\right\rfloor}\right\} \leq \frac{2d}{g(b)}\left\lfloor\left(\frac{\kappa_1 b}{\ln g(b)}\right)^{q/(q+2)}\right\rfloor. \tag{27}$$

If all errors in Assumption 4 are sub-Gaussian, then η_1^* *and* κ_1 *can be replaced by* η_2^* *and* κ_2*, respectively, in (26) and (27).*

Proof. To ensure the discretization and selection error terms decay to zero at the same rate, choose $\delta = \eta_1^*/\lfloor m^{1/q}\rfloor$ in Theorem 2. Then (24), $n = \lfloor b/m\rfloor$, and $\lfloor m^{1/q}\rfloor \geq 1$ imply

$$\mathrm{P}\left\{G_m(b) > \left(\frac{\tilde{c}_q\ell + 2\eta_1^*}{\lfloor m^{1/q}\rfloor}\right)d^{1/p}\right\} \leq 2dm\exp\left(-n\frac{\kappa_1}{\lfloor m^{1/q}\rfloor^2}\right). \tag{28}$$

Now, we must ensure the right-hand side of (28) goes to zero, ideally at a fast rate. Let

$$n = \lfloor b/m\rfloor \geq \kappa_1^{-1}\lfloor m^{1/q}\rfloor^2 \ln g(b) \tag{29}$$

so that we have

$$\mathrm{P}\left\{G_m(b) > \left(\frac{\tilde{c}_q\ell + 2\eta_1^*}{\lfloor m^{1/q}\rfloor}\right)d^{1/p}\right\} \leq \frac{2d}{g(b)}m. \tag{30}$$

To get an upper bound on m in the right side of (30), notice that (29) and $b/m \geq n = \lfloor b/m\rfloor$ imply $m\lfloor m^{1/q}\rfloor^2 \leq \kappa_1 b/\ln g(b)$. Thus, given feasible b, ensure $m\lfloor m^{1/q}\rfloor^2 \leq m^{1+2/q} \leq \kappa_1 b/\ln g(b)$ by selecting m as the largest integer that satisfies the equation $m \leq (\kappa_1 b/\ln g(b))^{q/(q+2)}$, implying the choice of m in (26). The inequality in (27) results from plugging this value for m into (30). The result for sub-Gaussian errors follows by repeating the analysis and noticing that the same steps hold, except with updated constants. □

The result in Lemma 6 relies on the function g. The following Lemma 7 provides bounds on feasible choices for this function.

Lemma 7. *Given real constants* $\beta \in (0,1)$ *and* $\theta_1, \theta_2 > 0$*, any positive-valued monotone increasing function* g *satisfying*

$$g_{\mathrm{L}}(b;\theta_1) \coloneqq (\theta_1 b)^{q/(q+2)} \leq g(b) \leq \exp(\theta_2 b^\beta) =: g_{\mathrm{U}}(b;\theta_2,\beta) \text{ for all } b > 1 \tag{31}$$

also satisfies both conditions in (25).

Proof. Suppose $g(b)$ satisfies (31). Then applying conditions on the constants β, θ_1, θ_2,

$$\begin{aligned}
\lim_{b\to\infty}\left(\frac{q\ln(\theta_1 b)}{(q+2)b}\right)^{1/(q+2)} &\leq \lim_{b\to\infty}\left(\frac{\ln g(b)}{b}\right)^{1/(q+2)} \leq \lim_{b\to\infty}\left(\frac{\theta_2}{b^{1-\beta}}\right)^{1/(q+2)} = 0,\\
\lim_{b\to\infty}\frac{1}{\exp(\theta_2 b^\beta)}\left(\frac{b^{1-\beta}}{\theta_2}\right)^{q/(q+2)} &\leq \lim_{b\to\infty}\frac{1}{g(b)}\left(\frac{b}{\ln g(b)}\right)^{q/(q+2)} \leq \lim_{b\to\infty}\left(\frac{q+2}{q\ln(\theta_1 b)}\right)^{q/(q+2)} = 0,
\end{aligned}$$

which implies the result. □

There are at least two tractable ways to use Lemmas 6 and 7 to determine implementable values of b, m, and n: (i) specify $g(b)$ that satisfies Lemma 7 in advance, or (ii) determine values of b, m, and n that ensure $\mathrm{P}\{G_m(b) > \tau\} \leq \alpha$ for user-specified τ and α. In the second case, the choice of τ and α implicitly determine the function $g(b)$. We analyze the second case first in the following Theorem 3. Then, we use the insights gained from Theorem 3 to select a suitable $g(b)$ in Theorem 4.

Theorem 3. *Let the postulates and conditions of Lemma 6 hold, define $c_i \coloneqq \tilde{c}_q \ell + 2\eta_i^*, i = 1, 2$, and let $\tau > 0$, $\alpha \in (0, 1)$, and $\epsilon \in (0, 1)$ be given, fixed constants. Let*

$$b_{\alpha,\tau} \coloneqq \frac{1}{\kappa_1}\Big[\Big(\frac{c_1 d^{1/p}}{\tau} + 1\Big)^q + 1\Big]^{(q+2)/q} \ln\Big(\frac{2d}{\alpha}\Big[\Big(\frac{c_1 d^{1/p}}{\tau} + 1\Big)^q + 1\Big]\Big), \tag{32}$$

$$b_\alpha \coloneqq \min\Big\{b\colon b \geq b_{\alpha,\tau},\ \kappa_1 b^{1-\epsilon/q} - \frac{\epsilon}{q+2}\ln b \geq \ln\Big(\frac{2d}{\alpha}\Big)\Big\}, \tag{33}$$

$$b_\tau \coloneqq \min\Big\{b\colon b \geq \max\{b_{\alpha,\tau}, b_\alpha\},\ \frac{b}{\ln b} \geq \frac{q}{\kappa_1(q+2)}\Big[\Big(\frac{c_1 d^{1/p}}{\tau} + 1\Big)^q + 1\Big]^{(q+2)/q}\Big\}. \tag{34}$$

If $b^* = \lceil \max\{b_{\alpha,\tau}, b_\alpha, b_\tau\} \rceil$ and m and n are calculated as

$$m = \Big\lfloor \max\Big\{(b^*)^{\epsilon/(q+2)}, \Big(\frac{c_1 d^{1/p}}{\tau} + 1\Big)^q + 1\Big\}\Big\rfloor, \qquad n = \Big\lfloor \frac{b^*}{m} \Big\rfloor, \tag{35}$$

then using m and n from (35), $b = mn$ in Algorithm 1, and defining $\tau^* \coloneqq c_1 d^{1/p}/\lfloor m^{1/q} \rfloor$, implies

$$\mathrm{P}\big\{G_m(b) > \tau\big\} \leq \mathrm{P}\big\{G_m(b) > \tau^*\big\} = \mathrm{P}\Big\{G_m(b) > \frac{c_1 d^{1/p}}{\lfloor m^{1/q} \rfloor}\Big\} \leq \alpha. \tag{36}$$

If all errors in Assumption 4 are sub-Gaussian, then κ_1, c_1 can be replaced by κ_2, c_2 in the calculations of b^*, m, τ^*, and the guarantee in (36).

Proof. For any $\tau > 0$, define $\tilde{m}_\tau$ such that

$$\tilde{m}_\tau \coloneqq \Big(\frac{\kappa_1 b}{\ln g(b)}\Big)^{q/(q+2)}, \qquad \tau \geq \frac{c_1 d^{1/p}}{(\tilde{m}_\tau - 1)^{1/q} - 1}, \qquad \tilde{m}_\tau \geq \Big(\frac{c_1 d^{1/p}}{\tau} + 1\Big)^q + 1. \tag{37}$$

Using Lemma 6 and (37), it follows that

$$\begin{aligned} \mathrm{P}\big\{G_m(b) > \tau\big\} &\leq \mathrm{P}\Big\{G_m(b) > \frac{c_1 d^{1/p}}{(\tilde{m}_\tau - 1)^{1/q} - 1}\Big\} \leq \mathrm{P}\Big\{G_m(b) > \frac{c_1 d^{1/p}}{\lfloor \lfloor \tilde{m}_\tau \rfloor^{1/q} \rfloor}\Big\} \\ &\leq \frac{2d\,\tilde{m}_\tau}{g(b)} = 2d\,\tilde{m}_\tau \exp\Big(-\frac{\kappa_1 b}{\tilde{m}_\tau^{(q+2)/q}}\Big) \leq \alpha, \end{aligned} \tag{38}$$

where the inequalities leading to (38) shall ultimately imply the inequalities in (36). Solving for b on the right side of (38) implies

$$b \geq \kappa_1^{-1} \tilde{m}_\tau^{(q+2)/q} \ln\big(2d\,\alpha^{-1}\,\tilde{m}_\tau\big) \tag{39}$$

where ensuring b is small implies making $\tilde{m}_\tau$ as small as possible in (39).

Now, we must ensure that when choosing b and $\tilde{m}_\tau$, which implicitly determines $g(b)$ through

the definition of $\tilde{m}_\tau$, the conditions of Lemma 7 hold. Let us choose $\theta_1 = 1$, $\theta_2 = \kappa_1$, and $\beta = \frac{q-\epsilon}{q}$ in (31). Then, (31) and the definition of $\tilde{m}_\tau$ on the left side of (37) imply

$$\frac{q}{q+2}\ln(b) \leq \ln g(b) = \frac{\kappa_1 b}{\tilde{m}_\tau^{(q+2)/q}} \leq \kappa_1 b^{(q-\epsilon)/q}.$$

Rearranging terms above and applying the inequality on the right side of (37) yields

$$\max\left\{\left(\frac{c_1 d^{1/p}}{\tau}+1\right)^q + 1,\, b^{\epsilon/(q+2)}\right\} \leq \tilde{m}_\tau \leq \left(\frac{\kappa_1 b}{(q/(q+2))\ln b}\right)^{q/(q+2)}. \tag{40}$$

Since $\tilde{m}_\tau$ should be small according to (39), take $\tilde{m}_\tau$ from the left side of (40),

$$\tilde{m}_\tau = \max\{(c_1 d^{1/p}\tau^{-1}+1)^q + 1,\, b^{\epsilon/(q+2)}\}. \tag{41}$$

Then plugging the two possible values of $\tilde{m}_\tau$ from (41) into (39) implies that we require

$$b \geq \kappa_1^{-1}[(c_1 d^{1/p}\tau^{-1}+1)^q+1]^{(q+2)/q}\ln(2d\alpha^{-1}\left[(c_1 d^{1/p}\tau^{-1}+1)^q+1\right]),$$
$$b \geq \left(\kappa_1^{-1}\ln\left(2d\alpha^{-1}\, b^{\epsilon/(q+2)}\right)\right)^{q/(q-\epsilon)}.$$

