

The Exclusion–Inclusion Gap for Evaluating Finite Pareto Approximations in Multiobjective Stochastic Optimization

Burla E. Ondes¹ and Susan R. Hunter¹

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

Abstract

We introduce the exclusion–inclusion gap for evaluating finite Pareto approximations constructed from stochastic objective estimates. The gap measures the maximum of two objective-space errors associated with excluding Pareto points from the returned set’s true upper image and including dominated points in the returned set’s true lower image. The terms in its approximation–estimation error decomposition admit general bounds that allow deterministic approximation guarantees to be combined with componentwise stochastic-error control. The gap’s inclusion component is invariant to the return of dominated but weakly nondominated points, enabling performance analysis without positive separation assumptions. Under consistent estimation on a fixed evaluated set, the apparent advantages of such points vanish asymptotically. The results hold across continuous, integer, and mixed-integer settings. We apply the results by performing an approximation–estimation error analysis of a nonadaptive, equal-allocation algorithm for global multiobjective simulation optimization with continuous objective functions and a compact feasible set. The analysis yields finite-sample probabilistic error bounds, budget-allocation rules, sample-complexity results, and convergence rates. Under the infinity norm, we identify conditions under which the sufficient simulation budget grows logarithmically in the number of objectives. For a fixed problem instance, we show that the exclusion–inclusion gap converges at the same total-budget rate as the corresponding single-objective optimality gap, despite accounting for both exclusion and inclusion errors.

1 Introduction

Many real-world decisions involve multiple stakeholders with conflicting objectives and are made under uncertainty. The conflict may reflect different stakeholder perspectives, such as the desire to minimize operating cost while maximizing system effectiveness. Uncertainty arises because the system performance measures depend on random inputs or unknown quantities and, therefore, can only be estimated from data or a stochastic oracle. Such decision settings arise across a

Preprint: Wednesday October 14, 2026 at 12:24pm.

Corresponding author: susanhunter@purdue.edu.

Funding: 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.

wide variety of application areas including agriculture, energy systems, healthcare operations, humanitarian logistics and disaster response, machine learning, manufacturing, military operations, and transportation systems (Hunter et al. 2019, Hunter 2026).

Multiobjective stochastic optimization (MOSO) provides mathematical models to aid decision-making under conflicting objectives and uncertainty. Rather than prescribing a single compromise solution, solving a MOSO problem identifies decisions that represent the best attainable trade-offs among the objectives. These decisions form the efficient set, $\mathcal{E}$, and their images form the Pareto set, $\mathcal{P}$. Characterizing the Pareto set, and identifying efficient solutions that map into it, enables decision-makers to compare the trade-offs among high-quality alternatives. These alternatives may then serve as input to a subsequent optimization problem that selects a single optimal solution (Hunter and McClosky 2016), or they may provide decision-makers with a structured view of the attainable trade-offs to support informed decision-making (Gupta and Hunter 2025).

Because exact descriptions of the efficient and Pareto sets may be impossible to identify computationally or too large to be useful in practice, multiobjective optimization algorithms often seek a finite approximation of these sets (Sayin 2000, Ruzika and Wiecek 2005, Herzel et al. 2021). Thus, an algorithm may generate an approximation by evaluating only finitely many feasible points in a continuous, integer, or mixed-integer problem. In MOSO, the objective values corresponding to this finite evaluated set are unknown and must be estimated. Consequently, the quality of the returned approximation depends on both the deterministic error arising from the finite approximation and the estimation error arising from stochastic objective vector estimates.

1.1 Error Analysis for Finite Pareto Approximations Under Uncertainty

A central challenge in MOSO is to quantify how finite approximation and stochastic estimation jointly affect the quality of an algorithm's returned solution set. In single-objective stochastic optimization, it is well established that the optimality gap can be decomposed into approximation and estimation errors, which enables the combination of a deterministic approximation error guarantee with stochastic error control to yield finite-sample performance and convergence results; see, for example, results in Yakowitz et al. (2000), Shapiro et al. (2009), and Fu (2015). *To extend this optimality-gap-based approximation–estimation error analysis strategy to MOSO, we seek an optimality gap analogue.* Here, the optimality gap analogue is a scalar measure of the objective-space error between the true Pareto set, $\mathcal{P}$, and the true image of the algorithm's returned solution set, $f(\hat{\mathcal{E}}_m(b))$, which is a finite random set specifying the algorithm's true objective-space performance. The latter set is constructed by estimating the objective vector values at m feasible points using a total estimation budget b , thus obtaining the finite estimated solution set $\hat{\mathcal{E}}_m(b)$, and then applying the unknown vector-valued mapping f to obtain $f(\hat{\mathcal{E}}_m(b))$, which is unobserved. To analyze the error in the unobserved set $f(\hat{\mathcal{E}}_m(b))$, *the optimality gap analogue should admit an approximation–estimation error decomposition whose terms can be controlled*

through corresponding deterministic and stochastic error guarantees.

Assessing the output of a MOSO algorithm differs fundamentally from assessing the output of a deterministic multiobjective optimization algorithm that approximates the Pareto set. Because the objective vectors are known exactly, a deterministic multiobjective optimization algorithm can identify and remove dominated points. Therefore, it would typically return an *approximation set* in which no element is weakly dominated by another (Zitzler et al. 2003). In a MOSO setting, although the returned solutions in $\hat{\mathcal{E}}_m(b)$ are estimated as nondominated under the estimated objective vectors, their true images in $f(\hat{\mathcal{E}}_m(b))$ may exclude truly Pareto optimal points, include dominated points, or dominate one another. Thus, $f(\hat{\mathcal{E}}_m(b))$ is not necessarily an approximation set in the usual deterministic sense, and classical deterministic multiobjective optimization performance indicators, such as those discussed in Faulkenberg and Wiecek (2010), Li and Yao (2019), Audet et al. (2021), may not capture important properties of MOSO algorithm output. Moreover, objective-space uncertainty can make it difficult to specify the parameters required by some performance indicators, such as reference points. Accordingly, far fewer performance indicators are designed to assess the random sets returned by MOSO algorithms (Hunter and Ondes 2023, Branke 2025, Thengvall et al. 2025).

A natural starting point for the MOSO optimality gap analogue is the Hausdorff distance between $\mathcal{P}$ and $f(\hat{\mathcal{E}}_m(b))$ (Hunter et al. 2019). The Hausdorff distance accounts for the fact that $f(\hat{\mathcal{E}}_m(b))$ is not an approximation set, requires no reference point, and admits an approximation–estimation error decomposition through the triangle inequality. Specifically, we can use the *approximate Pareto set*, $\mathcal{P}_m$, as the intermediate deterministic object; this set is the Pareto set for the true MOSO problem restricted to the finite evaluated set. Then the Hausdorff distance can be decomposed as

$$\mathbb{H}(\mathcal{P}, f(\hat{\mathcal{E}}_m(b))) \leq \mathbb{H}(\mathcal{P}, \mathcal{P}_m) + \mathbb{H}(\mathcal{P}_m, f(\hat{\mathcal{E}}_m(b)))$$

where the first right-hand side term contains two one-way approximation error terms (coverage error and representation error), and the second right-hand side term contains two one-way estimation error terms (false exclusion error and false inclusion error). Within each pair, one term accounts for exclusion-related error (coverage error and false exclusion error), while the other term accounts for inclusion-related error (representation error and false inclusion error). However, not all terms admit bounds under comparably general conditions. The main analytical difficulties arise with respect to the inclusion-related errors. In particular, the false inclusion error term is difficult to analyze if the restricted finite image set contains dominated points that are also weakly nondominated. Such points have zero separation from at least one point in the approximate Pareto set, $\mathcal{P}_m$, on at least one objective and may remain statistically indistinguishable from truly nondominated points under finite sampling. These difficulties do not completely preclude Hausdorff-distance convergence analysis; indeed, convergence can be established under sufficient structural assumptions (Bonnel and

Collonge 2014). However, in general finite-evaluation settings where deterministic approximation and stochastic estimation errors coexist, controlling the false-inclusion error term may require positive-separation assumptions for dominated points, thus restricting the applicable problem and approximation method classes (Applegate et al. 2020, Ondes 2025).

To circumvent these analytical difficulties while retaining meaningful comparisons, we introduce the *exclusion–inclusion gap*, a performance indicator designed to support approximation–estimation error analysis for finite Pareto approximations constructed from stochastic objective vector estimates. Instead of directly comparing $\mathcal{P}$ with $f(\hat{\mathcal{E}}_m(b))$, the exclusion–inclusion gap uses upper- and lower-image constructions to separately measure errors associated with the exclusion of Pareto points and the inclusion of dominated points, and then takes the maximum. These constructions create nested sets that ultimately yield an approximation–estimation error decomposition with one approximation error term corresponding to exclusion and two estimation error terms corresponding to false exclusion and false inclusion. Unlike the Hausdorff distance, however, the false inclusion component is invariant to the return of weakly nondominated points whose classification may remain unresolved under finite sampling. Although a dominated but weakly nondominated point may remain intermittently classified as nondominated when a true objective tie prevents the estimated ordering from stabilizing, almost-sure estimator consistency implies that every apparent objective advantage supporting this misclassification converges to zero, while every strict objective disadvantage is eventually estimated with the correct ordering. Therefore, in the limit, the decision-maker is not misled by such points. Moreover, deterministic approximation error guarantees bound the approximation error term, while uniform componentwise stochastic error control bounds both estimation terms. These results apply across a broad class of finite-evaluation MOSO settings and do not depend on a particular finite-approximation method, objective estimator, or statistical functional used to resolve the objective-vector uncertainty, provided the estimators are consistent and admit suitable componentwise error control. For concreteness, we develop the results using the widely applicable expected-value formulation with sample-average estimators.

1.2 Application to Global Multiobjective Simulation Optimization

We demonstrate the utility of our results by analyzing the approximation and estimation errors produced by a simple nonadaptive, equal-allocation algorithm for global multiobjective simulation optimization with continuous objectives and a compact feasible set. The algorithm evaluates a prespecified space-filling design, allocates the same number of simulation replications to each design point, and returns the solutions that are nondominated under the estimated objective values. When the design points form a grid, this approach is called *grid search* (Ensor and Glynn 1997); more generally, the design may be determined by a low-dispersion point set (Yakowitz et al. 2000). To the best of our knowledge, this simple algorithm has not yet received a systematic finite-sample and convergence analysis in the multiobjective context.

Our motivation for studying this algorithm is threefold. First, it provides a theoretically informative setting in which to demonstrate how deterministic approximation and stochastic estimation guarantees combine under the exclusion–inclusion gap. Specifically, under a fixed total simulation budget $b = mn$, increasing the number of design points, m , improves coverage of the feasible set, thereby reducing approximation error, but it leaves fewer replications n for estimating the objective values at each point, thereby increasing estimation error. The exclusion–inclusion gap makes this trade-off explicit. The resulting analysis allows us to isolate how decision-space dimension, objective-space dimension, stochastic error, and total simulation budget jointly affect algorithm performance. Second, this algorithm’s nonadaptive, equal-allocation structure makes it transparent and straightforward to implement. When computing resources are plentiful, simulation runs are relatively inexpensive, or no additional problem structure is available to exploit, such an algorithm provides an accessible approach to global MOSO. It is also readily parallelizable (Heidelberger 1988, Glynn and Heidelberger 1990, 1991) and amenable to common random numbers because every evaluated point receives the same replication budget. Third, it provides a foundation for adaptive variants and a theoretically grounded benchmark for evaluating the benefits of other algorithmic mechanisms. The simulation optimization literature emphasizes the importance of common test problems, reference methods, and diagnostic tools for evaluating and comparing algorithms (Pasupathy and Henderson 2006, 2011, Eckman et al. 2023a,b). Characterizing the performance attainable through space-filling coverage and equal allocation helps distinguish the gains attributable to additional problem structure, adaptivity, or algorithmic mechanisms such as simulated annealing, probabilistic branch-and-bound, stochastic kriging, surrogate modeling, Bayesian modeling, and plausible inference (Alrefaei and Diabat 2009, Mattila et al. 2013, Andradóttir 2015, Huang and Zabinsky 2014, 2026, Zhang et al. 2017, Rojas-Gonzalez et al. 2020, Rojas-Gonzalez and Van Nieuwenhuyse 2020, Han and Ouyang 2022, Nika et al. 2025, Zhao and Eckman 2026). Together, these features make the simple nonadaptive, equal-allocation algorithm a useful setting for demonstrating analyses under the exclusion–inclusion gap and establishing a theoretically characterized baseline for future algorithmic developments.

1.3 Contributions

We introduce the exclusion–inclusion gap for evaluating finite Pareto approximations constructed from stochastic objective vector estimates. We establish its basic properties and approximation–estimation error decomposition (Theorem 1), as well as the inclusion-component’s invariance to the return of weakly nondominated points whose classification may remain unresolved under finite sampling (Proposition 1). Further, for a fixed finite evaluated set and assuming a consistent objective vector estimator, the estimated approximate Pareto set observed by the decision-maker converges almost surely to $f(\hat{\mathcal{E}}_m(b))$ regardless of whether the weakly dominated classifications stabilize (Proposition 2). Then, we show that deterministic coverage error and additive ε -approximation

guarantees bound the approximation error term (Lemma 1), and uniform componentwise control of the stochastic errors bounds both estimation error terms (Lemma 2). Ultimately, the exclusion–inclusion gap provides a selective relaxation of the Hausdorff distance that retains meaningful comparisons and supports approximation–estimation error analysis. It does so without requiring the resolution of weakly nondominated boundary classifications in the approximate problem or imposing a potentially strong separation assumption. Because the results apply under relatively general conditions, they support the combination of deterministic approximation methods with stochastic estimation procedures to obtain algorithmic guarantees under the exclusion–inclusion gap across a broad class of finite-evaluation settings.

We illustrate the application of these results by analyzing a nonadaptive, equal-allocation algorithm for global multiobjective simulation optimization with continuous objectives and a compact feasible set. Combining low-dispersion coverage guarantees with concentration bounds that control the stochastic error yields finite-sample probability guarantees (Theorem 2, Corollary 1), fixed-precision and fixed-budget allocation rules (Theorems 3 and 4), sample-complexity results (Corollary 2), and convergence rates (Corollary 3). Measuring the exclusion–inclusion gap under the infinity norm, and for fixed tolerance and confidence requirements with bounded problem parameters, the sufficient simulation budget grows logarithmically in the number of objectives. Moreover, both the fixed-precision and fixed-budget allocation rules attain the convergence rate $O_p((b/\ln b)^{-1/(q+2)})$, where q is the decision space dimension and $O_p(\cdot)$ denotes boundedness in probability; here, the exponent depends on the decision-space dimension but not on the objective space dimension. Thus, under analogous Lipschitz and low-dispersion assumptions, the exclusion–inclusion gap converges at the same rate as the classical single-objective optimality gap in Yakowitz et al. (2000), despite measuring the convergence of the true image of the MOSO algorithm’s returned solution set, $f(\hat{\mathcal{E}}_m(b))$, and accounting for both exclusion and inclusion errors.

1.4 Notation and Terminology

When comparing $y, y' \in \mathbb{R}^d$, we write $y' \leq y$ if $y'_k \leq y_k$ for all $k = 1, \dots, d$, and write $y' \leq y$ if $y' \leq y$ and $y' \neq y$. For $p \geq 1$, the p -norm is $\|y\|_p := (|y_1|^p + \dots + |y_d|^p)^{1/p}$, and the ∞ -norm is $\|y\|_\infty := \max\{|y_1|, \dots, |y_d|\}$. Define $\mathbb{R}_{\geq}^d := \{y \in \mathbb{R}^d : y \geq 0\}$ and $\mathbb{R}_{>}^d := \{y \in \mathbb{R}^d : y > 0\}$. For sets $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^d$, $\mathcal{A} + \mathcal{B} := \{x + y : x \in \mathcal{A}, y \in \mathcal{B}\}$ is the Minkowski sum. For nonempty sets $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^d$ and $x \in \mathbb{R}^d$, $\mathbb{D}_p(x, \mathcal{B}) := \inf_{y \in \mathcal{B}} \|x - y\|_p$ and $\mathbb{D}_p(\mathcal{A}, \mathcal{B}) := \sup_{x \in \mathcal{A}} \mathbb{D}_p(x, \mathcal{B})$. The Hausdorff distance is $\mathbb{H}_p(\mathcal{A}, \mathcal{B}) := \max\{\mathbb{D}_p(\mathcal{A}, \mathcal{B}), \mathbb{D}_p(\mathcal{B}, \mathcal{A})\}$. We allow $\mathbb{D}_p(\mathcal{A}, \mathcal{B})$ and $\mathbb{H}_p(\mathcal{A}, \mathcal{B})$ to take values in $[0, \infty]$. When the choice of norm is immaterial, we write $\mathbb{H}$ instead of $\mathbb{H}_p$. A set $\mathcal{S} \subseteq \mathbb{R}^d$ has complement $\mathcal{S}^c := \mathbb{R}^d \setminus \mathcal{S}$. The upper image of $\mathcal{S}$ is $\mathcal{S}^\uparrow := \text{cl}(\mathcal{S} + \mathbb{R}_{\geq}^d)$, and the lower image is $\mathcal{S}^\downarrow := \text{cl}(\mathcal{S} - \mathbb{R}_{\geq}^d)$; see, e.g., Löhne (2011). The set $\mathcal{S}^\uparrow$ is *upper closed* if $s \in \mathcal{S}^\uparrow$ and $r \in \mathbb{R}_{\geq}^d$ implies $s + r \in \mathcal{S}^\uparrow$. We say $\mathcal{S}$ is $\mathbb{R}_{\geq}^d$ -compact if for all $s \in \mathcal{S}$, the section $(s - \mathbb{R}_{\geq}^d) \cap \mathcal{S}$ is compact. Let $\mathcal{S}_N$ and $\mathcal{S}_{wN}$ be the sets of all nondominated and weakly nondominated points in

$\mathcal{S}$, respectively, also called the Pareto and weakly Pareto sets: $\mathcal{S}_N := \{s \in \mathcal{S} : \nexists s' \in \mathcal{S} \text{ with } s' \leq s\}$, $\mathcal{S}_{wN} := \{s \in \mathcal{S} : \nexists s' \in \mathcal{S} \text{ with } s' < s\}$. The nondominated set $\mathcal{S}_N$ is *externally stable* if for each $y \in \mathcal{S} \setminus \mathcal{S}_N$, there exists $y^* \in \mathcal{S}_N$ such that $y \in y^* + \mathbb{R}_{\geq}^d$ (Ehrgott 2005, p. 33). If $\mathcal{S} \subset \mathbb{R}^d$ is nonempty and $\mathbb{R}_{\geq}^d$ -compact, then $\mathcal{S}_N$ is externally stable and $\mathcal{S} \subset \mathcal{S}_N + \mathbb{R}_{\geq}^d$.

2 MOSO Problem Context

We define the MOSO problem under the widely-used expected-value formulation as

$$\begin{aligned} & \text{minimize} && (\mathbb{E}[F_1(x, \xi)], \dots, \mathbb{E}[F_d(x, \xi)])^\top \\ & \text{subject to} && x \in \mathcal{X}, \end{aligned} \tag{1}$$

where the feasible set $\mathcal{X} \subseteq \mathbb{R}^q$ is nonempty, $F_k: \mathcal{X} \times \Xi \rightarrow \mathbb{R}$, $k = 1, \dots, d$ are real-valued random functions, $\xi: \Omega \rightarrow \Xi$ is a random object defined in some probability space, and for each $x \in \mathcal{X}$, the expected value is $\mathbb{E}[F_k(x, \xi)] = \int_{\Xi} F_k(x, u) dP(u)$. The formulation in (1) includes both direct minimization of expected performance measures and minimization of expected losses defined from those measures. For notational convenience, let

$$\begin{aligned} f(x) &= (f_1(x), \dots, f_d(x))^\top = \mathbb{E}[F(x, \xi)] \\ &= (\mathbb{E}[F_1(x, \xi)], \dots, \mathbb{E}[F_d(x, \xi)])^\top. \end{aligned}$$

Thus, the unknown mapping $f: \mathcal{X} \rightarrow \mathbb{R}^d$ comprises $d \geq 1$ real-valued objectives $f_k: \mathcal{X} \rightarrow \mathbb{R}$, $k = 1, \dots, d$, and $F: \mathcal{X} \times \Xi \rightarrow \mathbb{R}^d$ denotes the corresponding vector-valued random function. The solution to (1) is the efficient set $\mathcal{E}$ and its image is the global Pareto set $\mathcal{P}$,

$$\begin{aligned} \mathcal{E} &:= \{x^* \in \mathcal{X} : \nexists x \in \mathcal{X} \text{ with } f(x) \leq f(x^*)\}, \\ \mathcal{P} &= f(\mathcal{E}) = \{f(x^*) \in \mathbb{R}^d : x^* \in \mathcal{E}\}. \end{aligned}$$

Because it characterizes the attainable trade-offs, our primary object of interest is the Pareto set $\mathcal{P}$. However, practical algorithms generally return a finite approximation constructed from a finite set of evaluated points. Consequently, in what follows, we distinguish between the true Pareto set $\mathcal{P}$, its finite deterministic approximation, and the relevant stochastic estimators. To begin formalizing this distinction, let $\mathcal{X}_m \subseteq \mathcal{X}$ contain m feasible points evaluated by an algorithm. Restricting (1) to $\mathcal{X}_m$ yields the approximate problem

$$\begin{aligned} & \text{minimize} && (\mathbb{E}[F_1(x, \xi)], \dots, \mathbb{E}[F_d(x, \xi)])^\top \\ & \text{subject to} && x \in \mathcal{X}_m. \end{aligned} \tag{2}$$

The solution to (2) is a finite approximation to the solution to (1). We refer to this quantity as the

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

$$\begin{aligned}\mathcal{E}_m &:= \{x \in \mathcal{X}_m : \nexists x' \in \mathcal{X}_m \text{ with } f(x') \leq f(x)\}, \\ \mathcal{P}_m &:= f(\mathcal{E}_m) = \{f(x) \in \mathbb{R}^d : x \in \mathcal{E}_m\}.\end{aligned}$$

The construction of $\mathcal{X}_m$ is generic and may be implicit. For example, $\mathcal{X}_m$ may be determined implicitly by an algorithm, or it may consist of solutions obtained by scalarizing (1) and solving the resulting single-objective problem for a predetermined finite set of scalarization parameter values.

