Constrained stochastic blackbox optimization using a progressive barrier and probabilistic estimates

Kwassi Joseph Dzahini[*] Michael Kokkolaras[†] Sébastien Le Digabel[*]

November 9, 2020

Abstract: This work introduces the StoMADS-PB algorithm for constrained stochastic blackbox optimization, which is an extension of the mesh adaptive direct-search (MADS) method originally developed for deterministic blackbox optimization under general constraints. The values of the objective and constraint functions are provided by a noisy blackbox, i.e., they can only be computed with random noise whose distribution is unknown. As in MADS, constraint violations are aggregated into a single constraint violation function. Since all functions values are numerically unavailable, StoMADS-PB uses estimates and introduces so-called probabilistic bounds for the violation. Such estimates and bounds obtained from stochastic observations are required to be accurate and reliable with high but fixed probabilities. The proposed method, which allows intermediate infeasible iterates, accepts new points using sufficient decrease conditions and imposing a threshold on the probabilistic bounds. Using Clarke nonsmooth calculus and martingale theory, Clarke stationarity convergence results for the objective and the violation function are derived with probability one.

[*] GERAD and Département de Mathématiques et de Génie Industriel, Polytechnique Montréal, C.P. 6079, Succ. Centre-ville, Montréal, Québec H3C 3A7, Canada (www.gerad.ca/fr/people/kwassi-joseph-dzahini, www.gerad.ca/Sebastien.Le.Digabel).

[†] GERAD and McGill University, Mechanical Engineering Department, 845 Rue Sherbrooke Ouest, Montréal, Québec H3A 0G4, Canada (www.mcgill.ca/mecheng/people/staff/michael-kokkolaras).

1 Introduction

Blackbox optimization (BBO) considers the development and analysis of algorithms designed for objectives and constraints functions that are given by a process called a blackbox which returns an output when provided an input but whose inner workings are analytically unavailable [12]. Mesh adaptive direct-search (MADS) [7, 8] with progressive barrier (PB) is an algorithm for deterministic BBO. This work considers the following constrained stochastic BBO problem

$$\min_{x\in\mathcal{D}} f(x) \tag{1}$$

where $\mathcal{D} = \{x \in \mathcal{X} : c(x) \leq 0\} \subset \mathbb{R}^n$ is the feasible region, $c = (c_1, c_2, \dots, c_m)^\top$, $\mathcal{X}$ is a subset of $\mathbb{R}^n$, $f(x) = \mathbb{E}_{\Theta_0}\left[f_{\Theta_0}(x)\right]$ with $f\colon \mathcal{X} \mapsto \mathbb{R}$, and $c_j(x) = \mathbb{E}_{\Theta_j}\left[c_{\Theta_j}(x)\right]$ with $c_j\colon \mathcal{X} \mapsto \mathbb{R}$ for all $j \in J := \{1, 2, \dots, m\}$. $\mathbb{E}_{\Theta_j}$ denotes the expectation with respect to the random variable Θ_j for all $j \in J \cup \{0\}$, which are supposed to be independent with unknown possibly different distributions. $f_{\Theta_0}(\cdot)$ denotes the noisy computable version of the numerically unavailable objective function $f(\cdot)$, while for all $j \in J$, $c_{\Theta_j}(\cdot)$ denotes the noisy computable version of the numerically unavailable constraint $c_j(\cdot)$. Note that the noisy objective function f_{Θ_0} and the constraints $c_{\Theta_j}, j \in J$, are typically the outputs of a blackbox. By means of some useful terminology, constraints that must always be satisfied, such as those defining $\mathcal{X}$, are differentiated from those that need only to be satisfied at the solution, such as $c(x) \leq 0$. The former will be called *unrelaxable* non-quantifiable constraints and the latter, *relaxable* quantifiable constraints [41].

Solving stochastic blackbox optimization problems such as Problem (1), which often arise in signal processing and machine learning [27], has recently been a topic of intense research. Most methods for solving such problems borrow ideas from the stochastic gradient method [49]. Several works have also attempted to transfer ideas from deterministic DFO methods to the stochastic context. However, most of such proposed methods are restricted to unconstrained optimization. Indeed, after [18] which is among the first to propose a stochastic variant of the deterministic Nelder-Mead (NM) method [47], [3] also considered the optimization of functions whose evaluations are subject to random noise and proposed an algorithm which is shown to have convergence properties, based on Markov chain theory [32]. Another stochastic variant of NM was recently proposed in [22] and was proved to have global convergence properties with probability one. Using elements from [17, 40], [23] proposed STORM, a trust-region algorithm designed for stochastic optimization problems, with almost sure global convergence results. Many other researches that extend the traditional deterministic trust-region method to stochastic setting have been conducted in [28, 52]. In [48], a classical backtracking Armijo line search method [5] has been adapted to the stochastic optimization setting and was shown to have first-order complexity bounds. Robust-MADS, a kernel smoothing-based variant of MADS [7], was proposed in [13] to approach the minimizer of an objective function whose values can only be computed with a random noise. It was shown to possess zeroth-order [9] convergence properties. Another stochastic variant of MADS was proposed in [2] for BBO, where the noise corrupting the blackbox was supposed to be Gaussian. Convergence results of the proposed method have been derived, making use of statistical inference techniques. [11] proposed another stochastic optimization approach using an algorithmic framework similar to that of MADS. StoMADS uses estimates of function values obtained from stochastic observations. By assuming that such estimates satisfy a variance condition and are sufficiently accurate with a large but fixed probability conditioned to the past, a Clarke [25] stationarity convergence result of StoMADS has been derived with probability one, using martingale theory. A general framework for stochastic directional direct-search [26] methods was introduced in [33] with expected complexity analysis.

All the above stochastic optimization methods are restricted to unconstrained problems and most of them use estimated gradient information when seeking for an optimal solution. When the gradient does not exist or is computationally expensive to estimate, heuristics such as simulated annealing methods, genetic algorithms [39], and tabu/scatter search [38], are also used for problems with noisy constraints but do not present any convergence theory. Surrogate model based methods for constrained stochastic BBO have also been a topic of intense research, including the response surface methodology with stochastic constraints [4] developed for expensive simulation. In [16], the capabilities of the deterministic constrained trust-region algorithm NOWPAC [15] are generalized for the optimization of blackboxes with inherently noisy evaluations of the objective and constraint functions. To mitigate the noise in the latter functions evaluations, the resulting gradient-free method SNOWPAC utilizes Gaussian process surrogate combined with local fully linear surrogate models. Another surrogate-based approach that has gained in increasing popularity in various research fields is Kriging, also known as Bayesian optimization [45]. Various Bayesian optimization methods for constrained stochastic BBO have been demonstrated to be efficient in practice [42, 54].

Developing direct-search methods for BBO has received renewed interest since such methods generally known to be reliable and robust in practice [6], appear to be the most promising approach in most of real applications where the gradient does not exist or is computationally expensive to estimate. However, there is relatively scarce research on developing direct-search methods for constrained stochastic BBO, especially when noise is present in the constraint functions. A pattern search and implicit filtering algorithm (PSIFA) [29, 30] was recently developed for linearly constrained problems with a noisy objective function, and was shown to have global convergence properties. A class of direct-search methods for solving smooth linearly constrained problems was also studied in [34] but even though using a probabilistic feasible descent based approach, this work assumes the objective and constraints function values to be exactly computed without noise.

The present work introduces StoMADS-PB, a stochastic variant of the mesh adaptive direct-search with progressive barrier [8], using elements from [7, 8, 11, 17, 23, 48] and is, to the best of our knowledge, the first to propose a directional direct-search [26] stochastic BBO algorithm, capable to handle general noisy constraints without requiring any feasible initial point. Its main contribution is the analysis of the resulting new framework with fully supported theoretical results. StoMADS-PB uses no gradient information to find descent directions or improve feasibility compared to prior work. Rather, it uses so-called probabilistic estimates [23] of the objective and constraint function values and also introduces probabilistic bounds on a constraint violation function values. The reliability of such bounds is assumed to hold with a high but fixed probability. Moreover, although no distributions are assumed for the estimates and no assumption is made about the way they are generated, they are required to be sufficiently accurate with large but fixed probabilities and satisfy some variance conditions.

The manuscript is organized as follows. Section 2 presents the general framework of the proposed StoMADS-PB algorithm. Section 3 explains how the proposed method results in a stochastic process and discusses requirements on random estimates to guarantee convergence. It also shows how such estimates can be constructed in practice. Section 4 presents the main convergence results. Computational results are reported in Section 5 followed by a discussion and suggestions for future work. Additional results are provided as an annex.

2 The StoMADS-PB algorithm

StoMADS-PB is based on an algorithmic framework similar to that of MADS with PB [8]. For the needs of the convergence analysis of Section 4, deterministic constraint violations are aggregated into a single function h called the constraint violation function, defined using the ℓ_1-norm for needs of convergence studies as opposed to [8] where an ℓ_2-norm has been favored

$$h(x) := \begin{cases} \displaystyle\sum_{j=1}^{m} \max\{c_j(x), 0\} & \text{if } x \in \mathcal{X} \\ +\infty & \text{otherwise.} \end{cases}$$

According to this definition, $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $x \in \mathcal{D}$, i.e., x is feasible with respect to the relaxable constraints if and only if $h(x) = 0$. Moreover, if $0 < h(x) < +\infty$, then x is called infeasible and satisfies the unrelaxable constraints but not the relaxable ones.

In MADS with PB, feasibility improvement is achieved by decreasing h, specifically by comparing its function value at a current point x^k to that of a trial point $x^k + s^k$, where s^k denotes a direction around x^k. Likewise, to decrease f, MADS with PB uses objective function values since they are available in the deterministic setting.

The main challenge here is to guarantee for StoMADS-PB such decreases as well in f as in h whereas their function values are unavailable numerically, using only information provided by the noisy blackbox outputs f_{Θ_0} and c_{Θ_j}, $j \in J$. This section shows how this can be achieved, making use of so called ε-accurate estimates introduced in [23] and then presents the general framework of the proposed method.

2.1 Feasibility and objective function improvements

At iteration k, let x^k and $x^k + s^k$ be two points of $\mathcal{X}$. Since the constraint function values $c_j(x^k)$ and $c_j(x^k + s^k)$, $j \in J = \{1, 2, \dots, m\}$, are numerically unavailable, their corresponding estimates are respectively constructed using evaluations of the noisy blackbox outputs c_{Θ_j}, $j \in J$. In general for the remainder of the manuscript, unless otherwise stated, given a function $g : \mathcal{X} \to \mathbb{R}$, an estimate of $g(x^k)$ is denoted by $g_0^k(x^k)$ (or simply by g_0^k if there is no ambiguity) while that of $g(x^k + s^k)$ is denoted by $g_s^k(x^k{+}s^k)$ or g_s^k. In StoMADS-PB, the violations of the estimates $c_{j,0}^k(x^k)$ and $c_{j,s}^k(x^k{+}s^k)$ of $c_j(x^k)$ and $c_j(x^k + s^k)$, respectively, are aggregated in so-called *estimated violations* $h_0^k(x^k)$ and $h_s^k(x^k + s^k)$ defined as follows

$$h_0^k(x^k) \quad = \quad \begin{cases} \displaystyle\sum_{j=1}^{m} \max\left\{c_{j,0}^k(x^k), 0\right\} & \text{if } x^k \in \mathcal{X} \\ +\infty & \text{otherwise} \end{cases} \tag{2}$$

$$\text{and} \quad h_s^k(x^k + s^k) \quad = \quad \begin{cases} \displaystyle\sum_{j=1}^{m} \max\left\{c_{j,s}^k(x^k + s^k), 0\right\} & \text{if } x^k + s^k \in \mathcal{X} \\ +\infty & \text{otherwise.} \end{cases} \tag{3}$$

In order for such estimated constraint violations to be reliable enough to determine whether $h(x^k + s^k) < h(x^k)$ or not, the estimates $c_{j,0}^k(x^k)$ and $c_{j,s}^k(x^k + s^k)$ need to be sufficiently accurate. The following definition similar to that of [11] is adapted from [23].

Definition 1. *Let $\varepsilon > 0$ be a fixed constant and $\{\delta_p^k\}_{k\in\mathbb{N}}$ be a sequence of nonnegative real numbers. For a given function $g\colon \mathcal{X} \mapsto \mathbb{R}$ and $y^k \in \mathcal{X}$, let g^k be an estimate of $g(y^k)$. Then g^k is said to be an ε-accurate estimate of $g(y^k)$ for the given δ_p^k, if*

$$\left|g^k - g(y^k)\right| \leq \varepsilon(\delta_p^k)^2.$$

As in [11], the role of δ_p^k will be played by the so-called *poll size* parameter introduced in Section 2.2. The following result provides bounds on $h(x^k)$ and $h(x^k + s^k)$, respectively, which will allow, in Proposition 2, to guarantee a decrease in the constraint violation function h by means of a sufficient decrease condition on the estimated violations h_0^k and h_s^k.

Proposition 1. *Let $c_{j,0}^k$ and $c_{j,s}^k$ be ε-accurate estimates of $c_j(x^k)$ and $c_j(x^k + s^k)$, respectively, with x^k and $x^k + s^k \in \mathcal{X}$. Then the followings hold:*

$$\ell_0^k(x^k) := \sum_{j=1}^{m} \max\left\{c_{j,0}^k - \varepsilon(\delta_p^k)^2, 0\right\} \leq h(x^k) \leq \sum_{j=1}^{m} \max\left\{c_{j,0}^k + \varepsilon(\delta_p^k)^2, 0\right\} =: u_0^k(x^k) \qquad (4)$$

and

$$\ell_s^k(x^k+s^k) := \sum_{j=1}^{m} \max\left\{c_{j,s}^k - \varepsilon(\delta_p^k)^2, 0\right\} \leq h(x^k+s^k) \leq \sum_{j=1}^{m} \max\left\{c_{j,s}^k + \varepsilon(\delta_p^k)^2, 0\right\} =: u_s^k(x^k+s^k)$$

Proof. The result is shown for $h(x^k)$ but the proof for $h(x^k + s^k)$ is the same. Since $c_{j,0}^k$ is an ε-accurate estimate of $c_j(x^k)$ for all $j \in J$, then it follows from Definition 1 that

$$c_{j,0}^k - \varepsilon(\delta_p^k)^2 \leq c_j(x^k) \leq c_{j,0}^k + \varepsilon(\delta_p^k)^2, \quad \text{for all } j \in J,$$

which implies that

$$\max\left\{c_{j,0}^k - \varepsilon(\delta_p^k)^2, 0\right\} \leq \max\left\{c_j(x^k), 0\right\} \leq \max\left\{c_{j,0}^k + \varepsilon(\delta_p^k)^2, 0\right\}. \qquad (5)$$

Finally, summing each term of (5) from $j = 1$ to m leads to (4). □

Definition 2. *The estimates $\ell_0^k(x^k)$ and $u_0^k(x^k)$ of* Proposition 1*, satisfying $\ell_0^k(x^k) \leq h(x^k) \leq u_0^k(x^k)$, are said to be ε-reliable bounds for $h(x^k)$. Similarly, the estimates $\ell_s^k(x^k + s^k)$ and $u_s^k(x^k + s^k)$ satisfying $\ell_s^k(x^k + s^k) \leq h(x^k + s^k) \leq u_s^k(x^k + s^k)$ are said to be ε-reliable bounds for $h(x^k + s^k)$.*

The following result provides sufficient information to identify a decrease in h and will be also useful to determine an iteration type in Section 2.2.

Proposition 2. *Let $\ell_0^k(x^k)$ and $u_0^k(x^k)$ be ε-reliable bounds for $h(x^k)$, and let $\ell_s^k(x^k+s^k)$ and $u_s^k(x^k+s^k)$ be ε-reliable bounds for $h(x^k + s^k)$. Let h_0^k and h_s^k be the estimated constraint violations at x^k and $x^k + s^k \in \mathcal{X}$, respectively. Let $\gamma > 2$ be a constant. Then the following holds:*

$$\text{if } h_s^k - h_0^k \leq -\gamma m\varepsilon(\delta_p^k)^2, \text{ then } h(x^k + s^k) - h(x^k) \leq -(\gamma - 2)m\varepsilon(\delta_p^k)^2 < 0. \qquad (6)$$

Proof. It follows from Proposition 1 that

$$h(x^k+s^k)-h(x^k)\leq\sum_{j=1}^{m}\max\left\{c_{j,s}^k+\varepsilon(\delta_p^k)^2,0\right\}-\sum_{j=1}^{m}\max\left\{c_{j,0}^k-\varepsilon(\delta_p^k)^2,0\right\}. \tag{7}$$

By noticing that

$$\sum_{j=1}^{m}\max\left\{c_{j,s}^k+\varepsilon(\delta_p^k)^2,0\right\}\leq\sum_{j=1}^{m}\max\left\{c_{j,s}^k,0\right\}+m\varepsilon(\delta_p^k)^2=h_s^k+m\varepsilon(\delta_p^k)^2$$

$$\sum_{j=1}^{m}\max\left\{c_{j,0}^k-\varepsilon(\delta_p^k)^2,0\right\}\geq\sum_{j=1}^{m}\max\left\{c_{j,0}^k,0\right\}-m\varepsilon(\delta_p^k)^2=h_0^k-m\varepsilon(\delta_p^k)^2,$$

then it follows from (7) that

$$h(x^k+s^k)-h(x^k)\leq h_s^k-h_0^k+2m\varepsilon(\delta_p^k)^2\leq-(\gamma-2)m\varepsilon(\delta_p^k)^2,$$

where the last inequality follows from the assumption that $h_s^k-h_0^k\leq-\gamma m\varepsilon(\delta_p^k)^2$. The proof is complete by noticing that $\gamma>2$. □

As in [8], the present research also introduces a nonnegative barrier threshold $h_{\max}^k=u_0^k(x_{\text{inf}}^k)$, where x_{inf}^k is a so-called ε-infeasible solution. Definition 3 presents ε-infeasible points and the updating rules of x_{inf}^k is presented in Section 2.2. While x_{inf}^k is updated at the end of each iteration of StoMADS-PB, $h_{\max}^k$ is rather computed at the beginning of iterations in order to avoid keeping its possibly inaccurate values from one iteration to another. In fact, estimates in StoMADS-PB are always computed at the beginning of the iterations and their accuracy is improved compared to previous iterations as seen in Section 3.2. Consequently, even though the sequence $\{h_{\max}^k\}_{k\in\mathbb{N}}$ has a globally decreasing tendency, it is not nonincreasing as in MADS with PB, but can possibly increase between successive iterations. The goal of StoMADS-PB is to accept only the trial points satisfying $h(x^k)\leq h_{\max}^k$, and any trial point x^k for which the inequality $u_0^k(x^k)\leq h_{\max}^k$ does not hold is discarded from consideration since such an inequality implies that $h(x^k)\leq h_{\max}^k$ due to (4). However, this is a sufficient acceptance condition since $u_0^k(x^k)>h_{\max}^k$ does not necessarily imply that $h(x^k)\leq h_{\max}^k$ does not hold, but rather leads to a situation of uncertainty which is not explicitly distinguished in the present manuscript for the sake of simplicity.

The ε-reliable upper bound $u_0^k(x^k)$ previously obtained for $h(x^k)$ also allows to determine the feasibility with respect to the relaxable constraints of a given trial point $x^k\in\mathcal{X}$. Indeed, it obviously follows from (4) that $h(x^k)=0$ if $u_0^k(x^k)=0$, which is satisfied provided that $c_{j,0}^k(x^k)\leq-\varepsilon(\delta_p^k)^2$, for all $j\in J$. This means that in order for $h(x^k)=0$ to hold, all the estimates of constraint function values must be sufficiently negative and not simply zero. By means of the following definition, StoMADS-PB partitions the trial points into so-called *ε-feasible* and *ε-infeasible* points.