We must also guarantee that the right side of (40) holds for any τ, which implies

$$b \geq \kappa_1^{-1}(q/(q+2))\ln b[(c_1 d^{1/p}\tau^{-1}+1)^q+1]^{(q+2)/q}.$$

These three inequalities and (41) imply the result. □

Theorem 3 identifies a lower bound on the simulation budget b and the allocation of the budget between m and n required for probabilistic guarantees on the optimality gap to hold. Specifically, Theorem 3 identifies three regimes for the budget and its allocation. The first regime, specified by the closed-form expression for $b_{\alpha,\tau}$, implicitly selects the best "middle" function $g(b)$ such that $g_{\mathrm{L}}(b;1) < g(b) < g_{\mathrm{U}}(b;\kappa_1,(q-\epsilon)/q)$ for $\epsilon \in (0,1)$ in (31). To see this fact, notice that selecting $\tilde{m}_\tau$ from the left side of (40), which makes $\tilde{m}_\tau$ small in (39), implies that we attempt to select a faster $g(b)$ in (31) subject to the inequality on $\tilde{m}_\tau$ in (37) by setting $\tilde{m}_\tau = (c_1 d^{1/p}\tau^{-1}+1)^q + 1$. If τ is in balance with α, that is, if τ, α, and $\epsilon \in (0,1)$ are chosen such that after calculating $b_{\alpha,\tau}$ we have

$$\tau \geq \frac{c_1 d^{1/p}}{\left(\left(\frac{\kappa_1 b_{\alpha,\tau}}{(q/(q+2))\ln b_{\alpha,\tau}}\right)^{q/(q+2)}-1\right)^{1/q}-1} \quad \text{and} \quad \alpha \geq \frac{2d}{\exp\left(\kappa_1 b_{\alpha,\tau}^{1-\epsilon/q} - \frac{\epsilon}{q+2}\ln b_{\alpha,\tau}\right)}, \tag{42}$$

then the resulting $b_{\alpha,\tau}$ will be the largest among $b_{\alpha,\tau}$, b_α, and b_τ. However, if α is too small relative to τ, the $g(b)$ implicitly chosen by $b_{\alpha,\tau}$ is too fast in Lemma 7. In such a case, the small α becomes binding and we require $g(b) = g_{\mathrm{U}}(b;\kappa_1,(q-\epsilon)/q)$ to yield the increased budget value $b_\alpha > b_{\alpha,\tau}$. Otherwise, if τ is too small relative to α, the $g(b)$ implicitly chosen by $b_{\alpha,\tau}$ is too slow in Lemma 7. Then, the small τ becomes binding such that we require $g(b) = g_{\mathrm{L}}(b;1)$ to yield the increased budget value $b_\tau > b_{\alpha,\tau}$. Since our bound regarding τ in (36) is somewhat conservative due to the equations in (37) that remove the double floor functions on the left side of (27), to ensure the tightest bound

possible, we report $\mathrm{P}\{G_m(b) > \tau^*\} \leq \alpha$ according to (36) in the numerical examples in Section 7, where the gap between τ and τ^* decreases as m increases.

The results in Theorem 3 lead to the following convergence rate results, presented in Corollary 1. Corollary 1 states how the required total simulation budget grows as various parameters tend to their respective limits. Importantly, Corollary 1 also states how the total simulation budget increases as the number of objectives increases.

Corollary 1. *Under the postulates of Theorem 3, if Algorithm 1 is implemented with m and n from (35) and $b = mn$, which implies $\mathrm{P}\{G_m(b) > \tau\} \leq \alpha$ for chosen $\tau > 0$, $\alpha \in (0,1)$, $\epsilon \in (0,1)$, then*

1. *For fixed τ and ϵ, $b = \Theta\big((\ln \alpha^{-1})^{q/(q-\epsilon)}\big)$ as $\alpha \to 0$;*
2. *For fixed α, ϵ and any $\delta > 0$, $b = \Omega\big(\tau^{-(q+2)}\big)$ and $b = O\big(\tau^{-(q+2)/(1-\delta)}\big)$ as $\tau \to 0$;*
3. *For fixed τ, α, ϵ, and holding q constant, if $G_m(b)$ is calculated under the p-norm, $1 \leq p < \infty$, then $b = \Theta\big(d^{(q+2)/p} \ln d\big)$ as $d \to \infty$. If $G_m(b)$ is calculated under the ∞-norm, then $b = \Theta\big((\ln d)^{q/(q-\epsilon)}\big)$ as $d \to \infty$.*

Proof. For Part 1, fix $\tau > 0$, $\epsilon \in (0,1)$. From (32), (33), and (34), $b^* = \lceil b_\alpha \rceil$ for small enough α. From (33), the right-side inequality is binding. For any $\delta \in (0, \kappa_1)$ and large enough b_α (due to small enough α),

$$\kappa_1 b_\alpha^{1-\epsilon/q} - \delta b_\alpha^{1-\epsilon/q} \leq \ln(2d\alpha^{-1}) = \kappa_1 b_\alpha^{1-\epsilon/q} - \ln(b_\alpha^{\epsilon/(q+2)}) \leq \kappa_1 b_\alpha^{1-\epsilon/q}. \tag{43}$$

For Part 2, fix $\alpha \in (0,1)$. From (32), (33), and (34), $b^* = \lceil b_\tau \rceil$ for small enough τ. From (34), the right-side inequality is binding so that for any $\delta > 0$ and large enough b_τ (due to small enough τ), we have $b_\tau^{1-\delta} \leq \frac{q}{\kappa_1(q+2)}[(\frac{c_1 d^{1/p}}{\tau} + 1)^q + 1]^{(q+2)/q} = \frac{b_\tau}{\ln b_\tau} \leq b_\tau$.

For Part 3, fix $\tau > 0$, $\alpha \in (0,1)$, $\epsilon \in (0,1)$, and q. First, let $G_m(b)$ be calculated under the p-norm for $1 \leq p < \infty$. From (32), (33), and (34), $b^* = \lceil b_{\alpha,\tau} \rceil$ for large enough d. Then, the result follows directly from the expression for $b_{\alpha,\tau}$ in (32). If $G_m(b)$ is calculated under the ∞-norm, then $b^* = \lceil b_\alpha \rceil$ for large enough d, and (43) implies the result (where large enough b_α is due to large enough d). □

Given the insights of Theorem 3, we now consider the implementable values of m and n that result from specifying $g(b)$ in advance, thus implicitly choosing α, τ for given b. Such values are useful for a user who does not know how to specify α and τ, or for whom the b resulting from Theorem 3 is impossibly large or exceeds some predetermined computing limit. We select $g(b) = b$ for the following two reasons: First, this selection prioritizes smaller τ over smaller α in (27) by selecting relatively larger m and smaller n. Our numerical experience in Section 7 indicates that the bound in (27) is quite conservative with respect to the right-side probability, and it may be especially conservative under common random numbers. Given this conservativeness, we choose to prioritize smaller τ over smaller α. Second, the selection of $g(b) = b$ has precedent, since it is the value chosen by Yakowitz et al. [2000] in the single-objective case. This choice leads to the following Theorem 4. We omit a proof because the result follows directly from Lemmas 6 and 7.

Theorem 4. *Let the postulates and conditions of Lemma 6 hold, where $\tilde{\varepsilon} > 0$. For given total simulation budget b, if*

$$m = \left\lfloor \left(\frac{\kappa_1 b}{\ln b} \right)^{q/(q+2)} \right\rfloor \geq (1+\tilde{\varepsilon})^q \quad \text{and} \quad n = \left\lfloor \frac{b}{m} \right\rfloor \geq 1, \tag{44}$$

then

$$\mathrm{P}\left\{ G_m(b) > \frac{c_1 d^{1/p}}{\lfloor m^{1/q} \rfloor} \right\} \leq \frac{2d}{n}. \tag{45}$$

If all errors in Assumption 4 are sub-Gaussian, then the constants c_1, κ_1 can be replaced by c_2, κ_2, respectively, in (44) and (45), where $c_i = \tilde{c}_q \ell + 2\eta_i^*, i = 1, 2$.

Theorem 4 provides an extremely simple rule for selecting m and n that ensures the optimality gap decays to zero at a fast rate. To ensure a meaningful probabilistic bound in (45), a user should ensure that the total budget b is large enough that the resulting number of simulation replications per point is larger than twice the number of objectives, $n > 2d$.

7 Numerical examples

We now demonstrate the results in Theorem 3 on several numerical examples. Our main objectives in this section are to demonstrate the convergence of the optimality gap $G_m(b)$ under the allocations in Theorem 3, to show how the allocations change as a function of problem type and dimension, and to test the bounds for conservativeness. Throughout, we calculate all distances in the decision space under the $r = 2$ norm and all distances in the objective space under the $p = \infty$ norm.

7.1 Nonlinear problems

First, we consider a series of problems in one, two, and three objectives inspired by the problem in Yakowitz et al. [2000, Ex. 5, p. 943]. This single-objective problem is

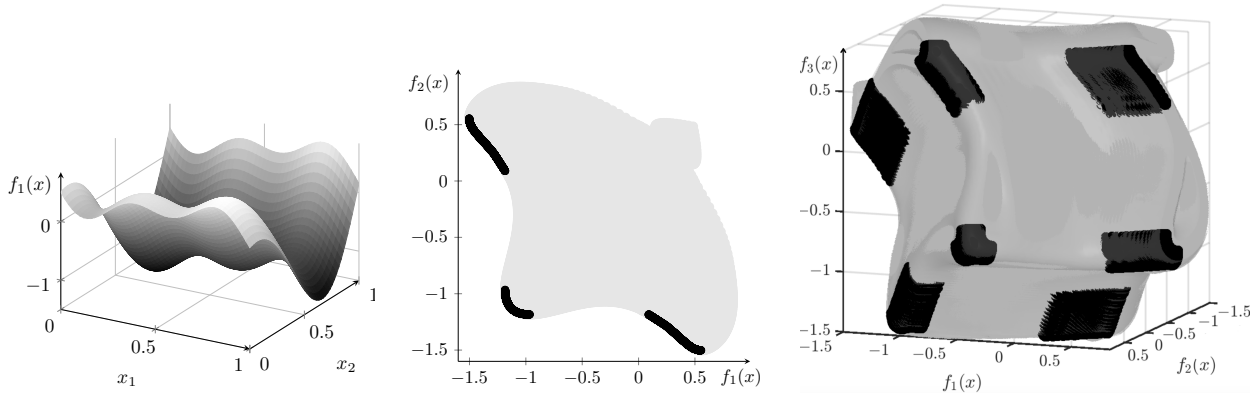
$$\begin{aligned} \text{minimize} \quad & f_1(x) = \mathrm{E}[(x_1 - .5)\sin(10x_1) + (x_2 + .5)\cos(5x_2) + Z_{11}(x)] \\ \text{s.t.} \quad & x \in \mathcal{X} = [0,1]^2 \end{aligned} \tag{N1}$$

where $Z_{11}(x)$ is a mean-zero random variable for all $x \in \mathcal{X}$. Figure 4(a) shows a plot of $f_1(x)$ for $x \in [0,1]^2$, where $x^* \approx (0.130640, 0.662397)$ is the solution and $f_1(x^*) \approx -1.502088$.

