ROBBO: An Efficient Method for Pareto Front Estimation With Guaranteed Accuracy

Roberto Boffadossi, Marco Leonesio, *Member, IEEE*, and Lorenzo Fagiano, *Senior Member, IEEE*

***Abstract*—A new method to estimate the Pareto front (PF) in bi-objective optimization problems (BOPs) is presented. Assuming a continuous PF, the approach, named ROBBO (RObust and Balanced BOP), needs to sample at most a finite, precomputed number of PF points. Upon termination, it guarantees that the worst case approximation error lies within a desired tolerance range, predefined by the decision maker (DM), for each of the two objective functions. Theoretical results are derived about the worst case number of PF samples required to guarantee the desired accuracy, both in general and for specific sampling methods from the literature. A comparative analysis, both theoretical and numerical, demonstrates the superiority of the proposed method with respect to popular ones. The approach is finally showcased in a constrained path-following problem for a 2-axis positioning system and in a steady-state optimization problem for a continuous-flow stirred tank reactor (CSTR). An open demo implementation of ROBBO is made available online.**

***Index Terms*—Bi-objective optimization (BOP), multiobjective optimization, optimal design, Pareto front (PF), set membership estimation, uncertainty quantification.**

I. Introduction

BI-OBJECTIVE optimization problems (BOPs) are ubiquitous in science, engineering, and economics [1], [2], [3]. In control, BOPs are extremely common because control tuning yields a wide range of possible tradeoffs. BOPs account for over 50% of the tuning problems, while about 30% pertain to three objectives, and 20% to more than three, see [4]. A typical approach to deal with a BOP is to approximate the set of all the nondominated feasible points in the objectives (or criteria) space, named Pareto front (PF), and let the decision maker (DM) manually select the desired one, based on various application-specific considerations. Such an a posteriori articulation of preferences [5] allows the DM to make informed choices, having a complete understanding of how the objectives conflict. It is also the most documented case for control problems, as shown in [4]. There are three categories of PF approximation strategies: pointwise, piecewise, and others.

The first class corresponds to computing a finite "solution set" that acts as a surrogate of the whole PF. The most common methods provide a point-cloud approximation by sampling the PF, each sample being the solution of an auxiliary single-objective optimization problem with scalarization techniques to find a specific tradeoff (see, for instance, [6], [7], [8], [9], [10]). Evolutionary strategies or meta-heuristics have been proposed as well ([11]; [12]), proving to be effective even for irregular PFs. Among these methods, it is common to use archiving techniques to preserve only high-quality solutions to improve performance indicators. Although they are important tools, no ideal archiver exists that simultaneously combines practical applicability and strong theoretical guarantees, see the survey [13]. The complete set of Pareto optimal solutions can involve an infinite number of potential tradeoffs, thus making it impossible to be fully described by a finite set of points. Moreover, reaching a good approximation is impractical when the sampling process is expensive, for example, requiring time-consuming simulations, global optimization, or experiments.

The second class, piecewise approximation methods, is a direct extension of the previous one, consisting of different interpolation/fitting strategies of the elements of the solution set: piecewise linear ([14]; [15]), outer, inner, and "sandwich" approximations ([16]; [17]), quadratic or cubic [18]. To refine the approximation, these methods exploit iterative techniques to sample the PF considering a chosen error metric.

The last category includes meta-models or higher order functions to provide a surrogate PF. Examples are enclosure or box coverage, like [19], [20], [21], and [22], which approximate the PF as the union of rectangles in between the sampled PF points and evaluate the approximation quality based on their size, or also continuous Pareto manifold approximation [23] for multiobjective reinforcement learning, or combination of local manifolds [24], which was shown to be particularly effective when dealing with complex, irregular PF. Moreover, in recent years, machine learning methods have been adopted to train models providing a complete representation of the PF in multiobjective optimization. The main approaches in this category are summarized in Table I.

Finding an accurate estimate of the PF with as little computational or experimental effort as possible, i.e., the smallest number of samples, is important for an effective

Received 16 November 2025; revised 12 February 2026, 6 May 2026, and 8 May 2026; accepted 8 May 2026. This work was supported in part by the European Union (EU) Project E2COMATION (H2020) under Grant 958410; in part by European Union–Next Generation EU in the Context of the Project Piano Nazionale di Ripresa e Resilienza (PNRR) M4C2, Investimento 1.3 DD. 341 del 15 marzo 2022–NEST—Network 4 Energy Sustainable Transition—Spoke 2-PE00000021 under Grant D43C22003090001; in part by the Fondazione Cariplo through the Project "NextWind-Advanced Control Solutions for Large Scale Airborne Wind Energy Systems" under Grant 2022-2005; and in part by the Italian Ministry of University and Research through the European Union–NextGenerationEU Fund, through the Project "DeepAirborne–Advanced Modeling, Control and Design Optimization Methods for Deep Offshore Airborne Wind Energy," under Grant P2022927H7. This article was recommended by Associate Editor M. Xia. *(Corresponding authors: Roberto Boffadossi; Lorenzo Fagiano.)*

Roberto Boffadossi and Marco Leonesio are with the Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, Italian National Research Council, 20133 Milan, Italy, and also with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milan, Italy (e-mail: roberto.boffadossi@stiima.cnr.it; marco.leonesio@stiima.cnr.it).

Lorenzo Fagiano is with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milan, Italy (e-mail: lorenzo.fagiano@polimi.it).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSMC.2026.3693888.

Digital Object Identifier 10.1109/TSMC.2026.3693888

TABLE I
OVERVIEW OF MACHINE LEARNING METHODS USED IN PF APPROXIMATION

Model	Methods	Sampling strategy	Quality measure
Gaussian process	Active learning of PF as a continuous function via GP regression in the criterion space [34]; PF estimation via Kriging model based on projections toward the origin onto the unit hyperplane [35]; PF approximated via Vorob'ev expectation using conditional simulations [36] of GP-based objective estimates.	Uncertainty sampling principle with ε-constraint scalarization [34]; comparison of simplex-lattice design, the incremental lattice design, Hammersley method. [35]; expected hypervolume improvement for objectives learning and uniform sampling and Sobol sequence for PF estimate [36].	GP prediction error [34], [35], inverted generational distance (for comparison) [34]; Vorob'ev deviation (uncertainty measure), hypervolume difference, epsilon and R2 (quality comparisons) [36].
Neural Network	Pareto Front Learning (PFL): a single hypernetwork is trained to map every preference vector to the corresponding PF point (Multi-objective optimization) or to the model weights (Multi-task learning) [37]–[40]; general PFL framework supporting completed scalarization functions [41]; transformer-based hypernetwork with split feasibility constraints to handle disconnected PFs [42]; PF learning treated as a distribution-transformation problem [43].	Uniform distribution of preference vectors [37]; preferences sampled randomly and iteratively from Dirichlet distribution [38], [41], [42]; learned preference distribution via a generative model aimed at producing evenly spaced PF points [40]; batch random sampling from subregions (Delaunay triangulation) or from Dirichlet distribution; [39]; Gaussian sampling and Latin hypercube sampling [43]	Hypervolume indicator [37]–[42], and Uniformity indicator [38]; Mean Euclid Distance, Hypervolume Difference [42]; Normalized Hypervolume gap [40]

and efficient decision-making process. At the same time, providing a guarantee on the estimation error is crucial to make an informed decision. Ideally, the DM would like to have a method that: 1) builds an estimated PF with proven guaranteed error with respect to the real one; 2) with the possibility to specify the error tolerances for each criterion; 3) knowing early on the worst case number of PF samples needed to guarantee the desired accuracy; and 4) paired with a "realization mechanism" that, once a point on the estimated PF has been selected, returns a point on the real PF that satisfies the tolerance requirements. Notwithstanding the large number of PF estimation approaches in the literature, to the best of our knowledge, none of them addresses all these requirements. We refer to such a problem as a robust and balanced PF estimation-realization: robust, because it provides accuracy guarantees, and balanced, because it constrains the ratio of the estimation errors of the two objectives to a desired value.

The main contribution of this article is to provide such a method. First, we introduce a linear, invertible transformation of the PF that casts the problem as that of approximating a conveniently chosen univariate scalar function. Such a transformation accounts for the error tolerances defined by the DM. Under a continuity assumption, we derive the set of all possible PFs (the feasible Pareto function set) compatible with the sampled points, along with its tight upper and lower bounds. A first theoretical result establishes the necessary and sufficient conditions on the PF samples to guarantee the desired accuracy, for any possible approximation belonging to the feasible set. Then, our new method, RObust and Balanced BOP (ROBBO), is introduced. It is an iterative algorithm that uses the point-wise distance between the bounds to select the most informative next sample, among a finite set of suitably chosen candidates, thus obtaining an efficient algorithm where the worst case uncertainty is quickly reduced to below the desired threshold. After presenting the method and deriving its theoretical properties, we compare it theoretically and numerically with popular alternative methods, demonstrating its superiority in terms of sample complexity. ROBBO optimally preallocates the sampling coordinates that guarantee meeting the required tolerances in the worst case, based on our theoretical results. Then, it explores these coordinates using an incremental learning strategy to adapt to each problem, potentially requiring fewer samples than the worst case theoretical bound. Furthermore, we showcase the performance of ROBBO on a constrained path-following problem for a 2-axis positioning system and on the computation of the economically optimal steady state of a continuous-flow stirred tank reactor (CSTR).

The article is organized as follows. Section II includes preliminary notions and problem formulation. Sections III–V present the theoretical results, the proposed approach, extensions, and comparisons. Section VI contains the simulation results, and Section VII concludes the article.

Notation: A bold-faced symbol like $\boldsymbol{z}$ denotes a vector, z_i its ith entry, and $\boldsymbol{z}^T$ its transpose. $\mathbb{R}^2_{\geq 0}$ and $\mathbb{R}^2_{\leq 0}$ are the set of vectors in $\mathbb{R}^2$ with non-negative and with nonpositive entries. The *Pareto Cone* is $\mathbb{R}^2_{\succeq 0} \doteq \mathbb{R}^2_{\geq 0} \setminus \{[0,0]^T\}$. Similarly, we define $\mathbb{R}^2_{\preceq 0} \doteq \mathbb{R}^2_{\leq 0} \setminus \{[0,0]^T\}$. The Minkowski sum of two sets A and B is $A \oplus B \doteq \{a + b \mid a \in A \wedge b \in B\}$. Throughout the article, for a continuous and uncertain function h, functions $\underline{h}$ and $\bar{h}$ denote its (tightest) lower and upper bounds, respectively.

II. PROBLEM STATEMENT

In this section, we introduce the BOP problem and the robust and balanced PF approximation problem, which is structured into the three subproblems P1–3.

We consider the BOP

$$\min_{\boldsymbol{x} \in S} \boldsymbol{f}(\boldsymbol{x}) \tag{1}$$

where the function $\boldsymbol{f}(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}^2$ returns the two objectives (or criteria) $z_{1,2} = f_{1,2}(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$, and the design variables are restricted to the feasible design set $S \subset \mathbb{R}^n$. We denote the feasible criterion set with $Z \doteq \{\boldsymbol{f}(\boldsymbol{x}), \boldsymbol{x} \in S\} \subset \mathbb{R}^2$. We further denote a specific sample in S, with an associated index j, as $\boldsymbol{x}^{(j)}$, and the corresponding objective vector as

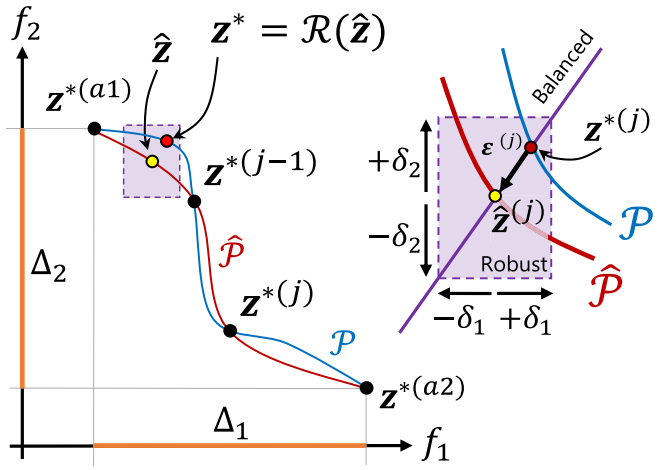

Fig. 1. Robust and balanced PF estimation-realization: PF (blue line) and its approximation (red line), candidate selection (yellow point) and its realization (red point) within the tolerance region (dashed violet box).

$\boldsymbol{z}^{(j)} = \boldsymbol{f}(\boldsymbol{x}^{(j)}) = [z_1^{(j)}, z_2^{(j)}]^T$. We recall a few fundamental definitions [35].

Definition 1: A point $\boldsymbol{z}^{(1)} \in Z$ dominates a point $\boldsymbol{z}^{(2)} \in Z$, with notation $\boldsymbol{z}^{(1)} \prec \boldsymbol{z}^{(2)}$, iff $\boldsymbol{z}^{(2)} \in \{\boldsymbol{z}^{(1)}\} \oplus \mathbb{R}^2_{>0}$. That is, $\boldsymbol{z}^{(1)}$ presents better performance in one objective and at least equal in the other.

Definition 2: The set of Pareto optimal decisions is $X^* \doteq \{\boldsymbol{x}^* \in S \mid \nexists \boldsymbol{x} \in S, \boldsymbol{x} \neq \boldsymbol{x}^* : \boldsymbol{f}(\boldsymbol{x}) \prec \boldsymbol{f}(\boldsymbol{x}^*)\}$.

Definition 3: The PF is the set of all nondominated points in Z, $\mathcal{P} \doteq \{\boldsymbol{z}^* = \boldsymbol{f}(\boldsymbol{x}^*) \ : \ \boldsymbol{x}^* \in X^*\}$.

We refer to $\boldsymbol{x}^*$ and $\boldsymbol{z}^*$ as a Pareto optimal point in the decision space and in the criterion space, respectively.