To ensure that the true and approximate Pareto sets from (1) and (2) are well-defined, and that our upper image constructions in Section 3 are valid, we impose the following standing assumption on the true image set $f(\mathcal{X})$ and the finite approximation set $\mathcal{X}_m$.

Assumption 1. *The following hold:*

1. *The efficient set $\mathcal{E} \subseteq \mathcal{X}$ is nonempty.*
2. *The image set $f(\mathcal{X})$ is $\mathbb{R}_{\geq}^d$ -compact.*
3. *The image set $f(\mathcal{X})$ is bounded below; that is, $y_* \leq f(x)$ for some $y_* \in \mathbb{R}^d$ and all $f(x) \in f(\mathcal{X})$.*
4. *For each finite approximation considered, $\mathcal{X}_m \subseteq \mathcal{X}$ is nonempty and finite with $|\mathcal{X}_m| = m \geq 1$.*

Assumption 1 Parts 1 and 2 ensure that $\mathcal{P}$ is nonempty and externally stable, so that every attainable objective vector is dominated by a Pareto point; see Subsection 1.4. Assumption 1 Part 3 ensures that the distances between the upper image of the Pareto set, $\mathcal{P}^\uparrow$, and the upper images based on the finite approximation, such as $\mathcal{P}_m^\uparrow$ and $f(\hat{\mathcal{E}}_m(b))^\uparrow$, remain bounded whenever $\mathcal{P}_m$ is nonempty. Finally, requiring $\mathcal{X}_m$ to be nonempty and finite implies that $\mathcal{E}_m$ and $\mathcal{P}_m$ are nonempty and finite.

Assumption 1 is stated in terms of the image set $f(\mathcal{X})$ so that the results apply under a variety of MOSO problem types. In particular, $\mathcal{P}$ may either be infinite or finite but too large to enumerate. If $\mathcal{P}$ is finite and sufficiently small that every feasible point can be evaluated, as in multi-objective ranking and selection, then the approximation error equals zero because $\mathcal{P} = \mathcal{P}_m$, and the analysis applies with estimation error only. One set of sufficient conditions for Assumption 1 Parts 1–3 is that $\mathcal{X} \subset \mathbb{R}^q$ is nonempty and compact and each $f_k : \mathcal{X} \rightarrow \mathbb{R}$ is continuous. Until Section 5, our results rely only on Assumption 1. Since Assumption 1 does not require an expected-value formulation in (1), our analysis applies under other appropriate statistical functionals used to resolve the uncertainty.

Because f is unknown, the objective values at evaluated feasible points must be estimated. For an evaluated point $x \in \mathcal{X}$, a common estimator under (1) is the sample average

$$\hat{F}(x) = n^{-1} \sum_{i=1}^n F(x, \xi_i) \tag{3}$$

where $\xi_1, \dots, \xi_n$ are random inputs or random objects and $\hat{F}$ is composed of individual estimators $\hat{F}_k$, $k = 1, \dots, d$. At this stage, we impose no particular assumptions on the random inputs. We neither require nor preclude the use of common random numbers; the random inputs $\xi_1, \dots, \xi_n$ may depend on x . For notational simplicity, we suppress the dependence of $\hat{F}$ on n until [Section 5](#) because up to that point, the sample size n may also depend on x . After estimating the objective vector at each point in $\mathcal{X}_m$ using a total budget of b samples from a dataset or simulation oracle, the *estimated approximate efficient set* is

$$\hat{\mathcal{E}}_m(b) := \{x \in \mathcal{X}_m : \nexists x' \in \mathcal{X}_m \text{ with } \hat{F}(x') \leq \hat{F}(x)\}. \quad (4)$$

The corresponding *estimated approximate Pareto set* is $\hat{\mathcal{P}}_m(b) = \{\hat{F}(x) : x \in \hat{\mathcal{E}}_m(b)\}$. Although $\hat{\mathcal{P}}_m(b)$ is returned by the algorithm and is the only set seen by a real-world decision-maker, our analysis focuses on the theoretical quantity

$$f(\hat{\mathcal{E}}_m(b)) := \{f(x) : x \in \hat{\mathcal{E}}_m(b)\},$$

which is the set containing the true objective vectors of the returned solutions. Until [Section 5](#), the total budget b used for estimation is defined generally and may, for example, equal n , mn , or some other value; [Section 5](#) specializes to equal allocation with $b = mn$. Finally, we use the sample average as a canonical estimator, but the results of [Sections 3](#) and [4](#) apply under other estimators that are consistent, as required by [Proposition 2](#), and for which componentwise stochastic-error control can be established, as required by [Lemma 2](#).

3 Performance Indicators for MOSO Under Finite Approximation

Under the finite approximation in [\(2\)](#), recall that we assess the quality of the returned estimated approximate efficient solutions from [\(4\)](#) through their true images, $f(\hat{\mathcal{E}}_m(b))$. The approximate Pareto set $\mathcal{P}_m$ provides an intermediate deterministic object: comparing $\mathcal{P}$ with $\mathcal{P}_m$ captures the approximation error, while comparing $\mathcal{P}_m$ with $f(\hat{\mathcal{E}}_m(b))$ captures the stochastic estimation error on the approximate problem. This separation is the multiobjective analogue of the approximation–estimation error decomposition used in single-objective stochastic optimization. To make the analogy precise, let $d = 1$, and let $X_m^*(b)$ be an estimated approximate solution that belongs to $\arg \min\{\hat{F}_1(x) : x \in \mathcal{X}_m\}$. Let x_m^* solve the corresponding true approximate problem, and let x^* solve the true problem. Then, the optimality gap admits the decomposition

$$f_1(X_m^*(b)) - f_1(x^*) = \underbrace{f_1(x_m^*) - f_1(x^*)}_{\text{approximation error}} + \underbrace{f_1(X_m^*(b)) - f_1(x_m^*)}_{\text{estimation error}}. \quad (5)$$

Thus, in the single-objective case, the optimality gap provides a natural and straightforward basis for separating the approximation and estimation errors (Yakowitz et al. 2000). In MOSO, however, the target and returned solutions are set-valued. A corresponding performance indicator intended to play the role of the optimality gap must compare $\mathcal{P}$ with $f(\hat{\mathcal{E}}_m(b))$ while remaining meaningful when stochastic error causes the latter to contain elements that weakly dominate others. In such a case, $f(\hat{\mathcal{E}}_m(b))$ is not an approximation set in the formal sense, and performance indicators designed for approximation sets may not directly capture some forms of inclusion-related nonconvergence.

In what follows, we first examine the Hausdorff distance as a natural candidate for the role of the optimality gap in MOSO. Although it provides an approximation–estimation error decomposition through the triangle inequality, the inclusion components present obstacles to obtaining general bounds. We then introduce the exclusion–inclusion gap, whose upper-image constructions yield approximation and estimation terms that admit general bounds.

3.1 The Hausdorff Distance

The Hausdorff distance between the true Pareto set $\mathcal{P}$ and the true image of the returned solutions $f(\hat{\mathcal{E}}_m(b))$, evaluated under the p -norm, is

$$\begin{aligned} \mathbb{H}_p(\mathcal{P}, f(\hat{\mathcal{E}}_m(b))) \\ = \max\{\mathbb{D}_p(\mathcal{P}, f(\hat{\mathcal{E}}_m(b))), \mathbb{D}_p(f(\hat{\mathcal{E}}_m(b)), \mathcal{P})\}. \end{aligned} \quad (6)$$

The Hausdorff distance is an attractive performance indicator because it equals zero if and only if the two sets have the same closure, requires no reference point or parameter apart from the norm, and admits an approximation–estimation error decomposition through the triangle inequality,

$$\begin{aligned} \mathbb{H}_p(\mathcal{P}, f(\hat{\mathcal{E}}_m(b))) \\ \leq \underbrace{\mathbb{H}_p(\mathcal{P}, \mathcal{P}_m)}_{\text{approximation error}} + \underbrace{\mathbb{H}_p(\mathcal{P}_m, f(\hat{\mathcal{E}}_m(b)))}_{\text{estimation error}} \\ = \max\left\{ \underbrace{\mathbb{D}_p(\mathcal{P}, \mathcal{P}_m)}_{\text{coverage error}}, \underbrace{\mathbb{D}_p(\mathcal{P}_m, \mathcal{P})}_{\text{representation error}} \right\} + \max\left\{ \underbrace{\mathbb{D}_p(\mathcal{P}_m, f(\hat{\mathcal{E}}_m(b)))}_{\text{false exclusion error}}, \underbrace{\mathbb{D}_p(f(\hat{\mathcal{E}}_m(b)), \mathcal{P}_m)}_{\text{false inclusion error}} \right\}. \end{aligned} \quad (7)$$

The first maximum in (7) measures the deterministic approximation error. Its terms are the coverage error, which measures how well $\mathcal{P}_m$ covers $\mathcal{P}$ (Sayin 2000) and can be considered a measure of exclusion, and the representation error, which measures how closely points in $\mathcal{P}_m$ represent $\mathcal{P}$ (Ruzika 2007, Hamacher et al. 2007) and can be considered a measure of inclusion. The second maximum in (7) measures the stochastic estimation error relative to the approximate problem. Its terms account for the false exclusion of points in $\mathcal{P}_m$ and the false inclusion of points in $f(\mathcal{X}_m) \setminus \mathcal{P}_m$.

Unfortunately, the inclusion terms in the decomposition in (7) are difficult to analyze. The

inclusion component of the approximation error term presents the first analysis difficulty. Coverage error guarantees control $\mathbb{D}_p(\mathcal{P}, \mathcal{P}_m)$ but do not automatically control the representation error term, $\mathbb{D}_p(\mathcal{P}_m, \mathcal{P})$. Some specialized constructions use additional structure to control both directions, as in [Kuhn and Ruzika \(2017\)](#) and [Ondes and Hunter \(2023\)](#). However, these results do not appear to extend directly to arbitrary finite sets $\mathcal{X}_m$ and general, possibly nonconvex, objective mappings. The inclusion component of the estimation error term presents a second analysis difficulty. If the restricted image set $f(\mathcal{X}_m)$ contains weakly nondominated points that do not belong to $\mathcal{P}_m$, such that $(f(\mathcal{X}_m))_{\text{wN}} \setminus \mathcal{P}_m$ is nonempty, then each of these points has zero separation from at least one point in $\mathcal{P}_m$ on at least one objective. Therefore, they may remain statistically indistinguishable from a nondominated point under finite sampling. Controlling the false inclusion error in (7) may require either assuming that such points do not exist or explicitly disregarding their misclassification; related approaches appear in [Feldman and Hunter \(2018\)](#), [Applegate et al. \(2020\)](#), [Cooper et al. \(2020\)](#), and [Ondes \(2025\)](#). Thus, although (7) separates the approximation and estimation errors, its four one-way distances do not provide a general path to the bounds we seek.

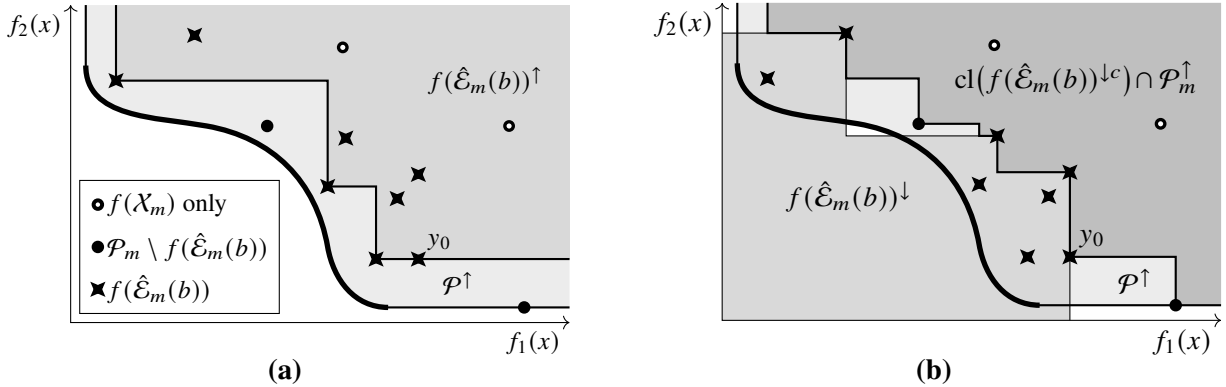
3.2 The Exclusion–Inclusion Gap

Motivated by the limitations of the Hausdorff decomposition in (7), we introduce a new performance indicator called the *exclusion–inclusion gap*, which combines the exclusion and inclusion components in a different way. Given the finite estimated approximate efficient set $\hat{\mathcal{E}}_m(b)$, the gap measures two errors relative to the upper image of the true Pareto set, $\mathcal{P}^\uparrow$: (i) the *exclusion error* due to excluding Pareto points from $f(\hat{\mathcal{E}}_m(b))^\uparrow$, whereby $f(\hat{\mathcal{E}}_m(b))^\uparrow$ fails to cover part of $\mathcal{P}^\uparrow$, and (ii) the *inclusion error* due to including non-Pareto points in $f(\hat{\mathcal{E}}_m(b))^\downarrow$, whereby the lower image set $f(\hat{\mathcal{E}}_m(b))^\downarrow$ extends further into the dominated region and shrinks its complement within $\mathcal{P}_m^\uparrow$. Each error is measured as a one-way set distance. The upper-image constructions yield nested comparison sets, so the relevant one-way set distances coincide with Hausdorff distances. The triangle inequality gives an approximation–estimation error decomposition whose terms admit the general bounds developed in [Section 4](#). Next, we discuss the exclusion and inclusion errors, define the exclusion–inclusion gap, and establish its main properties.

3.2.1 Exclusion Error

We measure the exclusion error as $\mathbb{D}_p(\mathcal{P}^\uparrow, f(\hat{\mathcal{E}}_m(b))^\uparrow)$. This distance measures the extent to which $f(\hat{\mathcal{E}}_m(b))^\uparrow$ fails to cover $\mathcal{P}^\uparrow$ because of evaluating only finitely many points or excluding points due to stochastic estimation error ([Hunter and Ondes 2023](#)). Since $\hat{\mathcal{E}}_m(b) \subseteq \mathcal{X}_m$ and $\mathcal{P}_m$ is externally stable relative to $f(\mathcal{X}_m)$, the upper image sets satisfy the nesting

$$f(\hat{\mathcal{E}}_m(b))^\uparrow \subseteq f(\mathcal{X}_m)^\uparrow = \mathcal{P}_m^\uparrow \subseteq \mathcal{P}^\uparrow. \quad (8)$$

Figure 1: The illustrations show upper- and lower-image constructions for an example with two objectives.

Note: Panel (a) shows $y_0 \in (f(\mathcal{X}_m))_{\text{wN}} \setminus \mathcal{P}_m$, which is weakly nondominated for the finite approximate problem. In Panel (b), $\text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow$ is the same regardless of whether or not y_0 belongs to $f(\hat{\mathcal{E}}_m(b))$.

Figure 1(a) illustrates the first and last sets in this nesting. Due to (8), each relevant Hausdorff distance reduces to a one-way set distance. Applying the triangle inequality therefore gives

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

3.2.2 Inclusion Error

To measure the inclusion error, we begin with the lower image set $f(\hat{\mathcal{E}}_m(b))^\downarrow$, which is illustrated in Figure 1(b). We begin with this set because a falsely returned point that is dominated relative to other points in $f(\hat{\mathcal{E}}_m(b))$ may enlarge $f(\hat{\mathcal{E}}_m(b))^\downarrow$ relative to the corresponding set without such points, $((f(\hat{\mathcal{E}}_m(b)))_{\text{N}})^\downarrow$. To compare $f(\hat{\mathcal{E}}_m(b))^\downarrow$ with $\mathcal{P}^\uparrow$, we take its complement, close the result, and restrict it to $\mathcal{P}_m^\uparrow$ to yield the inclusion comparison set $\text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow$. This set satisfies the nesting

$$\text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow \subseteq \mathcal{P}_m^\uparrow \subseteq \mathcal{P}^\uparrow. \tag{10}$$

Figure 1(b) illustrates the inclusion comparison set $\text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow$. Then, we measure the inclusion error as $\mathbb{D}_p(\mathcal{P}^\uparrow, \text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow)$. Here, a falsely included point that is strictly dominated with respect to (2) enlarges $f(\hat{\mathcal{E}}_m(b))^\downarrow$, thereby shrinking its complement within $\mathcal{P}_m^\uparrow$ and potentially increasing the distance from $\mathcal{P}^\uparrow$. If $f(\hat{\mathcal{E}}_m(b)) = f(\hat{\mathcal{E}}_m(b))^\uparrow \cap f(\hat{\mathcal{E}}_m(b))^\downarrow$, then $f(\hat{\mathcal{E}}_m(b))$ is an approximation set, as illustrated in Figure 2(a). As with the exclusion error, the nesting in (10) reduces the relevant Hausdorff distances to one-way set distances. The triangle inequality therefore gives

$$\begin{aligned}
\mathbb{D}_p(\mathcal{P}^\uparrow, \text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow) &= \mathbb{H}_p(\mathcal{P}^\uparrow, \text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow) \\
&\leq \mathbb{H}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow) + \mathbb{H}_p(\mathcal{P}_m^\uparrow, \text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow) \\
&= \mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow) + \mathbb{D}_p(\mathcal{P}_m^\uparrow, \text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow).
\end{aligned} \tag{11}$$

This construction of the inclusion error comparison set is especially useful when $f(X_m)$ contains weakly nondominated points outside $\mathcal{P}_m$. A point $y_0 \in (f(X_m))_{\text{wN}} \setminus \mathcal{P}_m$ may be falsely included in $f(\hat{\mathcal{E}}_m(b))$ because it has zero separation from a point in $\mathcal{P}_m$ on at least one objective. However, returning such a point does not change the comparison set, $\text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow$. Thus, the inclusion error does not penalize this type of boundary misclassification. It penalizes only geometrically consequential inclusion under the upper image construction, rather than all classification disagreement. Figure 1 illustrates this invariance, and later, Proposition 1 establishes it formally.

3.2.3 Gap Definition and Properties

Taking the maximum of the exclusion and inclusion errors yields the *exclusion–inclusion gap*, a performance indicator that plays the role of an optimality-gap analogue in our finite-approximation setting:

$$G_m(b) := \max \left\{ \underbrace{\mathbb{D}_p(\mathcal{P}^\uparrow, f(\hat{\mathcal{E}}_m(b))^\uparrow)}_{\text{exclusion error}}, \underbrace{\mathbb{D}_p(\mathcal{P}^\uparrow, \text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow)}_{\text{inclusion error}} \right\}. \tag{12}$$

Intuitively, the exclusion–inclusion gap measures the larger of two errors: the failure of $f(\hat{\mathcal{E}}_m(b))^\uparrow$ to cover $\mathcal{P}^\uparrow$, and the effect of returned dominated points that enlarge the lower image set $f(\hat{\mathcal{E}}_m(b))^{\downarrow}$ within $\mathcal{P}_m^\uparrow$ and thereby alter the inclusion comparison set $\text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow$. To state its approximation–estimation error decomposition, first, define the *false exclusion* (FE) and *false inclusion* (FI) errors relative to the approximate problem as

$$\begin{aligned}
G_m^{\text{FE}}(b) &:= \mathbb{D}_p(\mathcal{P}_m^\uparrow, f(\hat{\mathcal{E}}_m(b))^\uparrow), \\
G_m^{\text{FI}}(b) &:= \mathbb{D}_p(\mathcal{P}_m^\uparrow, \text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow).
\end{aligned}$$

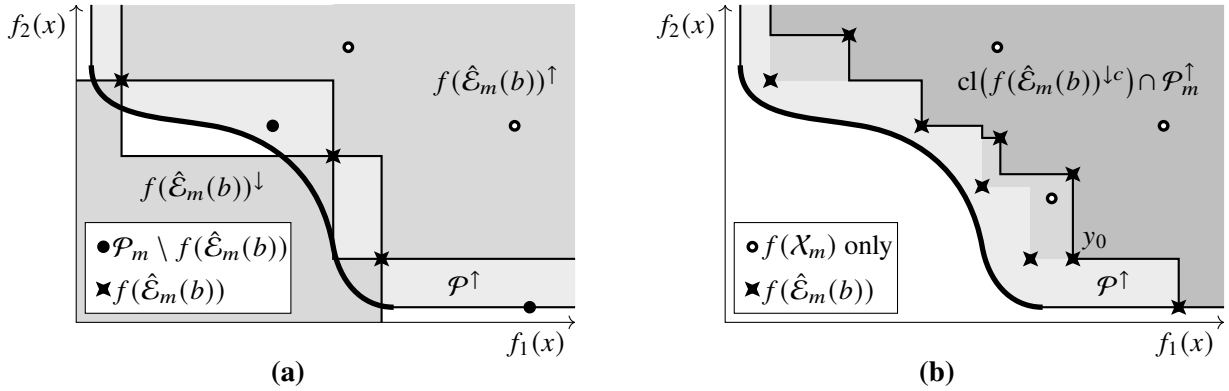
Then, combining (9) and (11) implies

$$G_m(b) \leq \underbrace{\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)}_{\text{approximation error}} + \underbrace{\max \{G_m^{\text{FE}}(b), G_m^{\text{FI}}(b)\}}_{\text{estimation error}}. \tag{13}$$

For simplicity, we suppress the dependence of $G_m(b)$, $G_m^{\text{FE}}(b)$, $G_m^{\text{FI}}(b)$ on the choice of norm p .

Theorem 1 states the basic properties of the exclusion–inclusion gap defined in (12). In particular, it characterizes when the gap vanishes, identifies its reductions when $f(\hat{\mathcal{E}}_m(b))$ is an approximation set or false exclusion is absent, and formalizes the decomposition in (13). Figure 2 illustrates these two cases. Finally, we remark that when $d = 1$, the exclusion–inclusion gap also recovers the usual single-objective optimality gap. If the estimated minimizer is a singleton $\{X_m^*(b)\}$, then

Figure 2: The illustrations show aspects of [Theorem 1](#) for two different examples, each having two objectives.