Definition 3. *Let $x^k\in\mathcal{X}$ be any trial point and $u_0^k(x^k)$ be an ε-reliable upper bound for $h(x^k)$. Then x^k is called ε-feasible if $u_0^k(x^k)=0$, and it is called ε-infeasible if $0<u_0^k(x^k)\leq h_{\max}^k$. Similarly, $x^k+s^k\in\mathcal{X}$ is called ε-feasible if $u_s^k(x^k+s^k)=0$, and it is called ε-infeasible if* $0<u_s^k(x^k+s^k)\leq h_{\max}^k$.

StoMADS-PB does not require that the starting point is ε-feasible. The algorithm can be applied to any problem satisfying only the following assumption adapted from [8].

Assumption 1. *There exists some point $x^0 \in \mathcal{X}$ such that $f_0^0(x^0)$ and $u_0^0(x^0)$ are both finite, and $u_0^0(x^0) \leq h_{\max}^0$.*

The next result similar to that in [11] provides a sufficient information to identify a decrease in f and also allows to determine an iteration type in Section 2.2.

Proposition 3. *Let f_0^k and f_s^k be ε-accurate estimates of $f(x^k)$ and $f(x^k+s^k)$, respectively, for x^k and $x^k+s^k \in \mathcal{X}$. Let $\gamma > 2$ be a constant. Then the following holds:*

$$\textit{if } f_s^k - f_0^k \leq -\gamma\varepsilon(\delta_p^k)^2, \textit{ then } f(x^k+s^k) - f(x^k) \leq -(\gamma-2)\varepsilon(\delta_p^k)^2 < 0. \tag{8}$$

Proof. The proof follows from Definition 1 and the next equality

$$f(x^k+s^k) - f(x^k) = f(x^k+s^k) - f_s^k + \left(f_s^k - f_0^k\right) + f_0^k - f(x^k).$$

□

2.2 The StoMADS-PB algorithm and parameter update

Recall first that MADS with PB is an iterative algorithm where every iteration comprises two main steps: an optional step called the SEARCH, and the POLL. The SEARCH which typically consists of a global exploration may use a plethora of strategies like those based on interpolatory models, heuristics and surrogate functions or simplified physics models [8] to explore the variables space. Each iteration of StoMADS-PB can also allow a SEARCH step, but it is not shown here for simplicity. Similarly to MADS with PB, the POLL step of StoMADS-PB is more rigidly defined unlike the freedom of the SEARCH and consists of a local exploration. During each of these two steps, a finite number of trial points is generated on an underlying *mesh* $\mathcal{M}^k$. The mesh is a discretization of the variables space, whose coarseness or fineness is controlled by a mesh size parameter δ_m^k thus deviating from the notation Δ_k^m from [8], since uppercase letters will be used to denote random variables. For the remainder of the manuscript, $s^k = \delta_m^k d^k$ where d^k is a nonzero direction around $x^k \in \mathcal{M}^k$. The POLL step is governed by the poll size parameter δ_p^k which is linked to δ_m^k by $\delta_m^k = \min\{\delta_p^k, (\delta_p^k)^2\}$ [12]. As specified earlier, $\{\delta_p^k\}_{k\in\mathbb{N}}$ will play the role of the sequence of nonnegative real numbers introduced in Definition 1. Let $\hat{z} \in \mathbb{N}$ be a large fixed integer and $\tau \in (0,1) \cap \mathbb{Q}$ be a fixed rational constant. For the needs of Section 4, note also that as in [11], δ_p^k is supposed to be bounded above by the positive and fixed constant $\tau^{-\hat{z}}$ in order for the random poll size parameter Δ_p^k introduced in Section 3.1 to be integrable. The definitions of the mesh $\mathcal{M}^k$ and the POLL set $\mathcal{P}^k$ inspired from [8] are given next.

Definition 4. *Let $\mathbf{D} \in \mathbb{R}^{n\times p}$ be a matrix, with columns denoted by the set $\mathbb{D}$ which form a positive spanning set. At the beginning of iteration k, let x_{inf}^k and x_{feas}^k denote respectively the ε-infeasible and the ε-feasible incumbent solutions (there might be only one), and let $\mathcal{V}^k := \{x_{\text{inf}}^k, x_{\text{feas}}^k\}$ be the set of such incumbents. The mesh $\mathcal{M}^k$ and the POLL set $\mathcal{P}^k$ are respectively*

$$\mathcal{M}^k := \{x^k + \delta_m^k d : x^k \in \mathcal{V}^k,\ d = \mathbf{D}y,\ y \in \mathbb{Z}^p\} \quad \textit{and} \quad \mathcal{P}^k := \mathcal{P}^k(x_{\text{inf}}^k) \cup \mathcal{P}^k(x_{\text{feas}}^k),$$

where $\forall x^k \in \mathcal{M}^k \cap \mathcal{X}$, $\mathcal{P}^k(x^k) = \{x^k + \delta_m^k d^k \in \mathcal{M}^k \cap \mathcal{X} : \delta_m^k \left\|d^k\right\|_\infty \leq \delta_p^k b,\ d^k \in \mathbb{D}_p^k(x^k)\}$ is called a frame around x^k, with $b = \max\{\|d'\|_\infty, d' \in \mathbb{D}\}$. $\mathbb{D}_p^k(x^k)$ is a positive spanning set which is said to be a set of frame directions around x^k. The set $\mathbb{D}_p^k$ of all polling directions at iteration k is defined by $\mathbb{D}_p^k := \mathbb{D}_p^k(x_{\text{inf}}^k) \cup \mathbb{D}_p^k(x_{\text{feas}}^k)$. When there is no incumbent ε-feasible solution x_{feas}^k, then the set $\mathcal{V}^k$ is reduced to $\{x_{\text{inf}}^k\}$, in which case $\mathcal{P}^k = \mathcal{P}^k(x_{\text{inf}}^k)$ and $\mathbb{D}_p^k = \mathbb{D}_p^k(x_{\text{inf}}^k)$.

After the POLL step is completed, StoMADS-PB computes not only estimates f_0^k, f_s^k, h_0^k and h_s^k of $f(x^k)$, $f(x^k+s^k)$, $h(x^k)$ and $h(x^k+s^k)$, respectively at trial points $x^k \in \mathcal{V}^k$ and $x^k+s^k \in \mathcal{P}^k$, but also upper bounds $u_s^k(x^k+s^k)$ and $u_0^k(x_{\text{inf}}^k)$, respectively for $h(x^k+s^k)$ and $h(x_{\text{inf}}^k)$. The values of such estimates and bounds determine the iteration type of the algorithm and govern also the way δ_p^k is updated. Recall Definition 3 of ε-feasible and ε-infeasible points at the beginning of iteration k. The incumbent solutions x_{inf}^k and x_{feas}^k are constructed by ranking trial mesh points of $\mathcal{X}$, making use of the dominance notion inspired from [8].

Definition 5. *The ε-feasible point x^k+s^k is said to dominate the ε-feasible point x^k, denoted $x^k+s^k \prec_{f;\varepsilon} x^k$, when $f_s^k - f_0^k \le -\gamma\varepsilon(\delta_p^k)^2$, with $u_s^k(x^k+s^k)=0$.*
The ε-infeasible point x^k+s^k is said to dominate the ε-infeasible point x^k, denoted $x^k+s^k \prec_{h;\varepsilon} x^k$, when $f_s^k - f_0^k \le -\gamma\varepsilon(\delta_p^k)^2$ and $h_s^k - h_0^k \le -\gamma m\varepsilon(\delta_p^k)^2$, with $0 < u_s^k(x^k+s^k) \le h_{\text{max}}^k$.

Adapting the terminologies from [8] and depending on the values of the aforementioned estimates and bounds, there are four StoMADS-PB iterations types: an iteration can be either f-Dominating, h-Dominating (the former and the latter are referred to as dominating iterations), Improving, or Unsuccessful. During a dominating iteration, either the algorithm has found a first ε-feasible iterate or a trial point that dominates an incumbent is generated. An iteration which is Improving is not dominating but it aims to improve the feasibility of the ε-infeasible incumbent. Unsuccessful iterations are neither dominating nor improving.

- At the beginning of iteration k, if there is no available ε-feasible solution, then the iteration is called f-Dominating if for $x^k \in \mathcal{V}^k$, a first trial point $x^k+s^k \in \mathcal{P}^k$ satisfying $u_s^k(x^k+s^k)=0$ is found, in which case $h(x^k+s^k)=0$ due to Proposition 1, meaning that x^k+s^k is ε-feasible. Otherwise, if an ε-feasible point that dominates the incumbent is generated, i.e., $x^k+s^k \prec_{f;\varepsilon} x_{\text{feas}}^k$ for some $x^k \in \mathcal{V}^k$, then the inequality $f_s^k(x^k+s^k) - f_0^k(x_{\text{feas}}^k) \le -\gamma\varepsilon(\delta_p^k)^2$ leads to a decrease in f due to Proposition 3. In either case, $x_{\text{feas}}^{k+1} := x^k+s^k$ and $\delta_p^{k+1} = \min\{\tau^{-1}\delta_p^k, \tau^{-\hat{z}}\}$. The ε-infeasible incumbent x_{inf}^k is not updated since there is no feasibility improvement.

- Iteration k is said to be h-Dominating whenever an ε-infeasible point that dominates the incumbent is generated, i.e., $x_{\text{inf}}^k + s^k \prec_{h;\varepsilon} x_{\text{inf}}^k$, which means that both inequalities $f_s^k(x_{\text{inf}}^k+s^k) - f_0^k(x_{\text{inf}}^k) \le -\gamma\varepsilon(\delta_p^k)^2$ and $h_s^k(x_{\text{inf}}^k+s^k) - h_0^k(x_{\text{inf}}^k) \le -\gamma m\varepsilon(\delta_p^k)^2$ hold. Consequently, it follows from Propositions 2 and 3 that decreases occur both in f and h. In this case, $x_{\text{feas}}^{k+1} = x_{\text{feas}}^k$ and since feasibility is improved, x_{inf}^{k+1} is set to equal $x_{\text{inf}}^k + s^k$ while the poll size parameter is updated as at f-Dominating iterations.

- Iteration k is said to be Improving if it is not dominating but there is at least one ε-infeasible point $x_{\text{inf}}^k+s^k$ satisfying $h_s^k(x_{\text{inf}}^k+s^k) - h_0^k(x_{\text{inf}}^k) \le -\gamma m\varepsilon(\delta_p^k)^2$. Indeed, this means that $x_{\text{inf}}^k+s^k$ improves the feasibility of the ε-infeasible incumbent x_{inf}^k since the previous inequality leads to a decrease in h due to Proposition 2. In this case, δ_p^k is updated as in dominating iterations, $x_{\text{feas}}^{k+1} = x_{\text{feas}}^k$ while the ε-infeasible incumbent is updated according to

$$x_{\text{inf}}^{k+1} \in \underset{x_{\text{inf}}^k+s^k}{\operatorname{argmin}} \left\{ u_s^k(x_{\text{inf}}^k+s^k) : h_s^k(x_{\text{inf}}^k+s^k) - h_0^k(x_{\text{inf}}^k) \le -\gamma m\varepsilon(\delta_p^k)^2 \right\}.$$

- Finally, an iteration is called Unsuccessful if it is neither dominating nor Improving. In this case, $\delta_p^{k+1} = \tau\delta_p^k$ while neither x_{inf}^k nor x_{feas}^k are updated.

Remark 1. *Denote by $t > 0$ the number of the first f-Dominating iteration of Algorithm 1 and assume that $t < +\infty$. Then it is easy to notice that $x^k_{\text{feas}} = x^0_{\text{inf}}$ for all $k = 0, 1, \dots, t$ while $x^{t+1}_{\text{feas}} \neq x^0_{\text{inf}}$. Moreover, even though estimates $f^k_0(x^k_{\text{feas}})$, $f^k_s(x^k_{\text{feas}} + s^k)$, $h^k_0(x^k_{\text{feas}})$ and $h^k_s(x^k_{\text{feas}} + s^k)$ are computed at x^k_{feas} and $x^k_{\text{feas}} + s^k \in \mathcal{P}^k$ respectively for all $k \leq t$, they are not used by the algorithm until the end of iteration t and it can also be noticed that no point in $\mathcal{P}^k$ that is generated using $\mathbb{D}^k_p(x^k_{\text{feas}})$ is evaluated until the end of iteration t. In fact, setting the initial ε-feasible guess to equal x^0_{inf} as it is in Algorithm 1 and then computing the latter estimates are not necessary in practice. However, doing so allows simply the aforementioned estimates to be defined for all $k \geq 0$ for theoretical needs, specifically the construction of the σ-algebra $\mathcal{F}^{C\cdot F}_{k-1}$ in Section 3.*

2.3 Frame center selection rule

Before describing the frame center selection rule, recall the set $\mathcal{V}^k$ of incumbent solutions introduced in Definition 4 and the fact that POLL trial points are generated inside frames around such incumbents At a given iteration, there are either one or two frame centers in $\mathcal{V}^k$. When $\mathcal{V}^k$ contains only one point, then using terminologies from [8], that point is called the primary frame center. In the event that there are two incumbent solutions x^k_{inf} and x^k_{feas}, one of them is chosen as the primary frame center while the other one is the secondary frame center. The primary frame center in [8] is chosen to be the infeasible incumbent solution while the secondary frame center is the feasible incumbent whenever $f^F_k - \rho > f^I_k$, where the positive scalar ρ is the so called frame center trigger, f^F_k and f^I_k are respectively the incumbent feasible and infeasible f-values at iteration k. Otherwise if the previous inequality does not hold, the primary and secondary frame centers are the feasible and infeasible incumbent solutions. Because of the unavailability of f function values for StoMADS-PB, a specific frame center selection strategy using estimates of such function values is proposed and relies on the following result.

Proposition 4. *Let $f^k_0(x^k_{\text{feas}})$ and $f^k_0(x^k_{\text{inf}})$ be ε-accurate estimates of $f(x^k_{\text{feas}})$ and $f(x^k_{\text{inf}})$ respectively. Let $\rho > 0$ be a scalar.*

$$\text{If } \; f^k_0(x^k_{\text{feas}}) - \rho > f^k_0(x^k_{\text{inf}}) + 2\varepsilon(\delta^k_p)^2, \;\; \text{then } \; f(x^k_{\text{feas}}) - \rho > f(x^k_{\text{inf}}). \tag{9}$$

Proof. Assume that $f^k_0(x^k_{\text{feas}}) - \rho > f^k_0(x^k_{\text{inf}}) + 2\varepsilon(\delta^k_p)^2$. Then, it follows from the ε-accuracy of $f^k_0(x^k_{\text{feas}})$ and $f^k_0(x^k_{\text{inf}})$ that

$$\begin{aligned} f(x^k_{\text{inf}}) - f(x^k_{\text{feas}}) &= \left[f(x^k_{\text{inf}}) - f^k_0(x^k_{\text{inf}})\right] + \left[f^k_0(x^k_{\text{inf}}) - f^k_0(x^k_{\text{feas}})\right] + \left[f^k_0(x^k_{\text{feas}}) - f(x^k_{\text{feas}})\right] \\ &< 2\varepsilon(\delta^k_p)^2 - (\rho + 2\varepsilon(\delta^k_p)^2) = -\rho. \end{aligned} \tag{10}$$

□

Thus according to Proposition 4, x^k_{feas} is always chosen as the StoMADS-PB primary frame center unless the estimates $f^k_0(x^k_{\text{feas}})$ and $f^k_0(x^k_{\text{inf}})$ satisfy a sufficient decrease condition leading to the inequality $f(x^k_{\text{feas}}) - \rho > f(x^k_{\text{inf}})$, which as in [8] allows the choice of the infeasible incumbent solution as primary frame center.

As in [8], StoMADS-PB as implemented for the computational study in Section 5 places less effort in polling around the secondary frame center than the primary one. Specifically, the default strategy is to use a maximal positive basis [12] for the primary frame center and only two directions with one being the negative of the first for the secondary frame center.

Algorithm 1: StoMADS-PB

```
[0] Initialization
    choose $x^0_{\text{inf}} \in \mathcal{X}$, $\delta^0_p > 0$, $\tau \in (0,1) \cap \mathbb{Q}$, $\varepsilon > 0$, $\gamma > 2$ and $\hat{z} \in \mathbb{N}^*$
    set the feasibility success $flag$ = FALSE, $\mathcal{V}^0 \leftarrow \{x^0_{\text{inf}}\}$ and $x^0_{\text{feas}} \leftarrow x^0_{\text{inf}}$
    set the iteration counter $k \leftarrow 0$
[1] Parameter Update
    set $\delta^k_m \leftarrow \min\{\delta^k_p, (\delta^k_p)^2\}$
[2] Poll
    generate a finite list $\mathcal{P}^k$ of candidates using the polling directions $\mathbb{D}^k_p(x^k_{\text{inf}}) \cup \mathbb{D}^k_p(x^k_{\text{feas}})$
    obtain estimates $f^k_0, f^k_s, h^k_0$ and $h^k_s$ of $f(x^k), f(x^k+s^k), h(x^k)$ and $h(x^k+s^k)$
    respectively, at $x^k \in \mathcal{V}^k \cup \{x^k_{\text{feas}}\},\ x^k+s^k \in \mathcal{P}^k$, then compute bounds $u^k_s(x^k+s^k)$
    and $u^k_0(x^k_{\text{inf}})$, using blackbox evaluations
    set the barrier threshold $h^k_{\max} \leftarrow u^k_0(x^k_{\text{inf}})$
    $f$-Dominating
    if $flag$ = FALSE and $u^k_s(x^k+s^k) = 0$ or $flag$ = TRUE and $x^k+s^k \prec_{f;\varepsilon} x^k_{\text{feas}}$
    for some $x^k \in \mathcal{V}^k$ and $s^k \in \{\delta^k_m d^k : d^k \in \mathbb{D}^k_p(x^k)\}$
    set $x^{k+1}_{\text{inf}} \leftarrow x^k_{\text{inf}}$, $x^{k+1}_{\text{feas}} \leftarrow x^k+s^k$ and $\delta^{k+1}_p \leftarrow \min\{\tau^{-1}\delta^k_p, \tau^{-\hat{z}}\}$
    reset the feasibility success $flag$ = TRUE, set $\mathcal{V}^{k+1} \leftarrow \{x^{k+1}_{\text{inf}}, x^{k+1}_{\text{feas}}\}$ and go to [4]
    $h$-Dominating
    else if $x^k_{\text{inf}}+s^k \prec_{h;\varepsilon} x^k_{\text{inf}}$ for some $s^k \in \{\delta^k_m d^k : d^k \in \mathbb{D}^k_p(x^k_{\text{inf}})\}$
    set $x^{k+1}_{\text{inf}} \leftarrow x^k_{\text{inf}}+s^k$, $x^{k+1}_{\text{feas}} \leftarrow x^k_{\text{feas}}$ and $\delta^{k+1}_p \leftarrow \min\{\tau^{-1}\delta^k_p, \tau^{-\hat{z}}\}$
    Improving
    else if $h^k_s(x^k_{\text{inf}}+s^k) - h^k_0(x^k_{\text{inf}}) \leq -\gamma m\varepsilon(\delta^k_p)^2$ for some previously evaluated $x^k_{\text{inf}}+s^k$
    set $x^{k+1}_{\text{inf}} \in \operatorname{argmin}_{x^k_{\text{inf}}+s^k}\{u^k_s(x^k_{\text{inf}}+s^k) : h^k_s(x^k_{\text{inf}}+s^k) - h^k_0(x^k_{\text{inf}}) \leq -\gamma m\varepsilon(\delta^k_p)^2\}$
    $x^{k+1}_{\text{feas}} \leftarrow x^k_{\text{feas}}$ and $\delta^{k+1}_p \leftarrow \min\{\tau^{-1}\delta^k_p, \tau^{-\hat{z}}\}$
    Unsuccessful
    otherwise, set $x^{k+1}_{\text{inf}} \leftarrow x^k_{\text{inf}}$, $x^{k+1}_{\text{feas}} \leftarrow x^k_{\text{feas}}$ and $\delta^{k+1}_p \leftarrow \tau\delta^k_p$
[3] Feasibility update
    if $flag$ = TRUE
    set $\mathcal{V}^{k+1} \leftarrow \{x^{k+1}_{\text{inf}}, x^{k+1}_{\text{feas}}\}$
    otherwise, $\mathcal{V}^{k+1} \leftarrow \{x^{k+1}_{\text{inf}}\}$
[4] Termination
    if no termination criterion is met
    set $k \leftarrow k+1$ and go to [1]
    otherwise stop
```

Figure 1: StoMADS-PB algorithm for constrained stochastic optimization.