We next introduce two specific points belonging to the PF, named anchor points, $\boldsymbol{z}^{*(a1)}$ and $\boldsymbol{z}^{*(a2)}$, obtained by solving the following auxiliary problems:

$$\boldsymbol{z}^{*(a1)} = \boldsymbol{f}\left(\boldsymbol{x}^{*(a1)}\right) : \boldsymbol{x}^{*(a1)} = \arg\min_{\boldsymbol{x} \in S} f_1(\boldsymbol{x}) \tag{2a}$$

$$\boldsymbol{z}^{*(a2)} = \boldsymbol{f}\left(\boldsymbol{x}^{*(a2)}\right) : \boldsymbol{x}^{*(a2)} = \arg\min_{\boldsymbol{x} \in S} f_2(\boldsymbol{x}). \tag{2b}$$

Assumption 1: $\mathcal{P}$ is a continuous and compact curve in $\mathbb{R}^2$.

Satisfaction of Assumption 1 depends on the application at hand and on the modeling choices made when defining $\boldsymbol{f}$ and S. Compactness is a very mild condition, since it holds whenever the anchor points (2) are finite, which is always the case in sensible engineering applications. Note that compactness of the PF does not require compactness of the feasible design set S, nor of the feasible criterion set Z. Also, continuity is a mild assumption in real-world engineering problems, where discontinuous PFs are not frequently encountered [36]. Furthermore, we discuss possible strategies to handle the presence of discontinuities in our approach at the end of Section IV-B. The entries of $\boldsymbol{z}^{*(a1)}$, $\boldsymbol{z}^{*(a2)}$ form the upper and lower extremes of the closed and bounded intervals obtained by projecting $\mathcal{P}$ on the Cartesian axes in $\mathbb{R}^2$. We indicate the lengths of such intervals with $\Delta_1 \doteq z_1^{*(a2)} - z_1^{*(a1)}$ and $\Delta_2 \doteq z_2^{*(a1)} - z_2^{*(a2)}$, see Fig. 1.

We now consider an estimate $\hat{\mathcal{P}} \approx \mathcal{P}$ (PF approximation). We denote a point in $\hat{\mathcal{P}}$ with $\hat{\boldsymbol{z}}$, and introduce the *realization* concept, which is a mapping $\mathcal{R} : \hat{\mathcal{P}} \to \mathcal{P}$. Thus, for any $\hat{\boldsymbol{z}}^{(j)} \in \hat{\mathcal{P}}$ we have that $\mathcal{R}(\hat{\boldsymbol{z}}^{(j)}) \in \mathcal{P}$. From a decision-making perspective we will refer to the point $\hat{\boldsymbol{z}}^{(j)}$ selected by the DM on the approximation $\hat{\mathcal{P}}$ as the *candidate*, and to $\boldsymbol{z}^{*(j)} = \mathcal{R}(\hat{\boldsymbol{z}}^{(j)})$ as the realization of the candidate, i.e. the related point on the actual PF. We denote the realization error as the vector $\boldsymbol{\varepsilon} = \hat{\boldsymbol{z}} - \boldsymbol{z}^*$, $\boldsymbol{\varepsilon} \in \mathbb{R}^2$.

We consider a scenario where the DM wants to derive $\hat{\mathcal{P}}$ and $\mathcal{R}$ such that the following property of the realization error holds:

$$\forall \hat{\boldsymbol{z}} \in \hat{\mathcal{P}}, \quad \boldsymbol{z}^* = \mathcal{R}(\hat{\boldsymbol{z}}), \ \ |\varepsilon_1| < \delta_1 \wedge |\varepsilon_2| < \delta_2 \tag{3}$$

where δ_1, δ_2 are defined by the DM according to the desired accuracy/tolerance for each criterion. As anticipated in the Introduction, we refer to a pair $(\hat{\mathcal{P}}, \mathcal{R})$ satisfying (3) as a *robust* PF estimate-realization.

We further consider that the approximated PF is obtained on the basis of a finite number M of PF samples, which include to a minimum the anchor points (thus, $M \geq 2$), forming the data-set $\mathcal{D} = \{\boldsymbol{z}^{*(a1)}, \boldsymbol{z}^{*(1)}, \ldots, \boldsymbol{z}^{*(M-2)}, \boldsymbol{z}^{*(a2)}\}$. Without loss of generality, we consider the elements $\boldsymbol{z}^{*(j)} \in \mathcal{D}$ to be ranked according to the f_1-coordinate so that $z_1^{*(j)} < z_1^{*(j+1)}$. A graphical illustration of the considered setup is presented in Fig. 1. Since obtaining each sample in $\mathcal{D}$ may require the solution of complex and time-consuming simulations, global optimization and/or experiments, it is of interest to the DM to have an estimate-realization that is not only robust, but also requires a "low" value of M. From the theoretical standpoint, our goal is thus to investigate what are the conditions on $\mathcal{D}$ to satisfy (3), whether there is a minimum number M of suitably chosen samples such that requirement (3) is surely satisfied, and whether there are estimation-realization algorithms that are particularly *efficient* in terms of worst case value of M, i.e. they require a low number of samples to provide the accuracy guarantees.

As we show in Section III, we can cast the PF estimation problem into a function approximation one. Then, given a data-set $\mathcal{D}$, one can derive the set of functions that are compatible with the observed PF samples. We name this set the feasible Pareto function set ($FPFS_{\mathcal{D}}$). Given this framework, we address the goal of efficiently deriving a robust reconstruction of the PF via three fundamental problems, P1–P3, formulated as follows.

1) *P1:* Derive necessary and sufficient conditions on $\mathcal{D}$ such that, for any PF $\mathcal{P}$ and any estimate $\hat{\mathcal{P}}$ belonging to the corresponding $FPFS_{\mathcal{D}}$, there exists a realization $\mathcal{R}$ such that (3) is satisfied.
2) *P2:* Derive the smallest number $\underline{M}$ of PF samples required to guarantee the conditions of P1.
3) *P3:* Obtain an algorithm that builds a specific estimate-realization pair $(\hat{\mathcal{P}}, \mathcal{R})$ satisfying (3) with a finite number of samples $M < \underline{M}$, which is the smallest possible in the worst case.

We remark that problems P1–P2 consider any possible (admissible, i.e., belonging to the $FPFS_{\mathcal{D}}$) estimate $\hat{\mathcal{P}}$ and, since the PF is not known except for its samples, any possible (admissible) PF. The value $\underline{M}$ shall thus be valid for all possible combinations of $\hat{\mathcal{P}}, \mathcal{P} \in FPFS_{\mathcal{D}}$.

On the other hand, in problem P3 we aim for a specific estimate $\hat{\mathcal{P}}$, thus the "worst case" scenario mentioned in P3

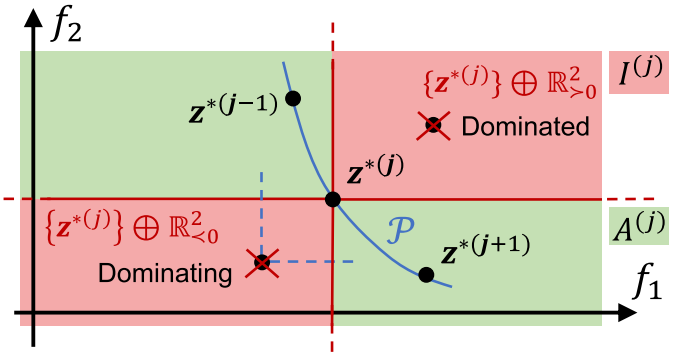

Fig. 2. Subsets $I^{(j)}$ (red) and $A^{(j)}$ (green) pertaining to a point $z^{*(j)}$.

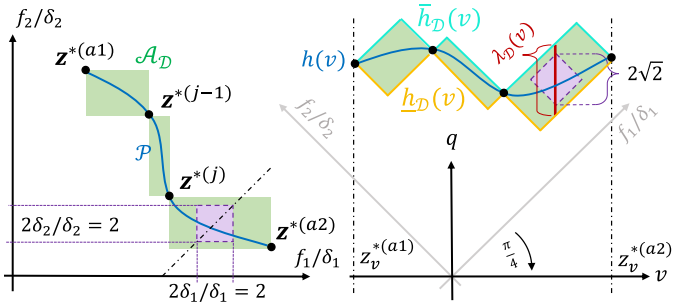

Fig. 3. Left figure: PF as a curve $\mathcal{P}$ in the normalized coordinates $(f_1/\delta_1, f_2/\delta_2)$. Right figure: PF as a function $h(v)$ in (v, q) coordinates. The dashed violet boxes represent the tolerance region.

refers to the uncertainty about $\mathcal{P}$ only. This is why the value of M in P3 is generally smaller than $\underline{M}$, and we look for the estimation approach that minimizes it. We provide the solutions to problems P1–P3 in Sections III and IV.

III. Robust PF Estimation—Theoretical Results

This section establishes the theoretical framework to address and solve problems P1 and P2, formulated in Section II.

A. Feasible Pareto Function Set and Its Optimal Bounds

According to the Pareto dominance relation, each PF point $z^{*(j)} \in \mathcal{D}$ defines the following two subsets:

$$I^{(j)} \doteq \left\{z^{*(j)}\right\} \oplus \mathbb{R}^2_{\succ 0} \cup \left\{z^{*(j)}\right\} \oplus \mathbb{R}^2_{\prec 0}, \quad A^{(j)} \doteq \mathbb{R}^2 \setminus I^{(j)}. \tag{4}$$

The set $I^{(j)}$ is the subset of the criterion space in which no points of the PF can be present, because they would either be dominated or dominate the point $z^{*(j)}$, contradicting the fact that the latter belongs to the PF, see Fig. 2. Thus, $I^{(j)} \cap \mathcal{P} = \emptyset$, $\forall j$.

Regarding the anchor points, the corresponding sets $I^{(a1)}$, $I^{(a2)}$ are larger, i.e.,

$$\begin{aligned} I^{(a1)} &\doteq \left\{z^{*(a1)}\right\} \oplus \mathbb{R}^2_{\succ 0} \cup \left\{z^{*(a1)}\right\} \oplus \mathbb{R}^2_{\prec 0} \\ &\cup \left\{z : z_1 < z_1^{*(a1)}\right\} \\ I^{(a2)} &\doteq \left\{z^{*(a2)}\right\} \oplus \mathbb{R}^2_{\succ 0} \cup \left\{z^{*(a2)}\right\} \oplus \mathbb{R}^2_{\prec 0} \\ &\cup \left\{z : z_2 < z_2^{*(a2)}\right\} \end{aligned} \tag{5}$$

while $A^{(a1)}, A^{(a2)}$ are still their complements, as in (4).

Considering now the set $\mathcal{D}$ altogether, we can partition the objective domain in the set $\mathcal{I}_\mathcal{D} \subset \mathbb{R}^2$ and its complement, $\mathcal{A}_\mathcal{D}$

$$\mathcal{I}_\mathcal{D} \doteq \bigcup_{z^{*(j)} \in \mathcal{D}} I^{(j)}, \quad \mathcal{A}_\mathcal{D} \doteq \mathbb{R}^2 \setminus \mathcal{I}_\mathcal{D}. \tag{6}$$

As presented in Fig. 3 (top), $\mathcal{A}_\mathcal{D}$ is the subset of the criterion space that is guaranteed to contain the actual PF, according to the information provided by the data $\mathcal{D}$. It also corresponds to the intersection $\mathcal{A}_\mathcal{D} = \bigcap_{z^{*(j)} \in \mathcal{D}} A^{(j)}$. By construction, we have that $\mathcal{P} \in \mathcal{A}_\mathcal{D}$ and $\mathcal{I}_\mathcal{D} \cap \mathcal{P} = \emptyset$.

The set $\mathcal{A}_\mathcal{D}$ is instrumental in defining what we call the feasible Pareto function set, $FPFS_\mathcal{D}$. We first convert the original coordinate system (f_1, f_2) into a new one, denoted by (v, q), by the following transformation matrix W:

$$\begin{bmatrix} v \\ q \end{bmatrix} = \underbrace{\frac{\sqrt{2}}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\delta_1} & 0 \\ 0 & \frac{1}{\delta_2} \end{bmatrix}}_{W} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}. \tag{7}$$

Namely, the transformation consists of normalizing the two objective functions by the corresponding desired error tolerances, and of rotating the scaled axes so that f_1/δ_1 overlaps with the bisector of the first quadrant of the new coordinates, see Fig. 3. This operation generalizes the mathematical abstraction of the problem, making it independent of the error bounds δ_1, δ_2. Moreover, the transformation matrix W allows us to reformulate the problem of estimating a continuous PF as that of estimating a black-box Lipschitz continuous function $q = h(v)$. Moreover, this specific transformation is such that, for a fixed value of v, the scalar error interval about the function h corresponds to error intervals on ε_1 and ε_2 that maintain the ratio δ_1/δ_2, as will be shown in the following. For any PF point $[z_1^*, z_2^*]^T$, we denote the corresponding point in the transformed coordinates with

$$\begin{bmatrix} z_v^* \\ z_q^* \end{bmatrix} = W \begin{bmatrix} z_1^* \\ z_2^* \end{bmatrix}. \tag{8}$$

The transformed anchor points are thus $[z_v^{*(a1)}, z_q^{*(a1)}]^T = W z^{*(a1)}$ and $[z_v^{*(a2)}, z_q^{*(a2)}]^T = W z^{*(a2)}$. Since the transformation corresponds to a scaling with positive coefficients and a clockwise rotation by $(\pi/4)$ of the objectives' axes (f_1, f_2), and recalling that the points in $\mathcal{D}$ are ordered by increasing f_1-values, note that the points in the rotated dataset are still ordered by increasing values of their first coordinate, i.e. $z_v^{*(a1)} < z_v^{*(1)} < \cdots < z_v^{*(M-2)} < z_v^{*(a2)}$.