Note: In panel (a), $f(\hat{E}_m(b)) = f(\hat{E}_m(b))^\uparrow \cap f(\hat{E}_m(b))^\downarrow$ implies that $f(\hat{E}_m(b))$ is an approximation set, and the exclusion–inclusion gap equals the exclusion error. In panel (b), $f(\hat{E}_m(b)) \supseteq \mathcal{P}_m$ implies no false exclusion and the exclusion–inclusion gap equals the inclusion error.

$G_m(b) = f_1(X_m^*(b)) - f_1(x^*)$. Without unique selection, $G_m(b)$ equals the largest true optimality gap among the returned estimated minimizers.

Theorem 1 (Basic Properties of the Exclusion–Inclusion Gap). *Let [Assumption 1](#) hold.*

1. The component errors and gap equal zero if and only if the relevant comparison sets are equal:

- (a) $\mathbb{D}_p(\mathcal{P}^\uparrow, f(\hat{E}_m(b))^\uparrow) = 0$ if and only if $\mathcal{P}^\uparrow = f(\hat{E}_m(b))^\uparrow$,
- (b) $\mathbb{D}_p(\mathcal{P}^\uparrow, \text{cl}(f(\hat{E}_m(b))^\downarrow) \cap \mathcal{P}_m^\uparrow) = 0$ if and only if $\mathcal{P}^\uparrow = \text{cl}(f(\hat{E}_m(b))^\downarrow) \cap \mathcal{P}_m^\uparrow$, and
- (c) $G_m(b) = 0$ if and only if $\mathcal{P}^\uparrow = f(\hat{E}_m(b))^\uparrow = \text{cl}(f(\hat{E}_m(b))^\downarrow) \cap \mathcal{P}_m^\uparrow$.

2. If $f(\hat{E}_m(b)) = f(\hat{E}_m(b))^\uparrow \cap f(\hat{E}_m(b))^\downarrow$, then $f(\hat{E}_m(b))$ is an approximation set, and the exclusion–inclusion gap equals the exclusion error, $G_m(b) = \mathbb{D}_p(\mathcal{P}^\uparrow, f(\hat{E}_m(b))^\uparrow)$.
3. If $f(\hat{E}_m(b)) \supseteq \mathcal{P}_m$, then there is no false exclusion and the exclusion–inclusion gap equals the inclusion error, $G_m(b) = \mathbb{D}_p(\mathcal{P}^\uparrow, \text{cl}(f(\hat{E}_m(b))^\downarrow) \cap \mathcal{P}_m^\uparrow)$.
4. The exclusion–inclusion gap admits the approximation–estimation error decomposition $G_m(b) \leq \mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow) + \max\{G_m^{\text{FE}}(b), G_m^{\text{FI}}(b)\}$.

Proof. Under [Assumption 1](#), $X_m \neq \emptyset$ is finite and $\mathcal{P} \neq \emptyset$ is externally stable. The finite image set $f(X_m)$ therefore has a nonempty nondominated set $\mathcal{P}_m$, which is externally stable relative to $f(X_m)$. Thus, the nesting relations in (8) and (10) hold. Parts 1, 3, and 4 follow from these relations, closedness of the comparison sets, the Hausdorff-distance equivalences in (9) and (11), and the definition of $G_m(b)$ in (12). For Part 2, the stated condition and finiteness imply that the points in $f(\hat{E}_m(b))$ are mutually nondominated, whereby $f(\hat{E}_m(b))^\uparrow \subseteq \text{cl}(f(\hat{E}_m(b))^\downarrow) \cap \mathcal{P}_m^\uparrow$. Thus, the inclusion error is no larger than the exclusion error, so the gap equals the exclusion error. $\square$

[Theorem 1](#) shows that the upper-image constructions preserve key exclusion and inclusion comparisons between $\mathcal{P}$, $\mathcal{P}_m$, and $f(\hat{E}_m(b))$, but avoid direct reliance on the four one-way terms

in the Hausdorff distance’s decomposition in (7). Next, [Proposition 1](#) formalizes the inclusion component’s invariance to the return of weakly nondominated points. Applying [Proposition 1](#) with $\mathcal{A} = f(\hat{\mathcal{E}}_m(b))$ demonstrates that returning any subset of weakly nondominated points belonging to $f(\mathcal{X}_m)$ changes neither the inclusion comparison set nor $G_m^{\text{FI}}(b)$.

Proposition 1 (False-Inclusion Invariance for Weakly Nondominated Points). *Under [Assumption 1](#), if $\mathcal{A} \subseteq f(\mathcal{X}_m)$ and $\mathcal{W} \subseteq (f(\mathcal{X}_m))_{\text{wN}} \setminus \mathcal{A}$, then $\text{cl}(\mathcal{A}^{\downarrow c}) \cap \mathcal{P}_m^\uparrow = \text{cl}((\mathcal{A} \cup \mathcal{W})^{\downarrow c}) \cap \mathcal{P}_m^\uparrow$.*

Proof. Since $\mathcal{A}^\downarrow \subseteq (\mathcal{A} \cup \mathcal{W})^\downarrow$, taking complements, closures, and intersections with $\mathcal{P}_m^\uparrow$ gives $\text{cl}(\mathcal{A}^{\downarrow c}) \cap \mathcal{P}_m^\uparrow \supseteq \text{cl}((\mathcal{A} \cup \mathcal{W})^{\downarrow c}) \cap \mathcal{P}_m^\uparrow$. To show the reverse inclusion, let $z \in \text{cl}(\mathcal{A}^{\downarrow c}) \cap \mathcal{P}_m^\uparrow$. Since $(\text{cl}(\mathcal{A}^{\downarrow c}))^c = \text{int}(\mathcal{A}^\downarrow)$, we have $z \notin \text{int}(\mathcal{A}^\downarrow)$. For a contradiction, suppose that $z \notin \text{cl}((\mathcal{A} \cup \mathcal{W})^{\downarrow c})$, which implies $z \in \text{int}((\mathcal{A} \cup \mathcal{W})^\downarrow)$. Since $\mathcal{A} \cup \mathcal{W}$ is finite, there exists $a \in \mathcal{A} \cup \mathcal{W}$ such that $z < a$. If $a \in \mathcal{A}$, then $z < a$ would imply $z \in \text{int}(\mathcal{A}^\downarrow)$, which is a contradiction. Therefore, $a \in \mathcal{W} \subseteq (f(\mathcal{X}_m))_{\text{wN}}$. Finally, since $z \in \mathcal{P}_m^\uparrow$, there exists $\tilde{y}^* \in \mathcal{P}_m$ such that $\tilde{y}^* \leq z$. Therefore, $\tilde{y}^* < a$, which contradicts $a \in (f(\mathcal{X}_m))_{\text{wN}}$. The reverse inclusion follows. $\square$

3.2.4 Statistical Interpretation of False-Inclusion Invariance

We emphasize that [Proposition 1](#) concerns classification stability rather than estimator consistency. A dominated but weakly nondominated point in $(f(\mathcal{X}_m))_{\text{wN}} \setminus \mathcal{P}_m$ may continue to enter $\hat{\mathcal{E}}_m(b)$ because the sign of an estimated difference on a tied objective may not stabilize. Nevertheless, fixing $\mathcal{X}_m$ and assuming $\hat{F}_k(x) \rightarrow f_k(x)$ w.p.1 as $b \rightarrow \infty$ for all $x \in \mathcal{X}_m$ and all $k = 1, \dots, d$, the estimated objective vectors of the returned solutions converge uniformly to their corresponding true values on the finite set $\mathcal{X}_m$. Furthermore, $\hat{\mathcal{P}}_m(b)$ and $f(\hat{\mathcal{E}}_m(b))$ are generated by applying $\hat{F}$ and f , respectively, to the same solution set $\hat{\mathcal{E}}_m(b)$. Thus, for every b and every $x \in \hat{\mathcal{E}}_m(b)$, the corresponding points $\hat{F}(x)$ and $f(x)$ are separated by at most $\Delta_m(b) := \max_{x \in \mathcal{X}_m} \|\hat{F}(x) - f(x)\|_p$, where $\Delta_m(b) \rightarrow 0$ w.p. 1 as $b \rightarrow \infty$. Thus, the decision-maker is not asymptotically misled about the objective values of the solutions that remain under consideration. [Proposition 2](#) formalizes this result without proof.

Proposition 2 (Value Consistency Under Weak-Dominance Ambiguity). *Let [Assumption 1](#) hold, fix $\mathcal{X}_m$, and suppose that for every $x \in \mathcal{X}_m$, $k = 1, \dots, d$, $\hat{F}_k(x) \rightarrow f_k(x)$ w.p. 1 as $b \rightarrow \infty$. Then $\mathbb{H}_p(\hat{\mathcal{P}}_m(b), f(\hat{\mathcal{E}}_m(b))) \leq \Delta_m(b)$ for every b , and $\mathbb{H}_p(\hat{\mathcal{P}}_m(b), f(\hat{\mathcal{E}}_m(b))) \rightarrow 0$ w.p. 1 as $b \rightarrow \infty$.*

To connect this result directly to [Proposition 1](#), consider $\tilde{x}, \tilde{x}' \in \mathcal{X}_m$ satisfying $f(\tilde{x}) \leq f(\tilde{x}')$. Every strict objective disadvantage is eventually estimated with the correct ordering. Moreover, for every objective k satisfying $f_k(\tilde{x}) = f_k(\tilde{x}')$, we have $|\hat{F}_k(\tilde{x}) - \hat{F}_k(\tilde{x}')| \leq 2\Delta_m(b)$. Therefore, if $\tilde{x}'$ remains estimated as nondominated relative to $\tilde{x}$, any apparent objective advantage supporting the unresolved classification converges to zero. Thus, [Proposition 2](#) demonstrates that the exclusion–inclusion gap disregards classification ambiguities among points whose apparent benefits vanish.

This observation also has a decision-theoretic interpretation. Let $s: \mathbb{R}^d \rightarrow \mathbb{R}$ be a scalarization function that is continuous and strictly increasing with respect to the componentwise order, so that if $y, y' \in \mathbb{R}^d$ and $y \leq y'$, then $s(y) < s(y')$. If $f(\tilde{x}) \leq f(\tilde{x}')$, then almost-sure estimator consistency and continuity of s imply $s(\hat{F}(\tilde{x})) \rightarrow s(f(\tilde{x}))$ and $s(\hat{F}(\tilde{x}')) \rightarrow s(f(\tilde{x}'))$ w.p. 1, where $s(f(\tilde{x})) < s(f(\tilde{x}'))$. Then, $s(\hat{F}(\tilde{x})) < s(\hat{F}(\tilde{x}'))$ for all sufficiently large b , w.p. 1. Therefore, if $\tilde{x}'$ continues to appear in $\hat{\mathcal{E}}_m(b)$ because its objective estimators fluctuate on objectives that are truly tied with those of points in $\mathcal{P}_m$, and $\tilde{x}$ is also returned, a decision-maker applying a scalarization function s over the estimated approximate Pareto set $\hat{\mathcal{P}}_m(b)$ eventually prefers the truly dominating solution $\tilde{x}$.

4 Approximation–Estimation Bounds for the Exclusion–Inclusion Gap

Given the approximation–estimation error decomposition in (13), we now derive general bounds on its terms. Together, these bounds allow deterministic finite-approximation error and stochastic estimation error guarantees to be combined in algorithm analysis, as demonstrated in Section 5. The section ends with commentary on the relationship to Hausdorff distance and the choice of norm.

4.1 Bounding the Approximation Error

Approximation methods have been studied in the deterministic multiobjective optimization literature for over forty years; see surveys by Ruzika and Wiecek (2005), Faulkenberg and Wiecek (2010), Herzel et al. (2021). Since our analysis uses finite point sets rather than interpolating approximations of $\mathcal{P}$, we consider deterministic guarantees for such finite point sets. Some methods provide a norm-based coverage error guarantee, as in Kirlik and Sayin (2018, 2025). Others provide an additive ε -approximation guarantee, under which every point in $\mathcal{P}$ is dominated within ε by an approximating point; see Ruzika and Wiecek (2005). Lemma 1 shows that either guarantee bounds the approximation error term $\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$ appearing in (13) by setting $\mathcal{A} = \mathcal{P}_m$.

Lemma 1 (Approximation Error from Coverage Error or Additive ε -Approximation). *Let Assumption 1 hold and $\mathcal{A} \subset \mathbb{R}^d$ be nonempty. Then the following hold.*

1. Coverage error: $\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{A}^\uparrow) \leq \mathbb{D}_p(\mathcal{P}, \mathcal{A})$.
2. Additive ε -approximation: *Let $\varepsilon \in \mathbb{R}_{\geq}^d$. If, for every $y^* \in \mathcal{P}$, there exists $a \in \mathcal{A}$ such that $a_k \leq y_k^* + \varepsilon_k$ for all $k = 1, \dots, d$, then $(\mathcal{P} + \varepsilon)^\uparrow \subseteq \mathcal{A}^\uparrow$ and $\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{A}^\uparrow) \leq \|\varepsilon\|_p$.*

Proof. For Part 1, let $\mathbb{D}_p(\mathcal{P}, \mathcal{A}) = u$, and suppose $u < \infty$ since otherwise, the result immediately follows. Assumption 1 implies that $\mathcal{P} \neq \emptyset$, and since $\mathcal{A} \subseteq \mathcal{A}^\uparrow$, we have $\mathbb{D}_p(\mathcal{P}, \mathcal{A}^\uparrow) \leq \mathbb{D}_p(\mathcal{P}, \mathcal{A}) = u < \infty$. Now, fix $y \in \mathcal{P}^\uparrow = \text{cl}(\mathcal{P} + \mathbb{R}_{\geq}^d)$. By the definition of closure, for any $\epsilon > 0$, there exists $y' \in \mathcal{P} + \mathbb{R}_{\geq}^d$ such that $\|y - y'\|_p < \epsilon$; further, $y' = y^* + r$ for some $y^* \in \mathcal{P}$ and $r \in \mathbb{R}_{\geq}^d$, whereby $\|y - (y^* + r)\|_p < \epsilon$. Since $\mathbb{D}_p(\mathcal{P}, \mathcal{A}^\uparrow) \leq u$, $y^* \in \mathcal{P}$ implies $\mathbb{D}_p(y^*, \mathcal{A}^\uparrow) \leq u$. Because $\mathcal{A}^\uparrow$ is nonempty and closed, there exists $a \in \mathcal{A}^\uparrow$ such that $\|y^* - a\|_p \leq u$. Now set $a' = a + r$,

where $a' \in \mathcal{A}^\uparrow$ because $\mathcal{A}^\uparrow$ is upper closed. Then $\mathbb{D}_p(y, \mathcal{A}^\uparrow) \leq \|y - a'\|_p = \|y - (a + r)\|_p \leq \|y - (y^* + r)\|_p + \|(y^* + r) - (a + r)\|_p \leq \epsilon + u$. Because $\epsilon > 0$ is arbitrary, letting $\epsilon \downarrow 0$ yields $\mathbb{D}_p(y, \mathcal{A}^\uparrow) \leq u$. Finally, because $y \in \mathcal{P}^\uparrow$ was arbitrary, the bound holds for every $y \in \mathcal{P}^\uparrow$. Taking the supremum yields the result.

For Part 2, for each $y^* \in \mathcal{P}$, we have $y^* + \varepsilon \in \mathcal{A}^\uparrow$. Because $\mathcal{A}^\uparrow$ is both closed and upper closed, $(\mathcal{P} + \varepsilon)^\uparrow \subseteq \mathcal{A}^\uparrow$. Thus, for any $y \in \mathcal{P}^\uparrow$, we have $y + \varepsilon \in (\mathcal{P} + \varepsilon)^\uparrow \subseteq \mathcal{A}^\uparrow$, which implies $\mathbb{D}_p(y, \mathcal{A}^\uparrow) \leq \|\varepsilon\|_p$. Taking the supremum over $y \in \mathcal{P}^\uparrow$ yields the result. $\square$

4.2 Bounding the Estimation Error

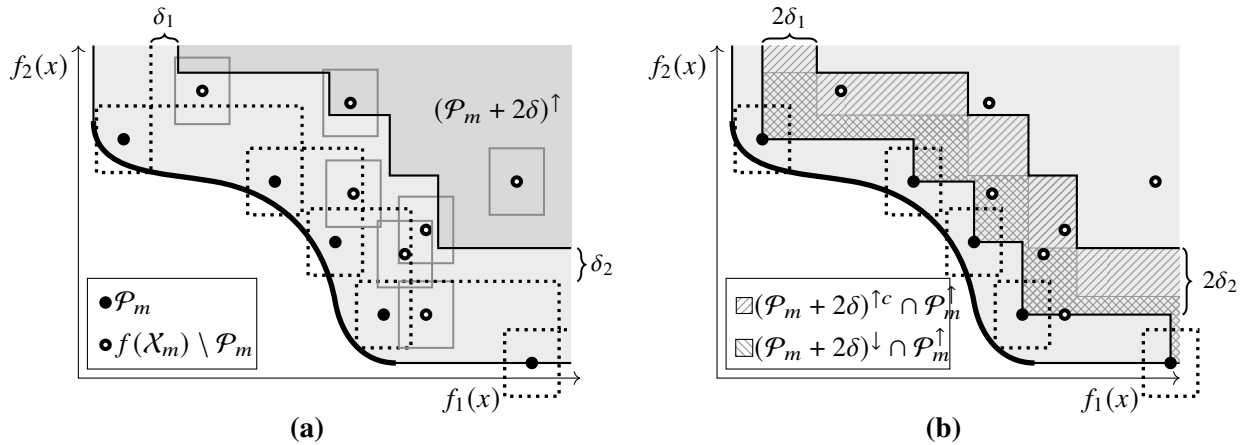
Next, we bound the estimation error arising from the stochastic objective estimates for fixed $\mathcal{X}_m$. The following lemma shows that uniform componentwise stochastic error control over $\mathcal{X}_m$ simultaneously bounds the false-exclusion and false-inclusion errors in (13). Figure 3 illustrates the sets used in the statement of Lemma 2, which involve the upper image of a 2δ shifted version of $\mathcal{P}_m$ for some $\delta \in \mathbb{R}_{>}^d$.

Lemma 2 (Estimation Error from Componentwise Stochastic Error). *Let Assumption 1 hold, and let $\delta = (\delta_1, \dots, \delta_d)^\top \in \mathbb{R}_{>}^d$. If $|\hat{F}_k(x) - f_k(x)| < \delta_k$ for all $x \in \mathcal{X}_m$, $k = 1, \dots, d$, then all returned points lie outside $(\mathcal{P}_m + 2\delta)^\uparrow$, thus,*

$$f(\hat{\mathcal{E}}_m(b)) \subseteq (\mathcal{P}_m + 2\delta)^{\uparrow c}, \quad (14)$$

and both estimation error comparison sets contain the 2δ -shifted upper image of $\mathcal{P}_m$, that is, $(\mathcal{P}_m + 2\delta)^\uparrow \subseteq f(\hat{\mathcal{E}}_m(b))^\uparrow \cap \text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow$. Consequently, the estimation error has upper bound $\max \{G_m^{\text{FE}}(b), G_m^{\text{FI}}(b)\} \leq 2\|\delta\|_p$.

Figure 3: The 2δ -shifted upper image of $\mathcal{P}_m$ and the resulting false-inclusion and false-exclusion regions.



Note: Panel (a) illustrates the componentwise error event $|\hat{F}_k(x) - f_k(x)| < \delta_k$ for all $x \in \mathcal{X}_m$, $k = 1, 2$. On this event, no point whose true image lies in $(\mathcal{P}_m + 2\delta)^\uparrow$ can be selected into $f(\hat{\mathcal{E}}_m(b))$. Panel (b) shows the induced regions inside $\mathcal{P}_m^\uparrow$. Points in $(\mathcal{P}_m + 2\delta)^{\uparrow c} \cap \mathcal{P}_m^\uparrow$ may be falsely included, while only points in $(\mathcal{P}_m + 2\delta)^\downarrow \cap \mathcal{P}_m^\uparrow$ may contribute to false exclusion.