For the second problem, we rotate f_1 to yield the second objective f_2,

$$\begin{aligned} \text{minimize} \quad & f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = \begin{bmatrix} \mathrm{E}[(x_1 - .5)\sin(10x_1) + (x_2 + .5)\cos(5x_2) + Z_{11}(x)] \\ \mathrm{E}[(x_2 - .5)\sin(10x_2) + (x_1 + .5)\cos(5x_1) + Z_{12}(x)] \end{bmatrix} \\ \text{s.t.} \quad & x \in \mathcal{X} = [0,1]^2 \end{aligned} \tag{N2}$$

where $Z_{11}(x)$ and $Z_{12}(x)$ are mean-zero random variables for all $x \in \mathcal{X}$. Figure 4(b) shows a plot of the objective space, $(f_1(x), f_2(x))$ for $x \in [0,1]^2$, where the Pareto set is shown in black and dominated points are shown in gray. Thus, problems (N1) and (N2) are directly comparable in all



(a) $f_1(x)$, $x \in [0,1]^2$ (b) $(f_1(x), f_2(x))$, $x \in [0,1]^2$ (c) $(f_1(x), f_2(x), f_3(x))$, $x \in [0,1]^3$

Figure 4: The objective spaces of the nonlinear problems (N1), (N2), and (N3), respectively, with the global Pareto sets in (N2) and (N3) shown in black.

respects except that we increase the number of objectives by one: (N1) has $d = 1$ objective while (N2) has $d = 2$ objectives.

To add a third objective, we increase the dimension of the decision space to $q = 3$ and rotate the objectives to yield the tri-objective problem,

$$\begin{aligned} \text{minimize} \quad & f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{bmatrix} = \begin{bmatrix} \mathrm{E}[(x_1 - .5)\sin(10x_1) + (x_2 + .5)\cos(5x_2) + Z_{11}(x)] \\ \mathrm{E}[(x_2 - .5)\sin(10x_2) + (x_3 + .5)\cos(5x_3) + Z_{12}(x)] \\ \mathrm{E}[(x_3 - .5)\sin(10x_3) + (x_1 + .5)\cos(5x_1) + Z_{13}(x)] \end{bmatrix} \\ \text{s.t.} \quad & x \in \mathcal{X} = [0,1]^3 \end{aligned} \tag{N3}$$

where $Z_{11}(x)$, $Z_{12}(x)$, and $Z_{13}(x)$ are mean-zero random variables for all $x \in \mathcal{X}$. Thus, problems (N2) and (N3) are directly comparable except that we increase both the decision space and objective space dimensions by one: (N2) has $q = 2, d = 2$ while (N3) has $q = 3, d = 3$. Figure 4(c) shows a plot of the objective space, $(f_1(x), f_2(x), f_3(x))$ for $x \in [0,1]^3$, where the Pareto set is shown in black and dominated points are shown in gray.

We use Algorithm 1 and Theorem 3 to solve the corresponding discretized estimated problems ($\hat{\mathrm{M}}_m$) as follows. First, we test two error types:

E1. Laplace error, where $Z_{jk}(x), j = 1, \ldots, n$, $k = 1, \ldots, d$, $x \in \mathcal{X}$ are iid mean-zero Laplace with scale parameter $\beta_k(x) = 1/\sqrt{2}$ and variance $\sigma_k^2(x) = \sqrt{2}\beta(x) = 1$ for all $k = 1, \ldots, d$, $x \in \mathcal{X}$;

E2. Normal error, where $Z_{jk}(x), j = 1, \ldots, n$, $k = 1, \ldots, d$, $x \in \mathcal{X}$ are iid mean-zero normal with scale parameter $\sigma_k^2(x) = 1$ for all $k = 1, \ldots, d$, $x \in \mathcal{X}$.

Both error types imply independent sampling across all $x \in \mathcal{X}_m$ since using common random numbers across points would render problems (N1), (N2), and (N3) deterministic problems under Algorithm 1. Under each error type, the constants η_1^*, κ_1, and η_2^*, κ_2 required by Theorem 3 are calculated according to the examples in Appendix A with the resulting numerical values reported in Table 1. Second, to ensure comparable values of the errors over different choices of $\mathcal{X}_m$, we discretize on the grid from Example 1 and ensure our sequence of grids contains subsequences that

share common points, as described in Appendix B. For example, in problems with decision space dimension $q = 2$, we choose the grid sequence $m_1 = 4$, $m_2 = 16$, $m_3 = 36$, $m_4 = 144$, and $m_5 = 324$, where $\mathcal{X}_{m_1} \subset \mathcal{X}_{m_3} \subset \mathcal{X}_{m_5}$ and $\mathcal{X}_{m_2} \subset \mathcal{X}_{m_4}$. In problems with $q = 3$, we choose $m_1 = 8$, $m_2 = 64$, $m_3 = 216$, and $m_4 = 1728$, where $\mathcal{X}_{m_1} \subset \mathcal{X}_{m_3}$ and $\mathcal{X}_{m_2} \subset \mathcal{X}_{m_4}$. On the subsequences that share points, the discretization error $\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}^\uparrow_m)$ is monotone decreasing. Since our values of m are fixed in advance, we also fix $\alpha = 0.5$ and then determine τ as the smallest value that satisfies the conditions of Theorem 3 for the given m and α; together, these values imply b and n. After determining this smallest τ and the corresponding b, m, and n, we calculate τ^*.

Table 1: Constants for nonlinear problems; see Appendix A.

	Error	d	q	r	p	ϵ	$\tilde{c}_q$	ℓ	η_1^*	κ_1	c_1	η_2^*	κ_2	c_2
(N1)	Laplace	1	2	2	∞	0.5	$\sqrt{2}/2$	9	$\sqrt{2}$	0.5	9.1924	–	–	–
(N1)	Normal	1	2	2	∞	0.5	$\sqrt{2}/2$	9	–	–	–	$\sqrt{8/3}$	4/3	9.6299
(N2)	Laplace	2	2	2	∞	0.5	$\sqrt{2}/2$	9	$\sqrt{2}$	0.5	9.1924	–	–	–
(N2)	Normal	2	2	2	∞	0.5	$\sqrt{2}/2$	9	–	–	–	$\sqrt{8/3}$	4/3	9.6299
(N3)	Laplace	3	3	2	∞	0.5	$\sqrt{3}/2$	9	$\sqrt{2}$	0.5	10.6227	–	–	–
(N3)	Normal	3	3	2	∞	0.5	$\sqrt{3}/2$	9	–	–	–	$\sqrt{8/3}$	4/3	11.0602

For each combination of error type and problem under the b, m, and n calculated as described above, we conduct $s = 10\,000$ independent macro replications of Algorithm 1 on problems (N1), (N2), and (N3) using the constants from Table 1. Each independent macro replication results in an optimality gap, which we denote as $G_{m,j}(b)$ for $j = 1, \ldots, s$. Because the bounds always hold, we report the maximum optimality gap observed across the s independent macro replications of Algorithm 1 for comparison with τ^*. Further, we calculate the sample average of $G_m(b)$ and its standard error. These values are denoted by

$$G_m^*(b) = \max_{j=1,\ldots,s} G_{m,j}(b), \quad \bar{G}_m(b) = \frac{1}{s}\sum_{j=1}^{s} G_{m,j}(b), \quad \hat{\text{se}}(\bar{G}_m(b)) = \sqrt{\frac{\frac{1}{s}\sum_{j=1}^{s}(G_{m,j}(b) - \bar{G}_m(b))^2}{s}},$$

respectively. The estimated probability that the optimality gap exceeds τ^* is $\hat{\alpha} = s^{-1}\,\mathbb{I}\{G_m(b) > \tau^*\}$; under Theorem 3, it is guaranteed to be smaller than α. Figure 5 provides plots of the results of these numerical experiments; detailed numerical results to accompany the plots appear in Tables C.1–C.3.

Figure 5 demonstrates the convergence of the optimality gap $G_m(b)$ for Algorithm 1 under m, n, and b specified by Theorem 3. By comparing the individual plots across the problem and error types, we observe the different effects of Laplace and normal errors on the threshold values of τ^* and on the maximum and average optimality gaps. Since Laplace error has a heavier tail than normal error, it decays at a slower rate. As compared to the normal error type, the Laplace error type tends to yield a slightly smaller τ^* but also yields a larger simulation budget b for the same number of discretization points m. Comparing the results for problems (N1) and (N2) in Figure 5, recall that the only change between these two problems is the increase in the number of objectives. From Corollary 1, this increase is slightly faster than logarithmic in d, which increases the required budget

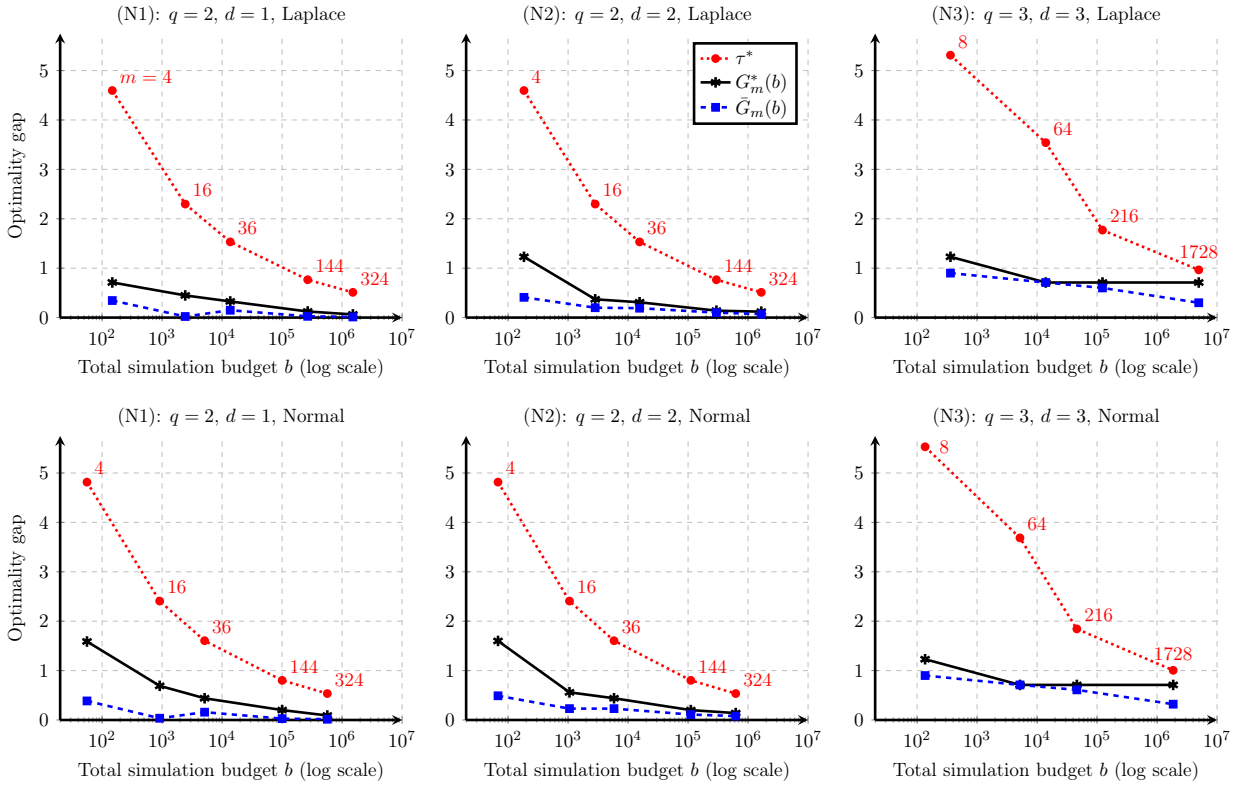


Figure 5: Results of running 10 000 independent macro replications of Algorithm 1 on (N1), (N2), and (N3) using the constants in Table 1, the values of m, n, and b from Theorem 3, and $\alpha = 0.5$. The corresponding m values appear in red next to the markers for τ^*. In all results, $\hat{\alpha}$ is estimated as zero to 4 decimal places and all values of $\hat{\text{se}}(\bar{G}_m(b))$ are estimated as less than 0.002. Detailed numerical results appear in Tables C.1–C.3.

b almost imperceptibly on the log scale. Otherwise, the results between these two problems are very similar. Comparing the results for problems (N1) and (N2) with the results for problem (N3) in Figure 5, we observe the effect of increasing q and d by one together in such a way that the number of points in the discretized objective space that are capable of creating a nonzero optimality gap also increases. In Figure 5, the reported average optimality gap $\bar{G}_m(b)$ is higher for problem (N3), which appears to result in part from a larger discretization error in the objective space (see Tables C.2 and C.3 for detailed comparisons regarding the discretization and selection errors). Finally, despite our relatively large choice of $\alpha = 0.5$ and conducting $s = 10\,000$ independent replications for each subplot in Figure 5, we do not observe any replications where the maximum observed optimality gap $G_m(b)$ exceeded the bound τ^*. Thus, we observe an estimate of $\hat{\alpha}$ which is equal to zero to four decimal places for all three problems under each of the two error types. These results point to some conservativeness in our bounds, which we investigate next.