3 Stochastic process generated by StoMADS-PB

The stochastic quantities in the present work are all defined on the same probability space $(\Omega, \mathcal{G}, \mathbb{P})$. The nonempty set Ω is referred to as the sample space and its subsets are called events. The collection $\mathcal{G}$ of such events is called a σ-algebra or σ-field and $\mathbb{P}$ is a finite measure satisfying $\mathbb{P}(\Omega) = 1$, referred to as probability measure and defined on the measurable space $(\Omega, \mathcal{G})$. Each element $\omega \in \Omega$ is referred to as a sample point or a possible outcome. Let $\mathcal{B}(\mathbb{R}^n)$ be the Borel σ-algebra of $\mathbb{R}^n$, i.e., the one generated by its open sets. A random variable X is a measurable map defined on $(\Omega, \mathcal{G}, \mathbb{P})$ into the measurable space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, where measurability means that each event $\{X \in B\} := X^{-1}(B)$ belongs to $\mathcal{G}$ for all $B \in \mathcal{B}(\mathbb{R}^n)$ [20, 33].

The estimates $f_0^k(x^k)$, $f_s^k(x^k+s^k)$, $c_{j,0}^k(x^k)$ and $c_{j,s}^k(x^k+s^k)$, for $j = 1, 2, \ldots, m$, $x^k \in \{x_{\text{inf}}^k, x_{\text{feas}}^k\}$ and $x^k + s^k \in \mathcal{P}^k$, of function values are computed at every iteration of Algorithm 1 using the noisy blackbox evaluations. Because of the randomness of the blackbox outputs, such estimates can respectively be considered as realizations of random estimates $F_0^k(X^k)$, $F_s^k(X^k + S^k)$, $C_{j,0}^k(X^k)$ and $C_{j,s}^k(X^k + S^k)$, for $j = 1, 2, \ldots, m$. Since each iteration k of Algorithm 1 is influenced by the randomness stemming from such random estimates, Algorithm 1 results in a stochastic process. For the remainder of the manuscript, uppercase letters will be used to denote random quantities while their realizations will be denoted by lowercase letters. Thus, $x^k = X^k(\omega)$, $x_{\text{inf}}^k = X_{\text{inf}}^k(\omega)$, $x_{\text{feas}}^k = X_{\text{feas}}^k(\omega)$, $s^k = S^k(\omega)$, $\delta_p^k = \Delta_p^k(\omega)$ and $\delta_m^k = \Delta_m^k(\omega)$ denote respectively realizations of X^k, X_{inf}^k, X_{feas}^k, S^k, Δ_p^k and Δ_m^k. Similarly, $f_0^k(x^k) = F_0^k(X^k)(\omega)$, $f_s^k(x^k + s^k) = F_s^k(X^k + S^k)(\omega)$, $c_{j,0}^k(x^k) = C_{j,0}^k(X^k)(\omega)$, $c_{j,s}^k(x^k+s^k) = C_{j,s}^k(X^k+S^k)(\omega)$, $h_0^k(x^k) = H_0^k(X^k)(\omega)$, $h_s^k(x^k+s^k) = H_s^k(X^k+S^k)(\omega)$, $\ell_0^k(x^k) = L_0^k(X^k)(\omega)$, $\ell_s^k(x^k + s^k) = L_s^k(X^k + S^k)(\omega)$, $u_0^k(x^k) = U_0^k(X^k)(\omega)$ and $u_s^k(x^k + s^k) = U_s^k(X^k + S^k)(\omega)$. When there is no ambiguity, F_0^k will be used instead of $F_0^k(X^k)$, etc. In general, following the notations in [11, 21, 23, 33, 48], F_0^k, F_s^k, H_0^k and H_s^k are respectively the estimates of $f(X^k)$, $f(X^k + S^k)$, $h(X^k)$ and $h(X^k + S^k)$. Moreover, as highlighted in [11], the notation "$f(X^k)$" is used to denote the random variable with realizations $f(X^k(\omega))$.

The present research aims to show that the stochastic process $\left\{X_{\text{inf}}^k, X_{\text{feas}}^k, \Delta_p^k, \Delta_m^k, F_0^k, F_s^k, H_0^k, H_s^k, L_0^k, U_0^k, L_s^k, U_s^k\right\}$ resulting from Algorithm 1 converges with probability one under some assumptions on the estimates $F_0^k, F_s^k, C_{j,0}^k, C_{j,s}^k, H_0^k, H_s^k$ and on the bounds $L_0^k, U_0^k, L_s^k, U_s^k$. In particular, the estimates $F_0^k, F_s^k, C_{j,0}^k$ and $C_{j,s}^k$ will be assumed to be accurate while the bounds will be assumed to be reliable, with sufficiently high but fixed probabilities, conditioned on the past.

3.1 Probabilistic bounds and probabilistic estimates

The previously mentioned notion of conditioning on the past is formalized following [11, 21, 23, 33, 48]. Denote by $\mathcal{F}_{k-1}^{C\cdot F}$ the σ-algebra generated by $F_0^\ell(X^\ell)$, $F_s^\ell(X^\ell + S^\ell)$, $C_{j,0}^\ell(X^\ell)$ and $C_{j,s}^\ell(X^\ell + S^\ell)$, for $j = 1, 2, \ldots, m$, for $X^\ell \in \left\{X_{\text{inf}}^\ell, X_{\text{feas}}^\ell\right\}$ and for $\ell = 0, 1, \ldots, k-1$. For completeness, $\mathcal{F}_{-1}^{C\cdot F}$ is set to equal $\sigma(x^0) = \sigma(x_{\text{inf}}^0)$. Thus, $\{\mathcal{F}_k^{C\cdot F}\}_{k\geq -1}$ is a filtration, i.e., a subsequence of increasing σ-algebras of $\mathcal{G}$.

Sufficient accuracy of functions estimates is measured using the poll size parameter and is formalized, following [11, 21, 23, 33, 48] by means of the definitions bellow.

Definition 6. *A sequence of random estimates $\{F_0^k, F_s^k\}$ is said to be β-probabilistically ε-accurate with respect to the corresponding sequence $\{X^k, S^k, \Delta_p^k\}$ if the events*

$$J_k = \{F_0^k, F_s^k,\ \textit{are } \varepsilon\textit{-accurate estimates of } f(x^k) \textit{ and } f(x^k + s^k),\ \textit{respectively for } \Delta_p^k\}$$

satisfy the following submartingale-like condition

$$\mathbb{P}\left(J_k \mid \mathcal{F}_{k-1}^{C\cdot F}\right) = \mathbb{E}\left(\mathbb{1}_{J_k} \mid \mathcal{F}_{k-1}^{C\cdot F}\right) \geq \beta,$$

where $\mathbb{1}_{J_k}$ denotes the indicator function of the event J_k, i.e., $\mathbb{1}_{J_k} = 1$ if $\omega \in J_k$ and $\mathbb{1}_{J_k} = 0$ otherwise. The estimates are called "good" if $\mathbb{1}_{J_k} = 1$. Otherwise they are called "bad".

Definition 7. *A sequence of random estimates $\{C_{j,0}^k, C_{j,s}^k\}$ is said to be $\alpha^{1/m}$-probabilistically ε-accurate for some $j = 1, 2, \ldots, m$ with respect to the corresponding sequence $\{X^k, S^k, \Delta_p^k\}$ if the events*

$$I_k^j = \{C_{j,0}^k, C_{j,s}^k, \textit{ are } \varepsilon\textit{-accurate estimates of } c_j(x^k) \textit{ and } c_j(x^k + s^k), \textit{ respectively for } \Delta_p^k\}$$

satisfy the following submartingale-like condition

$$\mathbb{P}\left(I_k^j \mid \mathcal{F}_{k-1}^{C\cdot F}\right) = \mathbb{E}\left(\mathbb{1}_{I_k^j} \mid \mathcal{F}_{k-1}^{C\cdot F}\right) \geq \alpha^{1/m}.$$

To formalize the sufficient reliability of random bounds in the present work, the following definition is introduced.

Definition 8. *A sequence of random bounds $\{L_0^k, U_0^k, L_s^k, U_s^k\}$ is said to be α-probabilistically ε-accurate with respect to the corresponding sequence $\{X^k, S^k, \Delta_p^k\}$ if the events*

$$\begin{aligned} I_k &= \{\text{"}L_0^k \textit{ and } U_0^k \textit{ are } \varepsilon\textit{-reliable bounds for } h(x^k)\text{", and "}L_s^k \textit{ and } U_s^k \textit{ are } \varepsilon\textit{-reliable bounds} \\ &\qquad \textit{for } h(x^k + s^k)\text{", respectively for } \Delta_p^k\} \end{aligned} \tag{11}$$

satisfy the following submartingale-like condition

$$\mathbb{P}\left(I_k \mid \mathcal{F}_{k-1}^{C\cdot F}\right) = \mathbb{E}\left(\mathbb{1}_{I_k} \mid \mathcal{F}_{k-1}^{C\cdot F}\right) \geq \mathbb{P}\left(\bigcap_{j=1}^{m} I_k^j \mid \mathcal{F}_{k-1}^{C\cdot F}\right) \geq \alpha,$$

The bounds are called "good" if $\mathbb{1}_{I_k} = 1$. Otherwise, $\mathbb{1}_{I_k} = 0$ and they are called "bad".

The p-integrability of random variables [11, 20] is defined below and will be useful for the analysis of Algorithm 1.

Definition 9. *Let $(\Omega, \mathcal{G}, \mathbb{P})$ be a probability space and $p \in [1, +\infty)$ be an integer. Then the Space $\mathbb{L}^p(\Omega, \mathcal{G}, \mathbb{P})$ of so-called p-integrable random variables is the set of all real-valued random variables X such that*

$$\|X\|_p := \left(\int_\Omega |X(\omega)|^p \, \mathbb{P}(d\omega)\right)^{\frac{1}{p}} = \left(\mathbb{E}\left(|X|^p\right)\right)^{\frac{1}{p}} < +\infty.$$

As in [11], the following is assumed in order for the random variables $f(X^k)$, $h(X^k)$ and $c_j(X^k)$, $j \in J$, to be integrable so that the conditional expectations $\mathbb{E}\left(f(X^k)|\mathcal{F}_{k-1}^{C\cdot F}\right)$, $\mathbb{E}\left(c_j(X^k)|\mathcal{F}_{k-1}^{C\cdot F}\right)$, $j \in J$ and $\mathbb{E}\left(h(X^k)|\mathcal{F}_{k-1}^{C\cdot F}\right)$ can be well defined [20].

Assumption 2. *The objective function f and the constraints violation function h are locally Lipschitz with constants $\lambda^f > 0$ and $\lambda^h > 0$, respectively. The constraint functions c_j, $j \in J$, are continuous on $\mathcal{X}$. The set $\mathcal{U} \subset \mathcal{X}$ containing all iterates realizations is compact.*

Local Lipschitz in the above assumption means, Lipschitz with a finite constant in some nonempty neighborhood intersected with $\mathcal{X}$ [8].

Proposition 5. *Under Assumption 2, there exists a finite constant $\kappa^f_{\max}$ satisfying $\left|f(x^k)\right| \leq \kappa^f_{\max}$ for all $x^k \in \mathcal{U}$. Moreover, the random variables $f(X^k)$, $h(X^k)$, $c_j(X^k)$ and Δ^k_p belong to $\mathbb{L}^1(\Omega, \mathcal{G}, \mathbb{P})$, for all $j \in J$ and for all $k \geq 0$.*

Proof. The proof is inspired from [11]. Since f is locally Lipschitz on the compact set $\mathcal{U}$, the it is bounded on $\mathcal{U}$. Consequently, there exists a finite constant $\kappa^f_{\max}$ such that $\left|f(x^k)\right| \leq \kappa^f_{\max}$ for all $x^k \in \mathcal{U}$. Similarly, there exist $\kappa^h_{\max}$ satisfying $\left|h(x^k)\right| \leq \kappa^h_{\max}$ and $\kappa^c_{\max}$ such that $\left|c_j(x^k)\right| \leq \kappa^c_{\max}$ for all $j \in J$ and all $x^k \in \mathcal{U}$, since h is locally Lipschitz and c_j is continuous on $\mathcal{U}$. Thus, $\mathbb{E}\left(\left|f(X^k)\right|\right) := \int_\Omega \left|f(X^k(\omega))\right| \mathbb{P}(d\omega) \leq \kappa^f_{\max} < +\infty$. Similarly, $\mathbb{E}\left(\left|h(X^k)\right|\right) \leq \kappa^h_{\max} \leq +\infty$ and for all $j \in J$, $\mathbb{E}\left(\left|c_j(X^k)\right|\right) \leq \kappa^c_{\max} \leq +\infty$. Finally, the integrability of Δ^k_p follows from the fact that $\Delta^k_p(\omega) \leq \tau^{-\hat{z}}$ for all $\omega \in \Omega$, which implies that $\mathbb{E}\left(\left|\Delta^k_p\right|\right) := \int_\Omega \left|\Delta^k_p(\omega)\right| \mathbb{P}(d\omega) \leq \tau^{-\hat{z}} < +\infty$. □

Next are stated some key assumptions on the nature of the stochastic information in Algorithm 1, some of which are made in [11] and which will be useful for the convergence analysis of Section 4.

Assumption 3. *For fixed α and $\beta \in (0,1)$, the followings hold for the random quantities generated by Algorithm 1.*

(i) The sequence of estimates $\{F^k_0, F^k_s\}$ generated by Algorithm 1 is β-probabilistically ε-accurate.

(ii) The sequence of estimates $\{F^k_0, F^k_s\}$ generated by Algorithm 1 satisfies the following variance condition for all $k \geq 0$,

$$\begin{aligned}
&\mathbb{E}\left(\left|F^k_s - f(X^k + S^k)\right|^2 \mid \mathcal{F}^{C\cdot F}_{k-1}\right) \leq \varepsilon^2(1-\sqrt{\beta})(\Delta^k_p)^4 \\
\textit{and}\quad &\mathbb{E}\left(\left|F^k_0 - f(X^k)\right|^2 \mid \mathcal{F}^{C\cdot F}_{k-1}\right) \leq \varepsilon^2(1-\sqrt{\beta})(\Delta^k_p)^4.
\end{aligned} \tag{12}$$

(iii) For all $j = 1,2,\ldots,m$, the sequence of estimates $\{C^k_{j,0}, C^k_{j,s}\}$ is $\alpha^{1/m}$-probabilistically ε-accurate.

(iv) For all $j = 1,2,\ldots,m$, the sequence of estimates $\{C^k_{j,0}, C^k_{j,s}\}$ satisfies the following variance condition for all $k \geq 0$,

$$\begin{aligned}
&\mathbb{E}\left(\left|C^k_{j,s} - c_j(X^k + S^k)\right|^2 \mid \mathcal{F}^{C\cdot F}_{k-1}\right) \leq \varepsilon^2\left(1-\alpha^{1/2m}\right)(\Delta^k_p)^4 \\
\textit{and}\quad &\mathbb{E}\left(\left|C^k_{j,0} - c_j(X^k)\right|^2 \mid \mathcal{F}^{C\cdot F}_{k-1}\right) \leq \varepsilon^2\left(1-\alpha^{1/2m}\right)(\Delta^k_p)^4.
\end{aligned} \tag{13}$$

(v) The sequence of random bounds $\{L^k_0, U^k_0, L^k_s, U^k_s\}$ is α-probabilistically ε-reliable.

(vi) The sequence of random estimated violations $\{H^k_0, H^k_s\}$ satisfies

$$\begin{aligned}
&\mathbb{E}\left(\left|H^k_s - h(X^k + S^k)\right| \mid \mathcal{F}^{C\cdot F}_{k-1}\right) \leq m\varepsilon(1-\alpha)^{1/2}(\Delta^k_p)^2 \\
\textit{and}\quad &\mathbb{E}\left(\left|H^k_0 - h(X^k)\right| \mid \mathcal{F}^{C\cdot F}_{k-1}\right) \leq m\varepsilon(1-\alpha)^{1/2}(\Delta^k_p)^2.
\end{aligned} \tag{14}$$

An iteration k for which $\mathbb{1}_{I_k}\mathbb{1}_{J_k} = 1$, i.e., for which the events I_k and J_k both occur, will be called "*true*". Otherwise, it will be called "*false*". Even though the present algorithmic framework does not allow to determine which iterations are true or false, Theorem 1 shows that true iterations occur infinitely often for convergence to hold, provided that estimates and bounds are sufficiently accurate. Theorem 1 will also be useful for the convergence analysis of Algorithm 1, more precisely in Subsection 4.3.

Theorem 1. *Assume that Assumption 3 holds for $\alpha\beta \in (1/2, 1)$. Then true iterations of Algorithm 1 occur infinitely often.*

Proof. Consider the following random walk

$$W_k = \sum_{i=0}^{k} (2 \cdot \mathbb{1}_{I_i} \mathbb{1}_{J_i} - 1). \tag{15}$$

Then, the result easily follows from the fact that $\left\{ \limsup_{k \to +\infty} W_k = +\infty \right\}$ almost surely, the proof of which can be derived from that of Theorem 4.16 in [23] (using $\mathcal{F}_{k-1}^{C \cdot F}$ instead of $\mathcal{F}_{k-1}^{I \cdot J}$), where a similar random walk was studied. Indeed, the latter result means that

$$\mathbb{P}\left(\left\{ \omega \in \Omega : \exists K(\omega) \subset \mathbb{N} \text{ such that } \lim_{k \in K(\omega)} W_k(\omega) = +\infty \right\} \right) = 1,$$

which implies that $\mathbb{1}_{I_i} \mathbb{1}_{J_i} = 1$ infinitely often. □

3.2 Computation of probabilistically accurate estimates and reliable bounds

This section discusses approaches for computing accurate random estimates and reliable bounds satisfying Assumption 3 in a simple random noise framework, and hence how corresponding deterministic estimates can be obtained using evaluations of the stochastic blackbox. Such approaches strongly rely on the computation of $\alpha^{1/m}$-probabilistically ε-accurate estimates $\{C_{j,0}^k, C_{j,s}^k\}$, using techniques derived in [23].

Consider the following typical noise assumption often made in stochastic optimization literature:

$$\begin{aligned} \mathbb{E}_{\Theta_0}\left[f_{\Theta_0}(x)\right] &= f(x) \quad \text{and} \quad \mathbb{V}_{\Theta_0}\left[f_{\Theta_0}(x)\right] \le V_0 < +\infty \ \text{ for all } x \in \mathcal{X} \\ \mathbb{E}_{\Theta_j}\left[c_{\Theta_j}(x)\right] &= c_j(x) \quad \text{and} \quad \mathbb{V}_{\Theta_j}\left[c_{\Theta_j}(x)\right] \le V_j < +\infty \ \text{ for all } x \in \mathcal{X} \text{ and for all } j \in J, \end{aligned}$$

where $V_i > 0$ is a constant for all $i = 0, 1, \ldots, m$. Let $V = \max\{V_0, V_1, \ldots, V_m\}$.