Proof. We prove the two stated inclusions in three steps. First, we show (14). Let $x' \in \hat{\mathcal{E}}_m(b)$. For a contradiction, suppose $f(x') \in (\mathcal{P}_m + 2\delta)^\uparrow$; see Figure 3(a). Then there exists $\tilde{x}^* \in \mathcal{E}_m$ such that $f(\tilde{x}^*) + 2\delta \leq f(x')$. By the componentwise error bound, $\hat{F}(\tilde{x}^*) \leq f(\tilde{x}^*) + \delta \leq f(x') - \delta \leq \hat{F}(x')$. Then $\hat{F}(\tilde{x}^*) \leq \hat{F}(x')$, which contradicts $x' \in \hat{\mathcal{E}}_m(b)$. Therefore, $f(x') \notin (\mathcal{P}_m + 2\delta)^\uparrow$. Since $x' \in \hat{\mathcal{E}}_m(b)$ was arbitrary, (14) follows. Second, we show that $(\mathcal{P}_m + 2\delta)^\uparrow \subseteq f(\hat{\mathcal{E}}_m(b))^\uparrow$. Let $y \in (\mathcal{P}_m + 2\delta)^\uparrow$. Then there exists $\tilde{x}^* \in \mathcal{E}_m$ such that $f(\tilde{x}^*) + 2\delta \leq y$. We show that $y \in f(\hat{\mathcal{E}}_m(b))^\uparrow$; there are two cases. Case (a): If $\tilde{x}^* \in \hat{\mathcal{E}}_m(b)$, then $f(\tilde{x}^*) \in f(\hat{\mathcal{E}}_m(b))$. Since $f(\tilde{x}^*) \leq f(\tilde{x}^*) + 2\delta \leq y$, we have $y \in f(\hat{\mathcal{E}}_m(b))^\uparrow$. Case (b): If $\tilde{x}^* \notin \hat{\mathcal{E}}_m(b)$, then finiteness of $\mathcal{X}_m$ and external stability of $\hat{\mathcal{P}}_m(b)$ relative to $\hat{F}(\mathcal{X}_m)$ implies there exists $x' \in \hat{\mathcal{E}}_m(b)$ such that $\hat{F}(x') \leq \hat{F}(\tilde{x}^*)$. Using the componentwise error bound, $f(x') \leq \hat{F}(x') + \delta \leq \hat{F}(\tilde{x}^*) + \delta \leq f(\tilde{x}^*) + 2\delta \leq y$. Thus, $f(x') \in f(\hat{\mathcal{E}}_m(b))$ and $f(x') \leq y$, which implies $y \in f(\hat{\mathcal{E}}_m(b))^\uparrow$. Since $y \in (\mathcal{P}_m + 2\delta)^\uparrow$ was arbitrary, the desired inclusion follows. Third, we show that $(\mathcal{P}_m + 2\delta)^\uparrow \subseteq \text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow$. Since $\delta > 0$, $(\mathcal{P}_m + 2\delta)^\uparrow \subseteq \mathcal{P}_m^\uparrow$. Then, it remains to show $(\mathcal{P}_m + 2\delta)^\uparrow \subseteq \text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c})$. By (14), $f(\hat{\mathcal{E}}_m(b)) \subseteq (\mathcal{P}_m + 2\delta)^{\uparrow c}$. Because $f(\hat{\mathcal{E}}_m(b))$ is finite and $(\mathcal{P}_m + 2\delta)^\uparrow$ is upper closed, it follows that $(\mathcal{P}_m + 2\delta)^\uparrow \subseteq f(\hat{\mathcal{E}}_m(b))^{\downarrow c}$. The desired containment follows. Finally, $\mathbb{D}_p(\mathcal{P}_m^\uparrow, (\mathcal{P}_m + 2\delta)^\uparrow) \leq 2\|\delta\|_p$ because for every $y \in \mathcal{P}_m^\uparrow$, the point $y + 2\delta$ belongs to $(\mathcal{P}_m + 2\delta)^\uparrow$ and satisfies $\|y - (y + 2\delta)\|_p = 2\|\delta\|_p$. Together with the inclusion results, we have $G_m^{\text{FE}}(b) \leq 2\|\delta\|_p$ and $G_m^{\text{FI}}(b) \leq 2\|\delta\|_p$, implying the final bound. $\square$

The use of a componentwise tolerance around a Pareto set is related to the ϵ -Pareto accuracy criterion used in Zuluaga et al. (2016). However, Lemma 2 establishes a different result with a useful geometric interpretation regarding how false exclusion and false inclusion events can happen. First, on the componentwise error event, no selected point in $f(\hat{\mathcal{E}}_m(b))$ has its true image in $(\mathcal{P}_m + 2\delta)^\uparrow$. Thus, false inclusions can arise only from points in $(\mathcal{P}_m + 2\delta)^{\uparrow c} \cap \mathcal{P}_m^\uparrow$. Second, the smaller region $(\mathcal{P}_m + 2\delta)^\downarrow \cap \mathcal{P}_m^\uparrow$ contains all points needed to determine the false-exclusion distance. These two regions are illustrated in Figure 3(b). At the same time, both estimation-error comparison sets contain $(\mathcal{P}_m + 2\delta)^\uparrow$, which yields the $2\|\delta\|_p$ upper bound.

4.3 Comparison with Hausdorff-Distance Bounds

The exclusion–inclusion gap decomposition in (13) relaxes specific requirements of the Hausdorff distance decomposition in (7), mainly regarding the inclusion components. For the exclusion components, by Lemma 1, bounds on the Hausdorff coverage error and false-exclusion error also bound the corresponding exclusion–inclusion gap terms through the upper-image construction. For the inclusion components, however, the representation-error and false-inclusion terms are the primary obstacles to obtaining general bounds under the Hausdorff decomposition. The exclusion–inclusion gap does not require control of the representation error $\mathbb{D}_p(\mathcal{P}_m, \mathcal{P})$ due to its upper-image constructions. Further, the two false-inclusion terms differ substantially. The Hausdorff term

$\mathbb{D}_p(f(\hat{\mathcal{E}}_m(b)), \mathcal{P}_m)$ measures the distance from every returned point in $f(\hat{\mathcal{E}}_m(b))$ to $\mathcal{P}_m$ and takes the maximum. In contrast, $G_m^{\text{FI}}(b) = \mathbb{D}_p(\mathcal{P}_m^\uparrow, \text{cl}(f(\hat{\mathcal{E}}_m(b))^{\downarrow c}) \cap \mathcal{P}_m^\uparrow)$ gives a measure of whether the returned points enlarge the lower image of $f(\hat{\mathcal{E}}_m(b))$ relative to $\mathcal{P}_m^\uparrow$. Consequently, weakly nondominated points outside $\mathcal{P}_m$ may increase the Hausdorff false-inclusion error without increasing $G_m^{\text{FI}}(b)$, as demonstrated by [Proposition 1](#).

The following [Proposition 3](#) states a more precise comparison between the two approximation–estimation decomposition bounds. For any nonempty, finite set $\mathcal{A} \subset f(\mathcal{X}_m)$, first, define the Hausdorff distance decomposition bound as $B_p^{\text{H}}(\mathcal{A}) := \mathbb{H}_p(\mathcal{P}, \mathcal{P}_m) + \mathbb{H}_p(\mathcal{P}_m, \mathcal{A})$ and the corresponding exclusion–inclusion gap decomposition bound as

$$B_p^{\text{EI}}(\mathcal{A}) := \mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow) + \max \{ \mathbb{D}_p(\mathcal{P}_m^\uparrow, \mathcal{A}^\uparrow), \mathbb{D}_p(\mathcal{P}_m^\uparrow, \text{cl}(\mathcal{A}^{\downarrow c}) \cap \mathcal{P}_m^\uparrow) \}.$$

Then, under the infinity norm and taking $\mathcal{A} = f(\hat{\mathcal{E}}_m(b))$, [Proposition 3](#) implies that the exclusion–inclusion gap decomposition bound is no larger than the corresponding Hausdorff bound.

Proposition 3. *Under [Assumption 1](#), if $\mathcal{A} \subset f(\mathcal{X}_m)$ is nonempty and finite, then*

$$\begin{aligned} B_p^{\text{EI}}(\mathcal{A}) &\leq \mathbb{H}_p(\mathcal{P}, \mathcal{P}_m) + d^{1/p} \mathbb{H}_p(\mathcal{P}_m, \mathcal{A}) \\ &\leq d^{1/p} B_p^{\text{H}}(\mathcal{A}). \end{aligned}$$

In particular, if $p = \infty$, then $B_\infty^{\text{EI}}(\mathcal{A}) \leq B_\infty^{\text{H}}(\mathcal{A})$.

Proof. Let $u := \mathbb{H}_p(\mathcal{P}, \mathcal{P}_m)$ and $v := \mathbb{H}_p(\mathcal{P}_m, \mathcal{A})$, so that $B_p^{\text{H}}(\mathcal{A}) = u + v$. Applying [Lemma 1](#) to the original problem and then to the finite approximate problem gives $\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow) \leq \mathbb{D}_p(\mathcal{P}, \mathcal{P}_m) \leq u$ and $\mathbb{D}_p(\mathcal{P}_m^\uparrow, \mathcal{A}^\uparrow) \leq \mathbb{D}_p(\mathcal{P}_m, \mathcal{A}) \leq v$. It remains to bound the false-inclusion term of $B_p^{\text{EI}}(\mathcal{A})$. Let $w = \mathbb{D}_p(\mathcal{A}, \mathcal{P}_m)$, the largest distance from a point in $\mathcal{A}$ to its nearest point in $\mathcal{P}_m$. Fix $y \in \mathcal{P}_m^\uparrow$. Letting $1_d = (1, \dots, 1)$ be a d -dimensional vector of ones, we claim $y + w1_d \in \text{cl}(\mathcal{A}^{\downarrow c}) \cap \mathcal{P}_m^\uparrow$. First, since $\mathcal{P}_m^\uparrow$ is upper closed, $y \in \mathcal{P}_m^\uparrow$ implies $y + w1_d \in \mathcal{P}_m^\uparrow$. Then for a contradiction, suppose $y + w1_d \notin \text{cl}(\mathcal{A}^{\downarrow c})$, which implies $y + w1_d \in \text{int}(\mathcal{A}^{\downarrow})$. Because $\mathcal{A}$ is finite, there exists $a \in \mathcal{A}$ such that $y + w1_d < a$. By the definition of w , choose $p_a \in \mathcal{P}_m$ with $\|a - p_a\|_p \leq w$, which implies $\|a - p_a\|_\infty \leq w$. Then $y < a - w1_d \leq p_a$. Since $y \in \mathcal{P}_m^\uparrow$ and $\mathcal{P}_m$ is finite, there exists $p \in \mathcal{P}_m$ such that $p \leq y$. Therefore, $p < p_a$, contradicting nondominance of $p_a \in \mathcal{P}_m$. Then, it follows that $y + w1_d \in \text{cl}(\mathcal{A}^{\downarrow c}) \cap \mathcal{P}_m^\uparrow$, which implies $\mathbb{D}_p(y, \text{cl}(\mathcal{A}^{\downarrow c}) \cap \mathcal{P}_m^\uparrow) \leq \|y - (y + w1_d)\|_p \leq d^{1/p} w$. Taking the supremum over $y \in \mathcal{P}_m^\uparrow$ gives $\mathbb{D}_p(\mathcal{P}_m^\uparrow, \text{cl}(\mathcal{A}^{\downarrow c}) \cap \mathcal{P}_m^\uparrow) \leq d^{1/p} w$. Combining the three bounds yields the result. $\square$

4.4 The Choice of Norm

The choice of norm determines how the errors on each objective are aggregated in the set distances. Finite p -norms allow for compensation across the objectives, in the sense that a larger deviation in

one objective can be offset by smaller deviations in the others. Choosing the 1-norm emphasizes aggregate error, while the 2-norm facilitates intuition through familiar Euclidean distances. The infinity norm, however, emphasizes uniformity and worst-objective behavior, such that a large error on one objective cannot be offset by small errors in the other objectives. The acceptability of a given norm is context dependent and tied to preference modeling (Sayin 2000, Audet et al. 2021).

The infinity norm is especially useful for our analysis for two reasons. First, Proposition 3 shows that, under the infinity norm, the exclusion–inclusion gap decomposition bound is no larger than the corresponding Hausdorff distance decomposition bound. Second, Lemma 2 is based on componentwise stochastic-error control that defines a stochastic-error hyperrectangle with side lengths $2\delta_1, \dots, 2\delta_d$. Under a finite p -norm, the estimation bound depends on the aggregate size of the box, $2\|\delta\|_p$, whereas under the infinity norm, it depends only on the largest side length. Thus, in both the first and second cases, we avoid a $d^{1/p}$ growth factor which leads to a more favorable dimension dependence in the sample complexity; an example of this dimension dependence occurs in the analysis of Section 5. For these reasons, and because we emphasize uniformity and convergence guarantees, our numerical experiments in Subsection 5.4 use $p = \infty$.

5 Analysis of an Equal Allocation Algorithm with Low-Dispersion Design

We now specialize the general results from Sections 3 and 4 to analyze the global multiobjective simulation optimization algorithm detailed in Algorithm 1. The purpose of this section is not to propose a competitive algorithm, but rather to show how deterministic approximation error guarantees and stochastic error control combine to yield new finite-sample results and convergence rates. Here, the goal is to solve the problem in (1) to global optimality. We assume f is continuous but accessible only through pointwise stochastic estimators, as often occurs when the objectives are defined implicitly through a black-box Monte Carlo or stochastic simulation model. Given a finite discretization $\mathcal{X}_m \subset \mathcal{X}$ and the number of simulation replications per point n , Algorithm 1 solves

Algorithm 1 A simple algorithm for global multiobjective simulation optimization

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

Input: A finite set of feasible points $\mathcal{X}_m \subset \mathcal{X}$ chosen as a low-dispersion point set with cardinality $m \geq 1$; the number of simulation replications per point $n \geq 1$

- 1: $b \leftarrow mn$ ▷ determine the total simulation budget
 - 2: **for all** $x \in \mathcal{X}_m$ **do**
 - 3: $\bar{F}(x, n) \leftarrow \text{oracle}(x, n)$ ▷ estimate $f(x)$ using n simulation replications for each $x \in \mathcal{X}_m$
 - 4: **end for**
 - 5: $\hat{\mathcal{E}}_m(b) \leftarrow \{x \in \mathcal{X}_m : \nexists x' \in \mathcal{X}_m \text{ such that } \bar{F}(x', n) \leq \bar{F}(x, n)\}$
 - 6: $\hat{\mathcal{P}}_m(b) \leftarrow \bar{F}(\hat{\mathcal{E}}_m(b), n) = \{\bar{F}(x, n) : x \in \hat{\mathcal{E}}_m(b)\}$
 - 7: **return** $\hat{\mathcal{E}}_m(b), \hat{\mathcal{P}}_m(b)$ ▷ return estimated approximate efficient and Pareto sets
-

the resulting *approximate sample-path problem*,

$$\begin{aligned} & \text{minimize} && \bar{F}(x, n) = (\bar{F}_1(x, n), \dots, \bar{F}_d(x, n))^\top \\ & \text{subject to} && x \in \mathcal{X}_m, \end{aligned} \tag{15}$$

for sample-average estimator $\bar{F} = \hat{F}$, defined in (3) for each $x \in \mathcal{X}$ and n , and returns the estimated approximate efficient and Pareto sets. In (15), we denote the dependence of the objective estimator from (3) on n , which is now the same at every visited point. Thus, henceforth, we replace the notation $\hat{F}(x)$ with $\bar{F}(x, n)$.

In this setting, we derive finite-sample choices for the number of evaluated points m , the number of simulation replications per point n , and the total budget $b = mn$ that guarantee probabilistic control of the exclusion–inclusion gap by ensuring $\mathbb{P}\{G_m(b) > \tau\} \leq \alpha$ for some tolerance $\tau > 0$ and probability $\alpha \in (0, 1)$. Then, we use these allocations to derive corresponding sample complexity results and convergence rates. We first state the assumptions required for this analysis.

5.1 Assumptions

To control the approximation and estimation errors, we impose mild assumptions on the true problem and its discretization, and on the stochastic errors, respectively.

5.1.1 The True Problem and Its Discretization

The following hold for the true problem.

Assumption 2. *The feasible set $\mathcal{X} \subset \mathbb{R}^q$ is compact, and for given $r, p \geq 1$, 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 $\ell > 0$.*

We impose [Assumption 2](#) only for the analysis of [Algorithm 1](#). Compact $\mathcal{X}$ and continuous f imply that $f(\mathcal{X})$ is compact, so [Assumption 1](#) Parts 2 and 3 hold automatically in this setting.

How well the approximate 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 design $\mathcal{X}_m$ covers the feasible set $\mathcal{X}$. We follow the analysis of [Yakowitz et al. \(2000\)](#) and measure the quality of the discretization $\mathcal{X}_m$ by its dispersion under the r -norm, $t := \mathbb{D}_r(\mathcal{X}, \mathcal{X}_m)$. Equivalently, t is the smallest radius such that the closed r -norm balls centered at points in $\mathcal{X}_m$ cover $\mathcal{X}$. The Lipschitz condition gives objective-space coverage error of order ℓt .

Assumption 3. *The discretization $\mathcal{X}_m \subset \mathcal{X} \subset \mathbb{R}^q$ has $m \geq (1 + \tilde{\varepsilon})^q$ 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 does not depend 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](#)). 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 $\{(x_1, \dots, x_q) : x_j \in \{\frac{1}{2\nu}, \frac{3}{2\nu}, \dots, \frac{2\nu-1}{2\nu}\} \text{ for each } j\} \cup \Psi$, where Ψ is an arbitrary set of $m - \nu^q$ points in $\mathcal{X}$ outside the grid. Then under the $r \in \{1, 2, \dots\}$ and $r = \infty$ norms, the respective dispersions of $\mathcal{X}_m$ are $\mathbb{D}_r(\mathcal{X}, \mathcal{X}_m) \leq q^{1/r} / (2 \lfloor m^{1/q} \rfloor)$ and $\mathbb{D}_\infty(\mathcal{X}, \mathcal{X}_m) \leq 1 / (2 \lfloor m^{1/q} \rfloor)$. Therefore, $\tilde{c}_q = (1/2)q^{1/r}$ under the r -norm and $\tilde{c}_q = 1/2$ under the ∞ -norm.

Any quantity that depends on the dispersion t also depends on m , since $t \rightarrow 0$ implies $m \rightarrow \infty$. Any sequence of designs $\{\mathcal{X}_m, m \geq (1 + \tilde{\varepsilon})^q\}$ satisfying [Assumption 3](#) has $t \rightarrow 0$ as $m \rightarrow \infty$.

5.1.2 The Stochastic Error

Next, we specify the nature of the stochastic error model. For $j = 1, \dots, n$, define the vector-valued error random fields $Z_j(\cdot) := F(\cdot, \xi_j) - f(\cdot)$ where $Z_j(\cdot) = (Z_{j1}(\cdot), \dots, Z_{jd}(\cdot))^\top$. Then, the vector-valued objective function estimator is $\bar{F}(\cdot, n) := f(\cdot) + n^{-1} \sum_{j=1}^n Z_j(\cdot)$, where the values of $\bar{F}(\cdot, n)$ at points outside $\mathcal{X}_m$ are well-defined but remain unobserved in (15). To ensure the estimators are unbiased, we impose the following assumption.

Assumption 4. *The sequence $\xi_j, j = 1, \dots, n$ is independent and identically distributed (i.i.d.) with $\mathbb{E}[Z_{1k}(x)] = 0$ for each objective $k \in \{1, \dots, d\}$ and feasible point $x \in \mathcal{X}$.*

Under [Assumption 4](#), the error random fields $Z_j(\cdot), j = 1, \dots, n$ are i.i.d., which implies $\bar{F}(\cdot, n)$ is an unbiased estimator for $f(\cdot)$. This assumption permits common random numbers, where the same random inputs are used to estimate the objective function values at multiple feasible points, and allows the components of the vector $Z_j(\cdot)$ to be correlated across objectives.

Next, we ensure the unbiased estimator is of sufficient quality by requiring the stochastic error to be sub-exponential with uniformly bounded sub-exponential norms. We use standard concentration inequalities and norm definitions; see [Vershynin \(2018\)](#). For a sub-exponential random variable X , let the sub-exponential norm of X be $\|X\|_{\psi_1} := \inf\{s > 0 : \mathbb{E}[\exp(|X|/s)] \leq 2\}$.

Assumption 5. *The stochastic errors $Z_{jk}(x)$ are sub-exponential for all $x \in \mathcal{X}$, replications $j = 1, 2, \dots$, and objectives $k = 1, \dots, d$. Moreover, for each k , there exists a constant $\eta_{1k}^* > 0$ such that $\eta_{1k}^* := \sup \{\|Z_{jk}(x)\|_{\psi_1} : x \in \mathcal{X}, j = 1, 2, \dots\} < \infty$.*

[Assumption 5](#) is used to obtain concentration bounds uniformly over the finite set $\mathcal{X}_m$ under common error models such as normal, Laplace, and all error models specified by distributions with bounded support. Under [Assumptions 4](#) and [5](#), Bernstein's inequality implies that for each $x \in \mathcal{X}$, $k = 1, \dots, d$, and every $\delta > 0$, there exists an absolute constant $\kappa_1 > 0$ such that

$$\begin{aligned} & \mathbb{P}\left\{\left|n^{-1} \sum_{j=1}^n Z_{jk}(x)\right| \geq \delta\right\} \\ & \leq 2 \exp\left(-n\kappa_1 \min\left\{\frac{\delta^2}{\|Z_{1k}(x)\|_{\psi_1}^2}, \frac{\delta}{\|Z_{1k}(x)\|_{\psi_1}}\right\}\right). \end{aligned} \tag{16}$$

If the errors are known to be sub-Gaussian, then we can employ the sharper bound arising from Hoeffding’s inequality. For a sub-Gaussian random variable X , define the sub-Gaussian norm as $\|X\|_{\psi_2} := \inf\{s > 0: \mathbb{E}[\exp(X^2/s^2)] \leq 2\}$. Since X^2 is sub-exponential and $\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2$, sub-Gaussian errors are a special case of sub-exponential errors. If [Assumptions 4](#) and [5](#) hold and the stochastic errors are sub-Gaussian with $\eta_{2k}^* := \sup\{\|Z_{jk}(x)\|_{\psi_2}: x \in \mathcal{X}, j = 1, 2, \dots\} < \infty$ for every objective k , then for each $x \in \mathcal{X}$, $k = 1, \dots, d$, and every $\delta > 0$,

$$\mathbb{P}\left\{\left|n^{-1} \sum_{j=1}^n Z_{jk}(x)\right| \geq \delta\right\} \leq 2 \exp\left(\frac{-n\kappa_2\delta^2}{\|Z_{1k}(x)\|_{\psi_2}^2}\right) \quad (17)$$

where $\kappa_2 > 0$ is an absolute constant. The inequalities in [\(16\)](#) and [\(17\)](#) do not require the stochastic errors to have the same distributional family across the objectives k or across the values of $x \in \mathcal{X}$.

In what follows, we make the more general assumption of sub-exponential errors and provide explicit modifications to the results in case all errors are sub-Gaussian. For mixed error classes, we require [\(16\)](#). [Appendix EC.1](#) contains examples of how to compute the constants κ_1, η_{1k}^* and κ_2, η_{2k}^* . In [Subsection 5.4](#), we use $\kappa_1 = 1/2$ for Laplace error and $\kappa_2 = 4/3$ for normal error.

5.2 A Finite-Sample Bound for the Exclusion–Inclusion Gap

We now combine the deterministic approximation bound from [Lemma 1](#), the stochastic estimation error bound from [Lemma 2](#), and the assumptions in [Subsection 5.1](#). The result is a finite-sample bound on the exclusion–inclusion gap for [Algorithm 1](#), presented in [Theorem 2](#).