7.2 Linear problems

In this section, we consider problems that remove some potential sources of conservativeness in the bounds. First, to ensure the Lipschitz constant is binding for points in or near the Pareto set, we consider only linear problems. Second, we test error types which are selected for their potential to

induce events in which $G_m(b) > \tau^*$. Toward this end, we drop Laplace error, whose bound does not appear as tight as normal error in Subsection 7.1. (To see this fact more clearly, compare τ^* with $G_m^*(b)$ across the Laplace and normal error types in Tables C.1–C.3.) Then, we add a symmetric two-point error distribution, Rademacher, which we implement both with independent sampling and with various types of correlation. Finally, to facilitate understanding, we deliberately choose straightforward problems that simplify aspects of the bounds.

First, we consider the linear function in $d = 1$ objectives with $q = 1$,

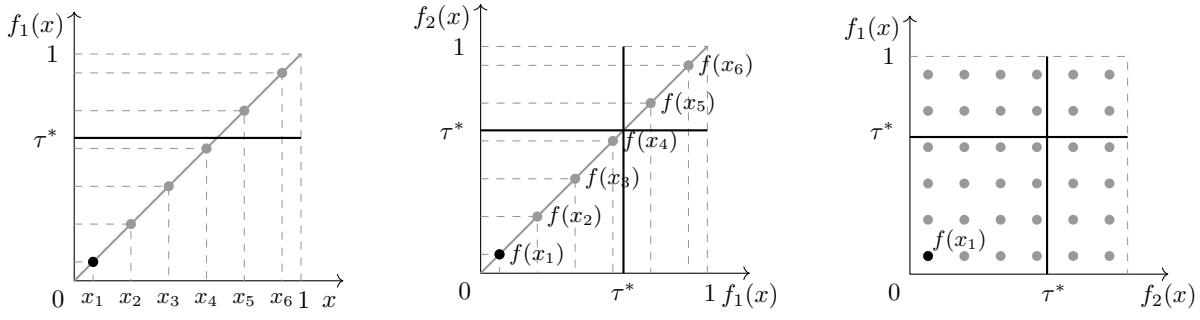
$$\text{minimize} \quad f_1(x) = \mathrm{E}[x + Z_{11}(x)] \quad \text{s.t.} \quad x \in [0, 1] \tag{L1}$$

where $Z_{11}(x)$ is a mean-zero random variable for all $x \in \mathcal{X}$. In (L1), the Lipschitz constant $\ell = 1$ for all $x \in \mathcal{X}$, and the solution is $x^* = 0$ with optimal value $f_1(x^*) = 0$. Figure 6(a) shows a plot of $f_1(x)$ for $x \in [0, 1]$ along with the discretization for $m = 6$ and the corresponding value of $\tau^* = c_2/m = 0.6277$ calculated for the normal error type from the previous section; in this version of the problem, only points labeled as x_5 and x_6 in Figure 6(a) can produce an event where the optimality gap exceeds the threshold, $G_m(b) > \tau^*$.

Then, we add an objective to (L1) to obtain a problem in $d = 2$ objectives with $q = 1$,

$$\text{minimize} \quad f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = \begin{bmatrix} \mathrm{E}[x + Z_{11}(x)] \\ \mathrm{E}[x + Z_{12}(x)] \end{bmatrix} \quad \text{s.t.} \quad x \in [0, 1], \tag{L2}$$

where $Z_{11}(x)$ and $Z_{12}(x)$ are mean-zero random variables for all $x \in \mathcal{X}$. We remark that while usually $q \geq d$ in convex multi-objective problems because otherwise, reduction techniques may apply [Ehrgott, 2005, Ward, 1989], our bounds do not require $q \geq d$. In problem (L2), the Lipschitz constant $\ell = 1$ for all $x \in \mathcal{X}$, the efficient set is $\mathcal{E} = \{x^*\}$ for $x^* = 0$, and the Pareto set is $\mathcal{P} = \{(0, 0)\}$. Figure 6(b) shows a plot of $f(x)$ for $x \in [0, 1]$ for discretization $m = 6$ and $\tau^* = c_2/m = 0.6277$ under normal error.



(a) Plot of (L1) with $m = 6$ and $\tilde{x}_m^* = x_1$ and $x^* = 0$.

(b) Plot of (L2) with $m = 6$ and $\mathcal{P}_m = \{f(x_1)\}$, $\mathcal{P} = \{(0, 0)\}$.

(c) Plot of (L3) with $m = 36$ and $\mathcal{P}_m = \{f(x_1)\}$, $\mathcal{P} = \{(0, 0)\}$.

Figure 6: The objective spaces of the linear problems, where τ^* is calculated under normal error. Only points above τ^* on at least one objective can produce an event where $G_m(b) > \tau^*$.

Finally, we increase the decision space dimension to obtain a problem with $d = 2$ and $q = 2$,

$$\text{minimize} \quad f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = \begin{bmatrix} \mathrm{E}[x_1 + Z_{11}(x)] \\ \mathrm{E}[x_2 + Z_{12}(x)] \end{bmatrix} \quad \text{s.t.} \quad x \in [0,1]^2, \tag{L3}$$

where $Z_{11}(x)$ and $Z_{12}(x)$ are mean-zero random variables for all $x \in \mathcal{X}$. Figure 6(c) shows a plot of the objective space for $m = 36$ with $\tau^* = 0.6622$ calculated for the normal error type. Considering each point in $\mathcal{P}_m^c$ in a pairwise comparison with $f(x_1)$ and holding all other points fixed, only points above τ^* on at least one objective can produce a false exclusion event resulting in $G_m(b) > \tau^*$. To do so, these points must be falsely estimated as better than $f(x_1)$ on all objectives. Likewise, only points that are above τ^* on all objectives can produce a false inclusion event resulting in $G_m(b) > \tau^*$. To do so, these points must be falsely estimated as better than $f(x_1)$ on at least one objective.

With all discretizations $\mathcal{X}_m$ calculated as described in Subsection 7.1, we implement three additional error types selected for their potential to induce events where $G_m(b) > \tau^*$:

E3. Rademacher error, where $Z_{jk}(x), j = 1, \ldots, n$, $k = 1, \ldots, d$, $x \in \mathcal{X}$ are iid Rademacher, that is, $Z_{jk}(x) = -1$ w.p. $1/2$ and $Z_{jk}(x) = 1$ w.p. $1/2$, and $\sigma_k^2(x) = 1$ for all $k = 1, \ldots, d$, $x \in \mathcal{X}$.

E4. Rademacher adversarial error with independent objectives, where for given $x \in \mathcal{X}$, $Z_{jk}(x), j = 1, \ldots, n$, $k \in \{1, \ldots, d\}$ are iid Rademacher, but we induce correlation across the points: For every objective k, let $\tilde{\mathcal{X}}_k(\tau^*) \coloneqq \{x \in \mathcal{X} \colon f_k(x) \le \tau^*\}$ be the set of points below τ^* on objective k. For every j and k, first set $Z_{jk}(x) = Z_{jk}(x')$ for all $x, x' \in \tilde{\mathcal{X}}_k(\tau^*)$ and all $x, x' \in \tilde{\mathcal{X}}_k(\tau^*)^c$; then, set $Z_{jk}(x) = -Z_{jk}(x')$ for all $x \in \tilde{\mathcal{X}}_1(\tau^*), x' \in \tilde{\mathcal{X}}_1(\tau^*)^c$.

E5. Rademacher adversarial error with correlated objectives, where for given $k \in \{1, \ldots, d\}$, $x \in \mathcal{X}$, $Z_{jk}(x), j = 1, \ldots, n$ are iid Rademacher, but we induce correlation across the points and objectives: For every j, first set $Z_{jk'}(x) = Z_{jk''}(x')$ for every pair $x, x' \in \cap_{k=1}^d \tilde{\mathcal{X}}_k(\tau^*)$ and every pair $k', k'' \in \{1, \ldots, d\}$. Then for a given $x' \in \cap_{k=1}^d \tilde{\mathcal{X}}_k(\tau^*)$ and $k'' \in \{1, \ldots, d\}$, set $Z_{jk'}(x) = -Z_{jk''}(x')$ for every $x \in \cup_{k=1}^d \tilde{\mathcal{X}}_k(\tau^*)^c$ and $k' \in \{1, \ldots, d\}$.

The Rademacher adversarial error types are omniscient in the sense that implementation requires advance knowledge of which points have objective values above the threshold τ^*. Also, there is extreme dependence between the points and objectives, which our bounds protect against. Constants for each problem appear in Table 2.

Figure 7 reports the results of running $s = 10^8$ independent macro replications of Algorithm 1

Table 2: Constants for linear problems; see Appendix A.