For some fixed $j \in J$, let Θ_j^0 and Θ_j^s be two independent random variables following the same distribution as Θ_j. Let $\Theta_{j,\ell}^0$, $\ell = 1, 2, \ldots, p_j^k$ and $\Theta_{j,\ell}^s$, $\ell = 1, 2, \ldots, p_j^k$ be independent random samples of Θ_j^0 and Θ_j^s respectively, where $p_j^k \ge 1$ is an integer denoting the sample size. In order to satisfy Assumption 3-(iii), define $C_{j,0}^k$ and $C_{j,s}^k$ respectively by

$$C_{j,0}^k = \frac{1}{p_j^k} \sum_{\ell=1}^{p_j^k} c_{\Theta_{j,\ell}^0}(x^k) \quad \text{and} \quad C_{j,s}^k = \frac{1}{p_j^k} \sum_{\ell=1}^{p_j^k} c_{\Theta_{j,\ell}^s}(x^k + s^k).$$

By noticing that $\mathbb{E}\left(C_{j,0}^k\right) = c_j(x^k)$ and that $\mathbb{V}\left(C_{j,0}^k\right) \le \frac{V}{p_j^k}$ for all j, then it follows from the Chebyshev inequality that

$$\mathbb{P}\left(\left|C_{j,0}^k - c_j(x^k)\right| > \varepsilon(\delta_p^k)^2\right) = \mathbb{P}\left(\left|C_{j,0}^k - \mathbb{E}\left(C_{j,0}^k\right)\right| > \varepsilon(\delta_p^k)^2\right) \le \frac{V}{p_j^k \varepsilon^2 (\delta_p^k)^4}. \tag{16}$$

Thus, choosing p_j^k such that

$$p_j^k \ge \frac{V}{\varepsilon^2 \left(1 - \alpha^{1/2m}\right) (\delta_p^k)^4} \tag{17}$$

ensures that $\frac{V}{p_j^k \varepsilon^2 (\delta_p^k)^4} \leq 1 - \alpha^{1/2m}$. Then, combining (16) and (17) yields for all $j \in J$,

$$\mathbb{P}\left(\left|C_{j,0}^k - c_j(x^k)\right| \leq \varepsilon(\delta_p^k)^2\right) \geq \alpha^{1/2m} \tag{18}$$

and similarly, $\mathbb{P}\left(\left|C_{j,s}^k - c_j(x^k + s^k)\right| \leq \varepsilon(\delta_p^k)^2\right) \geq \alpha^{1/2m}$. It follows from the independence of the random variables Θ_j^0 and Θ_j^s and both previous inequalities that

$$\mathbb{P}\left(\left\{\left|C_{j,0}^k - c_j(x^k)\right| \leq \varepsilon(\delta_p^k)^2\right\} \cap \left\{\left|C_{j,s}^k - c_j(x^k + s^k)\right| \leq \varepsilon(\delta_p^k)^2\right\}\right) \geq \alpha^{1/m}, \tag{19}$$

which means that Assumption 3-*(iii)* holds. Estimates $c_{j,0}^k = C_{j,0}^k(\omega)$ and $c_{j,s}^k = C_{j,s}^k(\omega)$, obtained by averaging p_j^k realizations of c_{Θ_j}, resulting from the evaluations of the stochastic blackbox, respectively at x^k and $x^k + s^k$, are obviously ε-accurate.

In order to satisfy Assumption 3-*(v)*, notice that the independence of the random variables $\Theta_j, j \in J$ combined with (18) implies

$$\mathbb{P}\left(\bigcap_{j=1}^m \left\{\left|C_{j,0}^k - c_j(x^k)\right| \leq \varepsilon(\delta_p^k)^2\right\}\right) = \prod_{j=1}^m \mathbb{P}\left(\left|C_{j,0}^k - c_j(x^k)\right| \leq \varepsilon(\delta_p^k)^2\right) \geq \alpha^{1/2} \tag{20}$$

$$\text{and similarly,} \quad \mathbb{P}\left(\bigcap_{j=1}^m \left\{\left|C_{j,s}^k - c_j(x^k + s^k)\right| \leq \varepsilon(\delta_p^k)^2\right\}\right) \geq \alpha^{1/2}. \tag{21}$$

Define the random bounds $L_0^k(x^k)$, $L_s^k(x^k + s^k)$, $U_0^k(x^k)$ and $U_s^k(x^k + s^k)$, respectively by

$$\begin{aligned} L_0^k(x^k) &= \sum_{j=1}^m \max\left\{C_{j,0}^k - \varepsilon(\delta_p^k)^2, 0\right\}, \qquad && U_0^k(x^k) = \sum_{j=1}^m \max\left\{C_{j,0}^k + \varepsilon(\delta_p^k)^2, 0\right\} \\ L_s^k(x^k + s^k) &= \sum_{j=1}^m \max\left\{C_{j,s}^k - \varepsilon(\delta_p^k)^2, 0\right\} \text{ and } && U_s^k(x^k + s^k) = \sum_{j=1}^m \max\left\{C_{j,s}^k + \varepsilon(\delta_p^k)^2, 0\right\}. \end{aligned}$$

Define the events E_0^k and E_s^k respectively by

$$E_0^k = \left\{L_0^k(x^k) \leq h(x^k) \leq U_0^k(x^k)\right\} \text{ and } E_s^k = \left\{L_s^k(x^k + s^k) \leq h(x^k + s^k) \leq U_s^k(x^k + s^k)\right\} \tag{22}$$

By noticing that

$$\bigcap_{j=1}^m \left\{\left|C_{j,0}^k - c_j(x^k)\right| \leq \varepsilon(\delta_p^k)^2\right\} = \bigcap_{j=1}^m \left\{C_{j,0}^k - \varepsilon(\delta_p^k)^2 \leq c_j(x^k) \leq C_{j,0}^k + \varepsilon(\delta_p^k)^2\right\} \subseteq E_0^k \tag{23}$$

$$\bigcap_{j=1}^m \left\{\left|C_{j,s}^k - c_j(x^k + s^k)\right| \leq \varepsilon(\delta_p^k)^2\right\} \subseteq E_s^k, \tag{24}$$

then combining respectively (20) and (23), and (21) and (24), yields

$$\mathbb{P}\left(E_0^k\right) \geq \mathbb{P}\left(\bigcap_{j=1}^m \left\{\left|C_{j,0}^k - c_j(x^k)\right| \leq \varepsilon(\delta_p^k)^2\right\}\right) \geq \alpha^{1/2} \tag{25}$$

$$\mathbb{P}\left(E_s^k\right) \geq \mathbb{P}\left(\bigcap_{j=1}^{m} \left\{\left|C_{j,s}^k - c_j(x^k + s^k)\right| \leq \varepsilon(\delta_p^k)^2\right\}\right) \geq \alpha^{1/2}. \tag{26}$$

It follows from the independence of the random variables $\Theta_{j,\ell}^0$ and $\Theta_{j,\ell}^s$, for all $j \in J$ and for all $\ell = 1, 2, \ldots, p_j^k$, that the events E_0^k and E_s^k are also independent. Hence, both inequalities (25) and (26) imply that

$$\begin{aligned}
\alpha &\leq \mathbb{P}\left(\bigcap_{j=1}^{m} \left\{\left|C_{j,0}^k - c_j(x^k)\right| \leq \varepsilon(\delta_p^k)^2\right\}\right) \times \mathbb{P}\left(\bigcap_{j=1}^{m} \left\{\left|C_{j,s}^k - c_j(x^k + s^k)\right| \leq \varepsilon(\delta_p^k)^2\right\}\right) \\
&= \mathbb{P}\left(\bigcap_{j=1}^{m} \left\{\left|C_{j,0}^k - c_j(x^k)\right| \leq \varepsilon(\delta_p^k)^2\right\} \cap \left\{\left|C_{j,s}^k - c_j(x^k + s^k)\right| \leq \varepsilon(\delta_p^k)^2\right\}\right) \\
&\leq \mathbb{P}\left(E_0^k\right) \times \mathbb{P}\left(E_s^k\right) = \mathbb{P}\left(E_0^k \cap E_s^k\right),
\end{aligned}$$

which shows that Assumption 3-*(v)* holds.

In order to show that Assumption 3-*(iv)* holds, notice that $\mathbb{E}\left(C_{j,0}^k - c_j(x^k)\right) = 0$ for all $j \in J$, which implies that for all $j \in J$,

$$\mathbb{E}\left(\left|C_{j,0}^k - c_j(x^k)\right|^2\right) = \mathbb{V}\left(C_{j,0}^k - c_j(x^k)\right) = \mathbb{V}\left(C_{j,0}^k\right) \leq \frac{V}{p_j^k} \leq \varepsilon^2\left(1 - \alpha^{1/2m}\right)(\delta_p^k)^4, \tag{27}$$

where the last inequality in (27) follows from (17). Similarly, since $\mathbb{E}\left(C_{j,s}^k - c_j(x^k + s^k)\right) = 0$ for all $j \in J$, then

$$\mathbb{E}\left(\left|C_{j,s}^k - c_j(x^k + s^k)\right|^2\right) \leq \varepsilon^2\left(1 - \alpha^{1/2m}\right)(\delta_p^k)^4, \tag{28}$$

which shows that Assumption 3-*(iv)* holds.

Before showing Assumption 3-*(vi)*, let first notice that

$$\begin{aligned}
\left|H_0^k - h(x^k)\right| &= \left|\sum_{j=1}^{m} \max\{C_{j,0}^k, 0\} - \sum_{j=1}^{m} \max\{c_j(x^k), 0\}\right| \\
&\leq \sum_{j=1}^{m} \left|\max\{C_{j,0}^k, 0\} - \max\{c_j(x^k), 0\}\right| \leq \sum_{j=1}^{m} \left|C_{j,0}^k - c_j(x^k)\right|,
\end{aligned} \tag{29}$$

where the last inequality in (29) follows from the inequality $|\max\{x, 0\} - \max\{y, 0\}| \leq |x - y|$, for all $x, y \in \mathbb{R}$. Moreover, it follows from the Cauchy-Schwarz inequality [20] that for all $j \in J$,

$$\mathbb{E}\left(\left|C_{j,0}^k - c_j(x^k)\right|\right) \leq \left[\mathbb{E}\left(\left|C_{j,0}^k - c_j(x^k)\right|^2\right)\right]^{1/2} \leq \varepsilon\left(1 - \alpha\right)^{1/2}(\delta_p^k)^2, \tag{30}$$

where the last inequality in (30) follows from (27). Thus, taking the expectation in (29), combined with (30) yields

$$\mathbb{E}\left(\left|H_0^k - h(x^k)\right|\right) \leq \sum_{j=1}^{m} \mathbb{E}\left(\left|C_{j,0}^k - c_j(x^k)\right|\right) \leq m\varepsilon\left(1 - \alpha\right)^{1/2}(\delta_p^k)^2,$$

$$\text{and similarly} \quad \mathbb{E}\left(\left|H_s^k - h(x^k+s^k)\right|\right) \leq m\varepsilon\,(1-\alpha)^{1/2}\,(\delta_p^k)^2,$$

which shows that Assumption 3-*(vi)* holds.

Finally, let compute estimates F_0^k and F_s^k that satisfy Assumption 3-*(i)* and *(ii)*. For that purpose, let Θ_0^0 and Θ_0^s be two independent random variables following the same distribution as Θ_0. Let $\Theta_{0,\ell}^0$, $\ell = 1, 2, \ldots, p_0^k$ and $\Theta_{0,\ell}^s$, $\ell = 1, 2, \ldots, p_0^k$ be independent random samples of Θ_0^0 and Θ_0^s respectively, where $p_0^k \geq 1$ denotes the sample size. Define F_0^k and F_s^k respectively by

$$F_0^k = \frac{1}{p_0^k}\sum_{\ell=1}^{p_0^k} f_{\Theta_{0,\ell}^0}(x^k) \quad \text{and} \quad F_s^k = \frac{1}{p_0^k}\sum_{\ell=1}^{p_0^k} f_{\Theta_{0,\ell}^s}(x^k+s^k).$$

Then $\mathbb{E}\left(F_0^k\right) = f(x^k)$, which implies that $\mathbb{V}\left(F_0^k\right) \leq \frac{V}{p_0^k}$. Thus, it is easy to notice that the proof of Assumption 3-*(i)* follows that of Assumption 3-*(iii)*. More precisely, the following inequality holds:

$$\mathbb{P}\left(\left\{\left|F_0^k - f(x^k)\right| \leq \varepsilon(\delta_p^k)^2\right\} \cap \left\{\left|F_s^k - f(x^k+s^k)\right| \leq \varepsilon(\delta_p^k)^2\right\}\right) \geq \beta, \tag{31}$$

provided that

$$p_0^k \geq \frac{V}{\varepsilon^2\left(1-\sqrt{\beta}\right)(\delta_p^k)^4} \tag{32}$$

Estimates $f_0^k = F_0^k(\omega)$ and $f_s^k = F_s^k(\omega)$, obtained by averaging p_0^k realizations of f_{Θ_0}, resulting from the evaluations of the stochastic blackbox, respectively at x^k and x^k+s^k, are obviously ε-accurate. It is also easy to notice that the proof of Assumption 3-*(ii)* follows that of Assumption 3-*(iv)*. Specifically,

$$\mathbb{E}\left(\left|F_0^k - f(x^k)\right|^2\right) \leq \varepsilon^2(1-\sqrt{\beta})(\delta_p^k)^4 \quad \text{and} \quad \mathbb{E}\left(\left|F_s^k - f(x^k+s^k)\right|^2\right) \leq \varepsilon^2(1-\sqrt{\beta})(\delta_p^k)^4,$$

provided that p_0^k is chosen according to (32).

4 Convergence analysis

Using ideas inspired by [8, 11, 23, 40, 48] this section presents convergence results of StoMADS-PB, most of which are stochastic variants of those in [8]. It introduces the random time T at which Algorithm 1 generates a first ε-feasible solution. Then assuming that T is either almost surely finite or almost surely infinite, a so-called zeroth-order result [10, 11] is derived showing that there exists a subsequence of Algorithm 1-generated random iterates with mesh realizations becoming infinitely fine and which converges with probability one to a limit. This is achieved by showing by means of Theorem 2 that the sequence of random poll size parameters converges to zero with probability one. Section 4.2 analyzes the function h and the random ε-infeasible iterates generated by Algorithm 1. In particular, it gives conditions under which an almost sure limit of a subsequence of such iterates is shown in Theorem 4 to satisfy a first-order necessary optimality condition via the Clarke generalized derivative of h with probability one. Then, a similar result for f and the sequence of ε-feasible iterates is derived in Theorem 6 of Section 4.3. Note finally that the proofs of the main results of this section are presented in the Appendix.

4.1 Zeroth-order convergence

Recall Remark 1 and denote by $\mathscr{S}_X^k = \{X_{\text{feas}}^\ell : X_{\text{feas}}^\ell \neq x_{\text{inf}}^0,\ \ell \leq k\}$ the set of all random ε-feasible iterates generated by Algorithm 1 until the beginning of iteration k. Consider the following random time T defined by

$$T := \inf\{k \geq 0 : \mathscr{S}_X^k \neq \emptyset\}. \tag{33}$$

Then it is easy to notice that $T \geq 1$ and that for all $k \geq 1$, the occurrence of the event $\{T \leq k\}$ is determined by observing the random quantities generated by Algorithm 1 until the iteration $k-1$, which means that T is a stopping time [32] for the stochastic process generated by Algorithm 1. The following is assumed for the remainder of the analysis.

Assumption 4. *The stopping time T associated to the stochastic process generated by Algorithm 1 is either almost surely finite or almost surely infinite.*

The next result implies that the sequence $\{\Delta_p^k\}_{k\in\mathbb{N}}$ of random poll size parameters converges to zero with probability one and will be useful for the Clarke stationarity results of Sections 4.2 and 4.3. It holds under the assumption below.

Assumption 5. *The objective function f is bounded from below, i.e., there exists $\kappa_{\min}^f \in \mathbb{R}$ such that $-\infty < \kappa_{\min}^f \leq f(x)$, for all $x \in \mathbb{R}^n$.*

Theorem 2. *Let Assumptions 2, 4 and 5 be satisfied. Let $\gamma > 2$ and $\tau \in (0,1) \cap \mathbb{Q}$. Let $\nu \in (0,1)$ be chosen such that*

$$\frac{\nu}{1-\nu} \geq \frac{2(\tau^{-2}-1)}{\gamma - 2} \tag{34}$$

and assume that Assumption 3 holds for α and β chosen such that

$$\alpha\beta \geq \frac{4\nu}{(1-\nu)(1-\tau^2)}\left[(1-\alpha)^{1/2} + 2(1-\beta)^{1/2}\right]. \tag{35}$$

Then, the sequence $\{\Delta_p^k\}_{k\in\mathbb{N}}$ of frame size parameters generated by Algorithm 1 satisfies

$$\sum_{k=0}^{+\infty} (\Delta_p^k)^2 < +\infty \quad \textit{almost surely}. \tag{36}$$

The following result is a simple consequence of Theorem 2. It shows that the sequences $\{\Delta_m^k\}_{k\in\mathbb{N}}$ and $\{\Delta_p^k\}_{k\in\mathbb{N}}$ converge to zero almost surely respectively.

Corollary 1. *The followings hold under all the assumptions made in Theorem 2*

$$\lim_{k\to+\infty} \Delta_m^k = 0 \textit{ almost surely} \quad \textit{and} \quad \lim_{k\to+\infty} \Delta_p^k = 0 \textit{ almost surely}.$$

The next result shows that with probability one, the difference between the estimates and their corresponding true function values converge to zero. This means that Algorithm 1 behaves like an exact deterministic method asymptotically. This result will be also useful in Subsection 4.3 for the proof of Theorem 5.

Corollary 2. *Let all assumptions that were made in Theorem 2 hold. Then,*

$$\lim_{k\to+\infty} \left|H_0^k - h(X^k)\right| = 0 \textit{ almost surely} \quad \textit{and} \quad \lim_{k\to+\infty} \left|F_0^k - f(X^k)\right| = 0 \textit{ almost surely}, \tag{37}$$

and the same result holds for $\left|H_s^k - h(X^k + S^k)\right|$ *and* $\left|F_s^k - f(X^k + S^k)\right|$ *respectively.*

Definition 10. *A convergent subsequence* $\{x^k\}_{k\in\mathcal{K}}$ *of Algorithm 1 iterates, for some subset of indices* $\mathcal{K}$*, is called a refining subsequence if and only if the corresponding subsequence* $\{\delta_m^k\}_{k\in\mathcal{K}}$ *converges to zero. The limit* $\hat{x}$ *is called a refined point.*

Combining the results of Corollary 1 and the compactness hypothesis of Assumption 2 was shown in [11] to be enough to ensure the existence of refining subsequences. Specifically the following holds.

Theorem 3. *Let the assumptions that were made in Corollary 1 hold. Then there exists at least one refining subsequence* $\{X^k\}_{k\in K}$ *(where* K *is a sequence of random variables) which converges almost surely to a refined point* $\hat{X}$*.*

4.2 Nonsmooth optimality conditions: Results for h

This subsection aims to show with probability one that Algorithm 1 generates a refining subsequence $\{X_{\text{inf}}^k\}_{k\in K}$ with refined point $\hat{X}_{\text{inf}}$ which satisfies a first-order necessary optimality condition via the Clarke generalized derivative of h. As in [11], this optimality result strongly relies on the requirement that the polling directions $d^k \in \mathbb{D}_p^k(x_{\text{inf}}^k)$ of Algorithm 1 are such that $\delta_p^k\left\|d^k\right\|_\infty$ never approaches zero for all k. The way such an expectation can be met is discussed in [11]. Indeed, by choosing the columns of the matrix $\mathbf{D}$ used in the definition of the mesh $\mathcal{M}^k$ to be the $2n$ positive and negative coordinate directions, $\delta_p^0 = 1$ and $\tau = 1/2$, the directions $\delta_p^k d^k$ were shown in [11] to satisfy $\delta_p^k\left\|d^k\right\|_\infty \geq 1$ whenever d^k is constructed by means of the so-called Householder matrix [12]. Thus, the following assumption is made for the remainder of the analysis.

