Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

ON THE HAUSDORFF DISTANCE BETWEEN A PARETO SET AND ITS DISCRETIZATION

Burla E. Ondes
Susan R. Hunter

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

ABSTRACT

Our broad goal is to derive bounds on the performance of bi-objective simulation optimization algorithms that seek the global efficient set on a compact feasible set. Toward this end, we bound the expected Hausdorff distance from the true efficient set to the estimated discretized efficient set by the sum of deterministic and stochastic error terms. We provide an upper bound on the deterministic error term in the context of bi-objective convex quadratic optimization with spherical level sets. Our bound implies that if t is the dispersion of the observed points measured in the decision space, then the Hausdorff distance between the Pareto set and its discretization is $O(\sqrt{t})$ as t decreases to zero.

1 INTRODUCTION

We consider the context of bi-objective simulation optimization problems,

$$\text{Problem } M\text{:}\quad \text{minimize } \{\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x})) := (\mathrm{E}[F_1(\mathbf{x},\xi)], \mathrm{E}[F_2(\mathbf{x},\xi)])\} \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{X}$$

where $\mathbf{f}\colon \mathcal{D} \to \mathbb{R}^2, \mathcal{D} \subseteq \mathbb{R}^q$ is a vector-valued objective function, the feasible set $\mathcal{X} \subseteq \mathcal{D}$ is compact, ξ is a random variable, and each objective function can only be observed with stochastic error, e.g., as the output from a Monte Carlo simulation oracle (Hunter et al. 2019). The solution to Problem M is the efficient set $\mathcal{E}$, that is, the set of all feasible points whose images are non-dominated; its image is the Pareto set $\mathcal{P}$.

Consider the following simple procedure to solve this problem. First, select m distinct points at which to observe the objective function values, $\mathcal{X}_m := \{\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_m\} \subset \mathcal{X}$, where $t > 0$ is the *dispersion* of the discretized feasible set (Yakowitz et al. 2000),

$$t := d(\mathcal{X}, \mathcal{X}_m) = \sup_{\mathbf{x}\in\mathcal{X}} \inf_{\tilde{\mathbf{x}}\in\mathcal{X}_m} \|\mathbf{x} - \tilde{\mathbf{x}}\| = \sup_{\mathbf{x}\in\mathcal{X}} \min_{1\le i\le m} \|\mathbf{x} - \tilde{\mathbf{x}}_i\|.$$

Then, observe n simulation replications at each point and estimate the solution to Problem M using

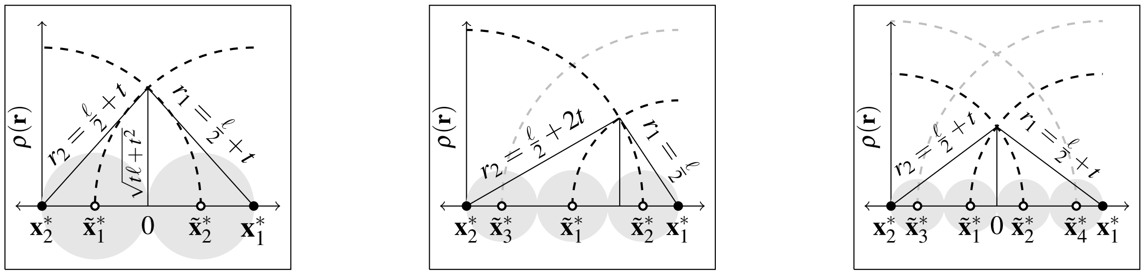
$$\text{Problem } \hat{M}_m(n)\text{:}\quad \text{minimize } \left\{\hat{\boldsymbol{F}}(\tilde{\mathbf{x}}, n) = \left(n^{-1}\textstyle\sum_{i=1}^n F_1(\tilde{\mathbf{x}}, \xi_i), n^{-1}\sum_{i=1}^n F_2(\tilde{\mathbf{x}}, \xi_i)\right)\right\} \quad \text{s.t.} \quad \tilde{\mathbf{x}} \in \mathcal{X}_m \subseteq \mathcal{X},$$

where $\xi_1, \ldots, \xi_n$ are i.i.d. copies of ξ. The solution to Problem $\hat{M}_m(n)$ is $\hat{\mathcal{E}}_m(n)$, which estimates $\mathcal{E}$.

Given a set of points $\mathcal{X}_m$ and simulation budget n, we wish to measure the quality of the estimated solution, $\hat{\mathcal{E}}_m(n)$, as a function of t and n. Let $\mathcal{E}_m$ and $\mathcal{P}_m$ be the respective efficient and Pareto sets for Problem M when solved over the discretized feasible set $\mathcal{X}_m$. Then we consider the following expected Hausdorff distances, where the bounds on the right hand sides follow from the triangle inequality,

$$\mathrm{E}[\mathbb{H}(\mathcal{E}, \hat{\mathcal{E}}_m(n))] \le \mathbb{H}(\mathcal{E}, \mathcal{E}_m) + \mathrm{E}[\mathbb{H}(\mathcal{E}_m, \hat{\mathcal{E}}_m(n))]; \quad \mathrm{E}[\mathbb{H}(\mathcal{P}, \mathbf{f}(\hat{\mathcal{E}}_m(n)))] \le \mathbb{H}(\mathcal{P}, \mathcal{P}_m) + \mathrm{E}[\mathbb{H}(\mathcal{P}_m, \mathbf{f}(\hat{\mathcal{E}}_m(n)))].$$

Toward obtaining upper bounds as a function of t and n, we provide a least upper bound on $\mathbb{H}(\mathcal{E}, \mathcal{E}_m)$ as a function of t, calculated across all possible discretizations, under the following simplifying assumption.