	Error	d	q	r	p	ϵ	$\tilde{c}_q$	ℓ	η_2^*	κ_2	c_2
(L1)	Normal	1	1	2	∞	0.5	1/2	1	$\sqrt{8/3}$	4/3	3.7660
(L1)	Rademacher (all types)	1	1	2	∞	0.5	1/2	1	$1/\sqrt{\ln 2}$	0.9999	2.9022
(L2)	Normal	2	1	2	∞	0.5	1/2	1	$\sqrt{8/3}$	4/3	3.7660
(L2)	Rademacher (all types)	2	1	2	∞	0.5	1/2	1	$1/\sqrt{\ln 2}$	0.9999	2.9022
(L3)	Normal	2	2	2	∞	0.5	$\sqrt{2}/2$	1	$\sqrt{8/3}$	4/3	3.9731
(L3)	Rademacher (all types)	2	2	2	∞	0.5	$\sqrt{2}/2$	1	$1/\sqrt{\ln 2}$	0.9999	3.1094

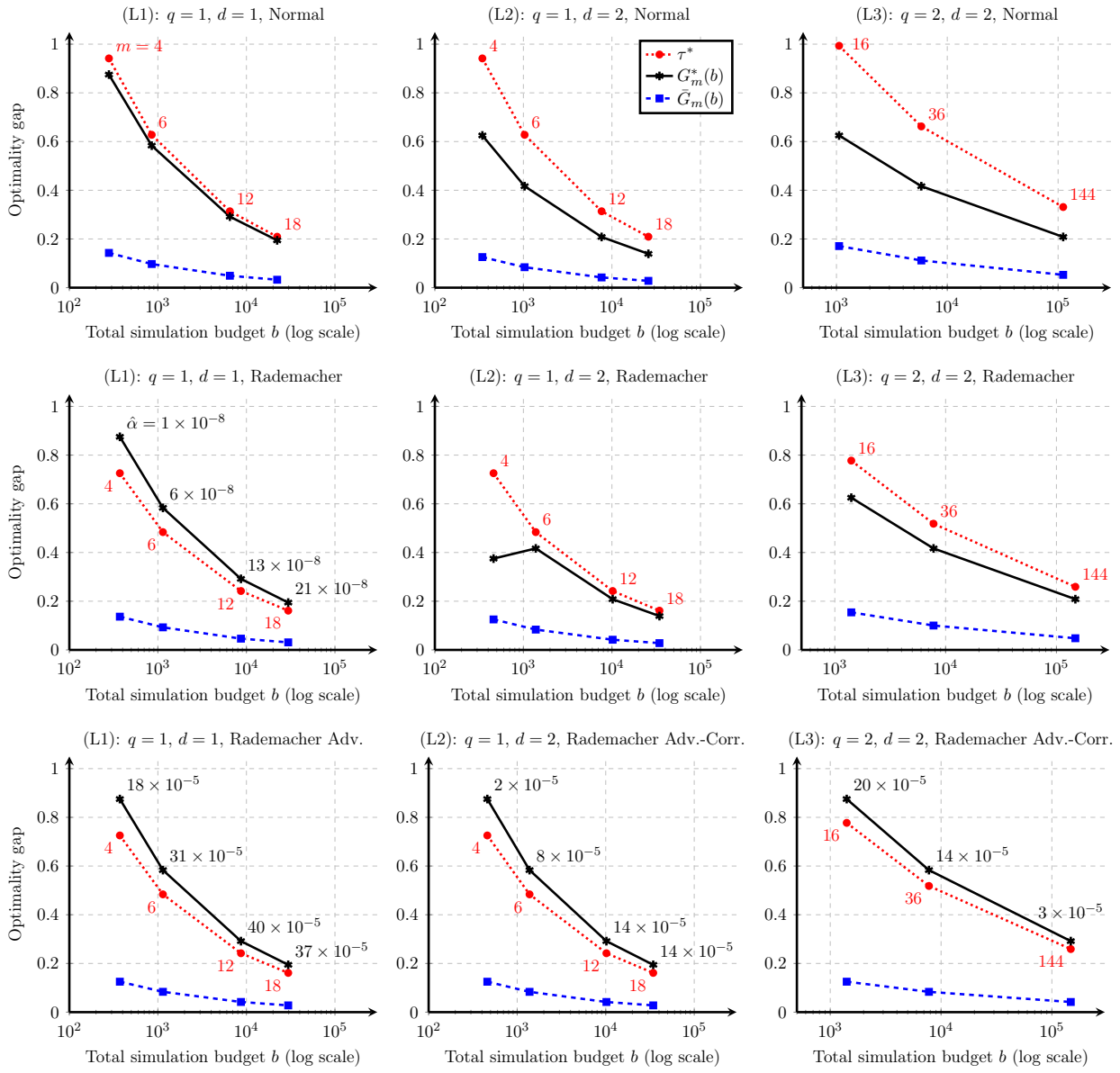


Figure 7: Results of running $s = 10^8$ independent macro replications of Algorithm 1 on (L1), (L2), and (L3) using the constants in Table 2, the values of m, n, and b from Theorem 3, and $\alpha = 0.5$. The m values appear in red next to the markers for τ^*, and nonzero $\hat{\alpha}$ values appear in black next to the markers for $G^*_m(b)$. (All $\hat{\alpha}$ values for Rademacher Adversarial error are rounded.) In all results, $\hat{\text{se}}(\bar{G}_m(b))$ is estimated as zero to 4 decimal places. Detailed numerical results for each error type appear in Tables D.4–D.6.

on problems (L1), (L2), and (L3) using the constants in Table 2 and the values of m, n, and b from Theorem 3. For brevity in the main paper, we only report the results for normal error, Rademacher error, and Rademacher adversarial error with correlated objectives. Complete results for all error types are included in Tables D.4–D.6. As in the previous section, all results demonstrate the convergence of $G_m(b)$ to zero with increasing m and n.

Comparing the results for normal error in Figure 5 with the results for normal error in Figure 7, and specifically comparing the results for problem (N2), $q = 2, d = 2$ with the results for problem

(L3) $q = 2, d = 2$ under normal error, we see that the red dotted line for the bound τ^* is closer to the black solid line for $G_m^*(b)$ under the linear problem structure. This result is expected, since the Lipschitz constant is the same for all points in the objective space of the linear problems, which results in a tighter bound for τ^*. However, as in Figure 5, none of the macro replications for the normal error type have an optimality gap which exceeds the bound τ^*; that is, the red dotted line for τ^* is always above the black solid line for $G_m^*(b)$ in Figures 5 and 7 under the normal error type.

However, for the Rademacher error type, we do observe macro replications in which the optimality gap exceeds the bound τ^* in Figure 7. Specifically, we observe the black solid line for $G_m^*(b)$ above the red dotted line for τ^* in all reported results for the Rademacher adversarial error with correlated objectives. Despite setting $\alpha = 0.5$, the reported values of $\hat{\alpha}$ are much smaller. The largest observed value of $\hat{\alpha}$ from the results in Figure 7 is approximately $\hat{\alpha} = 40 \times 10^{-5} = 0.0004$ with an estimated standard error of approximately $\hat{\text{se}}(\hat{\alpha}) = 2 \times 10^{-6}$. Thus, even with omniscient adversarial error, our probabilistic bound appears conservative by a few orders of magnitude. This insight implies that if the budget allocations in Theorem 3 and Theorem 4 were used to implement a simple algorithm on an amenable version of problem (M) in a practical application, knowing the constants only approximately, e.g., by estimating them through a pilot study or using a decision-maker's apriori approximation, is likely sufficient to retain the probabilistic guarantee.

8 Concluding remarks

We construct a new, tractable performance indicator that leads to finite-time probabilistic bounds on the optimality gap when implementing simple algorithms for global MOSO. The bounds protect against many possible scenarios that may rarely arise together in practice, such as omniscient adversarial noise and large Lipschitz constants that may occur far away from the true Pareto set. In this sense, the bounds are likely to be conservative for practical problems. The bounds, and the resulting convergence rates, may be able to be tightened under additional structural assumptions, or when using common random numbers as described in Ha et al. [2024]. Nevertheless, holding everything else fixed and using the infinity norm in the objective space, we find that the required total simulation budget grows only slightly faster than logarithmically in the number of objectives — a result that may be of interest to decision makers with many objectives [Fonseca et al., 2020]. Finally, there is a tradeoff between fully exploiting common random numbers by obtaining the same number of simulation replications at every point in the discretized feasible set and the efficiencies that arise from differential sampling. Future work may incorporate the differential sampling capabilities identified in the context of multi-objective ranking and selection [Rojas-Gonzalez et al., 2019, Applegate et al., 2020].

Code and data disclosure

A link to the code and data to support the numerical experiments in this paper can be found at `https://github.com/HunterResearch/GlobalMOSOErrorBounds`.

A Calculating constants in the concentration inequalities

In Examples A.1–A.3, we demonstrate how to determine good values for the constants in (14) and (17) when all errors share the same distributional family.

Example A.1 (Laplace error, also see Rigollet [2015, p. 22-25])**.** Given an objective k and feasible point $x \in \mathcal{X}$, suppose the errors $Z_{jk}(x), j = 1, \ldots, n$ are iid with mean-zero Laplace distribution having scale parameter $\beta_k(x) > 0$ and standard deviation $\sigma_k(x) = \sqrt{2}\beta_k(x)$. To compute a bound specific to the Laplace distribution, use a Chernoff bound to see that for any $\delta > 0$, $s > 0$,

$$\mathrm{P}\{\tfrac{1}{n}\textstyle\sum_{j=1}^n Z_{jk}(x) \geq \delta\} \leq e^{-sn\delta} \prod_{j=1}^n \mathrm{E}\left[e^{sZ_{jk}(x)}\right], \tag{46}$$

where for each $j = 1, \ldots, n$, the moment generating function of $Z_{jk}(x)$ has upper bound

$$\mathrm{E}\left[e^{sZ_{jk}(x)}\right] = \left(1 - s^2\beta_k^2(x)\right)^{-1} \leq e^{2s^2\beta_k^2(x)} \quad \text{for all} \quad |s| \leq (2\beta_k(x))^{-1}.$$

Choosing $s = \min\left\{1/(2\beta_k(x)), \delta/(2\beta_k(x))^2\right\}$ and using the fact that $Z_{jk}(x), j = 1, \ldots, n$ are iid random variables from a symmetric distribution, which implies the distribution of $\sum_{j=1}^n Z_{jk}(x)$ is also symmetric, it follows that

$$\begin{aligned}\mathrm{P}\{|\tfrac{1}{n}\textstyle\sum_{j=1}^n Z_{jk}(x)| \geq \delta\} &\leq 2\exp\left(-n\left(s\delta - (1/2)s^2(2\beta_k(x))^2\right)\right) \\ &\leq 2\exp\left(-(n/2)\min\left\{\delta/(2\beta_k(x)), \delta^2/(2\beta_k(x))^2\right\}\right).\end{aligned} \tag{47}$$

From (12), each $|Z_{jk}(x)|$ has an exponential distribution with parameter $1/\beta_k(x)$. Then the sub-exponential norm of $Z_{1k}(x)$ is $\|Z_{1k}(x)\|_{\psi_1} = \inf\{s > \beta(x) \colon s(s-\beta(x))^{-1} \leq 2\} = 2\beta_k(x)$. Substituting $2\beta_k(x) = \|Z_{1k}(x)\|_{\psi_1}$ into (47) and ensuring the right side of (47) is less than or equal to the right side of (14) yields that $\kappa_1 \leq 1/2$. Therefore, we set κ_1 as large as possible, so that $\kappa_1 = 1/2$.

Now we consider the bound across objectives $k = 1, \ldots, d$ and feasible points $x \in \mathcal{X}$. If all errors belong to the Laplace family, then we can apply $\kappa_1 = 1/2$ for all k and all x. Further, η_1^* is twice the largest scale parameter, $\eta_1^* = 2\sup\{\beta_k(x) \colon k \in \{1, \ldots, d\},\ x \in \mathcal{X}\}$.