Assumption 6. *Let* $d^k \in \mathbb{D}_p^k$ *be any polling direction used by Algorithm 1 at iteration* k*. Then there exists a constant* $d_{\min} > 0$ *such that* $\delta_p^k\left\|d^k\right\|_\infty \geq d_{\min}$ *for all* $k \geq 0$*.*

The main result of this subsection relies on the properties of the random function Ψ_k^h introduced next, a similar of which was used in [11].

Lemma 1. *Let the same assumptions that were made in Theorem 2 hold and assume in addition to (35) that* $\alpha\beta \in (1/2, 1)$*. Consider the random function* Ψ_k^h *with realizations* ψ_k^h *defined by*

$$\psi_k^h := \frac{h(x_{\text{inf}}^k) - h(x_{\text{inf}}^k + \delta_m^k d^k)}{\delta_p^k} \quad \textit{for all } k \geq 0,$$

where $d^k \in \mathbb{D}_p^k(x_{\text{inf}}^k)$ *denotes any available polling direction around* x_{inf}^k *at iteration* k*. Then the following holds,*

$$\liminf_{k\to+\infty} \Psi_k^h \leq 0 \textit{ almost surely.} \tag{38}$$

The following definition of refining directions [7, 12] will be useful in the analysis.

Definition 11. *Let $\hat{x}$ be the refined point associated to a convergent refining subsequence $\{x^k\}_{k\in\mathcal{K}}$. A direction v is said to be a refining direction for $\hat{x}$ if and only if there exists an infinite subset $\mathcal{L} \subseteq \mathcal{K}$ with polling directions $d^k \in \mathbb{D}_p^k(x^k)$ such that $v = \lim\limits_{k\in\mathcal{L}} \frac{d^k}{\left\|d^k\right\|_\infty}$.*

The analysis in this subsection also relies on the following definitions [8]. The Clarke generalized derivative $h^\circ(\hat{x}; v)$ of h at $\hat{x} \in \mathcal{X}$ in the direction $v \in \mathbb{R}^n$ is defined by

$$h^\circ(\hat{x}; v) := \limsup_{\substack{y\to\hat{x},\ y\in\mathcal{X} \\ t\searrow 0,\ y+tv\in\mathcal{X}}} \frac{h(y+tv) - h(y)}{t}. \tag{39}$$

As highlighted in [8], this definition from [36] is a generalization of the original one by Clarke [25] to the case where the constraints violation function h is not defined outside $\mathcal{X}$.

The analysis involves a specific cone $T_{\mathcal{X}}^H(\hat{x}_{\text{inf}})$ called the hypertangent cone [50] to $\mathcal{X}$ at $\hat{x}_{\text{inf}}$. The hypertangent cone to a subset $\mathcal{O} \subseteq \mathcal{X}$ at $\hat{x}$ is defined by

$$T_{\mathcal{O}}^H(\hat{x}) := \{v \in \mathbb{R}^n : \exists\bar{\epsilon} > 0 \text{ such that } y + tw \in \mathcal{O}\ \forall y \in \mathcal{O} \cap \mathcal{B}_{\bar{\epsilon}}(\hat{x}), w \in \mathcal{B}_{\bar{\epsilon}}(v) \text{ and } 0 < t < \bar{\epsilon}\}.$$

Next is stated a lemma [8] from elementary analysis, that will be useful latter in the present analysis.

Lemma 2. *If $\{a_k\}$ is a bounded real sequence and $\{b_k\}$ is a convergent real sequence, then*

$$\limsup_k (a_k + b_k) = \limsup_k a_k + \lim_k b_k.$$

The next result is a stochastic variant of Theorem 3.5 in [8]. Since the inequality $h(x_{\text{inf}}^k + \delta_m^k d^k) - h(x_{\text{inf}}^k) \geq 0$ on which relies the latter theorem does not hold in the present stochastic setting, then the proof of the result below is based on the random function Ψ_k^h $\liminf$-type result of Lemma 1.

Theorem 4. *Let Assumptions 1, 6 and all the assumptions made in Theorem 2 and Lemma 1 hold. Then Algorithm 1 generates a convergent ε-infeasible refining subsequence $\{X_{\text{inf}}^k\}_{k\in K}$, for some sequence $K \subseteq K'$ of random variables satisfying $\lim_{K'} \Psi_k^h \leq 0$ almost surely, such that if $\hat{x}_{\text{inf}} \in \mathcal{X}$ is a refined point for a realization $\{x_{\text{inf}}^k\}_{k\in\mathcal{K}}$ of $\{X_{\text{inf}}^k\}_{k\in K}$ for which the events $\Delta_p^k \to 0$ and $\lim_{K'} \Psi_k^h \leq 0$ both occur, and if $v \in T_{\mathcal{X}}^H(\hat{x}_{\text{inf}})$ is a refining direction for $\hat{x}_{\text{inf}}$, then $h^\circ(\hat{x}_{\text{inf}}; v) \geq 0$. In particular, this means that*

$$\begin{aligned}\mathbb{P}\Bigg(\Big\{\omega \in \Omega : \exists K(\omega) \subseteq \mathbb{N} \textit{ and } \exists\hat{X}_{\text{inf}}(\omega) &= \lim_{k\in K(\omega)} X_{\text{inf}}^k(\omega), \hat{X}_{\text{inf}}(\omega) \in \mathcal{X},\ \textit{such that} \\ &\forall V(\omega) \in T_{\mathcal{X}}^H(\hat{X}_{\text{inf}}(\omega)),\ h^\circ(\hat{X}_{\text{inf}}(\omega); V(\omega)) \geq 0\Big\}\Bigg) = 1.\end{aligned} \tag{40}$$

Next is stated a stochastic variant of a result in [8], showing that Clarke stationarity is ensured when the set of refining directions is dense in a nonempty hypertangent cone to $\mathcal{X}$.

Corollary 3. *Let all assumptions that were made in Theorem 4 hold. Let $\{X_{\text{inf}}^k\}_{k\in K}$ be the ε-infeasible refining subsequence of Theorem 4, with realizations $\{x_{\text{inf}}^k\}_{k\in\mathcal{K}}$ which converges to a refined point $\hat{x}_{\text{inf}} \in \mathcal{X}$. If the set of refining directions for $\hat{x}_{\text{inf}}$ is dense in $T_{\mathcal{X}}^H(\hat{x}_{\text{inf}}) \neq \emptyset$, then $\hat{x}_{\text{inf}}$ is a Clarke stationary point for the problem $\min\limits_{x\in\mathcal{X}} h(x)$.*

Proof. The proof of this result is almost identical to the proof of a similar result (Corollary 3.6) in [8] and hence will not be presented here again. □

4.3 Nonsmooth optimality conditions: Results for f

The analysis presented in this subsection assumes that Algorithm 1 generates infinitely many ε-feasible points. It aims to show with probability one that StoMADS-PB generates a refining subsequence $\{X^k_{\text{feas}}\}_{k\in K}$ with refined point $\hat{X}_{\text{feas}}$, which satisfies a first-order necessary optimality condition based on the Clarke derivative of f. The following lemma will be useful latter in the analysis.

Lemma 3. *Let the same assumptions that were made in Theorem 2 hold and assume in addition to (35) that $\alpha\beta \in (1/2, 1)$. Assume that the random time T with realizations t is finite almost surely. Consider the random function $\Psi^{f,T}_k$ with realizations $\psi^{f,t}_k$ defined by*

$$\psi^{f,t}_k := \frac{f(x^{k\vee t}_{\text{feas}}) - f(x^{k\vee t}_{\text{feas}} + \delta^k_m d^k)}{\delta^k_p} \quad \textit{for all } k \geq 0,$$

where $k \vee t := \max\{k, t\}$ and d^k denotes any available polling direction around $x^{k\vee t}_{\text{feas}}$ at iteration k. Then the following holds,

$$\liminf_{k\to+\infty} \Psi^{f,T}_k \leq 0 \textit{ almost surely.} \tag{41}$$

Now let prove that the almost sure limit $\hat{X}_{\text{feas}}$ of any convergent refining subsequence of ε-feasible iterates which drives the random estimated violations $H^k_0(X^k_{\text{feas}})$ to zero almost surely, satisfies $\mathbb{P}\left(\hat{X}_{\text{feas}} \in \mathcal{D}\right) = 1$. First, notice that the existence of such a refining subsequence can be assumed. Indeed, it is known from Theorem 1 that true iterations occur infinitely often provided that estimates and bounds are sufficiently accurate. In addition, every ε-feasible point x^k_{feas} newly accepted by Algorithm 1 satisfies $u^k_0(x^k_{\text{feas}}) = 0$, which implies that $h^k_0(x^k_{\text{feas}}) = 0$, thus leading to the overall conclusion that $\liminf_{k\to+\infty} H^k_0(X^k_{\text{feas}}) = 0$ almost surely, which is implicitly assumed next.

Theorem 5. *Let all the assumptions of Lemma 3 hold. Let $\hat{X}_{\text{feas}}$ be the almost sure limit of a convergent ε-feasible refining subsequence $\{X^{k\vee T}_{\text{feas}}\}_{k\in K}$ for which $\lim_{k\in K} H^k_0(X^{k\vee T}_{\text{feas}}) = 0$ almost surely. Then*

$$\mathbb{P}\left(\hat{X}_{\text{feas}} \in \mathcal{D}\right) = 1. \tag{42}$$

The following result is a stochastic variant of Theorem 3.3 in [8].

Theorem 6. *Let Assumptions 1, 6 and all assumptions that were made in Theorem 2 and Lemma 3 hold. Let $\{X^{k\vee T}_{\text{feas}}\}_{k\in K}$ be an almost surely convergent ε-feasible refining subsequence, for some sequence K of random variables satisfying $\lim_K \Psi^{f,T}_k \leq 0$ and $\lim_K H^k_0(X^{k\vee T}_{\text{feas}}) = 0$ almost surely. Then, if $\hat{x}_{\text{feas}} \in \mathcal{D}$ is a refined point for a realization $\{x^{k\vee t}_{\text{feas}}\}_{k\in\mathcal{K}}$ of $\{X^{k\vee T}_{\text{feas}}\}_{k\in K}$ for which the events $\Delta^k_p \to 0$, $\lim_K \Psi^{f,T}_k \leq 0$ and $\lim_K H^k_0(X^{k\vee T}_{\text{feas}}) = 0$ occur, and if $v \in T^H_{\mathcal{D}}(\hat{x}_{\text{feas}})$ is a refining direction for $\hat{x}_{\text{feas}}$, then $f^\circ(\hat{x}_{\text{feas}}; v) \geq 0$. In particular, this means that*

$$\begin{aligned}\mathbb{P}\Bigg(\Big\{\omega \in \Omega : \exists K(\omega) \subseteq \mathbb{N} \textit{ and } \exists \hat{X}_{\text{feas}}(\omega) &= \lim_{k\in K(\omega)} X^{k\vee T}_{\text{feas}}(\omega), \hat{X}_{\text{feas}}(\omega) \in \mathcal{D},\ \textit{such that} \\ &\forall V(\omega) \in T^H_{\mathcal{D}}(\hat{X}_{\text{feas}}(\omega)),\ f^\circ(\hat{X}_{\text{feas}}(\omega); V(\omega)) \geq 0\Big\}\Bigg) = 1.\end{aligned} \tag{43}$$

Corollary 4. *Let all assumptions that were made in Theorem 6 hold. Let $\{X^{k\vee T}_{\text{feas}}\}_{k\in K}$ be the ε-feasible refining subsequence of Theorem 6, with realizations $\{x^{k\vee t}_{\text{feas}}\}_{k\in\mathcal{K}}$ which converges to a refined point $\hat{x}_{\text{feas}} \in \mathcal{D}$. If the set of refining directions for $\hat{x}_{\text{feas}}$ is dense in $T^H_{\mathcal{D}}(\hat{x}_{\text{feas}}) \neq \emptyset$, then $\hat{x}_{\text{feas}}$ is a Clarke stationary point for (1).*

Proof. The proof of this result is almost identical to the proof of a similar result (Corollary 3.4) in [8] and hence will not be presented here again. □

5 Computational study

This section illustrates the performance and the efficiency of StoMADS-PB using noisy variants of 42 continuous analytical computational constrained problems from the optimization literature. The sources and characteristics of these problems are summarized in Table 1. The number of variables ranges from $n = 2$ to $n = 20$, where every problem has at least one constraint ($m > 0$) other than bound constraints. In order to show the capability of StoMADS-PB to cope with noisy constrained problems compared to MADS with PB [8] referred to as MADS-PB, the latter algorithm is compared to several variants of StoMADS-PB. For all numerical investigations of both algorithms, only the POLL step is used, i.e., no SEARCH step is involved. The OrthoMADS-$2n$ directions [1] are used for the POLL which is ordered by means of an opportunistic strategy [12]. MADS-PB and all the proposed variants of StoMADS-PB are implemented in MATLAB.

The stochastic variants of the 42 abovementioned deterministic constrained optimization problems are solved using three different infeasible initial points for a total of 126 problem instances. Inspired from [11], such stochastic variants are constructed by additively perturbing the objective f by a random variable Θ_0 and each constraint $c_j, j = 1, 2, \dots, m$ by a random variable Θ_j as follows

$$f_{\Theta_0}(x) = f(x) + \Theta_0 \quad \text{and} \quad c_{\Theta_j}(x) = c_j(x) + \Theta_j, \text{ for all } j \in J, \tag{44}$$

where Θ_0 is uniformly generated in the interval $I(\sigma, x^0, f) = [-\sigma\left|f(x^0) - f^*\right|, \sigma\left|f(x^0) - f^*\right|]$ and Θ_j is uniformly generated in $I(\sigma, x^0, c_j) = [-\sigma\left|c_j(x^0)\right|, \sigma\left|c_j(x^0)\right|]$. The scalar $\sigma > 0$ is used to define different noise levels, x^0 denotes an initial point and f^* is the best known feasible minimum value of f. The random variables $\Theta_0, \Theta_1, \dots, \Theta_m$ are independent. For the remainder of the study, the process which returns the vector $[f_{\Theta_0}(x), c_{\Theta_1}(x), c_{\Theta_2}(x), \dots, c_{\Theta_m}(x)]$ when provided the input x will be referred to as noisy blackbox.

The MADS-PB algorithm [8] of which StoMADS-PB is a stochastic variant and to which the latter is compared is an iterative direct-search method originally developed for deterministic constrained blackbox optimization. In MADS-PB, feasibility is sought by progressively decreasing in an adaptive manner a threshold imposed on a constraint violation function into which all the constraint violations are aggregated. Any trial point with a constraint violation value greater than that threshold is rejected out of hand. Full description of MADS-PB iterations and useful information for better understanding of the algorithm behavior can also be found in [12].

The relative performance and efficiency of algorithms are assessed by performance profiles [31, 46] and data profiles [46], which require to define for a given computational problem a convergence test. For each of the 126 problems, denote by x^N the best feasible iterate found after N evaluations of the noisy blackbox and let x^* be the best feasible point obtained by all tested algorithms on all run instances. Then, the convergence test from [14] used for the experiments is defined as follows:

$$f(x^N) \leq f(x^*) + \tau(\bar{f}_{\text{feas}} - f(x^*)), \tag{45}$$

where, $\tau \in [0, 1]$ is the convergence tolerance and $\bar{f}_{feas}$ is a reference value obtained by taking the average of the first feasible f function values over all run instances of a given computational problem for all algorithms. If no feasible point is found, then the convergence test fails. Otherwise, a problem is said to be successfully solved within the tolerance τ if (45) holds. As highlighted in [14], $\bar{f}_{feas} = f(x^0)$ for unconstrained computational problems, where x^0 denotes the initial point.

The horizontal axis of the performance profiles shows the ratio of the number of noisy objective function evaluations while the fraction of computational problems solved within the convergence tolerance τ is shown on the vertical axis. On the horizontal axis of the data profiles is shown the number of function calls to the noisy blackbox divided by $(n+1)$[1] while the vertical axis shows the proportion of computational problems solved by all run instances of a given algorithm within a tolerance τ. As emphasized in [12], performance profiles capture information on speed of convergence (i.e., the quality of a given algorithm's output in terms of the objective function evaluations) and robustness (i.e., the fraction of computational problems solved) in a compact graphical format, while data profiles also examine the robustness and efficiency from a different perspective.

Now recall that in StoMADS-PB, according to Section 3.2, the noisy blackbox needs to be evaluated many times at a given point in order to compute function estimates unlike the MADS-PB method where it is evaluated only once at each point. But since a limited budget of $1000(n+1)$ noisy blackbox evaluations is set in all the experiments, that is, since MADS-PB and all variants of StoMADS-PB stop as soon as the number of noisy blackbox evaluations reaches $1000(n+1)$, only few calls to the blackbox need to be used when computing StoMADS-PB function estimates. However, given that such estimates are required to be sufficiently accurate in order for the solutions to be satisfactory, a procedure inspired from [11] aiming at improving the estimates accuracy by making use of available samples at a given current point is proposed. Note in passing that the proposed computation procedure is very efficient in practice as highlighted in [11] even though it is inherently biased. The following computation scheme is described only for $f_0^k(x^k)$ but is the same for $f_s^k(x^k+s^k)$, $c_{j,0}^k(x^k)$ and $c_{j,s}^k(x^k+s^k)$, for all $j \in J$. First, let mention that during the optimization, all trial points x^k used by StoMADS-PB and all corresponding values $f_{\Theta_0}(x^k)$ are stored in a cache. When constructing an estimate of $f(x^k)$ at the iteration $k \geq 1$, denote by $a^k(x^k)$[2] the number of sample values of $f_{\Theta_0}(x^k)$ available in the cache from previous blackbox evaluations until iteration $k-1$. Since all the values of the noisy objective function f_{Θ_0} are always computed independently of each other, the aforementioned sample values can be considered as independent realizations $f_{\theta_{0,1}}(x^k), f_{\theta_{0,2}}(x^k), \dots, f_{\theta_{0,a^k(x^k)}}(x^k)$ of $f_{\Theta_0}(x^k)$, where for all $\ell = 1, 2, \dots, a^k(x^k)$, $\theta_{0,\ell}$ is a realization of the random variable $\Theta_{0,\ell}$ following the same distribution as Θ_0. Now let $n^k \geq 1$ be the number of blackbox evaluations at x^k and consider the following independent realizations $\theta_{0,a^k(x^k)+1}, \theta_{0,a^k(x^k)+2}, \dots, \theta_{0,a^k(x^k)+n^k}$ of Θ_0. Then, an estimate $f_0^k(x^k)$ of $f(x^k)$ is computed according to,

$$f_0^k(x^k) = \frac{1}{p^k}\sum_{\ell=1}^{p^k} f_{\theta_{0,\ell}}(x^k), \tag{46}$$

where $p^k = n^k + a^k(x^k)$ is the sample size.

Same values are used to initialize most of the common parameters to StoMADS-PB and MADS-PB. Specifically, the mesh refining parameter $\tau = 1/2$, the frame center trigger $\rho = 0.1$ and $\delta_m^0 = \delta_p^0 = 1$. Nevertheless in MADS-PB, the initial barrier threshold is set equal its default value, i.e., $h_{\max}^0 = +\infty$ [8] while in StoMADS-PB it equals $u_0^0(x_{\text{inf}}^0)$, with $u_0^k(x^k)$ defined in (4) for all $k \in \mathbb{N}$.

[1] $n+1$ is the number of evaluations required to construct a linear interpolant or a simplex gradient [12] in $\mathbb{R}^n$ [14, 46].

[2] It is implicitly assumed without any loss of generality that $a^k(x^k) \geq 1$.

The default values of Algorithm 1 parameters $\gamma > 2$ and $\varepsilon > 0$[3] are borrowed from [11] in which StoMADS, an unconstrained stochastic variant of MADS [7] is introduced. Specifically, $\gamma = 17$ and $\varepsilon = 0.01$.