Lemma 3. *Let [Assumptions 1–3](#) hold, and let $\delta \in \mathbb{R}_{>}^d$. If $|\bar{F}_k(x, n) - f_k(x)| < \delta_k$ for all $x \in \mathcal{X}_m$ and $k = 1, \dots, d$, then $G_m(b) \leq \ell t + 2\|\delta\|_p$.*

Proof. By [Theorem 1](#), the gap decomposition in [\(13\)](#) holds. For the approximation term, since $\mathcal{X}_m \neq \emptyset$ is finite, $\mathcal{P}_m$ is externally stable relative to $f(\mathcal{X}_m)$, so $\mathcal{P}_m^\uparrow = f(\mathcal{X}_m)^\uparrow$. For any $y^* = f(x^*) \in \mathcal{P}$, $t = \mathbb{D}_r(\mathcal{X}, \mathcal{X}_m)$ gives an $\tilde{x} \in \mathcal{X}_m$ such that $\|x^* - \tilde{x}\|_r \leq t$. By [Assumption 2](#), $\|y^* - f(\tilde{x})\|_p = \|f(x^*) - f(\tilde{x})\|_p \leq \ell t$. Thus, $\mathbb{D}_p(\mathcal{P}, f(\mathcal{X}_m)) \leq \ell t$. Applying [Lemma 1](#) with $\mathcal{A} = f(\mathcal{X}_m)$ yields $\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow) = \mathbb{D}_p(\mathcal{P}^\uparrow, f(\mathcal{X}_m)^\uparrow) \leq \ell t$. The estimation error bound follows by [Lemma 2](#). $\square$

Having bounded $G_m(b)$ on the componentwise estimation event, we now bound the probability of a false exclusion or false inclusion event corresponding to the complement of the componentwise estimation event. [Lemma 4](#) applies a union bound to the probability that at least one of the m sampled points is estimated outside its δ box on at least one of the d objectives, and then applies a concentration inequality to each resulting term. While this union bound is conservative, it also keeps the result distributionally and structurally simple. The bound remains valid under common random numbers and dependence across design points and objectives; further, no additional regularity conditions are required on the sample-path function $\bar{F}$.

Lemma 4. Let [Assumption 1](#) hold, and let the stochastic error be governed by [Assumptions 4](#) and [5](#). Suppose $\hat{\mathcal{E}}_m(b)$ is constructed using $n \geq 1$ replications at each point in $\mathcal{X}_m$ and $b \geq mn$. Then there exists an absolute constant $\kappa_1 > 0$ such that for all $n \geq 1$ and all $\delta \in \mathbb{R}_{>}^d$,

$$\begin{aligned} & \mathbb{P}\left\{\max\left\{G_m^{\text{FE}}(b), G_m^{\text{FI}}(b)\right\} > 2\|\delta\|_p\right\} \\ & \leq 2dm \exp\left(-n\kappa_1 \min\left\{\min_k \left(\frac{\delta_k}{\eta_{1k}^*}\right)^2, \min_k \frac{\delta_k}{\eta_{1k}^*}\right\}\right). \end{aligned}$$

If all errors in [Assumption 5](#) are sub-Gaussian, then there exists an absolute constant $\kappa_2 > 0$ such that the right side of the inequality can be replaced by $2dm \exp(-n\kappa_2 \min_k (\delta_k/\eta_{2k}^*)^2)$.

Proof. Let $n \geq 1$ and $\delta \in \mathbb{R}_{>}^d$. By [Lemma 2](#), if $\max\{G_m^{\text{FE}}(b), G_m^{\text{FI}}(b)\} > 2\|\delta\|_p$, then there exists $\tilde{x}_i \in \mathcal{X}_m, k \in \{1, \dots, d\}$ such that $|\bar{F}_k(\tilde{x}_i, n) - f_k(\tilde{x}_i)| \geq \delta_k$. Therefore, $\mathbb{P}\left\{\max\left\{G_m^{\text{FE}}(b), G_m^{\text{FI}}(b)\right\} > 2\|\delta\|_p\right\} \leq \mathbb{P}\left\{|\bar{F}_k(\tilde{x}_i, n) - f_k(\tilde{x}_i)| \geq \delta_k \text{ for some } i, k\right\}$. The union bound gives

$$\begin{aligned} & \mathbb{P}\left\{|\bar{F}_k(\tilde{x}_i, n) - f_k(\tilde{x}_i)| \geq \delta_k \text{ for some } i, k\right\} \\ & \leq \sum_{k=1}^d \sum_{i=1}^m \mathbb{P}\left\{\left|n^{-1} \sum_{j=1}^n Z_{jk}(\tilde{x}_i)\right| \geq \delta_k\right\}. \end{aligned}$$

Under [Assumptions 4](#) and [5](#), Bernstein's inequality in [\(16\)](#) applies to each summand; using $\|Z_{1k}\|_{\psi_1} \leq \eta_{1k}^*$ gives the result. If all errors are sub-Gaussian, the result follows from [\(17\)](#). $\square$

Combining [Lemmas 3](#) and [4](#) yields the following finite-sample bound for [Algorithm 1](#), where the ratios δ_k/η_{1k}^* normalize objectives with different variances.

Theorem 2 (Finite-Sample Gap Bound). Let [Assumptions 1, 2, 4, and 5](#) hold, let [Assumption 3](#) hold for $m \geq (1 + \tilde{\varepsilon})^q$, and recall that $\hat{\mathcal{E}}_m(b)$ is constructed using $n \geq 1$ replications at each point in $\mathcal{X}_m$ where $b \geq mn$. Then for all $n \geq 1$ and all $\delta \in \mathbb{R}_{>}^d$,

$$\begin{aligned} & \mathbb{P}\left\{G_m(b) > \frac{\tilde{c}_q \ell}{\lfloor m^{1/q} \rfloor} + 2\|\delta\|_p\right\} \\ & \leq 2dm \exp\left(-n\kappa_1 \min\left\{\min_k \left(\frac{\delta_k}{\eta_{1k}^*}\right)^2, \min_k \frac{\delta_k}{\eta_{1k}^*}\right\}\right). \end{aligned} \tag{18}$$

If all errors in [Assumption 5](#) are sub-Gaussian, then the right side of the inequality can be replaced by $2dm \exp(-n\kappa_2 \min_k (\delta_k/\eta_{2k}^*)^2)$.

Proof. Let $m \geq (1 + \tilde{\varepsilon})^q, n \geq 1, \delta \in \mathbb{R}_{>}^d$. By [Assumption 3](#), $\ell t \leq \tilde{c}_q \ell (\lfloor m^{1/q} \rfloor)^{-1}$. Using [Lemma 3](#), $\mathbb{P}\left\{G_m(b) > \tilde{c}_q \ell (\lfloor m^{1/q} \rfloor)^{-1} + 2\|\delta\|_p\right\} \leq \mathbb{P}\left\{G_m(b) > \ell t + 2\|\delta\|_p\right\} \leq \mathbb{P}\left\{\ell t + \max\{G_m^{\text{FE}}(b), G_m^{\text{FI}}(b)\} > \ell t + 2\|\delta\|_p\right\}$. The result follows by applying [Lemma 4](#) to the final probability. $\square$

5.3 Budget Allocation and Convergence Rates

We now employ [Theorem 2](#) to develop allocations for two use cases. In the *fixed-precision* setting, a user wants to ensure that $P\{G_m(b) > \tau\} \leq \alpha$ holds. Therefore, the user specifies either the number of design points m or the error tolerance τ and the confidence requirement α . Then, the analysis determines a sufficient simulation budget. In the *fixed-budget* setting, the user specifies the maximum simulation budget b^* , and the analysis determines how to allocate it between m and n . In what follows, we first balance the approximation and estimation terms appearing in [Theorem 2](#) and state the resulting finite-sample bound as [Corollary 1](#). Then, we derive the fixed-precision and fixed-budget allocations and establish the sample-complexity and convergence-rate guarantees.

5.3.1 Balancing Approximation and Estimation

To balance between the error types, first, recall that [Theorem 2](#) bounds the gap by the sum of a deterministic approximation term of order $m^{-1/q}$ and a stochastic estimation term governed by the componentwise tolerances δ_k . As in [Yakowitz et al. \(2000\)](#), we ensure these error terms go to zero at the same rate by choosing

$$\delta_k = \frac{\eta_{ik}^*}{\lfloor m^{1/q} \rfloor} \quad \text{for all } k = 1, \dots, d, \quad i = 1, 2 \quad (19)$$

where $i = 1, 2$ corresponds to the sub-exponential and sub-Gaussian error cases, respectively. For notational convenience, we set $\eta_i^* := (\eta_{i1}^*, \dots, \eta_{id}^*)$ and define the constant

$$c_i := \tilde{c}_q \ell + 2\|\eta_i^*\|_p \quad \text{for } i = 1, 2,$$

which may depend on the number of objectives d through the norm $\|\eta_i^*\|_p$. Then [Theorem 2](#) and the choice of δ in (19) directly lead to the following corollary.

Corollary 1 (Balanced Finite-Sample Bound). *Let the conditions of [Theorem 2](#) hold. Under the componentwise tolerance choice in (19), for every $m \geq (1 + \tilde{\varepsilon})^q$ and $n \geq 1$,*

$$P\left\{G_m(b) > \frac{c_i}{\lfloor m^{1/q} \rfloor}\right\} \leq 2dm \exp\left(-\frac{n\kappa_i}{\lfloor m^{1/q} \rfloor^2}\right), \quad (20)$$

where $i = 1, 2$ correspond to the sub-exponential and sub-Gaussian error models, respectively.

Proof. Choose δ according to (19). Then $\|\delta\|_p = \|\eta_i^*\|_p / \lfloor m^{1/q} \rfloor$ and $\delta_k / \eta_{ik}^* = 1 / \lfloor m^{1/q} \rfloor$ for all k . Since $\lfloor m^{1/q} \rfloor \geq 1$, substituting into [Theorem 2](#) implies the result. $\square$

5.3.2 Fixed Design Size or Error Tolerance with Fixed Confidence Requirement

We first apply [Corollary 1](#) when the number of evaluated points m and the confidence requirement α are specified in advance. We then invert the result to obtain an allocation for a fixed error tolerance

τ and confidence requirement α . These results are presented formally in the following [Theorem 3](#). For a fixed error model $i \in \{1, 2\}$, we suppress the notational dependence on i and define

$$\tau_m := \frac{c_i}{\lfloor m^{1/q} \rfloor}.$$

Theorem 3 (Fixed-Precision Allocations). *Let the conditions of [Corollary 1](#) hold, where $i = 1, 2$ correspond to the sub-exponential and sub-Gaussian error models, respectively.*

1. Fixed design size: Given $m \geq (1 + \tilde{\varepsilon})^q$ and $\alpha \in (0, 1)$, if

$$n = \lceil \kappa_i^{-1} \lfloor m^{1/q} \rfloor^2 \ln(2dm\alpha^{-1}) \rceil, \quad b = mn, \quad (21)$$

then [Algorithm 1](#) satisfies $\mathbb{P}\{G_m(b) > \tau_m\} \leq \alpha$.

2. Fixed tolerance: For given $\tau > 0$ and $\alpha \in (0, 1)$, if

$$m = \max \left\{ \lceil (1 + c_i\tau^{-1})^q \rceil, \lceil (1 + \tilde{\varepsilon})^q \rceil \right\} \quad (22)$$

and n, b are set by (21), then [Algorithm 1](#) satisfies $\mathbb{P}\{G_m(b) > \tau\} \leq \mathbb{P}\{G_m(b) > \tau_m\} \leq \alpha$.

Proof. For Part 1, by (21), we have $n \geq \kappa_i^{-1} \lfloor m^{1/q} \rfloor^2 \ln(2dm\alpha^{-1})$. Therefore, from the right side of (20), we have $\mathbb{P}\{G_m(b) > \tau_m\} \leq 2dm \exp(-n\kappa_i \lfloor m^{1/q} \rfloor^{-2}) \leq 2dm \exp(-\ln(2dm\alpha^{-1})) = \alpha$. For Part 2, by (22), we have $m \geq (1 + c_i\tau^{-1})^q$ and $m \geq (1 + \tilde{\varepsilon})^q$. Thus, the design-size requirement in [Assumption 3](#) holds. Further, $m^{1/q} \geq 1 + c_i\tau^{-1}$. Then $\lfloor m^{1/q} \rfloor \geq m^{1/q} - 1 \geq c_i\tau^{-1}$. Therefore, $\tau_m = c_i/\lfloor m^{1/q} \rfloor \leq \tau$. Applying Part 1 with n, b from (21) gives the result. $\square$

5.3.3 Allocation Under a Fixed Simulation Budget

We now consider values of m and n when an upper limit on the total simulation budget, b^* , is fixed in advance. Such values are useful for a user who does not know how to specify one of the values m, τ , or α required by [Theorem 3](#), or for whom the resulting b is impossibly large or exceeds some predetermined computing limit.

Theorem 4 (Fixed-Budget Allocation). *Let the conditions of [Corollary 1](#) hold, let $i = 1, 2$ correspond to the sub-exponential and sub-Gaussian error models, respectively, and let $\tilde{\varepsilon} > 0$. If*

$$m = \left\lceil \left(\frac{\kappa_i b^*}{2 \ln b^*} \right)^{q/(q+2)} \right\rceil \geq (1 + \tilde{\varepsilon})^q, \quad n = \left\lfloor \frac{b^*}{m} \right\rfloor \geq 1, \quad (23)$$

and $b = mn$ for sufficiently large b^* , then $b \leq b^*$ and [Algorithm 1](#) satisfies $\mathbb{P}\{G_m(b) > \tau_m\} \leq 2dn^{-1}$.

Proof. Since $n = \lfloor b^*/m \rfloor$, we have $b = mn \leq b^*$ and $n \geq b^*/m - 1$. Further, $\lfloor m^{1/q} \rfloor \leq m^{1/q}$. Then $\kappa_i n \lfloor m^{1/q} \rfloor^{-2} \geq \kappa_i n m^{-2/q} \geq \kappa_i b^* m^{-(q+2)/q} - \kappa_i m^{-2/q}$. The definition of m implies $\kappa_i b^* m^{-(q+2)/q} \geq 2 \ln b^*$. For a total simulation budget $b^* > 1$ such that (23) satisfies $m \geq (1 + \tilde{\varepsilon})^q$ and $\kappa_i m^{-2/q} \leq \ln b^*$,

we have $\kappa_i n \lfloor m^{1/q} \rfloor^{-2} \geq \ln b^*$. Applying [Corollary 1](#), $P\{G_m(b) > \tau_m\} \leq 2dm \exp(-\kappa_i n \lfloor m^{1/q} \rfloor^{-2}) \leq 2dm(b^*)^{-1} \leq 2dn^{-1}$, where the final inequality follows from $mn = b \leq b^*$. The result follows. $\square$

[Theorem 4](#) provides a simple rule for allocating a fixed available budget b^* between the number of design points m and the number of simulation replications n per point. The resulting expended budget $b = mn$ satisfies $b \leq b^*$, with unexpended budget arising from the integer allocation.

5.3.4 Asymptotic Results

We now study the sample complexity and convergence rates of the allocations specified in [Theorems 3](#) and [4](#); the proofs for results in this section appear in [Appendix A](#). First, [Theorem 3](#) yields the sample-complexity results in [Corollary 2](#).

Corollary 2 (Sample Complexity). *Under the conditions of [Theorem 3](#), suppose [Algorithm 1](#) is implemented using the fixed-tolerance allocation in Part 2. Then:*

1. *Given a problem and error model, consider any sequence of fixed-precision requirements along which $\tau \rightarrow 0$ where $\alpha \in (0, 1)$ may vary with τ . Then $b = \Theta(\tau^{-(q+2)} \ln(\tau^{-q} \alpha^{-1}))$.*
2. *Given a problem and error model, for any fixed $\alpha \in (0, 1)$, $b = \Theta(\tau^{-(q+2)} \ln(\tau^{-1}))$ as $\tau \rightarrow 0$.*
3. *Given a problem and error model, for any fixed $\tau > 0$, $b = \Theta(\ln \alpha^{-1})$ as $\alpha \rightarrow 0$.*
4. *Consider any sequence of problem instances along which $d \rightarrow \infty$ while τ, α , and q remain fixed. Suppose $\tilde{c}_q \ell$ and $\tilde{\varepsilon}$ remain uniformly bounded independently of d , and the applicable κ_i remains uniformly bounded away from zero and infinity. If $G_m(b)$ is calculated under the p -norm with $1 \leq p < \infty$ and $\|\eta_i^*\|_p = \Theta(d^{1/p})$, then $b = \Theta(d^{(q+2)/p} \ln d)$ as $d \rightarrow \infty$. If $G_m(b)$ is calculated under the infinity norm and $\|\eta_i^*\|_\infty = O(1)$, then $b = \Theta(\ln d)$ as $d \rightarrow \infty$.*

[Corollary 2](#) Part 1 characterizes the joint effect of increasingly stringent tolerance and confidence requirements under a fixed problem and error model. The budget depends polynomially on the inverse tolerance, up to a logarithmic factor, while the inverse confidence level enters logarithmically. Then, Parts 2 and 3 isolate the individual effects of the tolerance and confidence requirements, respectively. The asymptotic orders of both rates do not depend on d and, therefore, agree with their single-objective counterparts. Part 4 characterizes the effect of increasing the number of objectives d while holding the tolerance fixed. Under a finite p -norm, comparable objective-wise errors accumulate across objectives, and the assumption $\|\eta_i^*\|_p = \Theta(d^{1/p})$ implies that c_i grows as $\Theta(d^{1/p})$. Maintaining a fixed aggregate tolerance therefore requires the decision-space coverage radius to decrease as $d^{-1/p}$. Because this radius is of order $m^{-1/q}$, the number of evaluated points must grow as $m = \Theta(d^{q/p})$. The corresponding increase in both the design size and replication requirement yields total budget growth of $\Theta(d^{(q+2)/p} \ln d)$. Under the infinity norm, the fixed tolerance τ instead imposes a fixed worst-objective error requirement. When the objective-wise scaled stochastic errors remain uniformly bounded as $\|\eta_i^*\|_\infty = O(1)$, c_i remains bounded, so the

required design size m does not grow in d . The number of objectives d then affects the budget only through the simultaneous stochastic-error guarantee, so that b grows logarithmically in d .

Next, we show that both the fixed-precision and fixed-budget allocations in [Theorems 3](#) and [4](#) yield the same convergence rate for the exclusion–inclusion gap and the returned-image estimation error. For a fixed number of objectives d , the convergence rate exponent depends on the decision-space dimension q but not on d ; it is the same convergence rate shown for the single-objective case in [Yakowitz et al. \(2000\)](#). The dependence on d enters separately through the results in [Corollary 2](#).

Corollary 3 (Convergence Rates). *Given a problem and error model, consider any of:*

1. *The fixed-design-size allocation in [Theorem 3](#) Part 1, along a sequence with $m \rightarrow \infty$ and $\alpha = \tau_m^\gamma$ for all sufficiently large sequence indices and some fixed $\gamma > 0$;*
2. *The fixed-tolerance allocation in [Theorem 3](#) Part 2, along a sequence with $\tau \rightarrow 0$ and $\alpha = \tau^\gamma$ for all sufficiently large sequence indices and some fixed $\gamma > 0$; or*
3. *The fixed-budget allocation in [Theorem 4](#) as $b^* \rightarrow \infty$.*

In each case, $G_m(b) = O_p((b/\ln b)^{-1/(q+2)})$ and $\mathbb{H}_p(\hat{\mathcal{P}}_m(b), f(\hat{\mathcal{E}}_m(b))) = O_p((b/\ln b)^{-1/(q+2)})$ as $b \rightarrow \infty$; therefore, both quantities converge in probability to zero.

The exclusion–inclusion gap and the returned-image estimation error attain the same total-budget rate because the balanced allocation controls the uniform estimation error at order $m^{-1/q}$. Under each allocation in [Corollary 3](#), $m^{-1/q} = \Theta((b/\ln b)^{-1/(q+2)})$.

5.4 Numerical Examples

We now demonstrate our results on several numerical examples. Our main objectives are to demonstrate the convergence of $G_m(b)$ under the error-balancing choice used in [Subsection 5.3.1](#), to show how the allocations change with problem type, dimension, and error model, and to assess the bounds for conservativeness. Throughout, we use the $r = 2$ and $p = \infty$ norms.

5.4.1 Nonlinear Test Problems: Convergence Assessment

First, we implement [Algorithm 1](#) on three nonlinear problems inspired by [Yakowitz et al. \(2000, Example 5\)](#). For decision space dimension $q \in \{2, 3\}$, let $x \in [0, 1]^q$, set $x_{q+1} = x_1$, and for each $k = 1, \dots, q$ define

$$f_k^{(q)}(x) = \mathbb{E}[(x_k - .5) \sin(10x_k) + Z_{1k}^{(q)}(x) + (x_{k+1} + .5) \cos(5x_{k+1})],$$

where $Z_{1k}^{(q)}(x)$ is mean zero for every $x \in [0, 1]^q$. Suppressing the superscript q on both the objectives and error fields, we consider the following respective single-, bi-, and tri-objective problems,

$$\text{minimize}_{x \in [0, 1]^2} f_1(x), \tag{24}$$

$$\text{minimize}_{x \in [0,1]^2} \quad (f_1(x), f_2(x))^\top, \quad (25)$$

$$\text{minimize}_{x \in [0,1]^3} \quad (f_1(x), f_2(x), f_3(x))^\top. \quad (26)$$

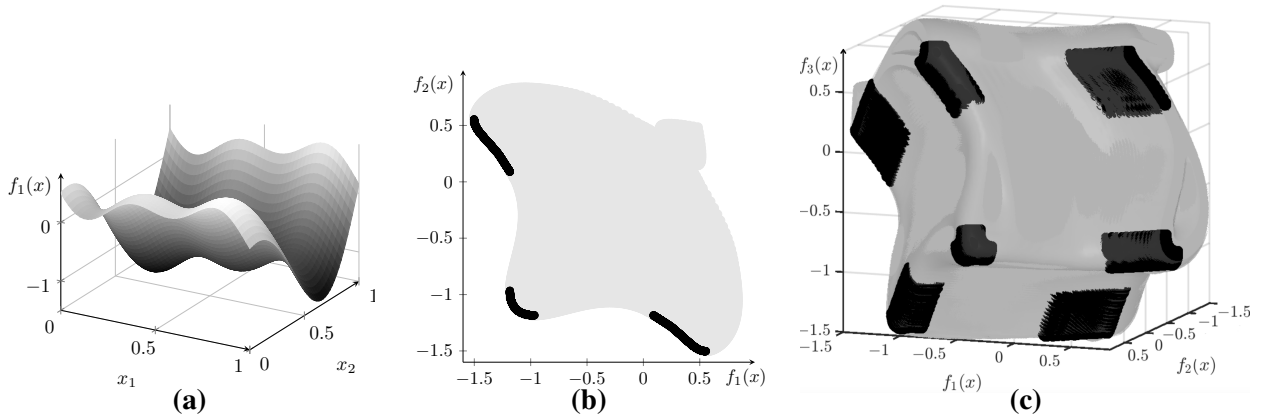
Figure 4 shows plots of the objective spaces. Problems (24) and (25) differ only in the number of objectives, while (25) and (26) differ in both the objective-space and decision-space dimensions.