Moreover, note that, by the Pareto dominance relation, in general for any two points $z^{*(j)}, z^{*(\ell)} \in \mathcal{P}$ we have $z_1^{*(j)} < z_1^{*(\ell)} \Leftrightarrow z_2^{*(j)} > z_2^{*(\ell)}$. Thus, $\mathcal{P}$ can be seen as a strictly monotonically decreasing function relating z_2^* with z_1^*. Since in (7) we apply a linear and invertible transformation, we can still interpret the transformed front as a function, denoted by h, such that $z_q^* = h(z_v^*)$. Under Assumption 1, h is continuous (i.e., $h \in C^0$). Its domain is given by the interval delimited by the transformed anchor points projected on the v-axis, i.e., $[z_v^{*(a1)}, z_v^{*(a2)}]$. Thus, the actual PF $\mathcal{P}$ can be also written as

$$\mathcal{P} = \left\{W^{-1}[v, h(v)]^T \quad \forall v \in [z_v^{*(a1)}, z_v^{*(a2)}]\right\} \tag{9}$$

which for simplicity we denote in short as $\mathcal{P} = W^{-1}h$. We are now in position to define the $FPFS_{\mathcal{D}}$, i.e., the set of all continuous functions h [i.e., all PF $\mathcal{P}$ according to (9)] compatible with the samples in $\mathcal{D}$

$$FPFS_{\mathcal{D}} \doteq \left\{h \in C^0, h : \left[z_v^{*(a1)}, z_v^{*(a2)}\right] \to \mathbb{R} \mid W^{-1}h \in \mathcal{A}_{\mathcal{D}}\right\}. \quad (10)$$

In this framework, estimating the PF is thus equivalent to deriving a function $\hat{h} \approx h$ in the rotated coordinates. For a given function $\hat{h}$, one can in fact obtain the corresponding estimated PF as $\hat{\mathcal{P}} = W^{-1}\hat{h}$.

With a slight abuse of notation, we indicate with $\mathcal{P}, \hat{\mathcal{P}} \in FPFS_{\mathcal{D}}$ a PF (or estimated PF) whose corresponding function $h, \hat{h} \in FPFS_{\mathcal{D}}$. Considering the function set $FPFS_{\mathcal{D}}$ allows us to compute tight bounds on the estimation error. We do this in a Set Membership function approximation framework [37], by first deriving the so-called optimal upper and lower bounds to the $FPFS_{\mathcal{D}}$, $\bar{h}_{\mathcal{D}}$ and $\underline{h}_{\mathcal{D}}$, defined as

$$\begin{aligned} \bar{h}_{\mathcal{D}} &: \left[z_v^{*(a1)}, z_v^{*(a2)}\right] \to \mathbb{R}, \bar{h}_{\mathcal{D}}(v) \doteq \sup_{h \in FPFS_{\mathcal{D}}} h(v) \\ \underline{h}_{\mathcal{D}} &: \left[z_v^{*(a1)}, z_v^{*(a2)}\right] \to \mathbb{R}, \underline{h}_{\mathcal{D}}(v) \doteq \inf_{h \in FPFS_{\mathcal{D}}} h(v). \end{aligned} \quad (11)$$

These functions are the tightest upper and lower bounds to the values that any function in $FPFS_{\mathcal{D}}$ can take. For a generic $v' \in [z_v^{*(a1)}, z_v^{*(a2)}]$, denote

$$\begin{aligned} \boldsymbol{z}^{*(j^-)}(v') &= \arg\min_{z^{*(j)} \in \mathcal{D}} \left|v' - z_v^{*(j)}\right| \text{ s.t. } z_v^{*(j)} \le v' \\ \boldsymbol{z}^{*(j^+)}(v') &= \arg\min_{z^{*(j)} \in \mathcal{D}} \left|v' - z_v^{*(j)}\right| \text{ s.t. } z_v^{*(j)} \ge v' \end{aligned} \quad (12)$$

i.e., the samples whose v-coordinates represent the extremes of the smallest segment, among those defined by each pair of samples, containing v'. These two samples always exist, since the anchor points themselves are assumed to be part of $\mathcal{D}$, and they coincide only if $v' = z_v^{*(j)}$ for some j. Otherwise, considering the ordering of the data-set by increasing v-values, we have that $\boldsymbol{z}^{*(j^+)}(v')$ is the point right after $\boldsymbol{z}^{*(j^-)}(v')$ in the dataset.

Theorem 1: Given the BOP (1) and a data-set $\mathcal{D}$, under Assumption 1 it holds

$$\begin{aligned} \bar{h}_{\mathcal{D}}(v) &\doteq \min_{z^{*(j^-)}(v),\, z^{*(j^+)}(v)} z_q{}^{*(j)} + \left|v - z_v{}^{*(j)}\right| \\ \underline{h}_{\mathcal{D}}(v) &\doteq \max_{z^{*(j^-)}(v),\, z^{*(j^+)}(v)} z_q{}^{*(j)} - \left|v - z_v{}^{*(j)}\right|. \end{aligned} \quad (13)$$

Proof: See the Appendix. ∎

Note that solving the optimization problems in (13) amounts to evaluating the upper and lower bounds contributed by just two points, $\boldsymbol{z}^{(j^-)}(v)$ and $\boldsymbol{z}^{(j^+)}(v)$, and taking the tightest ones.

Remark 1: From Theorem 1, it is evident that the upper and lower bounds (13) are Lipschitz continuous functions with Lipschitz constant equal to one. Therefore, by a local analysis in the neighborhood of a generic PF point $\boldsymbol{z}^*$, one can derive that, in the considered problem, $h(v)$ is a Lipschitz continuous function with Lipschitz constant smaller than one (see Fig. 3, bottom graph, for a visualization). This fact could be used to further refine the $FPFS_{\mathcal{D}}$ by restricting it to such a function class. However, this restriction would not change any of the findings and is thus omitted for simplicity.

B. Conditions for Robustness Guarantees

Equipped with the $FPFS_{\mathcal{D}}$ and its optimal bounds, let us now introduce the local, $\lambda_{\mathcal{D}}(v)$, and global, $\bar{\lambda}_{\mathcal{D}}$, worst case estimation errors on the basis of the available data $\mathcal{D}$ [37]

$$\lambda_{\mathcal{D}}(v) \doteq \sup_{\hat{h}, \tilde{h} \in FPFS_{\mathcal{D}}} \left|\hat{h}(v) - \tilde{h}(v)\right| = \bar{h}_{\mathcal{D}}(v) - \underline{h}_{\mathcal{D}}(v) \quad (14a)$$

$$\bar{\lambda}_{\mathcal{D}} \doteq \max_{v \in \left[z_v^{*(a1)}, z_v^{*(a2)}\right]} \lambda_{\mathcal{D}}(v). \quad (14b)$$

The local worst case error provides an upper bound to the difference, for a specific v, between any two functions in $FPFS_{\mathcal{D}}$, thus also including the rotated PF $h(v)$, since it belongs to this function set. Equation (14a) can be proven using the definition of the optimal bounds (see e.g., [37]). The global worst case error is the maximum local error overall v values in the functions' domain. In the BOP problem at hand, since we are dealing with a scalar function of a scalar variable, these quantities can always be computed analytically, using Theorem 1 and the compactness of the segment $[z_v^{*(a1)}, z_v^{*(a2)}]$. Indeed, computing $\lambda_{\mathcal{D}}(v)$ amounts to applying (13), while we provide next a result to derive $\bar{\lambda}_{\mathcal{D}}$. Let us denote with $R^{(\ell)}, \ell = 1, \dots, M-1$, the open intervals on the v-axis defined by each pair of consecutive points in $\mathcal{D}$, i.e., $R^{(1)} = (z_v^{*(a1)}, z_v^{*(1)})$, $R^{(2)} = (z_v^{*(1)}, z_v^{*(2)})$, $\dots, R^{(M-1)} = (z_v^{*(M-2)}, z_v^{*(a2)})$. For a given $\ell = 1, \dots, M-1$, let us denote generically with $\boldsymbol{z}^{*(\ell-)}$, $\boldsymbol{z}^{*(\ell+)}$ the samples whose v-coordinates define the extremes of $R^{(\ell)}$. Finally, let us denote

$$V^{(\ell)} = z_v^{*(\ell+)} - z_v^{*(\ell-)} \quad (15a)$$

$$Q^{(\ell)} = z_q^{*(\ell+)} - z_q^{*(\ell-)}. \quad (15b)$$

Proposition 1: Given the BOP problem (1) and a dataset $\mathcal{D}$, under Assumption 1 the global worst case error bound is

$$\bar{\lambda}_{\mathcal{D}} = \max_{\ell = 1, \dots, M-1} V^{(\ell)} - \left|Q^{(\ell)}\right|. \quad (16)$$

Proof: See the Appendix. ∎

The global bound $\bar{\lambda}_{\mathcal{D}}$ plays a key role in the solution to problem P1, which is provided by the next result. In the remainder, for a generic estimated Pareto point $\hat{\boldsymbol{z}}$ we will make use of the realization

$$\boldsymbol{z}^* = \mathcal{R}_{SM}(\hat{\boldsymbol{z}}) \doteq W^{-1}\begin{bmatrix} \hat{z}_v \\ h(\hat{z}_v) \end{bmatrix} \quad (17)$$

i.e., such that the v-coordinates of the estimated and of the actual Pareto point are the same. The value $h(\hat{z}_v)$ in (17) can be computed by solving a suitable auxiliary optimization problem, which we introduce in detail in Section IV-B.

Theorem 2: Given the BOP problem (1), a dataset $\mathcal{D}$, and a tolerance vector $\boldsymbol{\delta}$, under Assumption 1 it holds:

$$\begin{aligned} &\exists \mathcal{R} : \text{ Property (3) is satisfied} \quad \forall \hat{\mathcal{P}},\ \mathcal{P} \in FPFS_{\mathcal{D}} \\ &\iff \\ &\bar{\lambda}_{\mathcal{D}} \le \sqrt{2}. \end{aligned} \quad (18)$$

Proof: See the Appendix. ∎

We now address problem P2, i.e., to derive the minimum number of samples such that the condition laid out by

Theorem 2 is met, i.e., $\overline{\lambda}_{\mathcal{D}} \leq \sqrt{2}$. Recall that $\Delta_1 = z_1^{*(a2)} - z_1^{*(a1)}$ and $\Delta_2 = z_2^{*(a1)} - z_2^{*(a2)}$.

Proposition 2: Given the BOP problem (1) and a tolerance vector $\boldsymbol{\delta}$, under Assumption 1, the minimum number of samples required to satisfy condition (2) is

$$\underline{M} = \left\lceil \frac{1}{2}\left(\frac{\Delta_1}{\delta_1} + \frac{\Delta_2}{\delta_2}\right)\right\rceil + 1 \tag{19}$$

and the corresponding dataset $\mathcal{D}$ shall be evenly allocated in the segment $[z_v^{*(a1)}, z_v^{*(a2)}]$.

Proof: See the Appendix. ∎

Remark 2: Using Proposition 2, the DM can evaluate, after computing only the two anchor points, the minimum number of samples that surely (i.e., no matter the shape of the actual PF) attains the robustness conditions of Theorem 2. Note that the anchor points themselves are included in $\underline{M}$, and that the result also prescribes how these points should be sampled, i.e., such that they uniformly partition the interval $[z_v^{*(a1)}, z_v^{*(a2)}]$ into segments whose length is at most $\sqrt{2}$.

IV. Robust and Balanced BOP

We now address problem P3, i.e., we propose an algorithm, named ROBBO, to obtain an estimate-realization pair $(\hat{\mathcal{P}}, \mathcal{R})$ that attains the robustness guarantees with a much lower number of samples than what is obtained in the general case of Proposition 2. Moreover, the algorithm attains a balanced error vector, in a sense better specified below.

A. Central Estimate, Minimum Number of Samples and Balanced Error Property

ROBBO adopts the following estimate of the PF for a given set of samples $\mathcal{D}$:

$$\hat{\mathcal{P}}_c = W^{-1}\hat{h}_c, \quad \hat{h}_c = \frac{1}{2}\left(\bar{h}_{\mathcal{D}} + \underline{h}_{\mathcal{D}}\right). \tag{20}$$

The function $\hat{h}_c$ is the central estimate of h, and its worst case approximation error is the minimum among all possible estimates, considering the information that $h \in FPFS_{\mathcal{D}}$ [37], [38]. The global worst case error pertaining to $\hat{h}_c$ is equal to the so-called radius of information [38]

$$\max_{v \in \left[z_v^{*(a1)}, z_v^{*(a2)}\right]} \sup_{h \in FPFS_{\mathcal{D}}} \left|h(v) - \hat{h}_c(v)\right| = \frac{1}{2}\overline{\lambda}_{\mathcal{D}}. \tag{21}$$

Since the worst case approximation error obtained by $\hat{h}_c$ is the minimum one, this estimation approach results to be the most efficient in terms of number of samples required to guarantee the robustness condition (3) for any $\mathcal{P} \in \mathcal{A}_{\mathcal{D}}$ (i.e., any $h \in FPFS_{\mathcal{D}}$).

Proposition 3: Consider the BOP problem (1) and a tolerance vector $\boldsymbol{\delta}$. Under Assumption 1, if the central PF estimate $\hat{\mathcal{P}}_c$ (20) is adopted, then:

$$\begin{aligned} &\exists \mathcal{R}: \text{ property (3) is satisfied } \forall \mathcal{P} \in FPFS_{\mathcal{D}} \\ &\qquad\qquad \Longleftrightarrow \\ &\overline{\lambda}_{\mathcal{D}} \leq 2\sqrt{2}. \end{aligned} \tag{22}$$

Moreover, the corresponding minimum number of samples is

$$\underline{M}_c = \left\lceil \frac{1}{4}\left(\frac{\Delta_1}{\delta_1} + \frac{\Delta_2}{\delta_2}\right)\right\rceil + 1 \tag{23}$$

and the corresponding dataset $\mathcal{D}$ shall be evenly allocated in the segment $[z_v^{*(a1)}, z_v^{*(a2)}]$.