Table 1: Description of the set of 42 analytical problems.

No	Name	Source	n	m	Bnds	No	Name	Source	n	m	Bnds
1	ANGUN	[54]	2	1	Yes	22	MAD1	[43]	2	1	No
2	BARNES	[51]	2	3	Yes	23	MAD2	[43]	2	1	No
3	BERTSIMAS	[19]	2	2	No	24	MAD6	[43]	7	7	Yes
4	CHENWANG_F2	[24]	8	6	Yes	25	MEZMONTES	[44]	2	2	Yes
5	CHENWANG_F3	[24]	10	8	Yes	26	NEW-BRANIN	[54]	2	1	Yes
6	CONSTR-BRANIN	[54]	2	1	Yes	27	OPTENG-BENCH4	[37]	2	1	Yes
7	CRESCENT	[8]	10	2	No	28	OPTENG-BENCH5	[37]	2	3	Yes
8	DEMBO5	[43]	8	3	Yes	29	OPTENG-RBF	[37]	3	4	Yes
9	DISK	[8]	10	1	No	30	PENTAGON	[43]	6	15	No
10	G23	[9]	3	2	Yes	31	PRESSURE-VESSEL	[44]	4	4	Yes
11	G210	[9]	10	2	Yes	32	SASENA	[54]	2	1	Yes
12	G220	[9]	20	2	Yes	33	SNAKE	[8]	2	2	No
13	GOMEZ	[54]	2	1	Yes	34	SPEED-REDUCER	[44]	7	11	Yes
14	HS15	[35]	2	2	Yes	35	SPRING	[51]	3	4	Yes
15	HS19	[35]	2	2	Yes	36	TAOWANG_F1	[53]	2	2	Yes
16	HS22	[35]	2	2	No	37	TAOWANG_F2	[53]	7	4	Yes
17	HS23	[35]	2	5	Yes	38	WELDED-BEAM	[44]	4	7	Yes
18	HS29	[35]	3	1	No	39	WONG2	[43]	10	3	No
19	HS43	[35]	4	3	No	40	ZHAOWANG_F5	[55]	13	9	Yes
20	HS108	[35]	9	13	Yes	41	ZILONG_G4	[54]	5	1	Yes
21	HS114	[35]	10	5	Yes	42	ZILONG_G24	[54]	2	1	Yes

Table 2: Percentage of problems solved for each noise level σ within a convergence tolerance τ.

	$\tau = 10^{-1}$			$\tau = 10^{-3}$		
Algorithm	$\sigma = 0.01$	$\sigma = 0.03$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.03$	$\sigma = 0.05$
StoMADS-PB $n^k = 1$	74.6%	78.57%	73.02%	44.44%	45.24%	45.24%
StoMADS-PB $n^k = 2$	74.6%	76.98%	76.19%	47.62%	47.62%	50.79%
StoMADS-PB $n^k = 3$	76.19%	65.08%	66.67%	48.41%	41.27%	38.10%
MADS-PB	69.5%	64.29%	54.76%	41.27%	36.51%	29.37%

Three variants of StoMADS-PB corresponding to $n^k = 1, n^k = 2$ and $n^k = 3$ are compared to MADS-PB. The data and performance profiles used for the comparisons are depicted on Figures 2, 4 and 6 and Figures 3, 5 and 7. Three levels of noise are used during the experiments, which correspond to $\sigma = 0.01$, $\sigma = 0.03$ and $\sigma = 0.05$. For a given algorithm, the estimated percentages of problems

[3] The use of ε_f instead of ε is favored in [11].

solved after $1000(n+1)$ noisy blackbox evaluations for each noise level within a convergence tolerance τ are reported in Table 2. They are obtained based on the profiles graphs using MATLAB tools.

The data and performance profiles show that when given the time, StoMADS-PB eventually outperforms MADS-PB in general. Moreover as in [11], varying the value of the convergence tolerance τ in the data profiles does not significantly alter the conclusions drawn from the performance profiles. Indeed as expected, it can be easily observed from Table 2 that the higher the tolerance parameter τ, the larger the percentage of problems solved by all algorithms for a fixed noise level σ. Now notice that while for a given τ, the fraction of problems solved by MADS-PB decreases when the noise level increases from $\sigma = 0.01$ to $\sigma = 0.05$, this seems not to be the case for StoMADS-PB variants. Before giving an insight as to why, recall that in the present constrained framework, the success or failure of the convergence test (45) does not depend only on the values of the objective function f but also on whether a feasible point is found or not, unlike the framework of [11] where no constraints are involved. In fact, as highlighted in [11] from which is inspired the computation scheme (46), even though the robustness and efficiency of each StoMADS-PB variants depends on the number n^k of noisy blackbox evaluations which is constant for all k, the quality of the solutions is influenced by the sample size $p^k = n^k + a^k(x^k)$ which is not constant. On one hand, this is the reason why for $n^k = 1$, StoMADS-PB does not have the same behavior as MADS-PB. On the other hand, such computation scheme naturally favors StoMADS-PB by improving the accuracy of the estimates of its constraints function values, thus allowing it to find more feasible solutions than MADS-PB and consequently possibly solve larger fraction of problems when the noise level increases for a fixed tolerance parameter τ.

Finally, based on Table 2, it can be noticed that for a given convergence tolerance τ, varying σ seems not to have significant influences on the fractions of problems solved by StoMADS-PB variants corresponding to $n^k = 1$ and $n^k = 2$. Moreover, even though for the lowest noise level studied $\sigma = 0.01$, StoMADS-PB with $n^k = 3$ solved the most problems, the corresponding percentage is not significantly larger than that of StoMADS-PB with $n^k = 2$. For all these reasons, the latter variant seems preferable for constrained stochastic blackbox optimization problems.

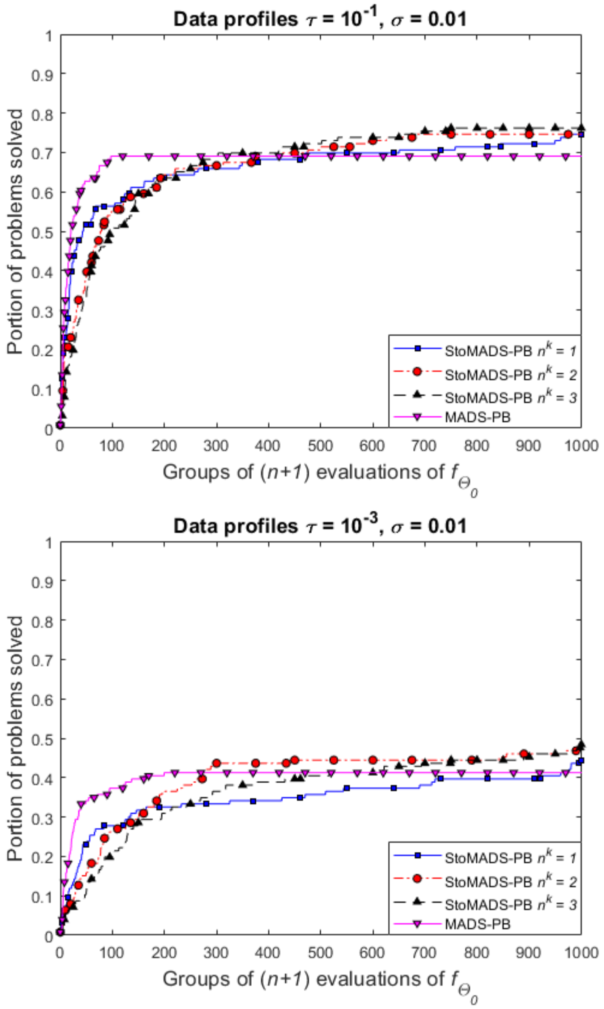

Figure 2: Data profiles for convergence tolerances $\tau = 10^{-1}$ and $\tau = 10^{-3}$, and noise level $\sigma = 0.01$ on 126 analytical constrained test problems additively perturbed in the intervals $I(\sigma, x^0, f)$ and $I(\sigma, x^0, c_j)$.

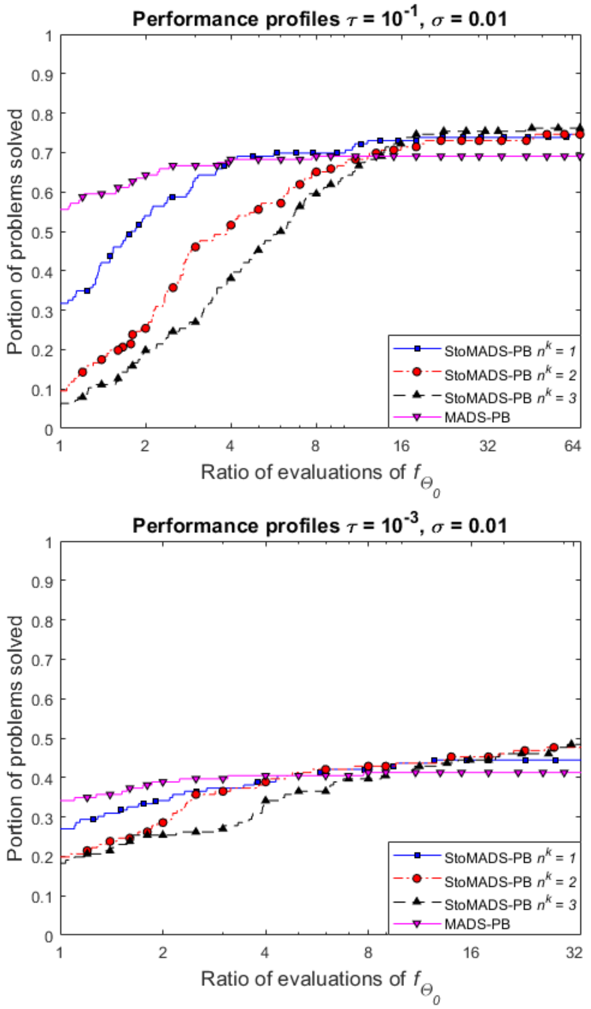

Figure 3: Performance profiles for convergence tolerances $\tau = 10^{-1}$ and $\tau = 10^{-3}$, and noise level $\sigma = 0.01$ on 126 analytical constrained test problems additively perturbed in the intervals $I(\sigma, x^0, f)$ and $I(\sigma, x^0, c_j)$.

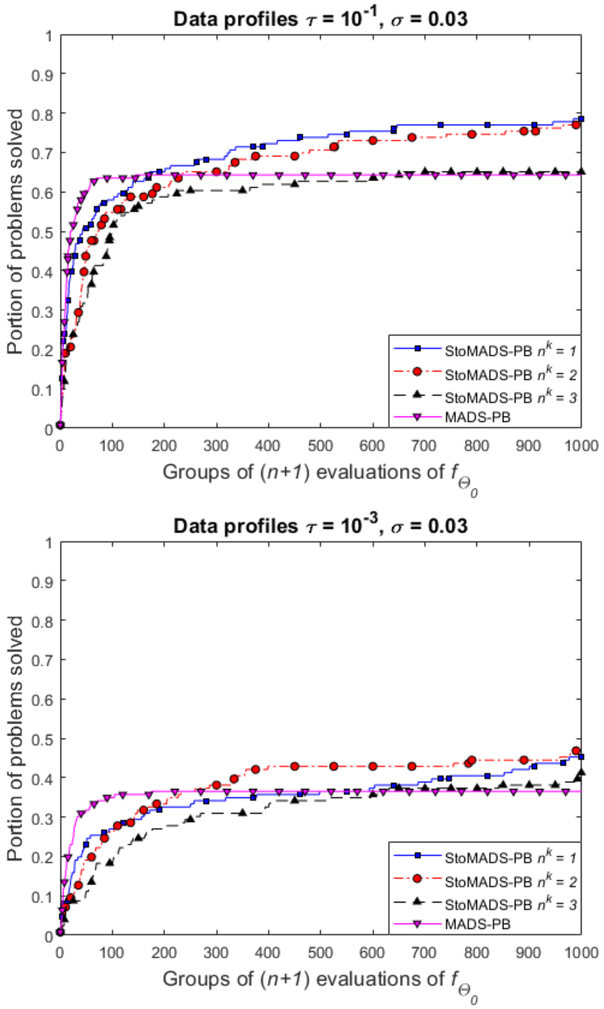

Figure 4: Data profiles for convergence tolerances $\tau = 10^{-1}$ and $\tau = 10^{-3}$, and noise level $\sigma = 0.03$ on 126 analytical constrained test problems additively perturbed in the intervals $I(\sigma, x^0, f)$ and $I(\sigma, x^0, c_j)$.

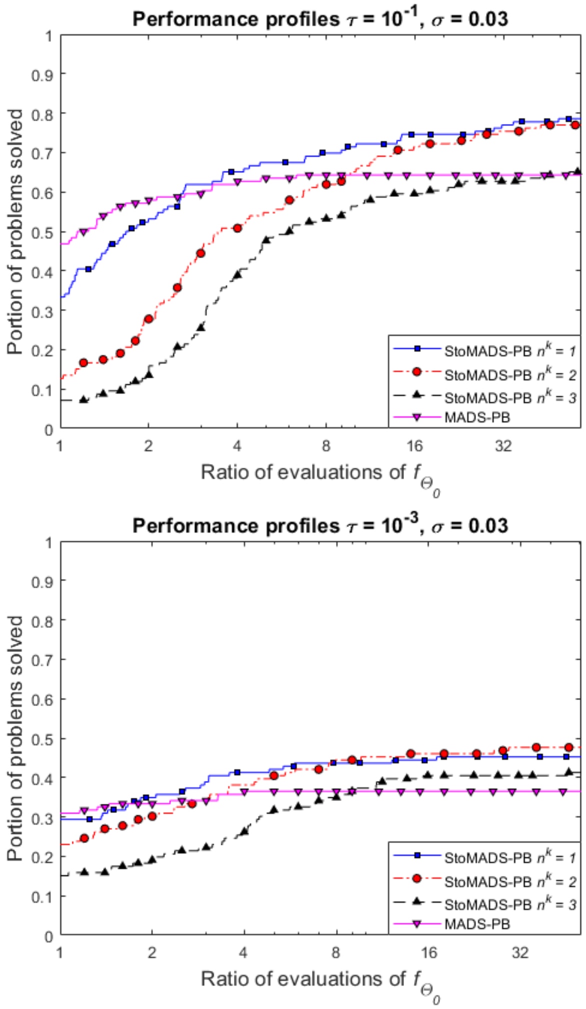

Figure 5: Performance profiles for convergence tolerances $\tau = 10^{-1}$ and $\tau = 10^{-3}$, and noise level $\sigma = 0.03$ on 126 analytical constrained test problems additively perturbed in the intervals $I(\sigma, x^0, f)$ and $I(\sigma, x^0, c_j)$.

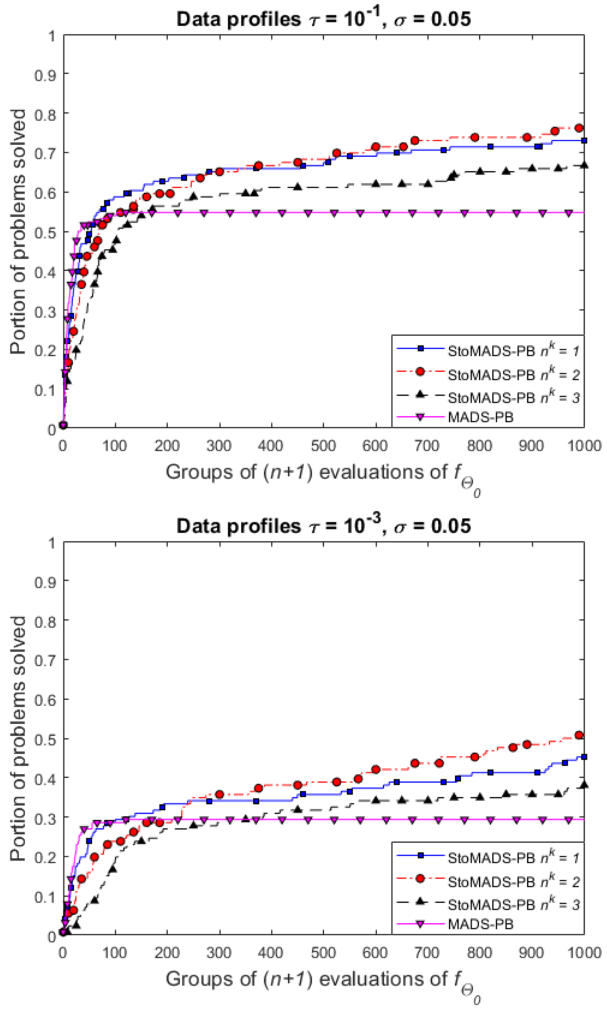

Figure 6: Data profiles for convergence tolerances $\tau = 10^{-1}$ and $\tau = 10^{-3}$, and noise level $\sigma = 0.05$ on 126 analytical constrained test problems additively perturbed in the intervals $I(\sigma, x^0, f)$ and $I(\sigma, x^0, c_j)$.

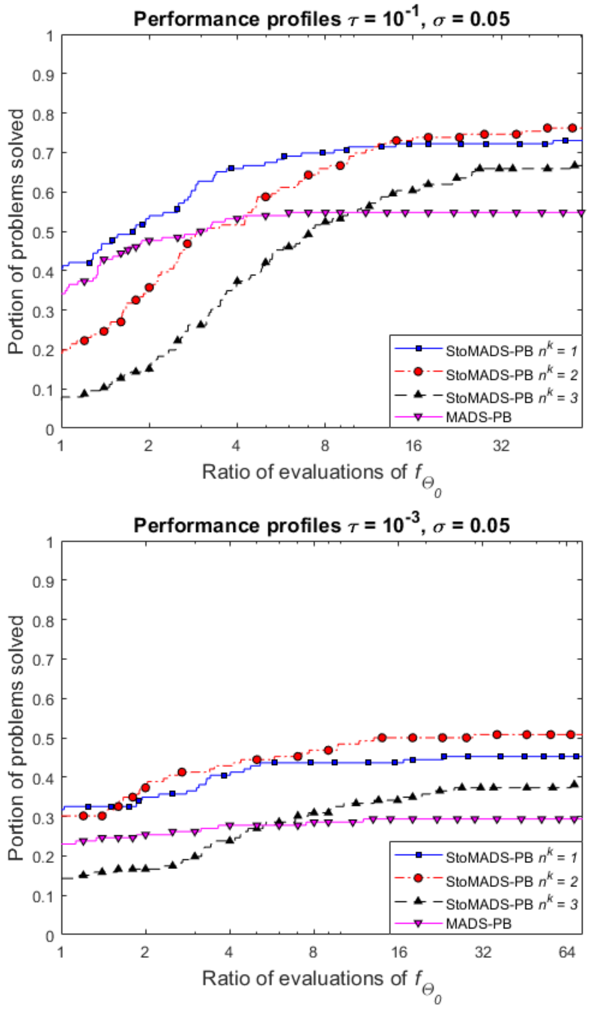

Figure 7: Performance profiles for convergence tolerances $\tau = 10^{-1}$ and $\tau = 10^{-3}$, and noise level $\sigma = 0.05$ on 126 analytical constrained test problems additively perturbed in the intervals $I(\sigma, x^0, f)$ and $I(\sigma, x^0, c_j)$.

Concluding remarks

This research proposes the StoMADS-PB algorithm for constrained stochastic blackbox optimization. The proposed method which uses an algorithmic framework similar to that of MADS considers the optimization of objective and constraints functions whose values can only be accessed through a stochastically noisy blackbox. It treats constraints using a progressive barrier approach, by aggregating their violations into a single function. It does not use any model or gradient information to find descent directions or improve feasibility unlike prior works, but instead, uses function estimates and introduces probabilistic bounds on which sufficient decrease conditions are imposed. By requiring the accuracy of such estimates and bounds to hold with sufficiently high but fixed probabilities, convergence results of StoMADS-PB are derived, most of which are stochastic variants of those of MADS.