We consider two error models. Under *Laplace error*, the $Z_{jk}(x)$ are i.i.d. mean-zero Laplace random variables with scale parameter $1/\sqrt{2}$ and variance 1 for all simulation replications j , objectives k , and feasible points x . Under *normal error*, they are i.i.d. standard normal random variables for all j , k , and x . Since common random numbers would make these test problems deterministic under Algorithm 1, sampling is independent across design points. Under each error type, the constants η_1^* , κ_1 , and η_2^* , κ_2 are calculated in Appendix EC.1 and reported in Table 1.

To make the discretization errors comparable across problem instances, we fix m first. We use the grid from Example 1 and select grid sizes that contain nested subsequences, as detailed in Appendix EC.2. For $q = 2$, we report results on the sequence $m \in \{16, 36, 144, 324\}$, where $\mathcal{X}_{16} \subset \mathcal{X}_{36} \subset \mathcal{X}_{144} \subset \mathcal{X}_{324}$. For $q = 3$, we report results on the sequence $m \in \{64, 216, 1728\}$, where $\mathcal{X}_{64} \subset \mathcal{X}_{216} \subset \mathcal{X}_{1728}$. Along nested subsequences, the discretization error $\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$ is nonincreasing.

For each combination of (24), (25), and (26) with error types Laplace and normal, we conduct s independent macro replications of Algorithm 1 under the allocation in Theorem 3 Part 1. To increase

Figure 4: Plots of the objective spaces for the nonlinear test problems in (24), (25), and (26), respectively.



Note: The global Pareto sets for the problems in (25) and (26) are shown in black in panels (b) and (c), respectively.

Table 1: Constants for Nonlinear Problems

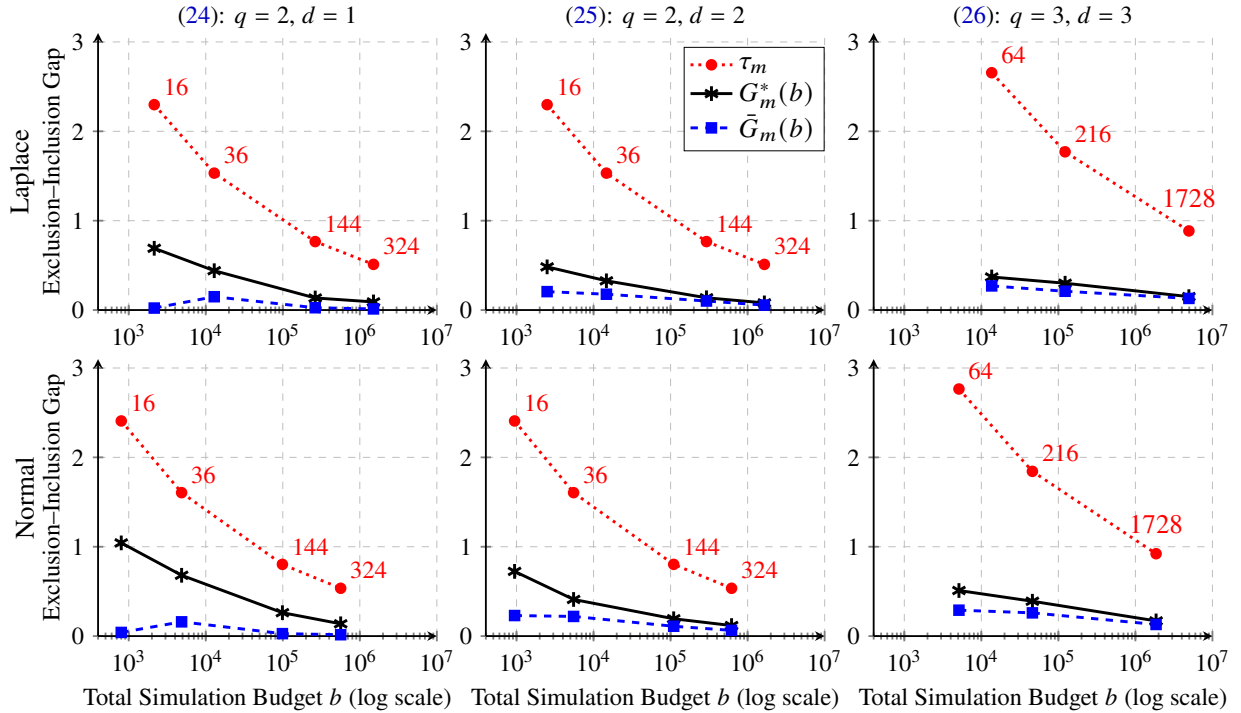
Problem	d	q	r	p	$\tilde{c}_q$	ℓ	Laplace			Normal		
							$\ \eta_1^*\ _\infty$	κ_1	c_1	$\ \eta_2^*\ _\infty$	κ_2	c_2
(24)	1	2	2	∞	$\sqrt{2}/2$	9	$\sqrt{2}$	0.5	9.1924	$\sqrt{8/3}$	4/3	9.6299
(25)	2	2	2	∞	$\sqrt{2}/2$	9	$\sqrt{2}$	0.5	9.1924	$\sqrt{8/3}$	4/3	9.6299
(26)	3	3	2	∞	$\sqrt{3}/2$	9	$\sqrt{2}$	0.5	10.6227	$\sqrt{8/3}$	4/3	11.0602

Note: Constants η_{ik}^* , $k = 1, \dots, d$, and κ_i for $i = 1, 2$ are calculated according to Appendix EC.1.

the chance of threshold exceedance, we set the confidence requirement as $\alpha = 0.5$. Letting $G_{m,j}(b)$ denote the exclusion–inclusion gap from macro replication $j = 1, \dots, s$, we report the maximum observed gap, $G_m^*(b) := \max_{j=1, \dots, s} G_{m,j}(b)$, the sample average gap, $\bar{G}_m(b) := s^{-1} \sum_{j=1}^s G_{m,j}(b)$, its estimated standard error, $\hat{\text{se}}(\bar{G}_m(b)) := \sqrt{s^{-1} \sum_{j=1}^s (G_{m,j}(b) - \bar{G}_m(b))^2 / \sqrt{s}}$, and the estimated probability the gap exceeds the threshold, $\hat{\alpha} := s^{-1} \sum_{j=1}^s \mathbb{I}\{G_{m,j}(b) > \tau_m\}$.

Figure 5 illustrates the convergence behavior of the exclusion–inclusion gap. Across all six problem and error-model combinations, the thresholds τ_m decrease as m and b increase, and $G_m^*(b)$ and $\bar{G}_m(b)$ generally become smaller at the larger design sizes. The results also illustrate the separate effects of the decision-space dimension q , the number of objectives d , and error model. Problems (24) and (25) have $q = 2$ and use the same design sizes, but (25) has one additional objective. The resulting increase in b is modest, consistent with the logarithmic dependence on d under the infinity norm. Problem (26) increases both q and d relative to (25). Its simulation budgets grow more rapidly as the design is refined, consistent with the stronger dependence of the convergence rate on q . The Laplace error models require more simulation replications than the corresponding normal error models because of the heavier-tailed errors. Nevertheless, no observed gap exceeded τ_m despite using the relatively permissive value $\alpha = 0.5$, suggesting conservativeness in the finite-sample guarantees.

Figure 5: Results from s independent macro replications of Algorithm 1 on the nonlinear problems, where $s = 10^8$ for (24) and $s = 10^4$ for (25), (26). Curves show τ_m , the maximum observed gap $G_m^*(b)$, and the average gap $\bar{G}_m(b)$.



Note: Red labels on red dots indicate the values of m . The confidence requirement is $\alpha = 0.5$. In all cases, $\hat{\alpha} = 0$ to four decimal places and $\hat{\text{se}}(\bar{G}_m(b)) < 0.001$. Detailed numerical results appear in Tables EC.1–EC.3.

5.4.2 Linear Test Problems: Conservativeness Assessment

We now directly assess the conservativeness of the allocation and bounds by considering linear problems for which the Lipschitz bound is tight. Further, we select error models that are likely to induce threshold exceedance events. For $q \in \{1, 2\}$, let $\mathcal{X} = [0, 1]^q$ and for each $x \in \mathcal{X}$ and $k = 1, 2$, let $Z_{1k}(x)$ be a mean-zero random variable. For $q = 1$, define $f_k(x) = E[x + Z_{1k}(x)]$, $k = 1, 2$. For $q = 2$, define $f_k(x) = E[x_k + Z_{1k}(x)]$, $k = 1, 2$. Then, we consider the following single- and bi-objective problems,

$$\text{minimize}_{x \in [0,1]} f_1(x), \quad (27)$$

$$\text{minimize}_{x \in [0,1]} (f_1(x), f_2(x))^\top, \quad (28)$$

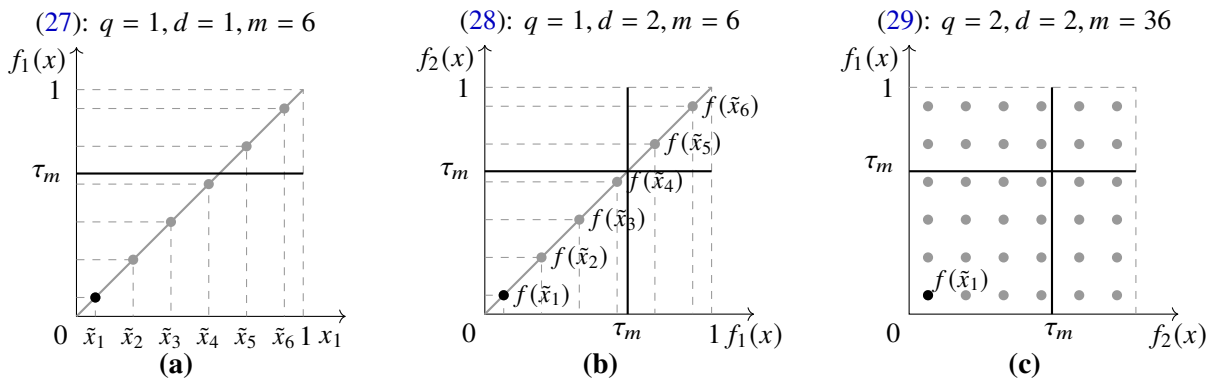
$$\text{minimize}_{x \in [0,1]^2} (f_1(x), f_2(x))^\top, \quad (29)$$

respectively. The transition from (27) to (28) increases the number of objectives d while keeping the decision-space dimension q fixed, while the transition from (28) to (29) increases q .

Figure 6 illustrates example objective spaces for (27)–(29) and identifies regions capable of producing gap values exceeding the threshold τ_m . Considering each non-Pareto point in a pairwise comparison with $f(\tilde{x}_1)$ and holding all other points fixed, only points above τ_m on at least one objective can produce a false exclusion event resulting in $G_m(b) > \tau_m$. To do so, these points must be falsely estimated as better than $f(\tilde{x}_1)$ on all objectives. Likewise, only points that are above τ_m on all objectives can produce a false inclusion event resulting in $G_m(b) > \tau_m$. To do so, these points must be falsely estimated as better than $f(\tilde{x}_1)$ on at least one objective.

As shown in Figure 6(c), every approximate problem for the linear test problem in (29) contains dominated points that are weakly nondominated. Here, convergence under the exclusion–inclusion gap means that the returned set’s upper image increasingly approximates the true Pareto set upper

Figure 6: The figure shows plots of the linear test problems under example values of m with τ_m calculated under the normal error model. Only points above τ_m on at least one objective can produce an event where $G_m(b) > \tau_m$.



Note: In Panel (a), $x_m^* = \tilde{x}_1$, $x^* = 0$, and $\tau_m = 0.6277$. In Panel (b), $\mathcal{P}_m = \{f(\tilde{x}_1)\}$, $\mathcal{P} = \{(0, 0)\}$, and $\tau_m = 0.6277$. In Panel (c), $\mathcal{P}_m = \{f(\tilde{x}_1)\}$, $\mathcal{P} = \{(0, 0)\}$, and $\tau_m = 0.6622$.

image, $(0, 0) + \mathbb{R}_{\geq}^2$, while the effect of returned dominated points on the corresponding inclusion comparison set vanishes. For finite m and n , the false-inclusion component of the exclusion–inclusion gap equals zero whenever $f(\hat{\mathcal{E}}_m(b))$ consists only of weakly nondominated points along the lower and left edges of the convex hull of $f(\mathcal{X}_m)$ in Figure 6(c). For a decision-maker observing only $\hat{\mathcal{P}}_m(b)$, these boundary points may appear to offer distinct trade-offs when they are falsely included under finite sampling; however, as the simulation budget increases, increasingly precise objective estimates reduce the apparent advantages responsible for their false inclusion.

In addition to the normal error model considered previously, we implement Algorithm 1 under three Rademacher-based error models designed to induce exceedance events. Under *Rademacher Error*, all $Z_{jk}(x)$ are i.i.d. Rademacher random variables; that is, $P(Z_{jk}(x) = -1) = P(Z_{jk}(x) = 1) = 1/2$ and $\text{Var}(Z_{jk}(x)) = 1$ for all simulation replications j , objectives k , and feasible points x . For the two adversarial models, let $\tilde{\mathcal{X}}_k(\tau_m) := \{x \in \mathcal{X} : f_k(x) \leq \tau_m\}$ be the set of decision points mapping to objective values within the threshold region on objective k . Under *Adversarial Rademacher Error with Independent Objectives*, for a given $x \in \mathcal{X}$, the $Z_{jk}(x)$ are i.i.d. Rademacher for all replications j and objectives k . Then, for given j and k , all $x \in \tilde{\mathcal{X}}_k(\tau_m)$ receive the same error value, either 1 or -1 , while all $x \in (\tilde{\mathcal{X}}_k(\tau_m))^c$ receive its antithetic value, -1 or 1, respectively. Thus, the errors are perfectly negatively correlated across the two regions. Finally, under *Adversarial Rademacher Error with Correlated Objectives*, for given objective k and $x \in \mathcal{X}$, the $Z_{jk}(x)$ are i.i.d. Rademacher for all replications j . Then, all $x \in \cap_{k=1}^d \tilde{\mathcal{X}}_k(\tau_m)$ receive the same error value, while points outside this intersection receive its antithetic value. Thus, the errors are perfectly correlated across both objectives and design points. Constants for each problem and error model appear in Table 2.

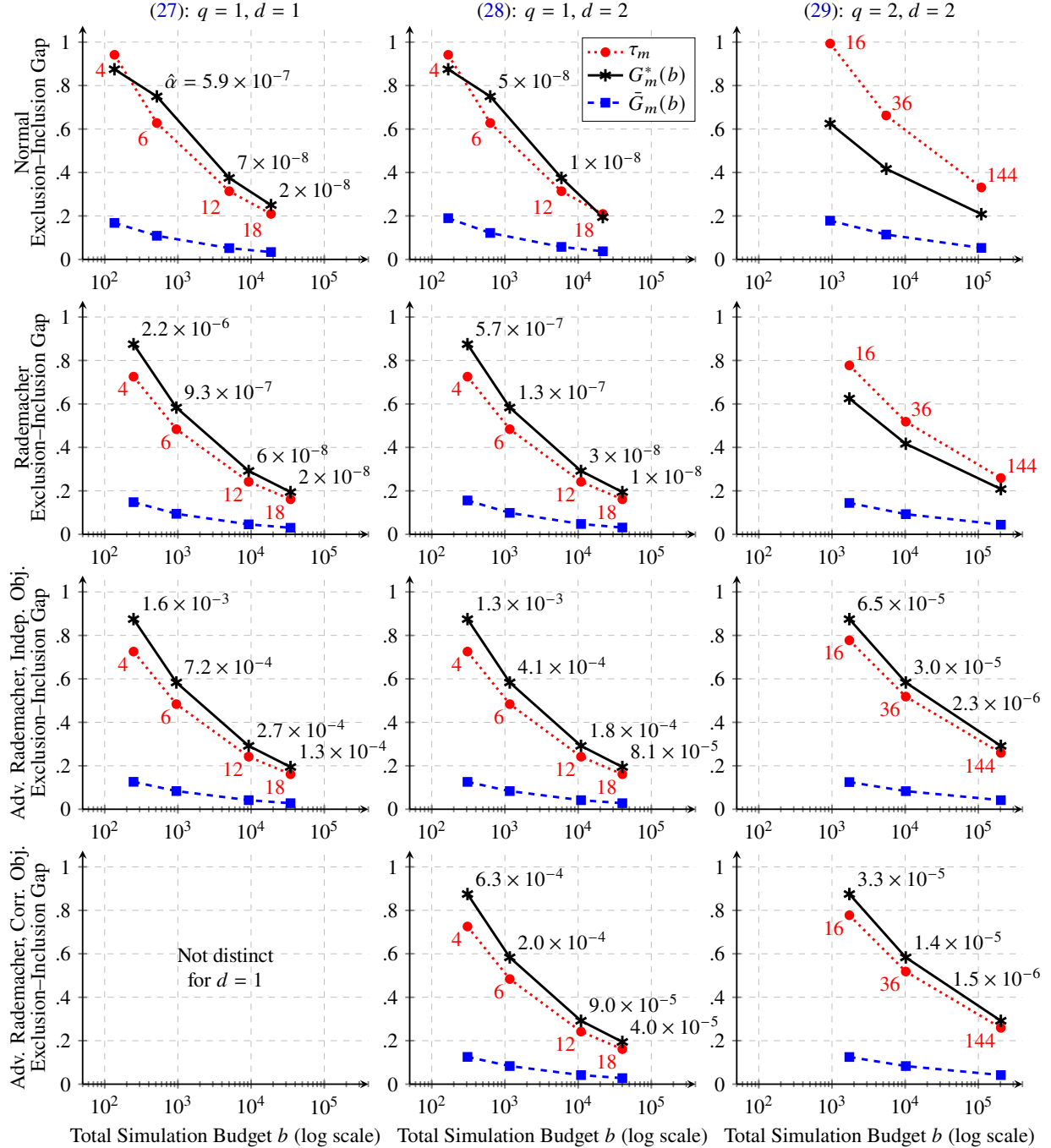
Figure 7 reports results for the linear problems in (27), (28), and (29) under the normal, Rademacher, and both adversarial Rademacher error models. Due to the computational speed of the linear problems, we perform $s = 10^8$ independent macro replications of Algorithm 1 under the allocation in Theorem 3 Part 1 for each problem and error type. Unlike the nonlinear problems in the previous section, the linear problems and error models produce observed threshold exceedance,

Table 2: Constants for Linear Problems

Problem	Error Model	d	q	r	p	$\tilde{c}_q$	ℓ	$\ \eta_2^*\ _\infty$	κ_2	c_2
(27)	Normal	1	1	2	∞	1/2	1	$\sqrt{8/3}$	4/3	3.7660
(27)	Rademacher (all types)	1	1	2	∞	1/2	1	$1/\sqrt{\ln 2}$	$1/(2 \ln 2)$	2.9022
(28)	Normal	2	1	2	∞	1/2	1	$\sqrt{8/3}$	4/3	3.7660
(28)	Rademacher (all types)	2	1	2	∞	1/2	1	$1/\sqrt{\ln 2}$	$1/(2 \ln 2)$	2.9022
(29)	Normal	2	2	2	∞	$\sqrt{2}/2$	1	$\sqrt{8/3}$	4/3	3.9731
(29)	Rademacher (all types)	2	2	2	∞	$\sqrt{2}/2$	1	$1/\sqrt{\ln 2}$	$1/(2 \ln 2)$	3.1094

Note: Constants η_{2k}^* , $k = 1, \dots, d$, and κ_2 are calculated according to the examples in Appendix EC.1.

Figure 7: Results from $s = 10^8$ independent macro replications of [Algorithm 1](#) on linear problems (27), (28), and (29). Curves show the threshold τ_m , the maximum observed gap $G_m^*(b)$, and the average gap $\tilde{G}_m(b)$.



Note: Red labels on red dots indicate the values of m . Black labels indicate (rounded) nonzero estimates of $\hat{\alpha}$ for comparison with $\alpha = 0.5$. For $m = 4$ on (27) and (28) under normal error, τ_m exceeds the largest possible gap and exceedance events are structurally impossible. In all cases, $\hat{\sigma}(\tilde{G}_m(b))$ is estimated as zero to 4 decimal places. Detailed numerical results appear in [Tables EC.4–EC.6](#).

permitting direct estimation of their probabilities. However, the estimated exceedance probabilities $\hat{\alpha}$ remain substantially below the nominal confidence requirement $\alpha = 0.5$, indicating that the guarantees are conservative by a few orders of magnitude. Finally, for the coarse-grid settings of $m = 4$ on problems (27) and (28) under normal error, the threshold τ_m exceeds the largest possible gap, making exceedance structurally impossible. Therefore, these cases demonstrate conservativeness that is distinct from the remaining experiments, in which exceedance is possible.

The two adversarial Rademacher error constructions also illustrate the role of dependence. The version with independent objectives produces the largest observed exceedance probabilities and creates additional opportunities for false inclusion by allowing multiple estimated efficient solutions with different true objective values to appear nondominated. Perfectly correlated objective errors instead constrain the estimated objective vectors and resulting estimated Pareto set. Thus, stronger objective-wise correlation does not necessarily produce larger exclusion–inclusion gaps. Ultimately, the experiments empirically confirm that the guarantees remain valid under extreme dependence.