Proof: See the Appendix. ∎

Regarding the realization $\mathcal{R}$, ROBBO employs $\mathcal{R}_{SM}$, already introduced in (17). The main reason is that this realization yields an additional property on the approximation error vector $\boldsymbol{\varepsilon}$: it is such that the ratio $|\varepsilon_1|/|\varepsilon_2|$ is always fixed and equal to the ratio of the prescribed error bounds

$$\forall \hat{\mathcal{P}}, \mathcal{P} \in FPFS_{\mathcal{D}} \quad \forall \hat{\boldsymbol{z}} \in \hat{\mathcal{P}}, \ \boldsymbol{z}^* = \mathcal{R}_{SM}(\hat{\boldsymbol{z}}), \quad \frac{|\varepsilon_1|}{|\varepsilon_2|} = \frac{\delta_1}{\delta_2}. \tag{24}$$

This property can be derived along the same lines as the proof of Theorem 2. We call an estimate that enjoys this property *balanced*. The balancing property ensures equal errors for the two criteria normalized to their respective tolerances, i.e., $(|\varepsilon_1|/|\varepsilon_2|) = (\delta_1/\delta_2) \rightarrow (|\varepsilon_1|/\delta_1) = (|\varepsilon_2|/\delta_2)$. Furthermore, balancing the error is consistent with the approximation requirements set by the DM through the choice of tolerances, since a larger tolerance is typically accompanied by a proportionally larger user-acceptable error. When a pair $(\hat{\mathcal{P}}, \mathcal{R})$ satisfies (3) and (24), we say that it is a *robust and balanced* PF estimate-realization pair, see Fig. 1 for visualization.

B. Iterative Algorithm

ROBBO builds iteratively the dataset $\mathcal{D}$ until the condition (3) is satisfied. Let us denote with $k \in \mathbb{N}$ the iteration number of the algorithm. Besides the anchor points, computed with the auxiliary problems (2), at each k ROBBO computes a new sample of the PF by choosing a suitable value $\tilde{v} \in [z_v^{*(a1)}, z_v^{*(a2)}]$ and solving the following auxiliary optimization problem:

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in S} \delta_1 f_1(\boldsymbol{x}) + \delta_2 f_2(\boldsymbol{x}) \tag{25a}$$

$$\text{s.t. } \delta_2 f_1(\boldsymbol{x}) - \delta_1 f_2(\boldsymbol{x}) - \frac{2\delta_1\delta_2}{\sqrt{2}}\tilde{v} = 0. \tag{25b}$$

The corresponding PF sample is $\boldsymbol{z}^* = f(\boldsymbol{x}^*)$. The equality constraint (25b) is designed to find the Pareto point $\boldsymbol{z}^*$ such that its v-coordinate is equal to $\tilde{v}$, and its q-coordinate is equal to $h(\tilde{v})$. In fact, we have

$$\begin{aligned} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = W^{-1}\begin{bmatrix} v \\ q \end{bmatrix} &= \begin{bmatrix} \delta_1 \Big/ \sqrt{2}(v+q) \\ \delta_2 \Big/ \sqrt{2}(q-v) \end{bmatrix} \\ q &= \frac{\sqrt{2}}{\delta_1} f_1 - v \\ \delta_1 f_2 = \frac{\delta_1\delta_2}{\sqrt{2}}(q-v) &= \delta_2 f_1 - \frac{2\delta_1\delta_2}{\sqrt{2}} v. \end{aligned} \tag{26}$$

Note that the gradient of the cost (25a) is perpendicular to the gradient of the equality constraint (25b), both seen as functions of f_1, f_2, so that ideally the point $[f_1(\boldsymbol{x}), f_2(\boldsymbol{x})]^T$ would slide along the equality constraint until reaching the desired PF point. Under Assumption 1, solving problem (25)

Algorithm 1 RObust and Balanced BOP

Inputs: $f_1, f_2, S, \boldsymbol{\delta}$ (or $\boldsymbol{\delta}^{\%}$)
Output: $\hat{\mathcal{P}}_c$
Algorithm:
a) Set $k \leftarrow 1$, compute anchor point $\boldsymbol{z}^{*(a1)}$ by solving (2a).
b) Set $k \leftarrow 2$, compute anchor point $\boldsymbol{z}^{*(a2)}$ by solving (2b).
c) Initialize the dataset $\mathcal{D}^{\langle 2\rangle} \leftarrow \{\boldsymbol{z}^{*(a1)}, \boldsymbol{z}^{*(a2)}\}$.
d) Preallocate the sampling coordinates:
  d1) compute $\Delta_1 \leftarrow z_1^{*(a2)} - z_1^{*(a1)}$, $\Delta_2 \leftarrow z_2^{*(a1)} - z_2^{*(a2)}$;
  d2) if tolerances as $\boldsymbol{\delta}^{\%}$, compute $\boldsymbol{\delta} \leftarrow [\Delta_1\delta_1^{\%}, \Delta_2\delta_2^{\%}]^T$;
  d3) compute the number of samples $\underline{M}_c$ from (23);
  d4) define a grid of strictly increasing v-values $\tilde{v}^{(r)}$, $r \in I = \{1, \ldots, \underline{M}_c - 2\}$ in the interval $[z_v^{*(a1)}, z_v^{*(a2)}]$, equally spaced one from the next and from the anchor points' coordinates $z_v^{*(a1)}, z_v^{*(a2)}$;
  d5) initialize the set of explored indices $E^{\langle k\rangle} \leftarrow \emptyset$.
e) Compute $\overline{\lambda}_{\mathcal{D}}^{\langle 2\rangle} \leftarrow V - |Q|$, from (16), with V and Q derived from (15), ($z_v^{*(\ell-)} = z_v^{*(a1)}$, $z_v^{*(\ell+)} = z_v^{*(a2)}$).
f) While $\overline{\lambda}_{\mathcal{D}}^{\langle k\rangle} > 2\sqrt{2}$:
  f1) set $k \leftarrow k + 1$;
  f2) select $\tilde{v}^{\langle k\rangle} \leftarrow \tilde{v}^{(r^*)}$, $r^* \leftarrow \arg\max_{r \in I \setminus E^{\langle k-1\rangle}} \bar{h}_{\mathcal{D}}(\tilde{v}^{(r)}) - \underline{h}_{\mathcal{D}}(\tilde{v}^{(r)})$, where $\bar{h}_{\mathcal{D}}$ and $\underline{h}_{\mathcal{D}}$ are computed using (13) with $\mathcal{D} = \mathcal{D}^{\langle k-1\rangle}$;
  f3) update the set $E^{\langle k\rangle} \leftarrow E^{\langle k-1\rangle} \cup \{r^*\}$;
  f4) solve (25) with $\tilde{v} = \tilde{v}^{\langle k\rangle}$, the new sample is $\boldsymbol{z}^{\langle k\rangle}$;
  f5) update the dataset $\mathcal{D}^{\langle k\rangle} \leftarrow \mathcal{D}^{\langle k-1\rangle} \cup \{\boldsymbol{z}^{\langle k\rangle}\}$;
  f6) compute $\overline{\lambda}_{\mathcal{D}}^{\langle k\rangle}$ as in (16) with $\mathcal{D} = \mathcal{D}^{\langle k\rangle}$.
g) Return the PF approximation:
  g1) compute $\bar{h}_{\mathcal{D}}$, $\underline{h}_{\mathcal{D}}$ from (13), with $\mathcal{D} = \mathcal{D}^{\langle k\rangle}$;
  g2) compute the central approximation $\hat{h}_c$ as in (20);
  g3) return $\hat{\mathcal{P}}_c \leftarrow W^{-1}\hat{h}_c$, where W^{-1} is the inverse of the linear transformation introduced in [7].

always yields a PF point, both for convex and nonconvex PFs. This auxiliary optimization problem also serves as a practical implementation of the realization $\boldsymbol{z}^* = \mathcal{R}_{SM}(\hat{\boldsymbol{z}})$ [see (17)]: in this case, one shall set $\tilde{v} = \hat{z}_v$, where $\hat{z}_v$ is the v-coordinate of $\hat{\boldsymbol{z}}$. The pseudo-code for ROBBO is outlined in Algorithm 1, where the notation $\cdot^{\langle k\rangle}$ denotes a quantity that is updated at each iteration k. In virtue of Proposition 3, Algorithm 1 is always guaranteed to end in at most $\underline{M}_c$ iterations (including the ones to compute the anchor points), satisfying the robustness guarantees (3). In practice, a much lower number of samples may be needed, depending on the actual shape of the PF. The algorithm also accepts relative tolerances $\delta_1^{\%} = \delta_1/\Delta_1$ and $\delta_2^{\%} = \delta_2/\Delta_2$ as an alternative option for the DM instead of absolute ones, see point d2). We present some benchmarks about the actual number of iterations in Section V, for different PF shapes. Regarding step f2) of the algorithm, if more than one value of $\tilde{v}(r)$ attains the same worst case error, then we select the one that is farthest, according to the v-coordinate, from the already collected samples. If there is more than one value with equal ranking, we pick the smallest one. ROBBO's computational complexity scales with the worst case axis-wise sample count $\lceil(\kappa/4)\rceil + 1$, where $\kappa = (1/\delta_1^{\%}) + (1/\delta_2^{\%})$ [see (23)], and with

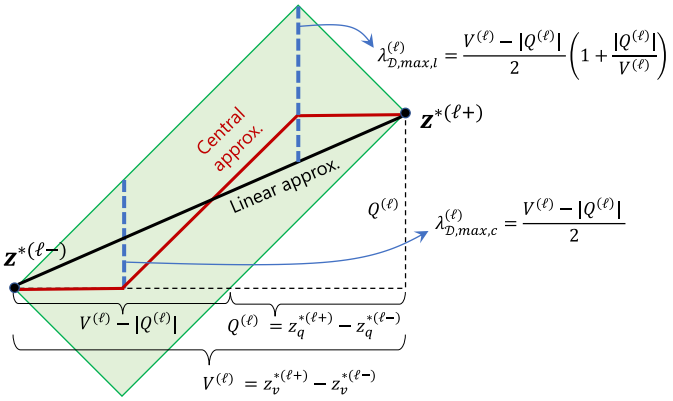

Fig. 4. Computation of $\lambda^{(\ell)}_{\mathcal{D},max,c}$, $\lambda^{(\ell)}_{\mathcal{D},max,l}$, for the central and linear approximation, in the generic interval $R^{(\ell)}$.

the features of the BOP problem. Specifically, the complexity of the algorithm is $O(\kappa C_s + \kappa^2)$. The first term is linear in κ and in the computational cost C_s of solving a scalarization problem, which is determined by the features of the BOP and of the chosen solver [39], typically representing the dominant computational burden. The second, quadratic term accounts for the loop searching for the new sampling coordinates, i.e., step f2) of Algorithm 1.

Remark 3: If Assumption 1 is violated, i.e., the PF is discontinuous, then the auxiliary problem (25) may either be infeasible or yield a solution outside the optimal bounds, i.e., not Pareto optimal. In the first case, one can adapt the problem according to Pascoletti–Serafini scalarization to always find a Pareto optimal solution, relaxing the linear constraint (25b), [40]. In the second, one can apply Pareto filters to refine the dataset $\mathcal{D}$, [41]. A rigorous extension of ROBBO to discontinuous PFs is the subject of current research and outside the scope of this work.

V. Sampling Strategy: Analysis and Comparisons

This section analyzes the sampling strategy adopted in the algorithm. First, a method to compute a priori the guaranteed accuracy as a function of the sampling budget is introduced (Section V-A. Then, central and linear approximations are compared (Section (V-B), while the bisection approach is assessed against uniform and greedy sampling (Sections V-C and V-D). Finally, the superiority of ROBBO in preallocating the number of samples over conventional scalarization methods is rigorously proved (Section V-E).

A. Limited Sampling Budget

A relevant situation in practice is when the DM has a limited budget of PF points to be sampled, denoted by n_B, and wants to know what are the attainable guaranteed worst case error bounds on each of the two objective functions, $\bar{\delta}_1$, $\bar{\delta}_2$, that comply with a desired ratio $\alpha = \bar{\delta}_1/\bar{\delta}_2$, representing the relative importance of the two criteria. Using the central estimate (which guarantees the smallest worst case error) and evenly distributed samples as done by ROBBO (which guarantees

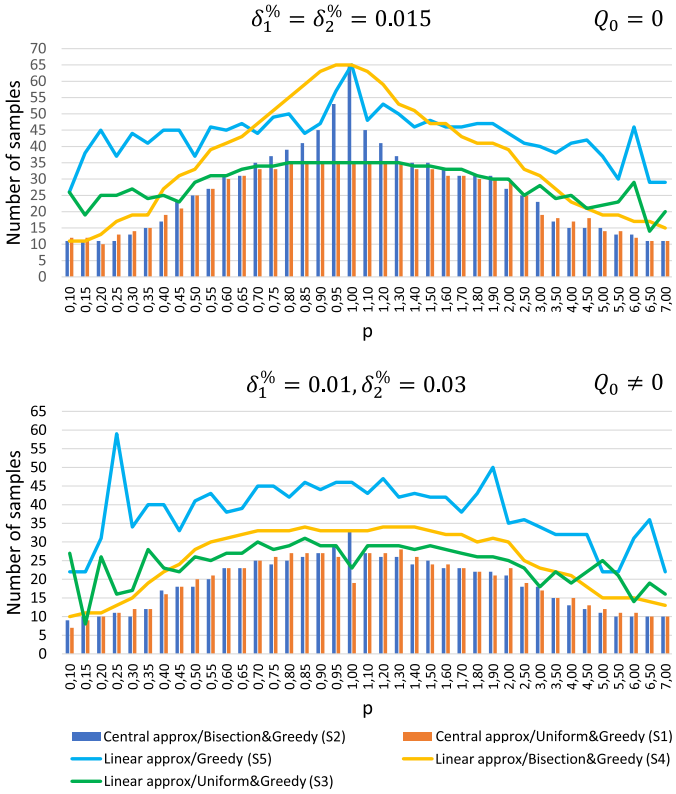

Fig. 5. Empirical comparison of the iterative sampling heuristics when varying the front shape.

satisfaction of property (3) with the smallest number of samples), we can impose the following condition:

$$\bar{\lambda}_{\mathcal{D}} = \frac{z_v^{*(a2)} - z_v^{*(a1)}}{n_B - 1} = \frac{1}{n_B - 1} \frac{\sqrt{2}}{2} \left(\frac{\Delta_1}{\bar{\delta}_1} + \frac{\Delta_2}{\bar{\delta}_2} \right) \tag{27}$$

which stems from Proposition 3 and includes the computation of the two anchor points as part of the budget n_B. Then, using $\bar{\delta}_1 = \alpha \bar{\delta}_2$, we obtain

$$\bar{\delta}_1 = \frac{\Delta_1 + \alpha \Delta_2}{4(n_B - 1)}, \quad \bar{\delta}_2 = \frac{\Delta_1 + \alpha \Delta_2}{4\alpha(n_B - 1)}. \tag{28}$$

This result can be used to precompute, after obtaining the anchor points, the attainable worst case accuracy guarantees as a function of the sampling budget n_B to be allocated. A case study for this approach is presented in Section VI-B.