(a) $m_t^* = 2, t = \ell/4$ (b) $m_t^* = 3, t = \ell/6$ (c) $m_t^* = 4, t = \ell/8$

Figure 1: Under Assumption 1, let $\ell = \|\mathbf{x}_1^* - \mathbf{x}_2^*\|$ be the length of $\mathcal{E}$ and let $\mathcal{E}_{m,t} = \mathcal{E}_m \cap B(\mathcal{E}, t)$ where $|\mathcal{E}_{m,t}| = m_t^*$. The gray circles represent t-radius balls around the m_t^* points, and $\rho(\mathbf{r}) \geq d(\mathcal{E}_m, \mathcal{E})$.

Assumption 1 We assume the following about Problem M:

1. The objective functions are convex quadratic $f_1(\mathbf{x}) = b_1 + a_1\|\mathbf{x} - \mathbf{x}_1^*\|^2$, $f_2(\mathbf{x}) = b_2 + a_2\|\mathbf{x} - \mathbf{x}_2^*\|^2$ for constants $a_1, a_2, b_1, b_2 \in \mathbb{R}$ with $a_1, a_2 > 0$, where $\mathbf{x}_1^* \neq \mathbf{x}_2^*$.
2. The feasible set $\mathcal{X} \subseteq \mathbb{R}^q$, $q \geq 2$ is compact, convex, and $\text{conv}\{\mathbf{x}_1^*, \mathbf{x}_2^*\} \subset \text{int}(\mathcal{X})$.
3. The t-expansion of $\mathcal{E}$ is feasible, that is, $B(\mathcal{E}, t) := \cup_{\mathbf{x}^* \in \mathcal{E}} \{\mathbf{x} \in \mathbb{R}^q : \|\mathbf{x} - \mathbf{x}^*\| \leq t\} \subset \mathcal{X}$.

2 THE MAIN RESULT

Recall that $\mathbb{H}(\mathcal{E}, \mathcal{E}_m) = \max\{d(\mathcal{E}, \mathcal{E}_m), d(\mathcal{E}_m, \mathcal{E})\}$, where under more general regularity conditions than ours, the result that $d(\mathcal{E}, \mathcal{E}_m) \leq t$ is provided by Pardalos et al. (2017). We know of no corresponding upper bound on $d(\mathcal{E}_m, \mathcal{E})$; Theorem 1 results from deriving a least upper bound on $d(\mathcal{E}_m, \mathcal{E})$ across all possible configurations of sampled points $\mathcal{X}_m$, which is always larger than t. Figure 1 illustrates the bounding configuration of points that maximizes the possible distance between $\mathcal{E}_m$ and $\mathcal{E}$.

Theorem 1 Let Assumption 1 hold. Let $\ell = \|\mathbf{x}_1^* - \mathbf{x}_2^*\|$ be the length of the efficient set. Then the least upper bound on the Hausdorff distance $\mathbb{H}(\mathcal{E}, \mathcal{E}_m)$, calculated across all possible discretizations, is $\mathbb{H}(\mathcal{E}, \mathcal{E}_m) \leq \sqrt{t\ell + t^2}$. Further, this result implies $\mathbb{H}(\mathcal{P}, \mathcal{P}_m) = O\left(\sqrt{t}\right)$ as $t \to 0$.

While Theorem 1 is interesting from a theoretical perspective, it has limited practical implications since we do not usually know the length of the efficient set ℓ. The following corollary presents a translation of these results into a more practical upper bound based on $m^* := |\mathcal{P}_m|$. This bound arises because $\ell \leq 2tm^*$.

Corollary 2 Let Assumption 1 hold. Then $\mathbb{H}(\mathcal{E}, \mathcal{E}_m) \leq t\sqrt{2m^* + 1}$.

3 CONCLUDING REMARKS

We provide an upper bound on the deterministic error under Assumption 1. Ongoing work includes relaxing Assumption 1, obtaining the required bound on the stochastic error term $\mathrm{E}[\mathbb{H}(\mathcal{E}_m, \hat{\mathcal{E}}_m(n))]$ as a function of n, and extending the results to a higher number of objectives. Such bounds can provide useful insight in algorithmic design and performance for bi-objective simulation optimization on a compact set.

REFERENCES

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

Pardalos, P. M., A. Žilinskas, and J. Žilinskas. 2017. *Non-Convex Multi-Objective Optimization*. Springer Optimization and Its Applications. Switzerland: Springer.

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