Example A.2 (normal error)**.** Given an objective k and feasible point $x \in \mathcal{X}$, let the errors $Z_{jk}(x), j = 1, \ldots, n$ be iid with mean-zero normal distribution having variance $\sigma_k^2(x) > 0$. Then using results in Vershynin [2018], it is straightforward to show that

$$\mathrm{P}\{|\tfrac{1}{n}\textstyle\sum_{j=1}^n Z_{jk}(x)| \geq \delta\} \leq 2\exp\left(-n\delta^2/(2\sigma_k^2(x))\right). \tag{48}$$

Use (15) together with (17) and (48) to determine possible values of the constants in (17). First, using (15), for a given $k \in \{1, \ldots, d\}$, $x \in \mathcal{X}$, the sub-Gaussian norm is

$$\|Z_{1k}(x)\|_{\psi_2} = \inf\{s > \sqrt{2}\sigma_k(x) \colon s(s^2 - 2\sigma_k(x)^2)^{-1/2} \leq 2\} = \sqrt{8/3}\,\sigma_k(x) \approx 1.6330\,\sigma_k(x).$$

Then, κ_2 is a constant which ensures the right side of (48) is less than or equal to the right side of (17). This inequality implies $\kappa_2 \leq \|Z_{1k}(x)\|_{\psi_2}^2/(2\sigma_k^2(x)) = 4/3$. Therefore, the bound in (17) is at least as tight as (48) if we set κ_2 as large as possible, so that $\kappa_2 = 4/3$.

Now we consider the bound across objectives $k = 1, \ldots, d$ and feasible points $x \in \mathcal{X}$. If all errors belong to the normal family, then we can apply $\kappa_2 = 4/3$ for all k and all x. Further, we can calculate η_2^* in (16) as $\sqrt{8/3}$ times the maximum standard deviation across all feasible x values, $\eta_2^* = \sqrt{8/3}\sup\{\sigma_k(x)\colon k \in \{1, \ldots, d\},\, x \in \mathcal{X}\}$.

Example A.3 (Rademacher error)**.** Given an objective k and feasible point $x \in \mathcal{X}$, let the errors $Z_{jk}(x), j = 1, \ldots, n$ be iid with mean-zero Rademacher distribution; that is, we have $P\{Z_{jk}(x) = -1\} = P\{Z_{jk}(x) = 1\} = 1/2$ for all $j = 1, \ldots, n$ and $\mathrm{Var}(Z_{jk}(x)) = 1$. First, from (15), the sub-Gaussian norm is $\|Z_{jk}(x)\|_{\psi_2} = \inf\{s > 0\colon \exp(1/s^2) \le 2\} = 1/\sqrt{\ln 2} \approx 1.2011$. Also, the moment generating function of $Z_{jk}(x)$ is $\mathrm{E}\left[e^{sZ_{jk}(x)}\right] = (1/2)(e^s + e^{-s}) = \cosh(s)$ for all $s \in \mathbb{R}$. Then from the Chernoff bound in (46), for any $\delta > 0$, $s > 0$, we have

$$\mathrm{P}\{\tfrac{1}{n}\textstyle\sum_{j=1}^n Z_{jk}(x) \ge \delta\} \le e^{-sn\delta}\cosh(s)^n = \exp(-sn\delta + n\ln\cosh(s)). \tag{49}$$

To get the tightest bound in (49), choose the smallest value of s by setting $s = \tanh^{-1}(\delta)$ for $\delta \in (0,1)$; otherwise, if $\delta \ge 1$, then $\mathrm{P}\{\frac{1}{n}\sum_{j=1}^n Z_{jk}(x) \ge \delta\} = 0$. Apply symmetry to yield

$$\mathrm{P}\{|\tfrac{1}{n}\textstyle\sum_{j=1}^n Z_{jk}(x)| \ge \delta\} \le 2\exp\big(-n\big[\delta\tanh^{-1}(\delta) - \ln\cosh(\tanh^{-1}(\delta))\big]\big)\mathbb{I}_{\{\delta\in(0,1)\}}. \tag{50}$$

Now, κ_2 ensures the right side of (50) is less than or equal to the right side of (17). If $\delta \ge 1$, then $\kappa_2 > 0$ can be any positive number, because the left side of (17) is zero. If $\delta \in (0,1)$, we require

$$\begin{aligned}\kappa_2 &\le \big(\|Z_{1k}(x)\|_{\psi_2}^2/\delta^2\big)\big[\delta\tanh^{-1}(\delta) - \ln\cosh(\tanh^{-1}(\delta))\big]\\ &= \big(1/(\delta^2\ln 2)\big)\big[\delta\tanh^{-1}(\delta) - \ln\cosh(\tanh^{-1}(\delta))\big] < 1,\end{aligned}$$

which implies the largest κ_2 satisfying this inequality, to four decimal places, is $\kappa_2 = .9999$.

Now we consider the bound across $k = 1, \ldots, d$ and $x \in \mathcal{X}$. If all errors are Rademacher, then we can apply $\kappa_2 = .9999$ for all k and all x. Further, we can calculate η_2^* in (16) as $\eta_2^* = 1/\sqrt{\ln 2}$.

For mixed error types, one may require the sub-exponential norm of a sub-Gaussian random variable for calculating η_1^* in (13). In this case, let $Z_{1k}(x)$ be sub-Gaussian for $k \in \{1, \ldots, d\}, x \in \mathcal{X}$ and let W be a random variable such that $W = 1$ w.p.1, which is sub-Gaussian. Then $\|Z_{1k}(x)W\|_{\psi_1} \le \|Z_{1k}(x)\|_{\psi_2}\|W\|_{\psi_2}$ [Vershynin, 2018, p. 31]. Since $\|W\|_{\psi_2} = 1/\sqrt{\ln 2} \approx 1.20$, we have $\|Z_{1k}(x)\|_{\psi_1} \le (1/\sqrt{\ln 2})\|Z_{1k}(x)\|_{\psi_2}$. If $Z_{1k}(x)$ is normal, then $\|Z_{1k}(x)\|_{\psi_1} \le \sqrt{8/(3\ln 2)}\sigma_k(x) \approx 1.96\sigma_k(x)$.

B Gridding sequences with common points

The discretization sequences used for the numerical results in Section 7 ensure that there exist subsequences sharing common points. These calculations rely on the following Proposition B.1; see Figure B.1 for a visual representation in $q = 2$ dimensions.

Proposition B.1. *Let $\mathcal{X} = [0,1]^q$ be the q-dimensional unit hypercube and set $\nu_1 = \lfloor m_1^{1/q}\rfloor$ and $\nu_2 = \lfloor m_2^{1/q}\rfloor$. As in Example 1, for each $i \in \{1,2\}$, define the sets*

$$\mathcal{X}_{m_i} = \Big\{(x_1, \ldots, x_q)\colon x_j \in \Big\{\frac{1}{2\nu_i}, \frac{3}{2\nu_i}, \ldots, \frac{2\nu_i - 1}{2\nu_i}\Big\} \text{ for each } j\Big\} \cup \Psi_i,$$

where Ψ_i is an arbitrary set of $m_i - \nu_i^q$ points in $\mathcal{X}$. If $3\nu_1 = \nu_2$, then $\mathcal{X}_{m_1} \setminus \Psi_1 \subset \mathcal{X}_{m_2} \setminus \Psi_2$.

Proof sketch. Suppose $3\nu_1 = \nu_2$. Then

$$\begin{aligned}
\mathcal{X}_{m_1} \setminus \Psi_1 &= \left\{(x_1, \ldots, x_q)\colon x_j \in \left\{\frac{1}{2\nu_1}, \frac{3}{2\nu_1}, \ldots, \frac{2\nu_1 - 1}{2\nu_1}\right\} \forall j\right\} \\
&= \left\{(x_1, \ldots, x_q)\colon x_j \in \left\{\frac{3}{2\nu_2}, \frac{9}{2\nu_2}, \ldots, \frac{2\nu_2 - 3}{2\nu_2}\right\} \forall j\right\} \\
&\subset \left\{(x_1, \ldots, x_q)\colon x_j \in \left\{\frac{1}{2\nu_2}, \frac{3}{2\nu_2}, \frac{5}{2\nu_2}, \ldots, \frac{2\nu_2 - 5}{2\nu_2}, \frac{2\nu_2 - 3}{2\nu_2}, \frac{2\nu_2 - 1}{2\nu_2}\right\} \forall j\right\} = \mathcal{X}_{m_2} \setminus \Psi_2. \quad \square
\end{aligned}$$

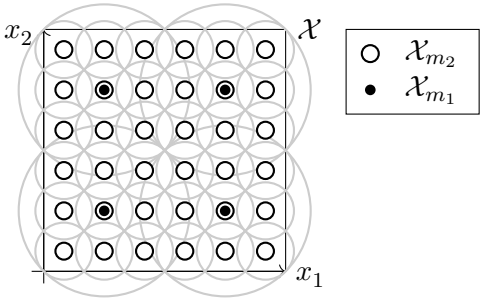


Figure B.1: By setting $\nu_2 = 3\nu_1$ in Proposition B.1, we have $\mathcal{X}_{m_1} \subset \mathcal{X}_{m_2}$.

C Detailed numerical results for nonlinear problems

Tables C.1–C.3 report the results of running $s = 10\,000$ independent macro replications of Algorithm 1 on (N1), (N2), and (N3) using the constants from Table 1 and the values of m, n, and b from Theorem 3. Discretization designs that follow a subset pattern share the same row shading in the tables. Given nonlinearities, across subsequences, results reported in the rows are not necessarily monotone decreasing. For each row, we start the random number stream at the same seed. Since false inclusion cannot occur without false exclusion, we do not report $G_m^{\mathrm{FI}}(b)$ when $d = 1$.

Table C.1: Results for $s = 10\,000$ independent macro replications of Algorithm 1 on problem (N1), $d = 1$, $q = 2$, using the constants in Table 1 and the values of m, n, and b from Theorem 3.

							Maximum Value					
m	τ	τ^*	α	b	n	$\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$	$G_m^{\mathrm{FE}}(b)$	$G_m^{\mathrm{FI}}(b)$	$G_m^*(b)$	$\bar{G}_m(b)$	$\hat{\mathrm{se}}(\bar{G}_m(b))$	$\hat{\alpha}$
Laplace error												
4	9.1924	4.5962	0.5	148	37	0.3268	0.3841	–	0.7109	0.3451	0.0008	0.0000
16	3.0642	2.2981	0.5	2432	152	0.0214	0.4273	–	0.4487	0.0218	0.0001	0.0000
36	1.8385	1.5321	0.5	13 680	380	0.1376	0.1891	–	0.3268	0.1484	0.0002	0.0000
144	0.8357	0.7660	0.5	267 408	1857	0.0214	0.1017	–	0.1231	0.0257	0.0001	0.0000
324	0.5408	0.5107	0.5	1 513 728	4672	0.0096	0.0543	–	0.0639	0.0125	0.0000	0.0000
Normal error												
4	9.6300	4.8150	0.5	56	14	0.3268	1.2622	–	1.5890	0.3859	0.0014	0.0000
16	3.2100	2.4075	0.5	912	57	0.0214	0.6697	–	0.6911	0.0357	0.0007	0.0000
36	1.9260	1.6050	0.5	5112	142	0.1376	0.3025	–	0.4402	0.1583	0.0004	0.0000
144	0.8755	0.8025	0.5	100 224	696	0.0214	0.1794	–	0.2008	0.0276	0.0001	0.0000
324	0.5665	0.5350	0.5	567 648	1752	0.0096	0.0813	–	0.0909	0.0152	0.0001	0.0000

Table C.2: Results for $s = 10\,000$ independent macro replications of Algorithm 1 on problem (N2), $d = 2$, $q = 2$ using the constants from Table 1 and the values of m, n, and b from Theorem 3.