B. Linear Interpolation

An alternative estimation technique, widely used in practice, is the piecewise linear approximation, denoted by $\hat{h}_l(v)$. For a generic interval $R^{(\ell)} = [z_v^{*(\ell-)}, z_v^{*(\ell+)}] \subset [z_v^{*(a1)}, z_v^{*(a2)}]$ (see Section III-B). This estimate is given by

$$\hat{h}_l(v) = z_q^{*(\ell-)} + \frac{z_q^{*(\ell+)} - z_q^{*(\ell-)}}{z_v^{*(\ell+)} - z_v^{*(\ell-)}} \left(v - z_v^{*(\ell-)} \right) \quad \forall v \in R^{(\ell)}. \tag{29}$$

For simplicity, we adopt a graphical interpretation to study the guaranteed accuracy properties of this estimation technique and compare them to those of the central estimate, see Fig. 4. We denote with $\lambda_{\mathcal{D},max,c}^{(\ell)}$ and with $\lambda_{\mathcal{D},max,l}^{(\ell)}$ the worst case error pertaining to the central and linear approximation, respectively. Regarding the central estimate, consistently with Proposition 1 and (21), we have $\lambda_{\mathcal{D},max,c}^{(\ell)} = ((V^{(\ell)} - |Q^{(\ell)}|)/2)$ and this value is attained for all v-values in the interval $[v'', v']$ (see Fig. 4 and (53) in the proof of Proposition 1). About the linear estimate, geometric considerations lead to $\lambda_{\mathcal{D},max,l}^{(\ell)} = ((V^{(\ell)} - |Q^{(\ell)}|)/2)(1 + (|Q^{(\ell)}|/V^{(\ell)}))$. It is immediate to note that the two worst case errors are the same only if the $q-$coordinates of $z^{*(\ell-)}$, $z^{*(\ell+)}$ coincide, i.e., $|Q^{(\ell)}| = 0$. This is the situation when the worst case error is largest in the interval $R^{(\ell)}$, as argued in the proof of Proposition 2. Otherwise, the central estimate has generally better worst case performance, depending on the difference $V^{(\ell)} - |Q^{(\ell)}|$. Note that, due to the Lipschitz continuity property of the rotated PF with a constant smaller than one (Remark 1), we always have $V^{(\ell)} > |Q^{(\ell)}|$. Finally, note that, for the linear approximation, the worst case error is attained only at coordinates v'' and v' and not in between them.

C. Greedy Bisection Sampling Strategy

It is interesting to compare the sampling approach adopted in ROBBO to a more common, greedy bisection method. In the latter, instead of picking a point from the uniform grid, at each iteration one selects, as value of $\tilde{v}$ in problem (25), the midpoint with the largest uncertainty among all intervals $R^{(\ell)}$, thus bisecting that interval. From the considerations of Section V-B, the midpoint always corresponds to the largest worst case error $\lambda_{\mathcal{D},max,c}^{(\ell)}$ for the central estimate, thus sampling it can generally bring a significant reduction of the uncertainty. For the linear approximation, to pick the point with maximum uncertainty one could sample at coordinate v' or v'', i.e. $z_v^{*(\ell)+} - (V^{(\ell)} - |Q^{(\ell)}|)/2$ or $z_v^{*(\ell-)} + (V^{(\ell)} - |Q^{(\ell)}|)/2$.

If the greedy bisection strategy is adopted, the algorithm still converges in finite iterations, and we can precompute the maximum number. This corresponds to $\underline{M}_c$ only in a very specific case; it is larger. In fact, if the bisection approach is used, in the worst case (i.e., when the actual PF is such that its q-coordinates are all equal, so that $|Q^{(\ell)}| = 0$, $\forall \ell$), the global worst case error bound $\max_\ell \lambda_{\mathcal{D},max,c}^{(\ell)}$ is halved only when all intervals have been halved. Each time that this happens, we say that one epoch has passed. We can then compute the worst case number of epochs n_e and the upper bound of samples, $\underline{M}_g$, for such a greedy strategy to guarantee satisfaction of (3), as a function of the distance V_a between the anchor points' v-coordinates

$$\frac{V_a}{2^{n_e}} \leq 2\sqrt{2} \Rightarrow n_e = \left\lceil \log_2 \left(\frac{V_a}{2\sqrt{2}} \right) \right\rceil$$
$$\underline{M}_g = 2 + \sum_{i=1}^{n_e} 2^{(i-1)} \tag{30}$$

where $\underline{M}_g$ takes into account the anchor points. We note that $\underline{M}_g$ of the greedy bisection strategy is equal to $\underline{M}_c$ of the uniform sampling only in the very special case that the ratio $V_a/(2\sqrt{2})$ is a power of 2, otherwise, it is larger.

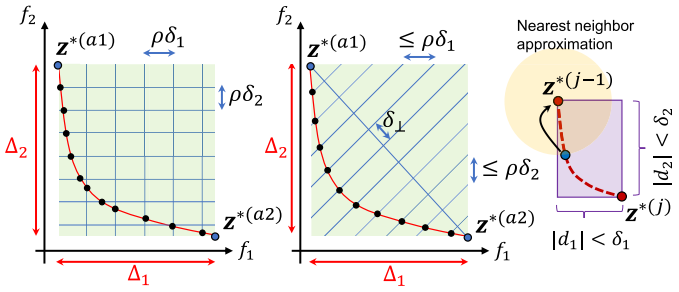

Fig. 6. Uniform sampling for EC (left plot), NBI (central plot) to satisfy the requirement (3), and nearest neighbor PF approximation (right).

D. Empirical Benchmark of Estimation and Sampling Strategies

While the theoretical findings of Sections V-B and V-C refer to worst case situations, it is important to assess the actual performances in different scenarios, i.e., the actual number of samples that ensures the robust accuracy requirement (3). We perform the analysis comparing the following alternatives: S1) central estimate with uniformly allocated points (i.e., ROBBO); S2) central estimate with greedy bisection sampling; S3) linear interpolation with uniform distribution of points; S4) linear interpolation with greedy bisection sampling; and S5) linear interpolation with greedy sampling of the points with largest uncertainty (i.e., points with coordinates v' or v'' mentioned in Section V-C).

We set the elements of the tolerance vector as percentages, and consider the following BOP:

$$\min_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}) = [x_1, x_2]^T \tag{31a}$$

$$\text{s.t.} \left(\sum_{i=1}^{2} {x_i}^p\right)^{\frac{1}{p}} \geq 10 \tag{31b}$$

$$x_1, x_2 \geq 0. \tag{31c}$$

The PF of this problem is given by $f_2 = (10^p - {f_1}^p)^{(1/p)}$, derived from constraint (31b). The anchor points are $\boldsymbol{z}^{*(a1)} = [0, 10]^T$, $\boldsymbol{z}^{*(a2)} = [10, 0]^T$. The PF is convex for $0 < p < 1$, it is linear for $p = 1$, and concave for $p > 1$. In addition to the shape, the value of $Q_0 = |z_q^{*(a2)} - z_q^{*(a1)}|$ can be controlled by varying $\delta_1^{\%}$, $\delta_2^{\%}$: since $\Delta_1 = z_1^{*(a2)}, \Delta_2 = z_2^{*(a1)}$, we have in fact $Q_0 = (\sqrt{2}/2)((1/\delta_1^{\%}) - (1/\delta_2^{\%}))$. Thus, when $\delta_1^{\%} = \delta_2^{\%}$ the two anchor points have the same q-coordinate. This allows us to also test the worst case situation, by setting $p = 1$ (linear front) and $\delta_1^{\%} = \delta_2^{\%}$ (the whole rotated front with the same q-coordinate). We present the results obtained in case of $Q_0 = 0$ with $\delta_1^{\%} = \delta_2^{\%} = 1.5\%$, and of $Q \neq 0$ with $\delta_1^{\%} = 1\%$, $\delta_2^{\%} = 3\%$. In both cases, we tested different values of $p \in [0.01, 7]$. The results are reported in Fig. 5.

Overall, we see that the central estimate (S1 and S2) performs better than the linear interpolation (S3, S4, and S5). As expected from the theory, when $Q = 0$, central and linear behave similarly when the shape of the front is close to linear (i.e., when $p \approx 1$), since both approximations match (S1 with S3 and S2 with S4). This comparison confirms that the central estimate combined with uniformly spaced samples (S1), as implemented in ROBBO, is overall the best option. A demo implementation of this example is openly available at [42], including an open implementation of the algorithm.

E. Theoretical Comparison With Popular PF Estimation Methods

We now compare the worst case required number of samples obtained by ROBBO with that of some of the most popular exact strategies for PF approximation. For this purpose, we have selected the ϵ-constraint approach [9], denoted by EC, and the Normal Boundary Intersection one [7], NBI. Because these methods deliver a point-wise approximation of the PF (i.e., nearest neighbor), we consider the distance between a noncomputed point and its nearest neighbor as the approximation error (see Fig. 6, right). Thus, to satisfy the robust accuracy condition (3), the distance between two sampled points $\boldsymbol{d} = \boldsymbol{z}^{*(j+1)} - \boldsymbol{z}^{*(j)}$ must satisfy

$$|d_1| \leq \delta_1 \wedge |d_2| \leq \delta_2. \tag{32}$$

Using the EC method, a point-wise approximation of PF is provided by solving several instances of the scalar problem that minimizes one of the two objectives, while constraining the other objective to be less than or equal to a value ϵ [9]. At each new instance, ϵ is increased by a fixed quantity to obtain a uniform distribution of points with respect to the constrained criterion. To satisfy condition (32), one needs to repeat the procedure for both objectives (see Fig. 6, left). The corresponding worst case number $\underline{M}_{EC}$ of samples is therefore

$$\underline{M}_{EC} = \left\lceil \frac{\Delta_1}{\delta_1} + \frac{\Delta_2}{\delta_2} \right\rceil + 1. \tag{33}$$

The NBI method [7] produces samples that are uniformly spaced when projected on the segment connecting the anchor points, see Fig. 6 (center). We denote with W such a segment, and with $\delta_\perp$ the distance between the projections of two subsequent samples on W. Let us consider the quantities $l_1 = \delta_1 \cos(\beta)$ and $l_2 = \delta_2 \sin(\beta)$, where $\beta = \tan^{-1}(\Delta_2/\Delta_1)$. In the worst case, to fulfill (32), it must hold $\delta_\perp \leq \min(l_1, l_2)$. Then, we can compute the worst case number of samples $\underline{M}_{NBI}$ by dividing the length of W by $\delta_\perp$ that satisfies this condition. Without loss of generality, we consider $\delta_1 = \alpha\delta_2$ and $\Delta_1 = \gamma\Delta_2$. Then, assuming $l_2 < l_1$ we have

$$\begin{aligned}
\frac{\Delta_2}{\delta_\perp \sin(\beta)} &= \frac{\Delta_2}{\delta_2 (\sin(\beta))^2} = \frac{\Delta_2}{\delta_2} \frac{1 + \tan(\beta)^2}{\tan(\beta)^2} \\
&= \frac{\Delta_2}{\delta_2} \frac{1 + (\Delta_2/\Delta_1)^2}{(\Delta_2/\Delta_1)^2} = \frac{{\Delta_1}^2 + {\Delta_2}^2}{\delta_2 \Delta_2} \\
&= \left(\frac{{\Delta_1}^2}{\delta_2 \Delta_2} + \frac{\Delta_2}{\delta_2}\right) = \left(\frac{{\Delta_1}^2}{\frac{\delta_1}{\alpha}\frac{\Delta_1}{\gamma}} + \frac{\Delta_2}{\delta_2}\right) \\
&= \left(\frac{\Delta_1}{\delta_1}\alpha\gamma + \frac{\Delta_2}{\delta_2}\right) \\
l_2 < l_1 &\Rightarrow \delta_2 \sin(\beta) < \delta_1 \cos(\beta) \Rightarrow \delta_2 \sin(\beta) \\
&< \alpha\delta_2 \cos(\beta) \\
&\Rightarrow \alpha > \frac{\sin(\beta)}{\cos(\beta)} \Rightarrow \alpha > \frac{\Delta_2}{\gamma\Delta_2} \Rightarrow \alpha\gamma > 1
\end{aligned} \tag{34}$$

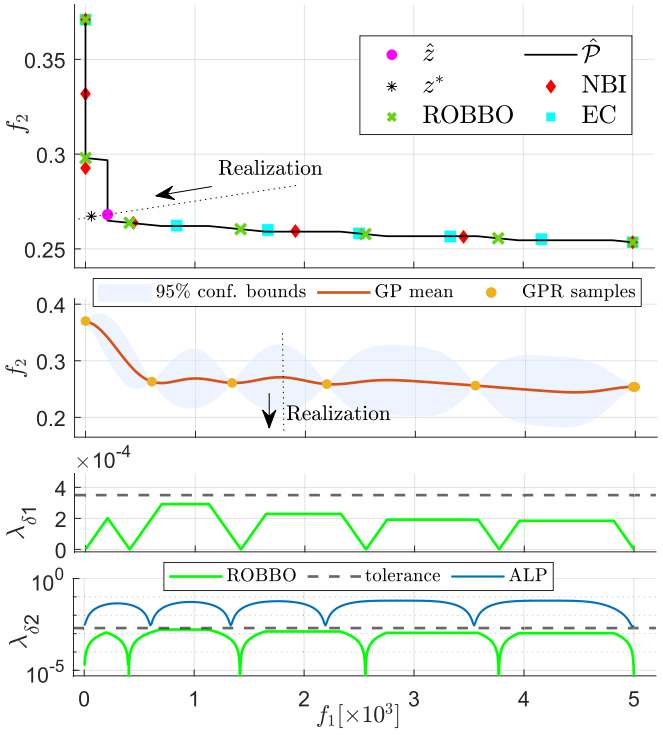

Fig. 7. Top plot: PF estimate from ROBBO (green samples), EC (light blue), and NBI (red). Second plot: performance of the active learning of the PF method (ALP, [25]) on the same problem. Lower plots: worst case approximation error λ_{δ_1} and λ_{δ_2} represented along the f_1 axis, for ROBBO and ALP, compared with the error tolerances. Note that the λ_{δ_2} plot is in logarithmic scale.