6 Concluding Remarks

This paper introduces the exclusion–inclusion gap for analyzing the joint effects of approximation and estimation error in finite Pareto approximations constructed from stochastic estimates. The gap penalizes exclusion and inclusion errors that alter the relevant image-space comparison sets while remaining invariant to weakly nondominated misclassifications that leave the inclusion comparison set unchanged. Estimator consistency ensures that any apparent advantage resulting in such a misclassification vanishes and that the displayed objective vectors converge to the true objective vectors of the returned solutions. The exclusion–inclusion gap’s approximation–estimation decomposition admits general bounds that allow deterministic approximation error guarantees to be combined with componentwise stochastic error control. Applying the analysis to a nonadaptive equal-allocation algorithm for global multiobjective simulation optimization results in a characterization of the required simulation budget growth with the number of objectives and establishes an explicit total-budget convergence rate corresponding to the single-objective rate. Future work may obtain sharper guarantees by exploiting stronger approximation and estimation methods, local Pareto geometry, differential sampling (Lee et al. 2010, Applegate et al. 2020), and dependence across objectives and evaluated solutions, such as dependence induced by common random numbers (Ha et al. 2025).

7 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 Proofs of Asymptotic Results

The following proofs were omitted from the main text of [Subsection 5.3.4](#).

A.1 Proof of [Corollary 2](#)

Proof. For Part 1, since $\tau \rightarrow 0$, the tolerance-driven term in (22) is eventually active, whereby $m = \Theta(\tau^{-q})$ and $\lfloor m^{1/q} \rfloor = \Theta(\tau^{-1})$. Also, since d is fixed and $m = \Theta(\tau^{-q})$, we have $\ln(2dm\alpha^{-1}) = \ln(\tau^{-q}\alpha^{-1}) + O(1)$. Because $\tau^{-q}\alpha^{-1} \geq \tau^{-q} \rightarrow \infty$, it follows that $\ln(2dm\alpha^{-1}) = \Theta(\ln(\tau^{-q}\alpha^{-1}))$. Using (21), $n = \Theta(\tau^{-2} \ln(\tau^{-q}\alpha^{-1}))$, where the integer rounding does not affect the asymptotic order. Since $b = mn$, the result follows. For Part 2, fix $\alpha \in (0, 1)$ and notice that Part 1 applies. Then, we have $\ln(\tau^{-q}\alpha^{-1}) = q \ln(\tau^{-1}) + \ln(\alpha^{-1}) = \Theta(\ln(\tau^{-1}))$. Substituting into Part 1 yields the result. For Part 3, since (22) does not depend on α , m and $\lfloor m^{1/q} \rfloor$ remain fixed. Using (21), $n = \Theta(\ln(2dm\alpha^{-1})) = \Theta(\ln(\alpha^{-1}))$. Since $b = mn$ and m is fixed, the result follows. For Part 4, first, suppose $1 \leq p < \infty$. By the assumptions of Part 4, $c_i = \tilde{c}_q \ell + 2 \|\eta_i^*\|_p = \Theta(d^{1/p})$. Thus, the tolerance-driven term in (22) is eventually active, and $m = \Theta(d^{q/p})$, $\lfloor m^{1/q} \rfloor = \Theta(d^{1/p})$. Because $m = \Theta(d^{q/p})$ and q, p , and α are fixed, $\ln(2dm\alpha^{-1}) = (1 + q/p) \ln d + O(1) = \Theta(\ln d)$. It follows from (21) and the uniform bounds on κ_i that $n = \Theta(d^{2/p} \ln d)$. Therefore, $b = mn = \Theta(d^{(q+2)/p} \ln d)$. Now suppose $p = \infty$, whereby $c_i = O(1)$. Since τ and q are fixed and $\tilde{\varepsilon}$ remains uniformly bounded, (22) gives $m = \Theta(1)$, $\lfloor m^{1/q} \rfloor = \Theta(1)$. It follows that $\ln(2dm\alpha^{-1}) = \ln d + O(1) = \Theta(\ln d)$. Therefore, (21) and the uniform bounds on κ_i give $n = \Theta(\ln d)$. Using $b = mn = \Theta(\ln d)$ yields the result. $\square$

A.2 Proof of [Corollary 3](#)

Proof. We first show that under each allocation, $b = \Theta(m^{(q+2)/q} \ln m)$, which ultimately implies

$$m^{-1/q} = \Theta((b/\ln b)^{-1/(q+2)}). \quad (30)$$

For Case 1, in [Theorem 3](#) Part 1, we have $\tau_m = \Theta(m^{-1/q})$ and since $\alpha = \tau_m^\gamma$, it follows that $\alpha^{-1} = \Theta(m^{\gamma/q})$. Therefore, $\ln(2dm\alpha^{-1}) = (1 + \gamma/q) \ln m + O(1) = \Theta(\ln m)$. Also, (21) gives $n = \Theta(m^{2/q} \ln m)$, whereby $b = mn$ yields $b = \Theta(m^{(q+2)/q} \ln m)$. For Case 2, using [Corollary 2](#) Part 1 with $\alpha = \tau^\gamma$ yields $b = \Theta(\tau^{-(q+2)} \ln(\tau^{-1}))$. The value of m in [Theorem 3](#) Part 2 gives $m = \Theta(\tau^{-q})$, whereby $m^{(q+2)/q} = \Theta(\tau^{-(q+2)})$ and $\ln m = \Theta(\ln(\tau^{-1}))$. Substitution gives $b = \Theta(m^{(q+2)/q} \ln m)$. For Case 3, [Theorem 4](#) gives $m = \Theta((b^*/\ln b^*)^{q/(q+2)})$, which implies $m^{(q+2)/q} = \Theta(b^*/\ln b^*)$ and $\ln m = \Theta(\ln b^*)$. Therefore, $b^* = \Theta(m^{(q+2)/q} \ln m)$. We now replace b^* with b . Since $b = m \lfloor b^*/m \rfloor$, $0 \leq b^* - b < m$ with $m/b^* = O((b^*)^{-2/(q+2)} (\ln b^*)^{-q/(q+2)}) \rightarrow 0$. Since $b^* - m < b \leq b^*$, we have $b/b^* \rightarrow 1$. Therefore, $b = \Theta(b^*)$, which implies $b = \Theta(m^{(q+2)/q} \ln m)$. Under all three allocations, $b = \Theta(m^{(q+2)/q} \ln m)$. Since $\ln b = \Theta(\ln m)$, we have $b/\ln b = \Theta(m^{(q+2)/q})$, and (30) follows.

Next, we show $G_m(b)$ and $\mathbb{H}_p(\hat{\mathcal{P}}_m(b), f(\hat{\mathcal{E}}_m(b)))$ are $O_p(m^{-1/q})$. For $G_m(b)$, in Case 1,

Theorem 3 Part 1 and $\alpha = \tau_m^\gamma$ give $P\{G_m(b) > \tau_m\} \leq \tau_m^\gamma \rightarrow 0$ because $\tau_m = c_i \lfloor m^{1/q} \rfloor^{-1} = \Theta(m^{-1/q}) \rightarrow 0$ as $m \rightarrow \infty$. In Case 2, **Theorem 3** Part 2 and $\alpha = \tau^\gamma$ give $P\{G_m(b) > \tau\} \leq \tau^\gamma \rightarrow 0$ as $\tau \rightarrow 0$. Moreover, $m = \Theta(\tau^{-q})$, and hence $\tau = \Theta(m^{-1/q})$. In Case 3, **Theorem 4** gives $P\{G_m(b) > \tau_m\} \leq 2d/n \rightarrow 0$ because d is fixed and $m/b^* \rightarrow 0$, so $b^*/m \rightarrow \infty$ which implies $n \rightarrow \infty$. Further, $\tau_m = \Theta(m^{-1/q})$. Therefore, in all three cases, $G_m(b) = O_p(m^{-1/q})$. For $\mathbb{H}_p(\hat{\mathcal{P}}_m(b), f(\hat{\mathcal{E}}_m(b)))$, consider the balanced componentwise error event $|\bar{F}_k(x, n) - f_k(x)| < \eta_{ik}^* \lfloor m^{1/q} \rfloor^{-1}$ for all $x \in \mathcal{X}_m, k = 1, \dots, d$. On this event, $\Delta_m(b) = \max_{x \in \mathcal{X}_m} \|\bar{F}(x, n) - f(x)\|_p \leq \|\eta_i^*\|_p \lfloor m^{1/q} \rfloor^{-1}$. Because $\hat{\mathcal{P}}_m(b)$ and $f(\hat{\mathcal{E}}_m(b))$ are obtained by applying $\bar{F}(\cdot, n)$ and f , respectively, to the same set $\hat{\mathcal{E}}_m(b)$, we have $\mathbb{H}_p(\hat{\mathcal{P}}_m(b), f(\hat{\mathcal{E}}_m(b))) \leq \Delta_m(b) \leq \|\eta_i^*\|_p \lfloor m^{1/q} \rfloor^{-1}$. The arguments leading to **Corollary 1** also imply $P\{\mathbb{H}_p(\hat{\mathcal{P}}_m(b), f(\hat{\mathcal{E}}_m(b))) > \|\eta_i^*\|_p \lfloor m^{1/q} \rfloor^{-1}\} \leq 2dm \exp(-\kappa_i n \lfloor m^{1/q} \rfloor^{-2})$. In Cases 1 and 2, the allocation in **Theorem 3** makes this bound no greater than α , which converges to zero under the respective choices $\alpha = \tau_m^\gamma$ and $\alpha = \tau^\gamma$. In Case 3, the calculation in **Theorem 4** makes the bound no greater than $2d/n$, which also converges to zero. Since the problem and error model are fixed, $\|\eta_i^*\|_p$ is constant and $\lfloor m^{1/q} \rfloor^{-1} = \Theta(m^{-1/q})$. Then, in all three cases, $\mathbb{H}_p(\hat{\mathcal{P}}_m(b), f(\hat{\mathcal{E}}_m(b))) = O_p(m^{-1/q})$.

Combining these two $O_p(m^{-1/q})$ results with (30) yields the stated rates. Because we have $(b/\ln b)^{-1/(q+2)} \rightarrow 0$, both quantities converge in probability to zero. $\square$

Acknowledgments

This work has benefited from Hunter's participation in Dagstuhl Seminar 20031 "Scalability in Multiobjective Optimization," <https://www.dagstuhl.de/20031>, Dagstuhl Seminar 23361 "Multiobjective Optimization on a Budget," <https://www.dagstuhl.de/23361>, and Dagstuhl Seminar 26041 "Uncertainty Quantification in Multiobjective Optimization," <https://www.dagstuhl.de/26041>. In the latter seminar, Hunter is especially grateful for the discussions and perspective of the second working group on representations.

References

- Alrefaei MH, Diabat AH (2009) A simulated annealing technique for multi-objective simulation optimization. *Applied Mathematics and Computation* 215(8):3029–3035.
- Andradóttir S (2015) A review of random search methods. Fu MC, ed., *Handbook of Simulation Optimization*, volume 216 of *International Series in Operations Research & Management Science*, 277–292 (New York: Springer), URL http://dx.doi.org/10.1007/978-1-4939-1384-8_10.
- Applegate EA, Feldman G, Hunter SR, Pasupathy R (2020) Multi-objective ranking and selection: Optimal sampling laws and tractable approximations via SCORE. *Journal of Simulation* 14(1):21–40, URL <http://dx.doi.org/10.1080/17477778.2019.1633891>.
- Audet C, Bignon J, Cartier D, Le Digabel S, Salomon L (2021) Performance indicators in multiobjective

- optimization. *European Journal of Operational Research* 292:397–422, URL <http://dx.doi.org/10.1016/j.ejor.2020.11.016>.
- Bonnell H, Collonge C (2014) Stochastic optimization over a Pareto set associated with a stochastic multi-objective optimization problem. *Journal of Optimization Theory and Applications* 162:405–427, URL <http://dx.doi.org/10.1007/s10957-013-0367-8>.
- Branke J (2025) Performance metrics for multi-objective optimisation under noise. *IEEE Transactions on Evolutionary Computation* 29(4):1449–1455, URL <http://dx.doi.org/10.1109/TEVC.2024.3438115>.
- Cooper K, Hunter SR, Nagaraj K (2020) Bi-objective simulation optimization on integer lattices using the epsilon-constraint method in a retrospective approximation framework. *INFORMS Journal on Computing* 32(4):1080–1100, URL <http://dx.doi.org/10.1287/ijoc.2019.0918>.
- Eckman DJ, Henderson SG, Shashaani S (2023a) Diagnostic tools for evaluating and comparing simulation-optimization algorithms. *INFORMS Journal on Computing* 35(2):350–367, URL <http://dx.doi.org/10.1287/ijoc.2022.1261>.
- Eckman DJ, Henderson SG, Shashaani S (2023b) SimOpt: a testbed for simulation-optimization experiments. *INFORMS Journal on Computing* 35(2):495–508, URL <http://dx.doi.org/10.1287/ijoc.2023.1273>.
- Ehrgott M (2005) *Multicriteria Optimization*, volume 491 of *Lecture Notes in Economics and Mathematical Systems* (Heidelberg: Springer), 2nd edition, URL <http://dx.doi.org/10.1007/3-540-27659-9>.
- Ensor KB, Glynn PW (1997) Stochastic optimization via grid search. *Lectures in Applied Mathematics* 33:89–100.
- Faulkenberg SL, Wiecek MM (2010) On the quality of discrete representations in multiple objective programming. *Optim Eng* 11:423–440, URL <http://dx.doi.org/10.1007/s11081-009-9099-x>.
- Feldman G, Hunter SR (2018) SCORE allocations for bi-objective ranking and selection. *ACM Transactions on Modeling and Computer Simulation* 28(1):7:1–7:28, URL <http://dx.doi.org/10.1145/3158666>.
- Fu MC, ed. (2015) *Handbook of Simulation Optimization*, volume 216 of *International Series in Operations Research & Management Science* (New York: Springer).
- Glynn PW, Heidelberger P (1990) Bias properties of budget constrained simulations. *Operations Research* 38(5):801–814.
- Glynn PW, Heidelberger P (1991) Analysis of parallel replicated simulations under a completion time constraint. *ACM Transactions on Modeling and Computer Simulation* 1(1):3–23.
- Gupta A, Hunter SR (2025) Properties of two-stage stochastic multi-objective linear programs. *Mathematical Programming* URL <http://dx.doi.org/10.1007/s10107-025-02280-7>.
- Ha Y, Shashaani S, Pasupathy R (2025) Complexity of zeroth- and first-order stochastic trust-region algorithms. *SIAM Journal on Optimization* 35(3):2098–2127, URL <http://dx.doi.org/10.1137/24M1664484>.
- Hamacher HW, Pedersen CR, Ruzika S (2007) Finding representative systems for discrete bicriterion optimization problems. *Operations Research Letters* 35.

- Han M, Ouyang L (2022) A novel Bayesian approach for multi-objective stochastic simulation optimization. *Swarm and Evolutionary Computation* 75.
- Heidelberger P (1988) Discrete event simulations and parallel processing: statistical properties. *Siam J. Stat. Comput.* 9(6):1114–1132.
- Herzel A, Ruzika S, Thielen C (2021) Approximation methods for multiobjective optimization problems: A survey. *INFORMS Journal on Computing* URL <http://dx.doi.org/10.1287/ijoc.2020.1028>.
- Huang H, Zabinsky ZB (2014) Multiple objective probabilistic branch and bound for Pareto optimal approximation. Tolk A, Diallo SY, Ryzhov IO, Yilmaz L, Buckley S, Miller JA, eds., *Proceedings of the 2014 Winter Simulation Conference*, 3916–3927 (Piscataway, NJ: IEEE), URL <http://dx.doi.org/10.1109/WSC.2014.7020217>.
- Huang H, Zabinsky ZB (2026) Non-linear multi-objective optimization with probabilistic branch and bound. *Asia-Pacific Journal of Operational Research* URL <http://dx.doi.org/10.1142/S0217595926400063>.
- Hunter SR (2026) Multi-objective stochastic optimization: Concepts, recent advances, and perspectives. Ramamohan V, Fakhimi M, Krejci C, Ruiz-Martin C, Onggo S, Mustafee N, eds., *Proceedings of the 2026 Winter Simulation Conference* (Piscataway, NJ: IEEE).
- Hunter SR, Applegate EA, Arora V, Chong B, Cooper K, Rincón-Guevara O, Vivas-Valencia C (2019) An introduction to multi-objective simulation optimization. *ACM Transactions on Modeling and Computer Simulation* 29(1):7:1–7:36, URL <http://dx.doi.org/10.1145/3299872>.
- Hunter SR, McClosky B (2016) Maximizing quantitative traits in the mating design problem via simulation-based Pareto estimation. *IIE Transactions* 48(6):565–578.
- Hunter SR, Ondes BE (2023) Properties of several performance indicators for global multi-objective simulation optimization. Corlu CG, Hunter SR, Lam H, Onggo BS, Shortle J, Biller B, eds., *Proceedings of the 2023 Winter Simulation Conference*, 3577–3588 (Piscataway, NJ: IEEE), URL <http://dx.doi.org/10.1109/WSC60868.2023.10408213>.
- Kirlik G, Sayin S (2018) Bilevel programming for generating discrete representations in multi-objective optimization. *Mathematical Programming* 169, URL <http://dx.doi.org/10.1007/s10107-017-1149-0>.
- Kirlik G, Sayin S (2025) Generating representative sets for multiobjective discrete optimization problems with specified coverage errors. *Computational Optimization and Applications* 90, URL <http://dx.doi.org/10.1007/s10589-024-00627-y>.
- Kuhn T, Ruzika S (2017) A coverage-based box-algorithm to compute a representation for optimization problems with three objective functions. *Journal of Global Optimization* 67, URL <http://dx.doi.org/10.1007/s10898-016-0425-5>.
- Lee LH, Chew EP, Teng S, Goldsman D (2010) Finding the non-dominated Pareto set for multi-objective simulation models. *IIE Transactions* 42:656–674, URL <http://dx.doi.org/10.1080/07408171003705367>.

- Li M, Yao X (2019) Quality evaluation of solution sets in multiobjective optimisation: A survey. *ACM Computing Surveys* 52(2), URL <http://dx.doi.org/10.1145/3300148>.
- Löhne A (2011) *Vector Optimization with Infimum and Supremum*. Vector Optimization (Springer Berlin, Heidelberg).
- Mattila V, Virtanen K, Hämäläinen RP (2013) A simulated annealing algorithm for noisy multiobjective optimization. *Journal of Multi-criteria Decision Analysis* 20:255–276.
- Nika A, Elahi S, Ararat C, Tekin C (2025) Beyond grids: Multi-objective bayesian optimization with adaptive discretization. *Transactions on Machine Learning Research* ISSN 2835-8856, URL <https://openreview.net/forum?id=Wq150HaRVE>.
- Ondes BE (2025) *Theory and methods for global multi-objective simulation optimization on compact sets*. Ph.D. thesis, Purdue University, West Lafayette, IN, USA, URL <http://dx.doi.org/10.25394/PGS.29377805>.
- Ondes BE, Hunter SR (2023) 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, URL <http://dx.doi.org/10.1007/s11590-022-01920-7>.
- Pasupathy R, Henderson SG (2006) A testbed of simulation-optimization problems. Perrone LF, Wieland FP, Liu J, Lawson BG, Nicol DM, Fujimoto RM, eds., *Proceedings of the 2006 Winter Simulation Conference*, 255–263 (Piscataway, NJ: IEEE), URL <http://dx.doi.org/10.1109/WSC.2006.323081>.
- Pasupathy R, Henderson SG (2011) SimOpt: A library of simulation optimization problems. Jain S, Creasey RR, Himmelspace J, White KP, Fu M, eds., *Proceedings of the 2011 Winter Simulation Conference*, 4075–4085 (Piscataway, NJ: IEEE), URL <http://dx.doi.org/10.1109/WSC.2011.6148097>.
- Rojas-Gonzalez S, Jalali H, Van Nieuwenhuysse I (2020) A multiobjective stochastic simulation optimization algorithm. *European Journal of Operational Research* 284:212–226, URL <http://dx.doi.org/10.1016/j.ejor.2019.12.014>.
- Rojas-Gonzalez S, Van Nieuwenhuysse I (2020) A survey on kriging-based infill algorithms for multiobjective simulation optimization. *Computers and Operations Research* 116, URL <http://dx.doi.org/10.1016/j.cor.2019.104869>.
- Ruzika S (2007) *On multiple objective combinatorial optimization*. Ph.D. thesis, Department of Mathematics, University of Kaiserslautern, Kaiserslautern, Germany.
- Ruzika S, Wiecek MM (2005) Approximation methods in multiobjective programming. *Journal of Optimization Theory and Applications* 126:473–501, URL <http://dx.doi.org/10.1007/s10957-005-5494-4>.
- Sayin S (2000) Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming. *Mathematical Programming* 87:543–560.
- Shapiro A, Dentcheva D, Ruszczyński A (2009) *Lectures on Stochastic Programming: Modeling and Theory*. MPS-SIAM Series on Optimization (Philadelphia, PA: Society for Industrial and Applied Mathematics).

- Thengvall BG, Hall SN, Deskevich MP (2025) Measuring the effectiveness and efficiency of simulation optimization metaheuristic algorithms. *Journal of Heuristics* 31.
- Vershynin R (2018) *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics* (Cambridge, UK: Cambridge University Press).
- Yakowitz S, L'Ecuyer P, Vázquez-Abad F (2000) Global stochastic optimization with low-dispersion point sets. *Operations Research* 48(6):939–950, URL <http://dx.doi.org/10.1287/opre.48.6.939.12393>.
- Zhang J, Ma Y, Yang T, Liu L (2017) Estimation of the Pareto front in stochastic simulation through stochastic Kriging. *Simulation Modelling Practice and Theory* 79:69–86, URL <http://dx.doi.org/10.1016/j.simpat.2017.09.006>.
- Zhao J, Eckman DJ (2026) Methods of plausible inference for multi-objective simulation optimization. *IIE Transactions* URL <http://dx.doi.org/10.1080/24725854.2026.2726218>.
- Zitzler E, Thiele L, Laumanns M, Fonseca CM, Grunert da Fonseca V (2003) Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation* 7(2):117–132.
- Zuluaga M, Krause A, Püschel M (2016) ϵ -PAL: An active learning approach to the multi-objective optimization problem. *Journal of Machine Learning Research* 17:1–32, URL <https://jmlr.csail.mit.edu/papers/volume17/15-047/15-047.pdf>.