							Maximum Value					
m	τ	τ^*	α	b	n	$\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$	$G_m^{\mathrm{FE}}(b)$	$G_m^{\mathrm{FI}}(b)$	$G_m^*(b)$	$\bar{G}_m(b)$	$\hat{\mathrm{se}}(\bar{G}_m(b))$	$\hat{\alpha}$
Laplace error												
4	9.1924	4.5962	0.5	184	46	0.39	0.43	0.88	1.23	0.41	0.0009	0.0000
16	3.0642	2.2981	0.5	2832	177	0.20	0.36	0.26	0.37	0.20	0.0002	0.0000
36	1.8385	1.5321	0.5	15 552	432	0.15	0.27	0.22	0.31	0.19	0.0003	0.0000
144	0.8357	0.7660	0.5	296 640	2060	0.09	0.10	0.13	0.14	0.10	0.0001	0.0000
324	0.5408	0.5107	0.5	1 660 176	5124	0.06	0.10	0.10	0.12	0.07	0.0001	0.0000
Normal error												
4	9.6300	4.8150	0.5	68	17	0.39	1.26	0.88	1.60	0.49	0.0018	0.0000
16	3.2100	2.4075	0.5	1056	66	0.20	0.36	0.51	0.56	0.23	0.0005	0.0000
36	1.9260	1.6050	0.5	5832	162	0.15	0.30	0.30	0.44	0.23	0.0005	0.0000
144	0.8755	0.8025	0.5	111 168	772	0.09	0.15	0.17	0.20	0.11	0.0002	0.0000
324	0.5665	0.5350	0.5	622 728	1922	0.06	0.12	0.12	0.14	0.08	0.0001	0.0000

Table C.3: Results for $s = 10\,000$ independent macro replications of Algorithm 1 on problem (N3), $d = 3$, $q = 3$ using the constants from Table 1 and the values of m, n, and b from Theorem 3.

							Maximum Value					
m	τ	τ^*	α	b	n	$\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$	$G_m^{\mathrm{FE}}(b)$	$G_m^{\mathrm{FI}}(b)$	$G_m^*(b)$	$\bar{G}_m(b)$	$\hat{\mathrm{se}}(\bar{G}_m(b))$	$\hat{\alpha}$
Laplace error												
8	10.6227	5.3113	0.5	360	45	0.90	0.40	0.88	1.23	0.90	0.0000	0.0000
64	3.5409	3.5409	0.5	13 952	218	0.71	0.51	0.51	0.71	0.71	0.0000	0.0000
216	2.1246	1.7704	0.5	123 120	570	0.60	0.40	0.40	0.71	0.60	0.0009	0.0000
1728	0.9657	0.9657	0.5	4 950 720	2865	0.27	0.37	0.56	0.71	0.30	0.0005	0.0000
Normal error												
8	11.0603	5.5301	0.5	136	17	0.90	0.40	0.88	1.23	0.90	0.0004	0.0000
64	3.6868	3.6867	0.5	5184	81	0.71	0.51	0.51	0.71	0.71	0.0000	0.0000
216	2.2121	1.8434	0.5	46 008	213	0.60	0.40	0.40	0.71	0.61	0.0009	0.0000
1728	1.0055	1.0055	0.5	1 855 872	1074	0.27	0.43	0.56	0.71	0.32	0.0006	0.0000

We report more digits in Table C.1 than in Tables C.2 and C.3 because we are able to calculate all quantities in Table C.1 with greater accuracy. Calculating the distances in Tables C.2 and C.3 relies on a fine mesh grid in the objective space that is independent of m. The maximum possible error due to this fine mesh grid in Tables C.2 and C.3 is 0.06; however, the actual error is likely smaller because we union the fine mesh grid with the discretized Pareto points before calculating the error. Nevertheless, we adjust the number of digits reported to reflect the lower (deterministic) numerical accuracy of these calculations.

D Detailed numerical results for linear problems

Tables D.4–D.6 report the results of running $s = 10^8$ independent macro replications of Algorithm 1 on (L1), (L2), and (L3) using the constants from Table 1 and the values of m, n, and b from Theorem 3. Discretization designs that follow a subset pattern share the same row shading in the tables. For each row, we start the random number stream at the same seed. For simplicity, we

Table D.4: Results for $s = 100\,000\,000 = 10^8$ independent replications of Algorithm 1 on problem (L1), $d = 1$, $q = 1$, using the constants in Table 2 and the values of m, n, and b from Theorem 3.

						Maximum Value					
m	$\tau^* = c_2/m$	α	b	n	$\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$	$G_m^{\mathrm{FE}}(b)$	$G_m^{\mathrm{FI}}(b)$	$G_m^*(b)$	$\bar{G}_m(b)$	$\hat{\mathrm{se}}(\bar{G}_m(b))$	$\hat{\alpha}$
Normal error											
4	*0.9415*	0.5	280	70	0.1250	0.7500	–	*0.8750*	0.1428	0.0000	0.000 000 00
6	*0.6277*	0.5	852	142	0.0833	0.5000	–	*0.5833*	0.0971	0.0000	0.000 000 00
12	*0.3138*	0.5	6504	542	0.0417	0.2500	–	*0.2917*	0.0490	0.0000	0.000 000 00
18	*0.2092*	0.5	22 248	1236	0.0278	0.1667	–	*0.1944*	0.0326	0.0000	0.000 000 00
Rademacher error											
4	*0.7256*	0.5	372	93	0.1250	0.7500	–	*0.8750*	0.1365	0.0000	0.000 000 01
6	*0.4837*	0.5	1140	190	0.0833	0.5000	–	*0.5833*	0.0927	0.0000	0.000 000 06
12	*0.2419*	0.5	8676	723	0.0417	0.2500	–	*0.2917*	0.0463	0.0000	0.000 000 13
18	*0.1612*	0.5	29 664	1648	0.0278	0.1667	–	*0.1944*	0.0310	0.0000	0.000 000 21
Rademacher adversarial error with one objective											
4	*0.7256*	0.5	372	93	0.1250	0.7500	–	*0.8750*	0.1251	0.0000	0.000 182 26
6	*0.4837*	0.5	1140	190	0.0833	0.5000	–	*0.5833*	0.0835	0.0000	0.000 305 13
12	*0.2419*	0.5	8676	723	0.0417	0.2500	–	*0.2917*	0.0418	0.0000	0.000 402 07
18	*0.1612*	0.5	29 664	1648	0.0278	0.1667	–	*0.1944*	0.0278	0.0000	0.000 372 06

Table D.5: Results for $s = 100\,000\,000 = 10^8$ independent replications of Algorithm 1 on problem (L2), $d = 2$, $q = 1$, using the constants in Table 2 and the values of m, n, and b from Theorem 3.

						Maximum Value					
m	$\tau^* = c_2/m$	α	b	n	$\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$	$G_m^{\mathrm{FE}}(b)$	$G_m^{\mathrm{FI}}(b)$	$G_m^*(b)$	$\bar{G}_m(b)$	$\hat{\mathrm{se}}(\bar{G}_m(b))$	$\hat{\alpha}$
Normal error											
4	*0.9415*	0.5	344	86	0.1250	0.5000	0.5000	*0.6250*	0.1256	0.0000	0.000 000 00
6	*0.6277*	0.5	1032	172	0.0833	0.3333	0.3333	*0.4167*	0.0840	0.0000	0.000 000 00
12	*0.3138*	0.5	7644	637	0.0417	0.1667	0.1667	*0.2083*	0.0421	0.0000	0.000 000 00
18	*0.2092*	0.5	25 812	1434	0.0278	0.1111	0.1111	*0.1389*	0.0280	0.0000	0.000 000 00
Rademacher error											
4	*0.7256*	0.5	460	115	0.1250	0.2500	0.2500	*0.3750*	0.1252	0.0000	0.000 000 00
6	*0.4837*	0.5	1380	230	0.0833	0.3333	0.3333	*0.4167*	0.0835	0.0000	0.000 000 00
12	*0.2419*	0.5	10 188	849	0.0417	0.1667	0.1667	*0.2083*	0.0418	0.0000	0.000 000 00
18	*0.1612*	0.5	34 416	1912	0.0278	0.1111	0.1111	*0.1389*	0.0279	0.0000	0.000 000 00
Rademacher adversarial error with independent objectives											
4	*0.7256*	0.5	460	115	0.1250	0.0000	0.0000	*0.1250*	0.1250	0.0000	0.000 000 00
6	*0.4837*	0.5	1380	230	0.0833	0.0000	0.0000	*0.0833*	0.0833	0.0000	0.000 000 00
12	*0.2419*	0.5	10 188	849	0.0417	0.2500	0.2500	*0.2917*	0.0417	0.0000	0.000 000 04
18	*0.1612*	0.5	34 416	1912	0.0278	0.1667	0.1667	*0.1944*	0.0278	0.0000	0.000 000 04
Rademacher adversarial error with correlated objectives											
4	*0.7256*	0.5	460	115	0.1250	0.7500	0.7500	*0.8750*	0.1250	0.0000	0.000 016 40
6	*0.4837*	0.5	1380	230	0.0833	0.5000	0.5000	*0.5833*	0.0834	0.0000	0.000 078 72
12	*0.2419*	0.5	10 188	849	0.0417	0.2500	0.2500	*0.2917*	0.0417	0.0000	0.000 135 32
18	*0.1612*	0.5	34 416	1912	0.0278	0.1667	0.1667	*0.1944*	0.0278	0.0000	0.000 136 36

suppress τ and only report τ^*.

Comparing Table D.4 with Table D.5, notice that the average optimality gap $\bar{G}_m(b)$ appears to decrease slightly with the transition from (L1) to (L2), which is the opposite of the pattern observed in Tables C.1 and C.2. This result likely occurs because in the transition from (N1) to (N2), we add

Table D.6: Results for $s = 100\,000\,000 = 10^8$ independent replications of Algorithm 1 on problem (L3), $d = 2$, $q = 2$, using the constants in Table 2 and the values of m, n, and b from Theorem 3.