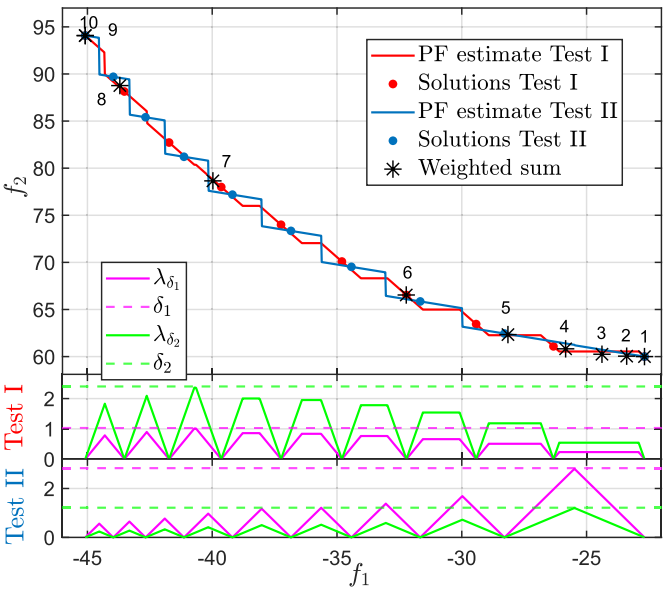

Fig. 8. CSTR example: Test I and test II results and scalarization by convex combination of the two objectives, with the same budget $n_B = 10$. Estimated PF of the economically optimal steady states for a CSTR (upper plot), comparison of the worst case approximation errors $\lambda_{\delta_{1,2}}$ represented along the f_1 axis, and the error tolerances $\delta_{1,2}$ (lower plots).

from which we derive

$$\underline{M}_{NBI} = \left\lceil \frac{\Delta_1}{\delta_1}\alpha\gamma + \frac{\Delta_2}{\delta_2} \right\rceil + 1\alpha\gamma > 1. \tag{35}$$

Assuming $l_1 < l_2$, with similar passages we obtain

$$\underline{M}_{NBI} = \left\lceil \frac{\Delta_1}{\delta_1} + \frac{\Delta_2}{\delta_2}\frac{1}{\alpha\gamma} \right\rceil + 1, \quad \frac{1}{\alpha\gamma} > 1. \tag{36}$$

Finally if $l_1 = l_2$ we find

$$\underline{M}_{NBI} = \underline{M}_{EC} = \left\lceil \frac{\Delta_1}{\delta_1} + \frac{\Delta_2}{\delta_2} \right\rceil + 1. \tag{37}$$

Now, comparing $\underline{M}_c$ (23), $\underline{M}_{EC}$ (33), and $\underline{M}_{NBI}$ (35)–(37), we see that NBI needs, in the worst case, at least the same number of samples as EC, which is approximately four times larger than in our method, where we exploit the optimal bounds and a continuous estimated PF instead of a nearest-neighbor one.

VI. Case Studies

To test the proposed approach, the ROBBO algorithm is adopted in this section to solve two bi-objective control problems. In the first case, the user defines the error tolerances a priori, whereas in the second, the thresholds are tuned according to the sampling budget and the error ratio between the two objectives.

A. Path-Following Positioning System

We present the application of ROBBO to a constrained-path following problem for a 2-axis positioning system, formulated as a finite-horizon optimal control problem (FHOCP). The system modeling and problem formulation are reported in [43]. The two computed anchor points are $\boldsymbol{z}^{*(a1)} = [10^{-8}, 3.71\,10^{-1}]$ and $\boldsymbol{z}^{*(a2)} = [5\,10^{-3}, 2.54\,10^{-1}]$. In this experiment, the objectives' tolerances are set to $\delta_1 = 3.5\,10^{-4}$ and $\delta_2 = 2\,10^{-3}$ (i.e., approx. 7% and 1.7% of the range of values given by the anchor points).

From the anchor points computation, we derive $\underline{M}_c = 20$, by (23), and ROBBO reaches the accuracy guarantees in 5 iterations after computing the anchor points, i.e., $k = 7$ samples in total. Fig. 7 shows that the algorithm sampled fewer points where f_1 is almost zero (i.e., very small tracking error) with large completion time, since the corresponding solutions are practically equivalent from the perspective of f_1.

Then, taking the role of the DM, we selected a candidate solution on the approximate PF, close to the knee, i.e., the point at which the completion time is greatly reduced but with a very low cost in terms of tracking error. The selected candidate is $\hat{\boldsymbol{z}} = [1.99\,10^{-4}, 2.683\,10^{-1}]$, and its corresponding realization $\boldsymbol{z}^*$ is $\mathcal{R}_{SM}(\hat{\boldsymbol{z}}) = [5\,10^{-5}, 2.674\,10^{-1}]$, both displayed in Fig. 7, upper plot. The two lower plots of the figure also show the value of the error bound along the approximation for both objectives and the related user-defined tolerances. Comparing the solution chosen by the DM and its realization, according to the theory we obtained an approximation error $\boldsymbol{\varepsilon} = [1.49\,10^{-4}, 8.5\,10^{-4}]$, such that $|\varepsilon_1| < \delta_1 \wedge |\varepsilon_2| < \delta_2$ and $(|\varepsilon_1|/|\varepsilon_2|) = (\delta_1/\delta_2) = 0.175$. The global worst case errors obtained by the approximation for objectives 1 and 2 are $\bar{\lambda}_{\delta_1} = 2.92\,10^{-4}$, $\bar{\lambda}_{\delta_2} = 1.67\,10^{-3}$, verifying that tolerances are met for any candidate. Finally, we compute a discrete PF approximation adopting the EC and NBI methods with the same number of samples as ROBBO. The results are displayed in Fig. 7. Neither EC nor NBI meets the condition (32) with the allotted number of samples. We obtain as highest gap

for EC $|d_1^{max}| = 8.34\,10^{-4}$, $|d_2^{max}| = 1.08\,10^{-1}$, and for NBI $|d_1^{max}| = 1.50\,10^{-3}$, $|d_2^{max}| = 3.92\,10^{-2}$. The same figure (second plot) also presents the results of a test using the method in [25], showing that the shape of the PF can be challenging to capture for learning approaches such as Gaussian processes. In this method, the realization is computed by selecting a coordinate along f_1. Hence, the uncertainty is measured only for the second objective, yielding a 95% confidence interval of $\pm 6.25\,10^{-2}$, see Fig. 7, lower plots.

B. Continuous Stirred-Tank Reactor

We use ROBBO to approximate the PF of a steady-state optimization problem for a CSTR, maximizing production rate versus minimizing cost. The system modeling and problem formulation are reported in [43]. We provide ROBBO with a budget of $n_B = 10$ samples, including anchor points, see Section V-A, and perform two tests. In test I, we set the tolerance ratio $\alpha = \bar{\delta}_1/\bar{\delta}_2 = 3/7$ to favor better accuracy for f_1, while in test II we set $\alpha = 7/3$. When adopting this strategy, step c) of Algorithm 1 presents the following changes: $\underline{M}_c$ is not computed but directly set equal to n_B, $\delta_1 = \bar{\delta}_1, \delta_2 = \bar{\delta}_2$ are computed using (28). Then, at step e), the algorithm is iterated as long as $k \leq n_B$, thus exploiting the whole sampling budget. In test I we obtain the guaranteed tolerances $\bar{\delta}_1 = 1.0284, \bar{\delta}_2 = 2.3997$, while for test II we obtain $\bar{\delta}_1 = 2.8314, \bar{\delta}_2 = 1.2134$. As expected from the theory, their ratios are equal to the imposed α values. The results are displayed in Fig. 8, where the upper plot presents the PF estimates for both test I and test II, in red and blue, respectively. The two lower plots present the worst case realization errors for each objective individually, computed as $\lambda_{\delta_1}(\hat{z}_1^{(j)}) = \delta_1(\sqrt{2}/2)\lambda_D(\hat{z}_v^{(j)})$ and $\lambda_{\delta_2}(\hat{z}_1^{(j)}) = \delta_2(\sqrt{2}/2)\lambda_D(\hat{z}_v^{(j)})$, and the error tolerances δ_1, δ_2. As shown in Fig. 8, the guaranteed error bounds λ_{δ_1} and λ_{δ_2} are always below the precomputed values of δ_1 and δ_2. The global worst case error bounds corresponds to $\bar{\lambda}_{\delta_1} = 1.012$, $\bar{\lambda}_{\delta_2} = 2.36$ for test I, and to $\bar{\lambda}_{\delta_1} = 2.806$, $\bar{\lambda}_{\delta_2} = 1.202$ for test II. It is interesting to notice how the samples' positions in terms of (f_1, f_2)-values change between the two tests, while in the (v, q)-coordinates the approach always adopts an even distribution of samples. This is due to the effect of the tolerance values on the linear transformation (7), recalling that the identification is performed in coordinates (v, q). Finally, Fig. 8 also presents a comparison between Test I, II, and the samples obtained by minimizing the convex combination $\beta f_1 + (1-\beta) f_2$ with the same budget $n_B = 10$ of β values evenly distributed in the interval $[0, 1]$. Such a scalarization method is a rather commonly used approach to attempt the PF approximation. It can be clearly noted that the scalarization results are extremely poor with respect to ROBBO, with a large part of the PF without any samples and corresponding high approximation error. The widest gap between two samples with the scalarization method measures $|d_1^{max}| = 7.75$, $|d_2^{max}| = 12.11$.

VII. Conclusion and Future Developments

We presented new results in the field of BOP, delivering the algorithm ROBBO, which builds an estimate of the PF with guaranteed accuracy, a low number of required samples, and a desired balancing between the estimation errors with respect to the two criteria. We proved these properties theoretically by assuming the PF to be a continuous and compact curve in $\mathbb{R}^2$, and compared the approach with other ones, both theoretically and empirically, showing the superiority of ROBBO in terms of the required number of samples to attain the desired guarantees. Thus, it is ideally suited for providing a comprehensive approximation of a continuous PF for bi-objective problems where computing nondominated solutions is computationally expensive. Future research will focus on problems with more than two objectives, discontinuous fronts, and uncertain fronts, i.e., with objective function outcomes affected by some intrinsic variability.

Appendix

Proof of Theorem 1: Consider a value $v \in [z_v^{*(a1)}, z_v^{*(a2)}]$, and take any $z^{*(j)} \in \mathcal{D}$. Assume that $z_v^{*(j)} < v$, i.e. $v = z_v^{*(j)} + \Delta v$ for some $\Delta v > 0$. Now consider a value $q = z_q^{*(j)} + |v - z_v^{*(j)}| + \Delta q$, where $\Delta q \in \mathbb{R}$. Since $\Delta v > 0$, we have that $q = z_q^{*(j)} + \Delta v + \Delta q$. Then, consider the point $[v, q]^T$ and apply the inverse transformation

$$z = W^{-1}\begin{bmatrix} v \\ q \end{bmatrix} = \frac{\delta_1\delta_2}{\sqrt{2}}\begin{bmatrix} \frac{v}{\delta_2} + \frac{q}{\delta_2} \\ -\frac{v}{\delta_1} + \frac{q}{\delta_1} \end{bmatrix}. \tag{38}$$

By substituting $v = z_v^{*(j)} + \Delta v$ and $q = z_q^{*(j)} + \Delta v + \Delta q$ and after a few manipulations, we obtain

$$z = z^{*(j)} + \begin{bmatrix} \frac{\delta_1}{\sqrt{2}}(2\Delta v + \Delta q) \\ \frac{\delta_2}{\sqrt{2}}\Delta q \end{bmatrix}. \tag{39}$$

Since $\Delta v > 0$, it is immediate to note that $z \notin \{z^{*(j)}\} \oplus \mathbb{R}^2_{>0} \iff \Delta q < 0$, which means

$$z \notin \left\{z^{*(j)}\right\} \oplus \mathbb{R}^2_{>0} \iff q < z_q^{*(j)} + \left|v - z_v^{*(j)}\right|. \tag{40}$$

If $\Delta v < 0$, i.e. $z_v^{*(j)} > v$, we have $q = z_q^{*(j)} - \Delta v + \Delta q$. Along the same lines, in this case, we obtain

$$z = z^{*(j)} + \begin{bmatrix} \frac{\delta_1}{\sqrt{2}}\Delta q \\ \frac{\delta_2}{\sqrt{2}}(2\,|\Delta v| + \Delta q) \end{bmatrix} \tag{41}$$

and we reach again the statement (40). Now, for z to be nondominated by any of the samples $z^{*(j)}$, the condition (40) must hold for all of them, i.e.,

$$z \notin \bigcup_{z^{*(j)} \in \mathcal{D}} \left\{z^{*(j)}\right\} \oplus \mathbb{R}^2_{>0} \iff q < \min_{z^{*(j)} \in \mathcal{D}} z_q^{*(j)} + \left|v - z_v^{*(j)}\right|. \tag{42}$$

Note that if $q = \min_{z^{*(j)} \in \mathcal{D}} z_q^{*(j)} + |v - z_v^{*(j)}|$, then we obtain either $z = z^{*(j)} + [0, (2\delta_2/\sqrt{2})|\Delta v|]^T$ or $z = z^{*(j)} + [(2\delta_1/\sqrt{2})|\Delta v|, 0]^T$ for some j: in both cases, the considered point cannot belong to the PF. Thus, the function

$$\min_{z^{*(j)} \in \mathcal{D}} z_q^{*(j)} + \left|v - z_v^{*(j)}\right| = \sup_{h \in FPFS_{\mathcal{D}}} h(v) = \bar{h}_{\mathcal{D}}(v) \tag{43}$$

is the optimal upper bound of $FPFS_{\mathcal{D}}$.