Computational experiments conducted on several variants of StoMADS-PB on a collection of constrained stochastically noisy problems showed the proposed method to eventually outperform MADS, and also showed some of its variants to be almost robust to random noise despite the use of very inaccurate estimates.

This research is to the best of our knowledge the first to propose a stochastic directional direct-search algorithm for BBO, developed to cope with a noisy objective and constraints that are also stochastically noisy.

future research could focus on improving the proposed method to handle large-scale machine learning problems, making use for example of parallel space decomposition.

Acknowledgments

The authors are grateful to Charles Audet from Polytechnique Montréal for valuable discussions and constructive suggestions. This work is supported by the NSERC CRD RDCPJ 490744-15 grant and by an InnovÉÉ grant, both in collaboration with Hydro-Québec and Rio Tinto, and by a FRQNT fellowship.

Appendix

Now we prove a sequence of convergence results of Section 4.

Proof of Theorem 2

Proof. This theorem is proved using ideas from [11, 21, 23, 33, 40, 48]. According to Assumptions 4, the proof considers two different parts: Part 1 assumes that $T = +\infty$ almost surely, i.e., no ε-feasible iterate is found by Algorithm 1, while Part 2 considers that $T < +\infty$ almost surely. Part 1 considers two separate cases: "good bounds" and "bad bounds", each of which is broken into whether an iteration is h-Dominating, Improving or Unsuccessful. Part 2 considers three separates cases: "good estimates and good bounds", "bad estimates and good bounds" and "bad bounds", each of which is broken into whether an iteration is f-Dominating, h-Dominating, Improving or Unsuccessful.

In order to show (36), the goal of Part 1 is to show that there exists a constant $\eta > 0$ such that conditioned on the almost sure event $\{T = +\infty\}$, the following holds for all $k \in \mathbb{N}$

$$\mathbb{E}\left(\Phi_{k+1} - \Phi_k | \mathcal{F}_{k-1}^{C\cdot F}\right) \leq -\eta(\Delta_p^k)^2, \tag{47}$$

where Φ_k is the random function defined by

$$\Phi_k := \frac{\nu}{m\varepsilon} h(X_{\text{inf}}^k) + (1-\nu)(\Delta_p^k)^2, \quad \text{for all } k \in \mathbb{N}. \tag{48}$$

Indeed, assume that (47) holds. Since $\Phi_k > 0$ for all $k \in \mathbb{N}$, then summing (47) over $k \in \mathbb{N}$ and taking expectations on both sides lead to

$$\mathbb{E}\left[\sum_{k=0}^{+\infty} (\Delta_p^k)^2\right] \leq \frac{\mathbb{E}(\Phi_0)}{\eta} = \frac{\Phi_0}{\eta}, \tag{49}$$

That is, (36) holds. Then, making use of the following random function

$$\Phi_k^T := \frac{\nu}{\varepsilon}(f(X_{\text{feas}}^{k\vee T}) - \kappa_{\min}^f) + \frac{\nu}{m\varepsilon} h(X_{\text{inf}}^k) + (1-\nu)(\Delta_p^k)^2, \quad \text{for all } k \in \mathbb{N}, \tag{50}$$

where $k \vee T := \max\{k, T\}$, Part 2 aims to show that for the same previous constant $\eta > 0$, then conditioned on the almost sure event $\{T < +\infty\}$, the following holds for all $k \in \mathbb{N}$

$$\mathbb{E}\left(\Phi_{k+1}^T - \Phi_k^T | \mathcal{F}_{k-1}^{C\cdot F}\right) \leq -\eta(\Delta_p^k)^2. \tag{51}$$

Indeed, assume that (51) holds. Since $\Phi_k^T > 0$ for all $k \geq 0$, then summing (51) over $k \in \mathbb{N}$ and taking expectations on both sides, yield

$$\begin{aligned}\mathbb{E}\left[\sum_{k=0}^{+\infty} (\Delta_p^k)^2\right] &\leq \frac{\mathbb{E}\left(\Phi_0^T\right)}{\eta} = \frac{1}{\eta}\left[\frac{\nu}{\varepsilon}\left(\mathbb{E}\left[f(X_{\text{feas}}^T)\right] - \kappa_{\min}^f\right) + \frac{\nu}{m\varepsilon} h(x_{\text{inf}}^0) + (1-\nu)(\delta_p^0)^2\right] \\ &\leq \frac{1}{\eta}\left[\frac{\nu}{\varepsilon}\left(\kappa_{\max}^f - \kappa_{\min}^f\right) + \frac{\nu}{m\varepsilon} h(x_{\text{inf}}^0) + (1-\nu)(\delta_p^0)^2\right] =: \mu,\end{aligned} \tag{52}$$

where the last inequality in (52) follows from the inequality $f(X_{\text{feas}}^k) \leq \kappa_{\max}^f$ for all $k \geq 0$, due to Proposition 5, and the fact that T is finite almost surely.

The remainder of the proof is devoted to showing that (47) and (51) hold. The following events are introduced for the sake of clarity in the analysis.

$\mathcal{D}_f$:= {The iteration is f-Dominating}, $\mathcal{D}_h$:= {The iteration is h-Dominating},
$\mathcal{I}$:= {The iteration is Improving}, $\mathcal{U}$:= {The iteration is Unsuccessful}.

Part 1 ($T = +\infty$ almost surely). The random function Φ_k defined in (48) will be shown to satisfy (47) with $\eta = \frac{1}{2}\alpha\beta(1-\nu)(1-\tau^2)$, no matter the change led in the objective function f by the ε-infeasible iterates encountered by Algorithm 1. Moreover, since T is infinite almost surely, then no iteration of Algorithm 1 can be f-Dominating. Two separate cases are distinguished and all that follows is conditioned on the almost sure event $\{T = +\infty\}$.

Case 1 (Good bounds, $\mathbb{1}_{I_k} = 1$). No matter the type of iteration which occurs, the random function Φ_k is shown to decrease and the smallest decrease is shown to happen on unsuccessful iterations, thus yielding the following conclusion

$$\mathbb{E}\left[\mathbb{1}_{I_k}(\Phi_{k+1} - \Phi_k) | \mathcal{F}_{k-1}^{C\cdot F}\right] \leq -\alpha(1-\nu)(1-\tau^2)(\Delta_p^k)^2. \tag{53}$$

(i) The iteration is h-Dominating ($\mathbb{1}_{\mathcal{D}_h} = 1$). The iteration is h-Dominating and the bounds are good, so a decrease occurs in h according to (6) as follows

$$\mathbb{1}_{I_k}\mathbb{1}_{\mathcal{D}_h}\frac{\nu}{m\varepsilon}(h(X_{\text{inf}}^{k+1}) - h(X_{\text{inf}}^k)) \leq -\mathbb{1}_{I_k}\mathbb{1}_{\mathcal{D}_h}\nu(\gamma - 2)(\Delta_p^k)^2 \tag{54}$$

The frame size parameter is updated according to $\Delta_p^{k+1} = \min\{\tau^{-1}\Delta_p^k, \delta_{\max}\}$, which implies that

$$\mathbb{1}_{I_k}\mathbb{1}_{\mathcal{D}_h}(1-\nu)[(\Delta_p^{k+1})^2 - (\Delta_p^k)^2] \leq \mathbb{1}_{I_k}\mathbb{1}_{\mathcal{D}_h}(1-\nu)(\tau^{-2} - 1)(\Delta_p^k)^2. \tag{55}$$

Then, by choosing ν according to (34), the right-hand side term of (54) dominates that of (55). Specifically, the following holds

$$-\nu(\gamma - 2)(\Delta_p^k)^2 + (1-\nu)(\tau^{-2} - 1)(\Delta_p^k)^2 \leq -\frac{1}{2}\nu(\gamma - 2)(\Delta_p^k)^2. \tag{56}$$

Then combining (54), (55) and (56) leads to

$$\mathbb{1}_{I_k}\mathbb{1}_{\mathcal{D}_h}(\Phi_{k+1} - \Phi_k) \leq -\mathbb{1}_{I_k}\mathbb{1}_{\mathcal{D}_h}\frac{1}{2}\nu(\gamma - 2)(\Delta_p^k)^2. \tag{57}$$

(ii) The iteration is Improving ($\mathbb{1}_{\mathcal{I}} = 1$). The iteration is Improving and the bounds are good, so again, a decrease occurs in h according to (6). Moreover, Δ_p^k is updated as at h-Dominating iterations. Thus, the change in Φ_k follows from (57) by replacing $\mathbb{1}_{\mathcal{D}_h}$ by $\mathbb{1}_{\mathcal{I}}$. Specifically,

$$\mathbb{1}_{I_k}\mathbb{1}_{\mathcal{I}}(\Phi_{k+1} - \Phi_k) \leq -\mathbb{1}_{I_k}\mathbb{1}_{\mathcal{I}}\frac{1}{2}\nu(\gamma - 2)(\Delta_p^k)^2. \tag{58}$$

(iii) The iteration is Unsuccessful ($\mathbb{1}_{\mathcal{U}} = 1$). There is a change of zero in h function values while the frame size parameter is decreased. Consequently,

$$\mathbb{1}_{I_k}\mathbb{1}_{\mathcal{U}}(\Phi_{k+1} - \Phi_k) = -\mathbb{1}_{I_k}\mathbb{1}_{\mathcal{U}}(1-\nu)(1-\tau^2)(\Delta_p^k)^2 \tag{59}$$

Then, the choice of ν according to (34) and the fact that $1 - \tau^2 < \tau^{-2} - 1$ ensures that unsuccessful iterations, more precisely (59), provide the worst case decrease when compared to (57) and (58). Specifically, the following holds

$$-\frac{1}{2}\nu(\gamma - 2)(\Delta_p^k)^2 \leq -(1-\nu)(1-\tau^2)(\Delta_p^k)^2. \tag{60}$$

Thus, it follows from (57), (58), (59) and (60) that the change in Φ_k is bounded as follows

$$\mathbb{1}_{I_k}(\Phi_{k+1} - \Phi_k) = \mathbb{1}_{I_k}(\mathbb{1}_{\mathcal{D}_h} + \mathbb{1}_{\mathcal{I}} + \mathbb{1}_{\mathcal{U}})(\Phi_{k+1} - \Phi_k) \leq -\mathbb{1}_{I_k}(1-\nu)(1-\tau^2)(\Delta_p^k)^2. \tag{61}$$

Since Assumption 3 holds, then taking conditional expectations with respect to $\mathcal{F}_{k-1}^{C\cdot F}$ on both sides of the inequality in (61) leads to (53).

Case 2 (Bad bounds, $\mathbb{1}_{\bar{I}_k} = 1$). Since the bounds are bad, Algorithm 1 can accept an iterate which leads to an increase in h and Δ_p^k, and hence in Φ_k. Such an increase in Φ_k is controlled making use of (14). Then, the probability of outcome (Part 1, Case 2) is adjusted to be sufficiently small so that Φ_k can be reduced sufficiently in expectation. More precisely, the following will be proved

$$\mathbb{E}\left[\mathbb{1}_{\bar{I}_k}(\Phi_{k+1} - \Phi_k)|\mathcal{F}_{k-1}^{C\cdot F}\right] \leq 2\nu(1-\alpha)^{1/2}(\Delta_p^k)^2. \tag{62}$$

(i) The iteration is h-Dominating ($\mathbb{1}_{\mathcal{D}_h} = 1$). The change in h is bounded as follows

$$\begin{aligned}
&\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{D}_h}\frac{\nu}{m\varepsilon}(h(X_{\text{inf}}^{k+1}) - h(X_{\text{inf}}^{k})) \\
&\qquad \leq \; \mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{D}_h}\frac{\nu}{m\varepsilon}\left[(H_s^k - H_0^k) + \left|h(X_{\text{inf}}^{k+1}) - H_s^k\right| + \left|h(X_{\text{inf}}^{k}) - H_0^k\right|\right] \\
&\qquad \leq \; \mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{D}_h}\nu\left[-\gamma(\Delta_p^k)^2 + \frac{1}{m\varepsilon}\left(\left|h(X_{\text{inf}}^{k+1}) - H_s^k\right| + \left|h(X_{\text{inf}}^{k}) - H_0^k\right|\right)\right], \qquad (63)
\end{aligned}$$

where (63) follows from $H_s^k - H_0^k \leq -\gamma m\varepsilon(\Delta_p^k)^2$ which is satisfied for every h-Dominating iteration. Moreover, the change in Δ_p^k can be obtained simply by replacing in (55) $\mathbb{1}_{I_k}$ by $\mathbb{1}_{\bar{I}_k}$ as follows

$$\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{D}_h}(1-\nu)[(\Delta_p^{k+1})^2 - (\Delta_p^k)^2] \leq \mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{D}_h}(1-\nu)(\tau^{-2}-1)(\Delta_p^k)^2. \tag{64}$$

Since choosing ν according to (34) ensures that $-\nu\gamma(\Delta_p^k)^2 + (1-\nu)(\tau^{-2}-1)(\Delta_p^k)^2 \leq 0$, then combining (63) and (64), yields

$$\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{D}_h}(\Phi_{k+1} - \Phi_k) \leq \mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{D}_h}\frac{\nu}{m\varepsilon}\left(\left|h(X_{\text{inf}}^{k+1}) - H_s^k\right| + \left|h(X_{\text{inf}}^{k}) - H_0^k\right|\right). \tag{65}$$

(ii) The iteration is Improving ($\mathbb{1}_{\mathcal{I}} = 1$). Δ_p^k is updated as at h-Dominating iterations and because of bad bounds, the increase in h is bounded following (63). Thus, the bound on the change in Φ_k can be obtained by replacing $\mathbb{1}_{\mathcal{D}_h}$ by $\mathbb{1}_{\mathcal{I}}$ in (65) as follows

$$\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{I}}(\Phi_{k+1} - \Phi_k) \leq \mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{I}}\frac{\nu}{m\varepsilon}\left(\left|h(X_{\text{inf}}^{k+1}) - H_s^k\right| + \left|h(X_{\text{inf}}^{k}) - H_0^k\right|\right). \tag{66}$$

(iii) The iteration is Unsuccessful ($\mathbb{1}_{\mathcal{U}} = 1$). The change in h is zero and Δ_p^k is decreased. Thus, the change in Φ_k follows from (59) by replacing $\mathbb{1}_{I_k}$ by $\mathbb{1}_{\bar{I}_k}$ and is trivially bounded as follows

$$\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{U}}(\Phi_{k+1} - \Phi_k) \leq \mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{U}}\frac{\nu}{m\varepsilon}\left(\left|h(X_{\text{inf}}^{k+1}) - H_s^k\right| + \left|h(X_{\text{inf}}^{k}) - H_0^k\right|\right). \tag{67}$$

Finally, it follows from (65), (66), (67) and the inequality $\mathbb{1}_{\bar{I}_k} \leq 1$, that

$$\mathbb{1}_{\bar{I}_k}(\Phi_{k+1} - \Phi_k) \leq \frac{\nu}{m\varepsilon}\left(\left|h(X_{\text{inf}}^{k+1}) - H_s^k\right| + \left|h(X_{\text{inf}}^{k}) - H_0^k\right|\right), \tag{68}$$

Then, taking conditional expectations with respect to $\mathcal{F}_{k-1}^{C\cdot F}$ on both sides of (68) and using the inequalities (14) of Assumption 3, lead to (62).

Now, combining (53) and (62) yields,

$$\begin{aligned}
\mathbb{E}\left(\Phi_{k+1} - \Phi_k | \mathcal{F}_{k-1}^{C\cdot F}\right) \;&=\; \mathbb{E}\left[(\mathbb{1}_{I_k} + \mathbb{1}_{\bar{I}_k})(\Phi_{k+1} - \Phi_k)|\mathcal{F}_{k-1}^{C\cdot F}\right] \\
&\leq\; \left[-\alpha(1-\nu)(1-\tau^2) + 2\nu(1-\alpha)^{1/2}\right](\Delta_p^k)^2. \qquad (69)
\end{aligned}$$

Then, choosing α according to (35) implies that $\alpha \geq \dfrac{4\nu(1-\alpha)^{1/2}}{(1-\nu)(1-\tau^2)}$, which ensures

$$-\alpha(1-\nu)(1-\tau^2) + 2\nu(1-\alpha)^{1/2} \leq -\frac{1}{2}\alpha(1-\nu)(1-\tau^2) \leq -\frac{1}{2}\alpha\beta(1-\nu)(1-\tau^2). \tag{70}$$

Thus, (47) follows from (69) and (70) with $\eta = \frac{1}{2}\alpha\beta(1-\nu)(1-\tau^2)$.

Part 2 ($T < +\infty$ almost surely). In order to show that the random function Φ_k^T defined by

$$\Phi_k^T = \frac{\nu}{\varepsilon}(f(X_{\text{feas}}^{k\vee T}) - \kappa_{\text{min}}^f) + \frac{\nu}{m\varepsilon}h(X_{\text{inf}}^k) + (1-\nu)(\Delta_p^k)^2$$

satisfies (51) with the same constant η derived in Part 1, notice that whenever the event $\{T > k\}$ occurs, then $f(X_{\text{feas}}^{(k+1)\vee T}) - f(X_{\text{feas}}^{k\vee T}) = 0$ since $\max\{k, T\} := k \vee T = (k+1)\vee T = T$. Thus, on the event $\{T > k\}$, the random function Φ_k used in Part 1 has the same increments as Φ_k^T. Specifically,

$$\mathbb{1}_{\{T<+\infty\}}\mathbb{1}_{\{T>k\}}(\Phi_{k+1}^T - \Phi_k^T) = \mathbb{1}_{\{T<+\infty\}}\mathbb{1}_{\{T>k\}}(\Phi_{k+1} - \Phi_k).$$

Moreover, it follows from the definition of the stopping time T that no iteration can be f-Dominating as in Part 1 when the event $\{T > k\}$ occurs. Consequently, it easily follows from the analysis in Part 1 and the fact that the random variable $\mathbb{1}_{\{T>k\}}$ is $\mathcal{F}_{k-1}^{C\cdot F}$-measurable that,

$$\mathbb{1}_{\{T>k\}}\mathbb{E}\left(\Phi_{k+1}^T - \Phi_k^T | \mathcal{F}_{k-1}^{C\cdot F}\right) \leq -\eta(\Delta_p^k)^2\mathbb{1}_{\{T>k\}}. \tag{71}$$

The remainder of the proof is devoted to showing that the following holds

$$\mathbb{1}_{\{T\leq k\}}\mathbb{E}\left(\Phi_{k+1}^T - \Phi_k^T | \mathcal{F}_{k-1}^{C\cdot F}\right) \leq -\eta(\Delta_p^k)^2\mathbb{1}_{\{T\leq k\}}, \tag{72}$$

since combining (71) and (72) leads to (51), which is the remaining overall goal. In all that follows, it is assumed that the event $\{T \leq k\}$ occurs.