						Maximum Value					
m	τ^*	α	b	n	$\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$	$G_m^{\mathrm{FE}}(b)$	$G_m^{\mathrm{FI}}(b)$	$G_m^*(b)$	$\bar{G}_m(b)$	$\hat{\mathrm{se}}(\bar{G}_m(b))$	$\hat{\alpha}$
Normal error											
16	*0.9933*	0.5	1056	66	0.1250	0.5000	0.5000	*0.6250*	0.1708	0.0000	0.000 000 00
36	*0.6622*	0.5	5832	162	0.0833	0.3333	0.3333	*0.4167*	0.1117	0.0000	0.000 000 00
144	*0.3311*	0.5	111 168	772	0.0417	0.1667	0.1667	*0.2083*	0.0527	0.0000	0.000 000 00
Rademacher error											
16	*0.7773*	0.5	1408	88	0.1250	0.5000	0.5000	*0.6250*	0.1542	0.0000	0.000 000 00
36	*0.5182*	0.5	7776	216	0.0833	0.3333	0.3333	*0.4167*	0.1001	0.0000	0.000 000 00
144	*0.2591*	0.5	148 176	1029	0.0417	0.1667	0.1667	*0.2083*	0.0479	0.0000	0.000 000 00
Rademacher adversarial error with independent objectives											
16	*0.7773*	0.5	1408	88	0.1250	0.7500	0.7500	*0.8750*	0.1253	0.0000	0.000 376 15
36	*0.5182*	0.5	7776	216	0.0833	0.5000	0.5000	*0.5833*	0.0834	0.0000	0.000 170 16
144	*0.2591*	0.5	148 176	1029	0.0417	0.2500	0.0000	*0.2917*	0.0417	0.0000	0.000 065 73
Rademacher adversarial error with correlated objectives											
16	*0.7773*	0.5	1408	88	0.1250	0.7500	0.0000	*0.8750*	0.1251	0.0000	0.000 188 58
36	*0.5182*	0.5	7776	216	0.0833	0.5000	0.0000	*0.5833*	0.0834	0.0000	0.000 144 91
144	*0.2591*	0.5	148 176	1029	0.0417	0.2500	0.0000	*0.2917*	0.0417	0.0000	0.000 033 19

points capable of creating false exclusion or false inclusion events to the discretized objective space. However, we add no such points in the transition from (L1) to (L2); see Figures 6(a) and 6(b). Thus, adding an objective as in (L2) appears to have made false inclusion and false exclusion events less likely; in particular, the nonzero $\hat{\alpha}$ values reported in Table D.4 are larger than the nonzero $\hat{\alpha}$ values reported in Table D.5. The $\hat{\alpha}$ values reported for the Rademacher adversarial error in Table D.4 are the largest seen in our numerical experiments so far. The only other similar values appear in the rows for Rademacher adversarial error types in Table D.6.

Acknowledgments

The authors gratefully acknowledge that this material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-23-1-0488. This work has also benefited from Hunter's participation in Dagstuhl Seminar 20031 "Scalability in Multiobjective Optimization," https://www.dagstuhl.de/20031 and Dagstuhl Seminar 23361 "Multiobjective Optimization on a Budget," https://www.dagstuhl.de/23361.

References

M. H. Alrefaei and A. H. Diabat. A simulated annealing technique for multi-objective simulation optimization. *Applied Mathematics and Computation*, 215(8):3029–3035, 2009. doi:10.1016/j.amc.2009.09.051.

S. Andradóttir. A review of random search methods. In M. C. Fu, editor, *Handbook of Simulation Optimization*, volume 216 of *International Series in Operations Research & Management Science*, pages 277–292. Springer, New York, 2015. doi:10.1007/978-1-4939-1384-8_10.

E. A. Applegate, G. Feldman, S. R. Hunter, and R. Pasupathy. Multi-objective ranking and selection: Optimal sampling laws and tractable approximations via SCORE. *Journal of Simulation*, 14(1):21–40, 2020. doi:10.1080/17477778.2019.1633891.

C. Audet, J. Bigeon, D. Cartier, S. Le Digabel, and L. Salomon. Performance indicators in multiobjective optimization. *European Journal of Operational Research*, 292:397–422, 2021. doi:10.1016/j.ejor.2020.11.016.

R. Boffadossi, M. Leonesio, and L. Fagiano. ROBBO: An efficient method for pareto front estimation with guaranteed accuracy. *arXiv*, 2025. doi:10.48550/arXiv.2506.18004.

J. Branke. Performance metrics for multi-objective optimisation under noise. *IEEE Transactions on Evolutionary Computation*, 2024. doi:10.1109/TEVC.2024.3438115.

K. Cooper, S. R. Hunter, and K. Nagaraj. 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, 2020. doi:10.1287/ijoc.2019.0918.

D. J. Eckman, S. G. Henderson, and S. Shashaani. Diagnostic tools for evaluating and comparing simulation-optimization algorithms. *INFORMS Journal on Computing*, 35(2):350–367, 2023a. doi:10.1287/ijoc.2022.1261.

D. J. Eckman, S. G. Henderson, and S. Shashaani. SimOpt: a testbed for simulation-optimization experiments. *INFORMS Journal on Computing*, 35(2):495–508, 2023b. doi:10.1287/ijoc.2023.1273.

M. Ehrgott. *Multicriteria Optimization*, volume 491 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Heidelberg, 2nd edition, 2005. doi:10.1007/3-540-27659-9.

K. B. Ensor and P. W. Glynn. Stochastic optimization via grid search. *Lectures in Applied Mathematics*, 33: 89–100, 1997.

S. L. Faulkenberg and M. M. Wiecek. On the quality of discrete representations in multiple objective programming. *Optim Eng*, 11:423–440, 2010. doi:10.1007/s11081-009-9099-x.

C. M. Fonseca, K. Klamroth, G. Rudolph, and M. M. Wiecek. Scalability in Multiobjective Optimization (Dagstuhl Seminar 20031). *Dagstuhl Reports*, 10(1):52–129, 2020. doi:10.4230/DagRep.10.1.52.

M. C. Fu, editor. *Handbook of Simulation Optimization*, volume 216 of *International Series in Operations Research & Management Science*. Springer, New York, 2015.

Y. Ha, S. Shashaani, and R. Pasupathy. Complexity of zeroth- and first-order stochastic trust-region algorithms. *arXiv*, 2024. URL `https://arxiv.org/abs/2405.20116`.

M. Han and L. Ouyang. A novel bayesian approach for multi-objective stochastic simulation optimization. *Swarm and Evolutionary Computation*, 75, 2022. ISSN 2210-6502. doi:https://doi.org/10.1016/j.swevo.2022.101192.

H. Huang and Z. B. Zabinsky. Multiple objective probabilistic branch and bound for Pareto optimal approximation. In A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, editors, *Proceedings of the 2014 Winter Simulation Conference*, pages 3916–3927, Piscataway, NJ, 2014. IEEE. doi:10.1109/WSC.2014.7020217.

H. Huang and Z. B. Zabinsky. Non-linear multi-objective optimization with probabilistic branch and bound, 2025. URL `https://arxiv.org/abs/2506.04554`.

S. R. Hunter and B. L. Nelson. Parallel ranking and selection. In A. Tolk, J. Fowler, G. Shao, and E. Yücesan, editors, *Advances in Modeling and Simulation: Seminal Research from 50 Years of Winter Simulation Conferences*, Simulation Foundations, Methods and Applications, chapter 12, pages 249–275. Springer International, Switzerland, 2017. doi:10.1007/978-3-319-64182-9.

S. R. Hunter and B. E. Ondes. Properties of several performance indicators for global multi-objective simulation optimization. In C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, editors, *Proceedings of the 2023 Winter Simulation Conference*, Piscataway, NJ, 2023. IEEE.

S. R. Hunter and R. Pasupathy. Central Limit Theorems for constructing confidence regions in strictly convex multi-objective simulation optimization. In B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C. G. Corlu, L. H. Lee, E. P. Chew, T. Roeder, and P Lendermann, editors, *Proceedings of the 2022 Winter Simulation Conference*, pages 3015–3026, Piscataway, NJ, 2022. IEEE. doi:10.1109/WSC57314.2022.10015390.

S. R. Hunter, E. A. Applegate, V. Arora, B. Chong, K. Cooper, O. Rincón-Guevara, and C. Vivas-Valencia. An introduction to multi-objective simulation optimization. *ACM Transactions on Modeling and Computer Simulation*, 29(1):7:1–7:36, 2019. doi:10.1145/3299872.

A. M. Law. *Simulation Modeling and Analysis*. McGraw Hill Education, New York, 5 edition, 2015.

M. Li and X. Yao. Quality evaluation of solution sets in multiobjective optimisation: A survey. *ACM Computing Surveys*, 52(2), 2019. doi:10.1145/3300148.

V. Mattila, K. Virtanen, and R. P. Hämäläinen. A simulated annealing algorithm for noisy multiobjective optimization. *Journal of Multi-criteria Decision Analysis*, 20:255–276, 2013.

H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992. doi:10.1137/1.9781611970081.

B. E. Ondes and S. R. Hunter. An upper bound on the Hausdorff distance between a Pareto set and its discretization in bi-objective convex quadratic optimization. *Optimization Letters*, 17:45–74, 2023. doi:10.1007/s11590-022-01920-7.

R. Pasupathy and S. G. Henderson. A testbed of simulation-optimization problems. In L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, editors, *Proceedings of the 2006 Winter Simulation Conference*, pages 255–263, Piscataway, NJ, 2006. IEEE. doi:10.1109/WSC.2006.323081.

R. Pasupathy and S. G. Henderson. SimOpt: A library of simulation optimization problems. In S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, editors, *Proceedings of the 2011 Winter Simulation Conference*, pages 4075–4085, Piscataway, NJ, 2011. IEEE. doi:10.1109/WSC.2011.6148097.

P. K. Ragavan, S. R. Hunter, R. Pasupathy, and M. R. Taaffe. Adaptive sampling line search for local stochastic optimization with integer variables. *Mathematical Programming*, 196:775–804, 2022. doi:10.1007/s10107-021-01667-6.

P. Rigollet. 18.S997: High dimensional statistics, lecture notes. Massachusetts Institute of Technology: MIT OpenCourseWare, 2015. URL `https://ocw.mit.edu/`. License: Creative Commons BY-NC-SA.

S. Rojas-Gonzalez and I. Van Nieuwenhuyse. A survey on kriging-based infill algorithms for multiobjective simulation optimization. *Computers and Operations Research*, 116, 2020. doi:0.1016/j.cor.2019.104869.

S. Rojas-Gonzalez, J. Branke, and I. Van Nieuwehuyse. Multiobjective ranking and selection with correlation and heteroscedastic noise. In N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, editors, *Proceedings of the 2019 Winter Simulation Conference*, pages 3392–3403, Piscataway, NJ, 2019. IEEE.

S. Rojas-Gonzalez, H. Jalali, and I. Van Nieuwenhuyse. A multiobjective stochastic simulation optimization algorithm. *European Journal of Operational Research*, 284:212–226, 2020. doi:10.1016/j.ejor.2019.12.014.

S. Sayin. Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming. *Mathematical Programming*, 87:543–560, 2000. doi:10.1007/s101070050011.

B. G. Thengvall, S. N. Hall, and M. P. Deskevich. Measuring the effectiveness and efficiency of simulation optimization metaheuristic algorithms. *Journal of Heuristics*, 31, 2025.

R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, Cambridge, UK, 2018.

J. Ward. Structure of efficient sets for convex objectives. *Mathematics of Operations Research*, 14(2):249–257, May 1989. doi:10.1287/moor.14.2.249.

S. Yakowitz, P. L'Ecuyer, and F. Vázquez-Abad. Global stochastic optimization with low-dispersion point sets. *Operations Research*, 48(6):939–950, 2000. doi:10.1287/opre.48.6.939.12393.

J. Zhang, Y. Ma, T. Yang, and L. Liu. Estimation of the Pareto front in stochastic simulation through stochastic Kriging. *Simulation Modelling Practice and Theory*, 79:69–86, 2017. doi:10.1016/j.simpat.2017.09.006.

E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7 (2):117–132, 2003. doi:10.1109/TEVC.2003.810758.