To find the optimal lower bound, we repeat the reasoning by taking this time $q = z_q^{*(j)} - |v - z_v^{*(j)}| + \Delta q$ for each j. Considering

$v = z_v^{*(j)} + \Delta v$, $\Delta v > 0$ and applying the inverse transformation to $[v, q]^T$ we obtain

$$z = z^{*(j)} + \begin{bmatrix} \frac{\delta_1}{\sqrt{2}} \Delta q \\ \frac{\delta_2}{\sqrt{2}} (-2\Delta v + \Delta q) \end{bmatrix}. \tag{44}$$

Since $\Delta v > 0$, we note that $z \notin \{z^{*(j)}\} \oplus \mathbb{R}^2_{<0} \iff \Delta q > 0$, which means

$$z \notin \left\{z^{*(j)}\right\} \oplus \mathbb{R}^2_{<0} \iff q > z_q^{*(j)} - \left|v - z_v^{*(j)}\right|. \tag{45}$$

The same conclusion is reached if $\Delta v < 0$; in fact, in this case, we have

$$z = z^{*(j)} + \begin{bmatrix} \frac{\delta_1}{\sqrt{2}} (-2|\Delta v| + \Delta q) \\ \frac{\delta_2}{\sqrt{2}} \Delta q \end{bmatrix}. \tag{46}$$

Thus, considering all samples, we obtain

$$z \notin \bigcup_{z^{*(j)} \in \mathcal{D}} \left\{z^{*(j)}\right\} \oplus \mathbb{R}^2_{<0} \iff q > \max_{z^{*(j)} \in \mathcal{D}} z_q^{*(j)} - \left|v - z_v^{*(j)}\right| \tag{47}$$

i.e., in order for z to be nondominating any of the PF samples in $\mathcal{D}$, the value of q must lie above the following function (optimal lower bound):

$$\max_{z^{*(j)} \in \mathcal{D}} z_q^{*(j)} - \left|v - z_v^{*(j)}\right| = \inf_{h \in FPFS_{\mathcal{D}}} h(v) = \underline{h}_{\mathcal{D}}(v). \tag{48}$$

What remains to be shown is that the optimal bounds can be evaluated by considering just the samples $z^{*(j^-)}(v)$, $z^{*(j^+)}(v)$ obtained from (12). Considering together conditions (42) and (47), we obtain [see (6)]:

$$z \in \mathcal{A}_{\mathcal{D}} \iff \max_{z^{*(j)} \in \mathcal{D}} z_q^{*(j)} - \left|v - z_v^{*(j)}\right| < q < \min_{z^{*(j)} \in \mathcal{D}} z_q^{*(j)} + \left|v - z_v^{*(j)}\right|. \tag{49}$$

Consider now the sample $z^{*(j^-)}(v)$. Since it belongs to the PF, it also belongs to $\mathcal{A}_{\mathcal{D}}$. Thus, according to (49), we have

$$z_q^{*(j^-)}(v) < z_q^{*(j)} + \left|z_v^{*(j^-)} - z_v^{*(j)}\right| \quad \forall z^{*(j)} \in \mathcal{D} \setminus \left\{z^{*(j^-)}(v)\right\} \tag{50}$$

which implies, for any j such that $z_v^{*(j)} < z_v^{*(j^-)}$

$$\begin{aligned} z_q^{*(j^-)}(v) + \left|v - z_v^{*(j^-)}\right| &< z_q^{*(j)} + \left|z_v^{*(j^-)} - z_v^{*(j)}\right| + \left|v - z_v^{*(j^-)}\right| \\ &= z_q^{*(j)} + \left|v - z_v^{*(j)}\right|. \end{aligned} \tag{51}$$

The result above derives from the fact that $z_v^{*(j)} < z_v^{*(j^-)} < v$ by the definition of $z^{*(j^-)}(v)$ (12). Thus, the upper bound $z_q^{*(j^-)}(v) + |v - z_v^{*(j^-)}|$ is lower than the one contributed by any other sample with v-coordinate smaller than $z_v^{*(j^-)}$, which can then be disregarded in the computation of (43). A similar reasoning applies to $z^{*(j^+)}(v)$, which rules out all samples with larger v-coordinate, and to the lower bounds. Thus, the optimal bounds can be obtained by considering $z^{*(j^-)}(v)$, $z^{*(j^+)}(v)$ only, proving the Theorem. ∎

Proof of Proposition 1: The nonoverlapping intervals $R^{(\ell)}$, $\ell = 1, \ldots, M-1$ cover the whole domain $[z_v^{*(a1)}, z_v^{*(a2)}]$ except for the points corresponding to the samples, where however the worst case error is zero because the upper and lower bounds coincide. Denoting $\lambda^{(\ell)}_{\mathcal{D},max} = \max_{v \in R^{(\ell)}} \bar{h}_{\mathcal{D}}(v) - \underline{h}_{\mathcal{D}}(v)$, we can then express the global worst case error bound as

$$\overline{\lambda}_{\mathcal{D}} = \max_{\ell=1,\ldots,M-1} \lambda^{(\ell)}_{\mathcal{D},max}. \tag{52}$$

For each $R^{(\ell)}$, considering Theorem 1, we see that the upper and lower bounds for any $v \in R^{(\ell)}$ depend only on the points defining its extremes, i.e. $z^{*(\ell-)}$, $z^{*(\ell+)}$. The value of $\lambda^{(\ell)}_{\mathcal{D},max}$ can be computed analytically on the basis of the (v, q)-coordinates of these points, as follows. First, we compute the coordinates v', $v'' \in R^{(\ell)}$ such that

$$\begin{aligned} z_q^{*(\ell-)} + \left(v' - z_v^{*(\ell-)}\right) &= z_q^{*(\ell+)} + \left(z_v^{*(\ell+)} - v'\right) \\ z_q^{*(\ell-)} - \left(v'' - z_v^{*(\ell-)}\right) &= z_q^{*(\ell+)} - \left(z_v^{*(\ell+)} - v''\right) \end{aligned}$$

i.e., the v-values at which the upper (respectively lower) bounds, contributed by the two samples delimiting the interval $R^{(\ell)}$, coincide. With a few manipulations, one obtains

$$\begin{aligned} v' &= \frac{z_v^{*(\ell+)} + z_v^{*(\ell-)}}{2} + \frac{Q^{(\ell)}}{2} \\ v'' &= \frac{z_v^{*(\ell+)} + z_v^{*(\ell-)}}{2} - \frac{Q^{(\ell)}}{2}. \end{aligned} \tag{53}$$

Thus, if the q-coordinates of $z^{*(\ell-)}$, $z^{*(\ell+)}$ are the same ($Q^{(\ell)} = 0$), v' and v'' coincide and are equal to the middle point of $R^{(\ell)}$, otherwise they differ and we can have either $v' > v''$ or $v' < v''$, depending on the sign of $Q^{(\ell)} = z_q^{*(\ell+)} - z_q^{*(\ell-)}$. Assume that $Q^{(\ell)} < 0$, which implies $v' < v''$. Then, the interval $R^{(\ell)}$ can be split in three subintervals: $I = (z_v^{*(\ell-)}, v']$, $II = (v', v'')$, $III = [v'', z_v^{*(\ell+)})]$. In intervals I and III, applying (13) one can find that both the upper and lower bounds are determined either by the sample $z^{*(\ell-)}$ (in I) or $z^{*(\ell+)}$ (in III), their maximum distance is attained when $v = v'$ (respectively $v = v''$) and it is equal to (using (53) and considering that we are assuming $Q^{(\ell)} < 0$)

$$\lambda^{(\ell)}_{\mathcal{D},max} = V^{(\ell)} - \left|Q^{(\ell)}\right|. \tag{54}$$

In interval II, the upper bound is determined by $z^{*(\ell+)}$, and the lower one by $z^{*(\ell-)}$. Using again (53) and the assumption $Q^{(\ell)} < 0$, one can see that the distance between the bounds is constant in the whole interval and equal to (54). The same conclusions are reached if $Q^{(\ell)} > 0$; thus the quantity $\lambda^{(\ell)}_{\mathcal{D},max}$ (54) is the worst case error bound pertaining to the interval $R^{(\ell)}$. By combining (52) and (54), the result is proven. ∎

Proof of Theorem 2: ($\Leftarrow$) For any $\hat{\mathcal{P}} \in FPFS_{\mathcal{D}}$ and any $\hat{z} \in \hat{\mathcal{P}}$, we consider the corresponding transformed point

$$\begin{bmatrix} \hat{z}_v \\ \hat{z}_q \end{bmatrix} = W \hat{z} \tag{55}$$

where $\hat{z}_v \in [z_v^{*(a1)}, z_v^{*(a2)}]$. Consider the realization $z^* = \mathcal{R}_{SM}(\hat{z})$ (17). The corresponding approximation error is

$$\boldsymbol{\varepsilon} = \hat{z} - z^* = \underbrace{\frac{\delta_1 \delta_2}{\sqrt{2}} \begin{bmatrix} \frac{1}{\delta_2} & \frac{1}{\delta_2} \\ -\frac{1}{\delta_1} & \frac{1}{\delta_1} \end{bmatrix}}_{W^{-1}} \begin{bmatrix} 0 \\ \hat{z}_q - h(\hat{z}_v) \end{bmatrix} \tag{56}$$

thus,

$$|\varepsilon_1| = \frac{\delta_1}{\sqrt{2}} \left|\hat{z}_q - h\left(\hat{z}_v\right)\right|; \quad |\varepsilon_2| = \frac{\delta_2}{\sqrt{2}} \left|\hat{z}_q - h(\hat{z}_v)\right|. \tag{57}$$

According to Theorem 1, both $h(\hat{z}_v)$ and $\hat{z}_q$ belong to the interval $(\underline{h}_{\mathcal{D}}(\hat{z}_v), \bar{h}_{\mathcal{D}}(\hat{z}_v))$, therefore $|\hat{z}_q - h(\hat{z}_v)| < \bar{h}_{\mathcal{D}}(\hat{z}_v) - \underline{h}_{\mathcal{D}}(\hat{z}_v)$. By (14) and the Theorem's necessary condition, we further have $\bar{h}_{\mathcal{D}}(\hat{z}_v) - \underline{h}_{\mathcal{D}}(\hat{z}_v) \leq \bar{\lambda}_{\mathcal{D}} \leq \sqrt{2}$. Using these inequalities in (57) we obtain $|\varepsilon_1| < \delta_1$ and $|\varepsilon_2| < \delta_2$, thus proving the condition.

($\Rightarrow$) We prove the sufficient condition by contradiction. First, note that (3) can be reformulated, considering the transformed error coordinates $\boldsymbol{\varepsilon}_{vq} = W\boldsymbol{\varepsilon} = [\varepsilon_v, \varepsilon_q]^T$, as

$$\|\boldsymbol{\varepsilon}_{vq}\|_1 < \sqrt{2} \tag{58}$$

where $\|\cdot\|_1$ is the ℓ_1 vector norm. This condition can be derived from the inverse transformation $\boldsymbol{\varepsilon} = W^{-1}\boldsymbol{\varepsilon}_{vq}$

$$\boldsymbol{\varepsilon} = \frac{\delta_1\delta_2}{\sqrt{2}} \begin{bmatrix} \frac{1}{\delta_2} & \frac{1}{\delta_2} \\ -\frac{1}{\delta_1} & \frac{1}{\delta_1} \end{bmatrix} \begin{bmatrix} \varepsilon_v \\ \varepsilon_q \end{bmatrix} = \begin{bmatrix} \frac{\delta_1\left(\varepsilon_v + \varepsilon_q\right)}{\sqrt{2}} \\ \frac{\delta_2\left(\varepsilon_q - \varepsilon_v\right)}{\sqrt{2}} \end{bmatrix}$$

$$|\varepsilon_1| < \delta_1 \Leftrightarrow \frac{\delta_1\left|\varepsilon_v + \varepsilon_q\right|}{\sqrt{2}} < \delta_1 \Leftrightarrow |\varepsilon_v + \varepsilon_q| < \sqrt{2}$$

$$|\varepsilon_2| < \delta_2 \Leftrightarrow \frac{\delta_2|\varepsilon_q - \varepsilon_v|}{\sqrt{2}} < \delta_2 \Leftrightarrow |\varepsilon_q - \varepsilon_v| < \sqrt{2}$$

$$\left|\varepsilon_q \pm \varepsilon_v\right| < \sqrt{2} \Leftrightarrow \|\boldsymbol{\varepsilon}_{vq}\|_1 < \sqrt{2}. \tag{59}$$

Suppose, for the sake of contradiction, that $\bar{\lambda}_{\mathcal{D}} > \sqrt{2}$. Then, we can select $\hat{\mathcal{P}}, \mathcal{P} \in FPFS_{\mathcal{D}}$ and two points $\hat{\boldsymbol{z}} \in \hat{\mathcal{P}}$, $\boldsymbol{z}^* \in \mathcal{P}$, such that:

$$\hat{z}_v = z_v^* \wedge \left|\hat{z}_q - z_q^*\right| > \sqrt{2}. \tag{60}$$

Assume now that there is still a realization $\mathcal{R}$ such that property (3) holds, and denote $\tilde{\boldsymbol{z}}^* = \mathcal{R}(\hat{\boldsymbol{z}})$. Further denote $\boldsymbol{r} = \tilde{\boldsymbol{z}}^* - \boldsymbol{z}^*$ and $[r_v, r_q]^T = W\boldsymbol{r}$. According to (58), it shall hold that $|\hat{z}_v - \tilde{z}_v^*| + |\hat{z}_q - \tilde{z}_q^*| < \sqrt{2}$. Then, using (60) we obtain

$$\begin{aligned} \|\varepsilon_{vq}\|_1 = \left|\hat{z}_v - \tilde{z}_v^*\right| + \left|\hat{z}_q - \tilde{z}_q^*\right| &< \sqrt{2} \\ \left|\hat{z}_v - z_v^* - r_v\right| + \left|\hat{z}_q - z_q^* - r_q\right| &< \sqrt{2} \\ |r_v| + \sqrt{2} - |r_q| &< \sqrt{2} \\ |r_v| - |r_q| < 0 \Rightarrow |r_v| < |r_q| \Rightarrow \left|\tilde{z}_v^* - z_v^*\right| &< \left|\tilde{z}_q^* - z_q^*\right|. \end{aligned} \tag{61}$$