Electronic Companion

EC.1 Calculating Constants in the Concentration Inequalities

In Examples EC.1–EC.3, we show how to determine values for the constants in (16) and (17).

Example EC.1 (Laplace Error). Given an objective k and feasible point $x \in \mathcal{X}$, suppose the errors $Z_{jk}(x)$, $j = 1, \dots, n$ are i.i.d. with mean-zero Laplace distribution having scale parameter $\beta_k(x) > 0$ and standard deviation $\sigma_k(x) = \sqrt{2}\beta_k(x)$. Using a Chernoff bound, for any $\delta > 0$, $s > 0$,

$$\mathbb{P}\left\{\frac{1}{n} \sum_{j=1}^n Z_{jk}(x) \geq \delta\right\} \leq e^{-sn\delta} \prod_{j=1}^n \mathbb{E}\left[e^{sZ_{jk}(x)}\right], \quad (31)$$

where for each $j = 1, \dots, n$, the moment generating function (mgf) of $Z_{jk}(x)$ has upper bound $\mathbb{E}\left[e^{sZ_{jk}(x)}\right] = (1 - s^2\beta_k^2(x))^{-1} \leq e^{2s^2\beta_k^2(x)}$ for all $|s| \leq (2\beta_k(x))^{-1}$. Choosing $s = \min\{1/(2\beta_k(x)), \delta/(2\beta_k(x))^2\}$ and using the fact that $Z_{jk}(x)$, $j = 1, \dots, n$ are i.i.d. random variables from a symmetric distribution, whereby the distribution of $\sum_{j=1}^n Z_{jk}(x)$ is also symmetric, it follows that

$$\begin{aligned} \mathbb{P}\left\{\left|\frac{1}{n} \sum_{j=1}^n Z_{jk}(x)\right| \geq \delta\right\} &\leq 2 \exp\left(-n(s\delta - (1/2)s^2(2\beta_k(x))^2)\right) \\ &\leq 2 \exp\left(- (n/2) \min\left\{\frac{\delta}{2\beta_k(x)}, \frac{\delta^2}{(2\beta_k(x))^2}\right\}\right). \end{aligned} \quad (32)$$

Each $|Z_{jk}(x)|$ has an exponential distribution with parameter $1/\beta_k(x)$. Then the sub-exponential norm of $Z_{1k}(x)$ is $\|Z_{1k}(x)\|_{\psi_1} = \inf\{s > \beta_k(x) : s(s - \beta_k(x))^{-1} \leq 2\} = 2\beta_k(x)$. Substituting $2\beta_k(x) = \|Z_{1k}(x)\|_{\psi_1}$ into (32) and ensuring the right side of (32) is less than or equal to the right side of (16) yields $\kappa_1 \leq 1/2$. Therefore, we set $\kappa_1 = 1/2$. If all errors belong to the Laplace family, then $\kappa_1 = 1/2$ for all k, x . Further, $\eta_{1k}^* = 2 \sup\{\beta_k(x) : x \in \mathcal{X}\}$.

Example EC.2 (Normal Error). Given an objective k and feasible point $x \in \mathcal{X}$, let the errors $Z_{jk}(x)$, $j = 1, \dots, n$ be i.i.d. with mean-zero normal distribution having variance $\sigma_k^2(x) > 0$. Then it is straightforward to show that

$$\mathbb{P}\left\{\left|\frac{1}{n} \sum_{j=1}^n Z_{jk}(x)\right| \geq \delta\right\} \leq 2 \exp\left(-\frac{n\delta^2}{2\sigma_k^2(x)}\right). \quad (33)$$

Moreover, $\|Z_{1k}(x)\|_{\psi_2} = \inf\{s > \sqrt{2}\sigma_k(x) : s(s^2 - 2\sigma_k^2(x))^{-1/2} \leq 2\} = \sqrt{8/3}\sigma_k(x)$. Therefore, $\kappa_2 = 4/3$ makes the right side of (17) equal to the tail bound in (33). If all errors belong to the normal family, then we apply $\kappa_2 = 4/3$ for all k, x . Further, $\eta_{2k}^* = \sqrt{8/3} \sup\{\sigma_k(x) : x \in \mathcal{X}\}$.

Example EC.3 (Rademacher Error). Given an objective k and feasible point $x \in \mathcal{X}$, let the errors $Z_{jk}(x)$, $j = 1, \dots, n$ be i.i.d. with mean-zero Rademacher distribution; that is, we have $P\{Z_{jk}(x) = -1\} = P\{Z_{jk}(x) = 1\} = 1/2$ and $\text{Var}(Z_{jk}(x)) = 1$ for all $j = 1, \dots, n$. First, because

$Z_{jk}(x)^2 = 1$ w.p. 1, the sub-Gaussian norm is $\|Z_{jk}(x)\|_{\psi_2} = \inf\{s > 0: \exp(1/s^2) \leq 2\} = 1/\sqrt{\ln 2}$. Also, the mgf of $Z_{jk}(x)$ is $E[e^{sZ_{jk}(x)}] = \frac{1}{2}(e^s + e^{-s})$ for all $s \in \mathbb{R}$. From (31), for any $\delta > 0$, $s > 0$, $P\{\frac{1}{n} \sum_{j=1}^n Z_{jk}(x) \geq \delta\} \leq \exp(-n[s\delta - \ln(\frac{1}{2}(e^s + e^{-s}))])$. For $\delta \in (0, 1)$, the Chernoff exponent is maximized at $s^* = \frac{1}{2} \ln(\frac{1+\delta}{1-\delta})$. Then apply symmetry to yield that

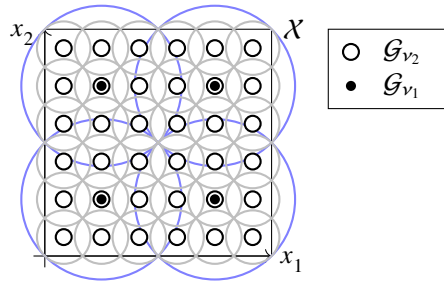
$$P\{|\frac{1}{n} \sum_{j=1}^n Z_{jk}(x)| \geq \delta\} \leq 2 \exp(-nI(\delta)) \quad (34)$$

where $I(\delta) := \frac{1+\delta}{2} \ln(1+\delta) + \frac{1-\delta}{2} \ln(1-\delta)$ for $\delta \in (0, 1)$. Because $\|Z_{jk}(x)\|_{\psi_2}^2 = 1/\ln(2)$, (17) holds for every $\delta \in (0, 1)$ provided that $\kappa_2 \leq \inf_{\delta \in (0,1)} I(\delta)(\delta^2 \ln 2)^{-1} = (2 \ln 2)^{-1}$, where the last equality holds because $I'(\delta) = \frac{1}{2} \ln(\frac{1+\delta}{1-\delta}) \geq \delta$, so $I(\delta) \geq \delta^2/2$ and $I(\delta)/\delta^2 \rightarrow 1/2$ as $\delta \downarrow 0$. Therefore, the largest valid constant is $\kappa_2 = 1/(2 \ln 2) \approx 0.7213$. The quadratic bound holds at $\delta = 1$, and the left side of (17) equals zero for $\delta > 1$. If all errors are Rademacher, then we apply this value for all k, x . Further, $\eta_{2k}^* = 1/\sqrt{\ln 2}$.

EC.2 Nested Grid Designs

The discretization sequences used in Subsection 5.4 ensure that certain grids share common points, which facilitates comparisons across different values of m . Let $\mathcal{X} = [0, 1]^q$, and for positive integer ν , define the grid $\mathcal{G}_\nu = \{(x_1, \dots, x_q) : x_j \in \{\frac{1}{2\nu}, \frac{3}{2\nu}, \dots, \frac{2\nu-1}{2\nu}\} \forall j\}$, which is the structured portion of the grid design used in Example 1. If $\nu_2 = 3\nu_1$, then $\mathcal{G}_{\nu_1} \subset \mathcal{G}_{\nu_2}$. To see this, let $x_j = (2i-1)/(2\nu_1)$ for some $i \in \{1, \dots, \nu_1\}$. Since $\nu_2 = 3\nu_1$, $x_j = 3(2i-1)/(2\nu_2)$. The numerator $3(2i-1)$ is odd and satisfies $1 \leq 3(2i-1) \leq 2\nu_2 - 3$, so x_j is also one of the grid coordinates defining $\mathcal{G}_{\nu_2}$. Applying this argument componentwise yields $\mathcal{G}_{\nu_1} \subset \mathcal{G}_{\nu_2}$. Figure EC.1 illustrates the construction for $q = 2$.

Figure EC.1: Example of nested grids. Setting $\nu_2 = 3\nu_1$ produces sequences with common design points.



EC.3 Detailed Numerical Results for Nonlinear Problems

Tables EC.1–EC.3 report the numerical values underlying Figure 5. Given the nonlinearity in the objectives, deterministic results reported in the rows are not necessarily monotone decreasing except within a subsequence of nested design points.

Each experiment for problems (25) and (26) uses $s = 10^4$ independent macro replications. The true Pareto upper image is approximated using a finite decision-space reference grid augmented with all reported experimental design grids. The reference grid contains 512^2 points for problem (25) and 144^3 points for problem (26), before augmentation. The distances are calculated directly from finite sets of nondominated boundary points that generate the relevant upper images under the componentwise order. Resolution studies compared reported distances under alternative decision-space reference grids. For (25), comparison of the 512^2 and 1024^2 point reference grids supports reporting distances to three decimal places. For (26), comparison of the 144^3 point reference grid with a reference formed from the union of the 96^3 and 144^3 point reference grids supports reporting

Table EC.1: Results for $s = 10^8$ independent replications of Algorithm 1 on (24) with $d = 1$, $q = 2$.

m	n	b	$\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$	Max Across Replications			τ_m	$\bar{G}_m(b)$	$\widehat{\text{se}}(\bar{G}_m(b))$	$\hat{\alpha}$
				$G_m^{\text{FE}}(b)$	$G_m^{\text{FI}}(b)$	$G_m^*(b)$				
Laplace Error										
16	134	2144	0.0214	0.6697	–	0.6911	2.2981	0.0222	0.0000	0.000 000 00*
36	358	12 888	0.1376	0.3025	–	0.4402	1.5321	0.1490	0.0000	0.000 000 00
144	1831	263 664	0.0214	0.1136	–	0.1350	0.7660	0.0257	0.0000	0.000 000 00
324	4645	1 504 980	0.0096	0.0813	–	0.0909	0.5107	0.0125	0.0000	0.000 000 00
Normal Error										
16	50	800	0.0214	1.0214	–	1.0428	2.4075	0.0410	0.0000	0.000 000 00*
36	135	4860	0.1376	0.5429	–	0.6805	1.6050	0.1596	0.0000	0.000 000 00
144	687	98 928	0.0214	0.2406	–	0.2619	0.8025	0.0275	0.0000	0.000 000 00
324	1742	564 408	0.0096	0.1280	–	0.1376	0.5350	0.0154	0.0000	0.000 000 00

Note: Given m and $\alpha = 0.5$, the results use n, b from Theorem 3 Part 1, constants in Table 1, and report only false exclusion since $d = 1$. *For the indicated rows, $\hat{\alpha} = 0$ because τ_m exceeds the largest attainable exclusion–inclusion gap.

Table EC.2: Results for $s = 10^4$ independent replications of Algorithm 1 on (25) with $d = 2$, $q = 2$.

m	n	b	$\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$	Max Across Replications			τ_m	$\bar{G}_m(b)$	$\widehat{\text{se}}(\bar{G}_m(b))$	$\hat{\alpha}$
				$G_m^{\text{FE}}(b)$	$G_m^{\text{FI}}(b)$	$G_m^*(b)$				
Laplace Error										
16	156	2496	0.202	0.352	0.318	0.483	2.2981	0.206	0.0002	0.0000
36	408	14 688	0.159	0.189	0.173	0.327	1.5321	0.177	0.0003	0.0000
144	2031	292 464	0.096	0.102	0.085	0.137	0.7660	0.101	0.0001	0.0000
324	5094	1 650 456	0.049	0.071	0.064	0.080	0.5107	0.055	0.0001	0.0000
Normal Error										
16	59	944	0.202	0.670	0.511	0.724	2.4075	0.230	0.0005	0.0000*
36	153	5508	0.159	0.272	0.301	0.410	1.6050	0.220	0.0005	0.0000
144	762	109 728	0.096	0.154	0.177	0.194	0.8025	0.110	0.0001	0.0000
324	1911	619 164	0.049	0.092	0.094	0.118	0.5350	0.064	0.0001	0.0000

Note: Given m and $\alpha = 0.5$, the results use n, b from Theorem 3 Part 1 and constants in Table 1. The reference Pareto image uses a 512^2 -point decision grid augmented with the experimental grids; this resolution supports reporting distances to three decimal places. *For the indicated rows, $\hat{\alpha} = 0$ because τ_m exceeds the largest attainable exclusion–inclusion gap.

Table EC.3: Results for $s = 10^4$ independent replications of [Algorithm 1](#) on (26) with $d = 3$, $q = 3$.

m	n	b	$\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$	Max Across Replications			τ_m	$\bar{G}_m(b)$	$\widehat{\text{se}}(\bar{G}_m(b))$	$\hat{\alpha}$
				$G_m^{\text{FE}}(b)$	$G_m^{\text{FI}}(b)$	$G_m^*(b)$				
Laplace Error										
64	213	13 632	0.27	0.35	0.24	0.37	2.6557	0.27	0.0001	0.0000*
216	566	122 256	0.20	0.16	0.17	0.30	1.7704	0.21	0.0002	0.0000
1728	2863	4 947 264	0.12	0.10	0.08	0.15	0.8852	0.13	0.0000	0.0000
Normal Error										
64	80	5120	0.27	0.35	0.32	0.51	2.7651	0.29	0.0004	0.0000*
216	213	46 008	0.20	0.27	0.30	0.39	1.8434	0.26	0.0003	0.0000
1728	1074	1 855 872	0.12	0.15	0.15	0.17	0.9217	0.13	0.0001	0.0000

Note: Given m and $\alpha = 0.5$, the results use n, b from [Theorem 3](#) Part 1 and constants in [Table 1](#). The reference Pareto image uses a 144^3 -point decision grid augmented with the experimental grids; this resolution supports reporting distances to two decimal places. *For the indicated rows, $\hat{\alpha} = 0$ because τ_m exceeds the largest attainable exclusion–inclusion gap.

distances to two decimal places.

EC.4 Detailed Numerical Results for Linear Problems

[Tables EC.4–EC.6](#) report the numerical values underlying [Figure 7](#).

Table EC.4: Results for $s = 10^8$ independent replications of [Algorithm 1](#) on (27) with $q = 1$, $d = 1$.

m	n	b	$\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$	Max Across Replications			τ_m	$\bar{G}_m(b)$	$\widehat{\text{se}}(\bar{G}_m(b))$	$\hat{\alpha}$
				$G_m^{\text{FE}}(b)$	$G_m^{\text{FI}}(b)$	$G_m^*(b)$				
Normal Error										
4	34	136	0.1250	0.7500	–	0.8750	0.9415	0.1679	0.0000	0.000 000 00*
6	86	516	0.0833	0.6667	–	0.7500	0.6277	0.1087	0.0000	0.000 000 59
12	419	5028	0.0417	0.3333	–	0.3750	0.3138	0.0518	0.0000	0.000 000 07
18	1040	18 720	0.0278	0.2222	–	0.2500	0.2092	0.0338	0.0000	0.000 000 02
Rademacher Error										
4	62	248	0.1250	0.7500	–	0.8750	0.7256	0.1479	0.0000	0.000 002 20
6	159	954	0.0833	0.5000	–	0.5833	0.4837	0.0944	0.0000	0.000 000 93
12	773	9276	0.0417	0.2500	–	0.2917	0.2419	0.0458	0.0000	0.000 000 06
18	1921	34 578	0.0278	0.1667	–	0.1944	0.1612	0.0301	0.0000	0.000 000 02
Adversarial Rademacher Error with One Objective										
4	62	248	0.1250	0.7500	–	0.8750	0.7256	0.1262	0.0000	0.001 572 98
6	159	954	0.0833	0.5000	–	0.5833	0.4837	0.0837	0.0000	0.000 715 11
12	773	9276	0.0417	0.2500	–	0.2917	0.2419	0.0417	0.0000	0.000 273 70
18	1921	34 578	0.0278	0.1667	–	0.1944	0.1612	0.0278	0.0000	0.000 129 51

Note: Given m and $\alpha = 0.5$, the results use n, b from [Theorem 3](#) Part 1, constants from [Table 2](#), and report only false exclusion because $d = 1$. *For the indicated rows, $\hat{\alpha} = 0$ because τ_m exceeds the largest attainable exclusion–inclusion gap.

Table EC.5: Results for $s = 10^8$ independent replications of [Algorithm 1](#) on (28) with $q = 1$, $d = 2$.

m	n	b	$\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$	Max Across Replications			τ_m	$\bar{G}_m(b)$	$\hat{s}(\bar{G}_m(b))$	$\hat{\alpha}$
				$G_m^{\text{FE}}(b)$	$G_m^{\text{FI}}(b)$	$G_m^*(b)$				
Normal Error										
4	42	168	0.1250	0.7500	0.7500	0.8750	0.9415	0.1896	0.0000	0.000 000 00*
6	105	630	0.0833	0.3333	0.6667	0.7500	0.6277	0.1217	0.0000	0.000 000 05
12	493	5916	0.0417	0.1667	0.3333	0.3750	0.3138	0.0575	0.0000	0.000 000 01
18	1208	21 744	0.0278	0.1111	0.1667	0.1944	0.2092	0.0373	0.0000	0.000 000 00
Rademacher Error										
4	77	308	0.1250	0.5000	0.7500	0.8750	0.7256	0.1558	0.0000	0.000 000 57
6	194	1164	0.0833	0.3333	0.5000	0.5833	0.4837	0.0987	0.0000	0.000 000 13
12	912	10 944	0.0417	0.1667	0.2500	0.2917	0.2419	0.0475	0.0000	0.000 000 03
18	2233	40 194	0.0278	0.0556	0.1667	0.1944	0.1612	0.0311	0.0000	0.000 000 01
Adversarial Rademacher Error with Independent Objectives										
4	77	308	0.1250	0.7500	0.7500	0.8750	0.7256	0.1259	0.0000	0.001 262 26
6	194	1164	0.0833	0.5000	0.5000	0.5833	0.4837	0.0835	0.0000	0.000 408 17
12	912	10 944	0.0417	0.2500	0.2500	0.2917	0.2419	0.0417	0.0000	0.000 179 77
18	2233	40 194	0.0278	0.0000	0.1667	0.1944	0.1612	0.0278	0.0000	0.000 080 53
Adversarial Rademacher Error with Correlated Objectives										
4	77	308	0.1250	0.7500	0.7500	0.8750	0.7256	0.1255	0.0000	0.000 629 09
6	194	1164	0.0833	0.5000	0.5000	0.5833	0.4837	0.0834	0.0000	0.000 203 91
12	912	10 944	0.0417	0.2500	0.2500	0.2917	0.2419	0.0417	0.0000	0.000 089 95
18	2233	40 194	0.0278	0.1667	0.1667	0.1944	0.1612	0.0278	0.0000	0.000 040 37

Note: Given m and $\alpha = 0.5$, the results use n, b from [Theorem 3](#) Part 1 and constants from [Table 2](#). *For the indicated rows, $\hat{\alpha} = 0$ because τ_m exceeds the largest attainable exclusion–inclusion gap.

Table EC.6: Results for $s = 10^8$ independent replications of [Algorithm 1](#) on (29) with $q = 2$, $d = 2$.

m	n	b	$\mathbb{D}_p(\mathcal{P}^\uparrow, \mathcal{P}_m^\uparrow)$	Max Across Replications			τ_m	$\bar{G}_m(b)$	$\hat{\text{s}}\hat{\text{e}}(\bar{G}_m(b))$	$\hat{\alpha}$
				$G_m^{\text{FE}}(b)$	$G_m^{\text{FI}}(b)$	$G_m^*(b)$				
Normal Error										
16	59	944	0.1250	0.5000	0.5000	0.6250	0.9933	0.1782	0.0000	0.000 000 00*
36	153	5508	0.0833	0.3333	0.3333	0.4167	0.6622	0.1143	0.0000	0.000 000 00
144	762	109 728	0.0417	0.1667	0.1667	0.2083	0.3311	0.0529	0.0000	0.000 000 00
Rademacher Error										
16	108	1728	0.1250	0.5000	0.5000	0.6250	0.7773	0.1443	0.0000	0.000 000 00
36	283	10 188	0.0833	0.3333	0.3333	0.4167	0.5182	0.0928	0.0000	0.000 000 00
144	1408	202 752	0.0417	0.1667	0.1667	0.2083	0.2591	0.0443	0.0000	0.000 000 00
Adversarial Rademacher Error with Independent Objectives										
16	108	1728	0.1250	0.7500	0.0000	0.8750	0.7773	0.1250	0.0000	0.000 065 13
36	283	10 188	0.0833	0.5000	0.0000	0.5833	0.5182	0.0833	0.0000	0.000 030 03
144	1408	202 752	0.0417	0.2500	0.0000	0.2917	0.2591	0.0417	0.0000	0.000 002 34
Adversarial Rademacher Error with Correlated Objectives										
16	108	1728	0.1250	0.7500	0.0000	0.8750	0.7773	0.1250	0.0000	0.000 033 28
36	283	10 188	0.0833	0.5000	0.0000	0.5833	0.5182	0.0833	0.0000	0.000 014 49
144	1408	202 752	0.0417	0.2500	0.0000	0.2917	0.2591	0.0417	0.0000	0.000 001 46

Note: Given m and $\alpha = 0.5$, the results use n, b from [Theorem 3](#) Part 1 and constants in [Table 2](#). *For the indicated rows, $\hat{\alpha} = 0$ because τ_m exceeds the largest attainable exclusion–inclusion gap.