The inequality above implies

$$\tilde{z}_q^* > z_q^* + \left|\tilde{z}_v^* - z_v^*\right| \ \vee \ \tilde{z}_q^* < z_q^* - \left|\tilde{z}_v^* - z_v^*\right| \tag{62}$$

which, in either case, shows that the point $\tilde{z}^*$ is not admissible, since it violates the Lipschitz continuity property of the transformed PF (see Remark 1). In the nonrotated coordinates, it means that $\tilde{z}^*$ either dominates or is dominated by $\boldsymbol{z}^*$, thus completing the contradictory argument and showing that the condition $\bar{\lambda}_{\mathcal{D}} \leq \sqrt{2}$ is necessary for the existence of a realization that satisfies (3) $\forall \hat{\mathcal{P}}, \mathcal{P} \in FPFS_{\mathcal{D}}$ and, conversely, the latter condition is sufficient to have $\bar{\lambda}_{\mathcal{D}} \leq \sqrt{2}$. ■

Proof of Proposition 2: From Proposition 1, we need to have $V^{(\ell)} - |Q^{(\ell)}| \leq \sqrt{2}$, $\ell = 1, \ldots, M-1$. From the same result, we can see that the worst case with respect to the actual values of the samples is when $Q^{(\ell)} = 0, \forall \ell$, i.e., the rotated PF samples have all the same q-coordinate. Then, considering that the length of the whole domain is $V_a = z_v^{*(a2)} - z_v^{*(a1)}$, the condition (2) is satisfied if $M = \left\lceil V_a/\sqrt{2} \right\rceil + 1$ using evenly distributed points in the segment $[z_v^{*(a1)}, z_v^{*(a2)}]$ including the extremes. Note that any other allocation of samples would result in a larger worst case error. Finally, we rewrite V_a as a function of Δ_1, Δ_2, δ_1, δ_2 by applying the transformation W to the anchor points

$$\begin{aligned} V_a &= z_v^{*(a2)} - z_v^{*(a1)} \\ V_a &= \frac{\sqrt{2}}{2}\left(\frac{z_1^{*(a2)}}{\delta_1} - \frac{z_2^{*(a2)}}{\delta_2}\right) - \frac{\sqrt{2}}{2}\left(\frac{z_1^{*(a1)}}{\delta_1} - \frac{z_2^{*(a1)}}{\delta_2}\right) \\ \frac{V_a}{\sqrt{2}} &= \frac{1}{2}\left(\frac{\Delta_1}{\delta_1} + \frac{\Delta_2}{\delta_2}\right). \end{aligned} \tag{63}$$

■

Proof of Proposition 3: The first part of the result can be demonstrated as a corollary of Theorem 2: since the worst case error obtained by the central approximation is $(1/2)\bar{\lambda}_{\mathcal{D}}$, applying the same reasoning of the proof of Theorem 2 one can conclude that $\bar{\lambda}_{\mathcal{D}} \leq 2\sqrt{2}$ is necessary and sufficient for the robust error guarantees. The second part, pertaining to the number of samples, can be demonstrated with the same arguments as Proposition 2, considering that the value of $V^{(\ell)}$ shall now be smaller than $2\sqrt{2}$ for all $\ell = 1, \ldots, M-1$. ■

REFERENCES

[1] L. Tang, Y. Li, D. Bai, T. Liu, and L. C. Coelho, "Bi-objective optimization for a multi-period COVID-19 vaccination planning problem," *Omega*, vol. 110, Jul. 2022, Art. no. 102617.

[2] E. Demir, T. Bektaş, and G. Laporte, "The bi-objective pollution-routing problem," *Eur. J. Oper. Res.*, vol. 232, no. 3, pp. 464–478, Feb. 2014.

[3] S. Wang, M. Liu, F. Chu, and C. Chu, "Bi-objective optimization of a single machine batch scheduling problem with energy cost consideration," *J. Cleaner Prod.*, vol. 137, pp. 1205–1215, Nov. 2016.

[4] A. Rodríguez-Molina, E. Mezura-Montes, M. G. Villarreal-Cervantes, and M. Aldape-Pérez, "Multi-objective meta-heuristic optimization in intelligent control: A survey on the controller tuning problem," *Appl. Soft Comput.*, vol. 93, Aug. 2020, Art. no. 106342.

[5] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidisciplinary Optim.*, vol. 26, no. 6, pp. 369–395, Apr. 2004.

[6] I. Das and J. E. Dennis, "A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems," *Struct. Optim.*, vol. 14, no. 1, pp. 63–69, Aug. 1997.

[7] I. Das and J. E. Dennis, "Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems," *SIAM J. Optim.*, vol. 8, no. 3, pp. 631–657, Aug. 1998.

[8] A. Ismail-Yahaya and A. Messac, "Effective generation of the Pareto frontier using the normal constraint method," in *Proc. 40th AIAA Aerosp. Sci. Meeting Exhib.*, Jan. 2002, Paper AIAA 2002-0178.

[9] G. Chiandussi, M. Codegone, S. Ferrero, and F. E. Varesio, "Comparison of multi-objective optimization methodologies for engineering applications," *Comput. Math. Appl.*, vol. 63, no. 5, pp. 912–942, Mar. 2012.

[10] A. Aliano Filho, A. C. Moretti, M. V. Pato, and W. A. de Oliveira, "An exact scalarization method with multiple reference points for bi-objective integer linear optimization problems," *Ann. Oper. Res.*, vol. 296, nos. 1–2, pp. 35–69, Jan. 2021.

[11] M. T. M. Emmerich and A. H. Deutz, "A tutorial on multiobjective optimization: Fundamentals and evolutionary methods," *Natural Comput.*, vol. 17, no. 3, pp. 585–609, Sep. 2018.

[12] J. L. J. Pereira, G. A. Oliver, M. B. Francisco, S. S. Cunha, and G. F. Gomes, "A review of multi-objective optimization: Methods and algorithms in mechanical engineering problems," *Arch. Comput. Methods Eng.*, vol. 29, no. 4, pp. 2285–2308, Jun. 2022.

[13] M. Li, M. López-Ibáñez, and X. Yao, "Multi-objective archiving," *IEEE Trans. Evol. Comput.*, vol. 28, no. 3, pp. 696–717, Jun. 2024.
[14] J. L. Cohon, R. L. Church, and D. P. Sheer, "Generating multiobjective trade-offs: An algorithm for bicriterion problems," *Water Resour. Res.*, vol. 15, no. 5, pp. 1001–1010, Oct. 1979.
[15] I. Y. Kim and O. L. de Weck, "Adaptive weighted sum method for multiobjective optimization: A new method for Pareto front generation," *Struct. Multidisciplinary Optim.*, vol. 31, no. 2, pp. 105–116, Feb. 2006.
[16] K. Klamroth, J. Tind, and M. M. Wiecek, "Unbiased approximation in multicriteria optimization," *Math. Methods Oper. Res. (ZOR)*, vol. 56, no. 3, pp. 413–437, Jan. 2003.
[17] R. Bokrantz and A. Forsgren, "An algorithm for approximating convex Pareto surfaces based on dual techniques," *INFORMS J. Comput.*, vol. 25, no. 2, pp. 377–393, May 2013.
[18] S. Ruzika and M. M. Wiecek, "Approximation methods in multiobjective programming," *J. Optim. Theory Appl.*, vol. 126, no. 3, pp. 473–501, Sep. 2005.
[19] G. Eichfelder and L. Warnow, "Advancements in the computation of enclosures for multi-objective optimization problems," *Eur. J. Oper. Res.*, vol. 310, no. 1, pp. 315–327, Oct. 2023.
[20] G. Eichfelder and L. Warnow, "An approximation algorithm for multi-objective optimization problems using a box-coverage," *J. Global Optim.*, vol. 83, no. 2, pp. 329–357, Jun. 2022.
[21] A. Payne and E. Polak, "An interactive rectangle elimination method for biobjective decision making," *IEEE Trans. Autom. Control*, vol. AC-25, no. 3, pp. 421–432, Jun. 1980.
[22] A. N. Payne, "Efficient approximate representation of bi-objective trade-off sets," *J. Franklin Inst.*, vol. 330, no. 6, pp. 1219–1233, Nov. 1993.
[23] S. Parisi, M. Pirotta, and M. Restelli, "Multi-objective reinforcement learning through continuous Pareto manifold approximation," *J. Artif. Intell. Res.*, vol. 57, pp. 187–227, Oct. 2016.
[24] Y. Tian, L. Si, X. Zhang, K. C. Tan, and Y. Jin, "Local model-based Pareto front estimation for multiobjective optimization," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 1, pp. 623–634, Jan. 2023.
[25] H. Nakayama, Y. Yun, and M. Yoon, "Basic concepts of multi-objective optimization," in *Sequential Approximate Multiobjective Optimization Using Computational Intelligence*. Cham, Switzerland: Springer, 2009, pp. 1–15.
[26] F. Logist, B. Houska, M. Diehl, and J. Van Impe, "Fast Pareto set generation for nonlinear optimal control problems with multiple objectives," *Struct. Multidisciplinary Optim.*, vol. 42, no. 4, pp. 591–603, Oct. 2010.
[27] M. Milanese and C. Novara, "Set membership identification of nonlinear systems," *Automatica*, vol. 40, no. 6, pp. 957–975, Jun. 2004.
[28] J. Traub and J. F. Woźniakowski, *A General Theory of Optimal Algorithms*. New York, NY, USA: Academic, 1980.
[29] F. A. Potra and S. J. Wright, "Interior-point methods," *J. Comput. Appl. Math.*, vol. 124, no. 1, pp. 281–302, Dec. 2000.
[30] R. S. Burachik, C. Y. Kaya, and M. M. Rizvi, "A new scalarization technique to approximate Pareto fronts of problems with disconnected feasible sets," *J. Optim. Theory Appl.*, vol. 162, no. 2, pp. 428–446, Aug. 2014.
[31] A. Messac, A. Ismail-Yahaya, and C. A. Mattson, "The normalized normal constraint method for generating the Pareto frontier," *Struct. Multidisciplinary Optim.*, vol. 25, no. 2, pp. 86–98, Jul. 2003.
[32] R. Boffadossi. (2025). *ROBBO Demo Web Page*. [Online]. Available: https://www.sas-lab.deib.polimi.it/?p=1296
[33] R. Boffadossi, M. Leonesio, and L. Fagiano, "ROBBO: An efficient method for Pareto front estimation with guaranteed accuracy," 2025, *arXiv:2506.18004*.
[34] P. Campigotto, A. Passerini, and R. Battiti, "Active learning of Pareto fronts," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 506–519, Mar. 2014.
[35] T. Takagi, K. Takadama, and H. Sato, "Pareto front estimation using unit hyperplane," in *Evolutionary Multi-Criterion Optimization*. Cham, Switzerland: Springer, 2021, pp. 126–138.
[36] M. Binois, D. Ginsbourger, and O. Roustant, "Quantifying uncertainty on Pareto fronts with Gaussian process conditional simulations," *Eur. J. Oper. Res.*, vol. 243, no. 2, pp. 386–394, Jun. 2015.
[37] X. Lin, Z. Yang, Q. Zhang, and S. Kwong, "Controllable Pareto multi-task learning," 2020, *arXiv:2010.06313*.
[38] A. Navon, A. Shamsian, G. Chechik, and E. Fetaya, "Learning the Pareto front with hypernetworks," 2020, *arXiv:2010.04104*.
[39] L. P. Hoang, D. D. Le, T. A. Tuan, and T. N. Thang, "Improving Pareto front learning via multi-sample hypernetworks," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, no. 7, pp. 7875–7883.
[40] P. Huang et al., "Learning to generate preferences for multiobjective deep learning," *IEEE Trans. Ind. Informat.*, early access, Apr. 16, 2026, doi: 10.1109/TII.2026.3680097.
[41] T. A. Tuan, L. P. Hoang, D. D. Le, and T. N. Thang, "A framework for controllable Pareto front learning with completed scalarization functions and its applications," *Neural Netw.*, vol. 169, pp. 257–273, Jan. 2024.
[42] T. A. Tuan, N. V. Dung, and T. N. Thang, "A hyper-transformer model for controllable Pareto front learning with split feasibility constraints," *Neural Netw.*, vol. 179, Nov. 2024, Art. no. 106571.
[43] R. Ye, L. Chen, W.-B. Kou, J. Zhang, and H. Ishibuchi, "Pareto front shape-agnostic Pareto set learning in multi-objective optimization," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2024, pp. 4730–4736.

Roberto Boffadossi received the M.Sc. degree in automation and control engineering from the Politecnico di Milano, Milan, Italy, in 2020, where he is currently pursuing the Ph.D. degree with the Department of Electronics, Information, and Bioengineering.

From 2021 to 2025, he was a Research Assistant with the Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, Italian National Research Council, Milan. In 2025, he was a Visiting Researcher with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K. His research interests include multiobjective optimization, discrete manufacturing, and decision support systems.

Marco Leonesio (Member, IEEE) received the M.Sc. degree in mechanical engineering with a specialization in automation and robotics from the Politecnico di Milano, Milan, Italy, in 2000.

Since 2008, he has been a Permanent Researcher at the Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, Italian National Research Council, Milan. His research interests include multibody modeling and simulation, geometrical calibration of machine tools and parallel manipulators, mechanism analysis, metal cutting analysis and simulation, process monitoring, and control exploiting machine learning techniques.

Lorenzo Fagiano (Senior Member, IEEE) received the Ph.D. degree in information and systems engineering from the Politecnico di Turin, Turin, Italy, in 2009.

He held positions at UC Santa Barbara, Santa Barbara, CA, USA, ETH Zürich, Zürich, Switzerland, and ABB Corporate Research, Baden, Switzerland. Since 2016, he has been with the Politecnico di Milano, Milan, Italy, where he is currently a Full Professor of Automation and Control Engineering. His research interests include constrained estimation and control, set membership methods, and applications to industrial, robotic, and energy systems.

Dr. Fagiano was a recipient of the 2019 European Control Award, the Mission Innovation Champion Award 2019 for Italy, and the 2010 ENI Award Debut in Research Prize.

Open Access funding provided by 'Politecnico di Milano' within the CRUI CARE Agreement