Case 1 (Good estimates and good bounds, $\mathbb{1}_{I_k}\mathbb{1}_{J_k} = 1$). Regardless of the iteration type, the smallest decrease in Φ_k^T is shown to happen on unsuccessful iterations, thus implying that

$$\mathbb{1}_{\{T\leq k\}}\mathbb{E}\left[\mathbb{1}_{I_k}\mathbb{1}_{J_k}(\Phi_{k+1}^T - \Phi_k^T) | \mathcal{F}_{k-1}^{C\cdot F}\right] \leq -\alpha\beta(1-\nu)(1-\tau^2)(\Delta_p^k)^2\mathbb{1}_{\{T\leq k\}}. \tag{73}$$

(i) The iteration is f-Dominating ($\mathbb{1}_{\mathcal{D}_f} = 1$). The iteration is f-Dominating and the estimates are good, so a decrease occurs in f according to (8) as follows

$$\begin{aligned}\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}\mathbb{1}_{\mathcal{D}_f}\frac{\nu}{\varepsilon}(f(X_{\text{feas}}^{(k+1)\vee T}) - f(X_{\text{feas}}^{k\vee T}))& \\ \leq\ -\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}\mathbb{1}_{\mathcal{D}_f}\nu(\gamma-2)(\Delta_p^k)^2.&\end{aligned} \tag{74}$$

Since the ε-infeasible iterate is not updated, then there is a change of zero in h. The frame size parameter is updated according to $\Delta_p^{k+1} = \min\{\tau^{-1}\Delta_p^k, \delta_{\max}\}$, thus implying that

$$\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}\mathbb{1}_{\mathcal{D}_f}(1-\nu)[(\Delta_p^{k+1})^2 - (\Delta_p^k)^2] \leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}\mathbb{1}_{\mathcal{D}_f}(1-\nu)(\tau^{-2}-1)(\Delta_p^k)^2. \tag{75}$$

Then, choosing ν according to (34) ensures that (56) holds, which implies that the right-hand side term of (74) dominates that of (75), thus leading to the inequality below

$$\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}\mathbb{1}_{\mathcal{D}_f}(\Phi_{k+1}^T - \Phi_k^T) \leq -\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}\mathbb{1}_{\mathcal{D}_f}\frac{1}{2}\nu(\gamma-2)(\Delta_p^k)^2. \tag{76}$$

(ii) The iteration is h-Dominating ($\mathbb{1}_{\mathcal{D}_h} = 1$). There is a change of zero in f since X_{feas}^k is not updated. Thus, the bound on the change in Φ_k^T follows from multiplying both sides of (57) by $\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{J_k}$, and replacing Φ_k by Φ_k^T as follows

$$\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}\mathbb{1}_{\mathcal{D}_h}(\Phi_{k+1}^T - \Phi_k^T) \leq -\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}\mathbb{1}_{\mathcal{D}_h}\frac{1}{2}\nu(\gamma-2)(\Delta_p^k)^2. \tag{77}$$

(iii) The iteration is Improving ($\mathbb{1}_{\mathcal{I}} = 1$). Again, there is a change of zero in f. Thus, the bound on the change in Φ_k^T easily follows from multiplying both sides of (58) by $\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{J_k}$, and replacing Φ_k by Φ_k^T as follows

$$\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}\mathbb{1}_{\mathcal{I}}(\Phi_{k+1}^T - \Phi_k^T) \leq -\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}\mathbb{1}_{\mathcal{I}}\frac{1}{2}\nu(\gamma-2)(\Delta_p^k)^2. \tag{78}$$

(iv) The iteration is Unsuccessful ($\mathbb{1}_{\mathcal{U}} = 1$). There is a change of zero in f and in h since no iterate is updated, while Δ_p^k is decreased. Consequently, the bound on the change in Φ_k^T follows from multiplying both sides of (59) by $\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{J_k}$, and replacing Φ_k by Φ_k^T as follows

$$\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}\mathbb{1}_{\mathcal{U}}(\Phi_{k+1}^T - \Phi_k^T) = -\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}\mathbb{1}_{\mathcal{U}}(1-\nu)(1-\tau^2)(\Delta_p^k)^2. \tag{79}$$

Then combining (76), (77), (78), (79) and using (60), yields

$$\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}(\Phi_{k+1}^T - \Phi_k^T) \leq -\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{J_k}(1-\nu)(1-\tau^2)(\Delta_p^k)^2. \tag{80}$$

Now, notice that under Assumption 3, simple calculations lead to $\mathbb{E}\left(\mathbb{1}_{I_k}\mathbb{1}_{J_k}|\mathcal{F}_{k-1}^{C\cdot F}\right) \geq \alpha\beta$. Then, taking expectations with respect to $\mathcal{F}_{k-1}^{C\cdot F}$ on both sides of (80) and using the $\mathcal{F}_{k-1}^{C\cdot F}$-measurability of the random variables $\mathbb{1}_{\{T\leq k\}}$ and Δ_p^k, lead to (73).
Case 2 (Bad estimates and good bounds, $\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k} = 1$). An increase in the difference of Φ_k^T may occurs since good bounds might not provide enough decrease to cancel the increase which occurs in f whenever Algorithm 1 wrongly accepts an iterate because of bad estimates. Specifically, the f-Dominating case dominates the worst-case increase in the change of Φ_k^T, thus leading to

$$\mathbb{1}_{\{T\leq k\}}\mathbb{E}\left[\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}(\Phi_{k+1}^T - \Phi_k^T)|\mathcal{F}_{k-1}^{C\cdot F}\right] \leq 2\nu(1-\beta)^{1/2}(\Delta_p^k)^2\mathbb{1}_{\{T\leq k\}}. \tag{81}$$

(i) The iteration is f-Dominating ($\mathbb{1}_{\mathcal{D}_f} = 1$). Whenever bad estimates occur and the iteration is f-Dominating, the change in f is bounded as follows

$$\begin{aligned}
&\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{D}_f}\frac{\nu}{\varepsilon}(f(X_{\text{feas}}^{(k+1)\vee T}) - f(X_{\text{feas}}^{k\vee T}))\\
&\quad\leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{D}_f}\frac{\nu}{\varepsilon}\left[(F_s^k - F_0^k) + \left|f(X_{\text{feas}}^{k+1}) - F_s^k\right| + \left|f(X_{\text{feas}}^k) - F_0^k\right|\right]\\
&\quad\leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{D}_f}\nu\left[-\gamma(\Delta_p^k)^2 + \frac{1}{\varepsilon}\left(\left|f(X_{\text{feas}}^{k+1}) - F_s^k\right| + \left|f(X_{\text{feas}}^k) - F_0^k\right|\right)\right]
\end{aligned} \tag{82}$$

where the last inequality in (82) follows from $F_s^k - F_0^k \leq -\gamma\varepsilon(\Delta_p^k)^2$ which is satisfied for every f-Dominating iteration. While the change in h is zero since X_{inf}^k is not updated, that in Δ_p^k follows (75) by replacing $\mathbb{1}_{J_k}$ by $\mathbb{1}_{\bar{J}_k}$ as follows

$$\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{D}_f}(1-\nu)[(\Delta_p^{k+1})^2 - (\Delta_p^k)^2] \leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{D}_f}(1-\nu)(\tau^{-2}-1)(\Delta_p^k)^2. \tag{83}$$

Then, (82), (83) and the inequality $-\nu\gamma(\Delta_p^k)^2 + (1-\nu)(\tau^{-2}-1)(\Delta_p^k)^2 \leq 0$ due to (34) yield

$$\begin{aligned}
&\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{D}_f}(\Phi_{k+1}^T - \Phi_k^T)\\
&\qquad\leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{D}_f}\frac{\nu}{\varepsilon}\left(\left|f(X_{\text{feas}}^{k+1}) - F_s^k\right| + \left|f(X_{\text{feas}}^k) - F_0^k\right|\right).
\end{aligned} \tag{84}$$

(ii) The iteration is h-Dominating ($\mathbb{1}_{\mathcal{D}_h} = 1$). The bound on the change in Φ_k^T which can be obtained by replacing $\mathbb{1}_{J_k}$ by $\mathbb{1}_{\bar{J}_k}$ in (77) is trivially bounded as follows

$$\begin{aligned}&\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{D}_h}(\Phi_{k+1}^T - \Phi_k^T)\\ &\qquad\leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{D}_h}\frac{\nu}{\varepsilon}\left(\left|f(X_{\text{feas}}^{k+1}) - F_s^k\right| + \left|f(X_{\text{feas}}^k) - F_0^k\right|\right). \qquad (85)\end{aligned}$$

(iii) The iteration is Improving ($\mathbb{1}_{\mathcal{I}} = 1$). Again, the change in Φ_k^T which can be obtained by replacing $\mathbb{1}_{J_k}$ by $\mathbb{1}_{\bar{J}_k}$ in (78) is trivially bounded as follows

$$\begin{aligned}&\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{I}}(\Phi_{k+1}^T - \Phi_k^T)\\ &\qquad\leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{I}}\frac{\nu}{\varepsilon}\left(\left|f(X_{\text{feas}}^{k+1}) - F_s^k\right| + \left|f(X_{\text{feas}}^k) - F_0^k\right|\right). \qquad (86)\end{aligned}$$

(iv) The iteration is Unsuccessful ($\mathbb{1}_{\mathcal{U}} = 1$). Because of the decrease of the frame size parameter and hence that in Φ_k^T, the bound on the change in Φ_k^T is obviously as follows

$$\begin{aligned}&\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{U}}(\Phi_{k+1}^T - \Phi_k^T)\\ &\qquad\leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}\mathbb{1}_{\mathcal{U}}\frac{\nu}{\varepsilon}\left(\left|f(X_{\text{feas}}^{k+1}) - F_s^k\right| + \left|f(X_{\text{feas}}^k) - F_0^k\right|\right). \qquad (87)\end{aligned}$$

Then, combining (84), (85), (86) and $\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k} \leq 1$, yields

$$\begin{aligned}&\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}(\Phi_{k+1}^T - \Phi_k^T)\\ &\qquad\leq \mathbb{1}_{\{T\leq k\}}\frac{\nu}{\varepsilon}\left(\left|f(X_{\text{feas}}^{k+1}) - F_s^k\right| + \left|f(X_{\text{feas}}^k) - F_0^k\right|\right). \qquad (88)\end{aligned}$$

Since Assumption 3 holds, it follows from the conditional Cauchy-Schwarz inequality [20] that

$$\begin{aligned}\mathbb{E}\left(\left|f(X_{\text{feas}}^k) - F_0^k\right| |\mathcal{F}_{k-1}^{C\cdot F}\right) &\leq \mathbb{E}\left(1|\mathcal{F}_{k-1}^{C\cdot F}\right)^{1/2}\left[\mathbb{E}\left(\left|f(X_{\text{feas}}^k) - F_0^k\right|^2 |\mathcal{F}_{k-1}^{C\cdot F}\right)\right]^{1/2}\\ &\leq \varepsilon(1-\beta)^{1/2}(\Delta_p^k)^2, \qquad (89)\end{aligned}$$

where (89) follows from (12) and the fact that $\mathbb{E}\left(1|\mathcal{F}_{k-1}^{C\cdot F}\right) = 1$. Similarly, the following holds

$$\mathbb{E}\left(\left|f(X_{\text{feas}}^{k+1}) - F_s^k\right| |\mathcal{F}_{k-1}^{C\cdot F}\right) \leq \varepsilon(1-\beta)^{1/2}(\Delta_p^k)^2. \qquad (90)$$

Thus, taking expectations with respect to $\mathcal{F}_{k-1}^{C\cdot F}$ on both sides of (88) and then using (89), (90) and the $\mathcal{F}_{k-1}^{C\cdot F}$-measurability of the random variables $\mathbb{1}_{\{T\leq k\}}$ and Δ_p^k, lead to (81).

Case 3 (Bad bounds, $\mathbb{1}_{\bar{I}_k} = 1$). The difference in Φ_k^T may increase since even though good estimates of f values occur, they might not provide enough decrease to cancel the increase in h whenever Algorithm 1 wrongly accepts an iterate because of bad bounds. The following will be shown

$$\mathbb{1}_{\{T\leq k\}}\mathbb{E}\left[\mathbb{1}_{\bar{I}_k}(\Phi_{k+1}^T - \Phi_k^T)|\mathcal{F}_{k-1}^{C\cdot F}\right] \leq 2\nu\left[(1-\alpha)^{1/2} + (1-\beta)^{1/2}\right](\Delta_p^k)^2\mathbb{1}_{\{T\leq k\}}. \qquad (91)$$

(i) The iteration is f-Dominating ($\mathbb{1}_{\mathcal{D}_f} = 1$). The change in Φ_k^T is bounded, taking into account the possible aforementioned increase in f. Since the change in h is zero, then it is easy to notice that the bound on the change in Φ_k^T can be derived from (84) by replacing $\mathbb{1}_{I_k}\mathbb{1}_{\bar{J}_k}$ by $\mathbb{1}_{\bar{I}_k}$ as follows

$$\begin{aligned}&\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{D}_f}(\Phi_{k+1}^T - \Phi_k^T)\\ &\qquad\leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{D}_f}\frac{\nu}{\varepsilon}\left(\left|f(X_{\text{feas}}^{k+1}) - F_s^k\right| + \left|f(X_{\text{feas}}^k) - F_0^k\right|\right). \qquad (92)\end{aligned}$$

(ii) The iteration is h-Dominating ($\mathbb{1}_{\mathcal{D}_h} = 1$). Since the change in f is zero, the bound on the change in Φ_k^T is obtained by multiplying both sides of (65) by $\mathbb{1}_{\{T \leq k\}}$ and replacing Φ_k by Φ_k^T

$$\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{D}_h}(\Phi_{k+1}-\Phi_k) \leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{D}_h}\frac{\nu}{m\varepsilon}\left(\left|h(X_{\text{inf}}^{k+1})-H_s^k\right|+\left|h(X_{\text{inf}}^k)-H_0^k\right|\right). \quad (93)$$

(iii) The iteration is Improving ($\mathbb{1}_{\mathcal{I}} = 1$). The frame size parameter is updated as at h-Dominating iterations and the change in f is zero. Thus, the bound on the change in Φ_k^T follows from (93) by replacing $\mathbb{1}_{\mathcal{D}_h}$ by $\mathbb{1}_{\mathcal{I}}$ as follows

$$\mathbb{1}_{\{T\leq k\}}\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{I}}(\Phi_{k+1}-\Phi_k) \leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{I}}\frac{\nu}{m\varepsilon}\left(\left|h(X_{\text{inf}}^{k+1})-H_s^k\right|+\left|h(X_{\text{inf}}^k)-H_0^k\right|\right). \quad (94)$$

(iv) The iteration is Unsuccessful ($\mathbb{1}_{\mathcal{U}} = 1$). Because of the decrease of the frame size parameter and hence that in Φ_k^T, the bound on the change in Φ_k^T is obviously as follows

$$\begin{aligned}\mathbb{1}_{\{T\leq k\}}&\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{U}}(\Phi_{k+1}^T-\Phi_k^T)\\ &\leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{\bar{I}_k}\mathbb{1}_{\mathcal{U}}\nu\left[\frac{1}{\varepsilon}\left(\left|f(X_{\text{feas}}^{k+1})-F_s^k\right|+\left|f(X_{\text{feas}}^k)-F_0^k\right|\right)\right.\\ &\left.+\frac{1}{m\varepsilon}\left(\left|h(X_{\text{inf}}^{k+1})-H_s^k\right|+\left|h(X_{\text{inf}}^k)-H_0^k\right|\right)\right]\end{aligned} \quad (95)$$

Since (95) dominates (92), (93) and (94), then combining all four cases lead to

$$\begin{aligned}\mathbb{1}_{\{T\leq k\}}&\mathbb{1}_{\bar{I}_k}(\Phi_{k+1}^T-\Phi_k^T)\\ &\leq \mathbb{1}_{\{T\leq k\}}\mathbb{1}_{\bar{I}_k}\nu\left[\frac{1}{\varepsilon}\left(\left|f(X_{\text{feas}}^{k+1})-F_s^k\right|+\left|f(X_{\text{feas}}^k)-F_0^k\right|\right)\right.\\ &\left.+\frac{1}{m\varepsilon}\left(\left|h(X_{\text{inf}}^{k+1})-H_s^k\right|+\left|h(X_{\text{inf}}^k)-H_0^k\right|\right)\right]\end{aligned} \quad (96)$$

Now, taking expectations with respect to $\mathcal{F}_{k-1}^{C\cdot F}$ on both sides of (96) and using (14), (89) and (90) lead to (91). Then, by combining the main results of Case 1, Case 2 and Case 3 of Part 2, specifically (73), (81) and (91), the following holds

$$\begin{aligned}\mathbb{1}_{\{T\leq k\}}\mathbb{E}\left[\Phi_{k+1}^T-\Phi_k^T|\mathcal{F}_{k-1}^{C\cdot F}\right] \leq \big[&-\alpha\beta(1-\nu)(1-\tau^2)+2\nu(1-\alpha)^{1/2}\\ &+4\nu(1-\beta)^{1/2}\big](\Delta_p^k)^2\mathbb{1}_{\{T\leq k\}}.\end{aligned} \quad (97)$$

Finally, choosing α and β according to (35) ensures that

$$-\alpha\beta(1-\nu)(1-\tau^2)+2\nu(1-\alpha)^{1/2}+4\nu(1-\beta)^{1/2} \leq -\frac{1}{2}\alpha\beta(1-\nu)(1-\tau^2), \quad (98)$$

and (72) obviously follows from (97) and (98) with the same constant $\eta = \frac{1}{2}\alpha\beta(1-\nu)(1-\tau^2)$ as Part 1, which achieves the proof. □

Proof of Corollary 2

Proof. Only (37) is proved but the proof also applies for $\left|H_s^k - h(X^k + S^k)\right|$ and $\left|F_s^k - f(X^k + S^k)\right|$. According to Assumption 3*(vi)*, $\mathbb{E}\left(\left|H_0^k - h(X^k)\right| \mid \mathcal{F}_{k-1}^{C\cdot F}\right) \leq m\varepsilon(1-\alpha)^{1/2}(\Delta_p^k)^2$, which implies that

$$\mathbb{E}\left(\left|H_0^k - h(X^k)\right|\right) \leq m\varepsilon(1-\alpha)^{1/2}\mathbb{E}\left[(\Delta_p^k)^2\right]. \tag{99}$$

By summing each side of (99) over k from 0 to N, and observing that

$$0 \leq S_N^h := \sum_{k=0}^{N}\left|H_0^k - h(X^k)\right| \nearrow \sum_{k=0}^{+\infty}\left|H_0^k - h(X^k)\right|, \quad \text{and} \quad 0 \leq S_N^\Delta := \sum_{k=0}^{N}(\Delta_p^k)^2 \nearrow \sum_{k=0}^{+\infty}(\Delta_p^k)^2,$$

then, it follows from the monotone convergence theorem [32] that

$$\begin{aligned}
\mathbb{E}\left(\sum_{k=0}^{+\infty}\left|H_0^k - h(X^k)\right|\right) &= \mathbb{E}\left(\lim_{N\to+\infty} S_N^h\right) = \lim_{N\to+\infty}\mathbb{E}\left(S_N^h\right) = \sum_{k=0}^{+\infty}\mathbb{E}\left(\left|H_0^k - h(X^k)\right|\right) \\
&\leq m\varepsilon(1-\alpha)^{1/2}\sum_{k=0}^{+\infty}\mathbb{E}\left[(\Delta_p^k)^2\right] = m\varepsilon(1-\alpha)^{1/2}\lim_{N\to+\infty}\mathbb{E}\left(S_N^\Delta\right) \\
&= m\varepsilon(1-\alpha)^{1/2}\mathbb{E}\left(\lim_{N\to+\infty} S_N^\Delta\right) = m\varepsilon(1-\alpha)^{1/2}\mathbb{E}\left[\sum_{k=0}^{+\infty}(\Delta_p^k)^2\right] \\
&\leq \mu \times m\varepsilon(1-\alpha)^{1/2} < +\infty,
\end{aligned}$$

where μ is the constant of (52). This means that $\sum_{k=0}^{+\infty}\left|H_0^k - h(X^k)\right| < +\infty$ almost surely, which implies the first result of (37). The proof for $\left|F_0^k - f(X^k)\right|$ is similar by observing that (see (89))

$$\mathbb{E}\left(\left|F_0^k - f(X^k)\right| |\mathcal{F}_{k-1}^{C\cdot F}\right) \leq \varepsilon(1-\beta)^{1/2}(\Delta_p^k)^2.$$

□

Proof of Lemma 1

Proof. The proof uses ideas derived in [11, 23]. The result is proved by contradiction conditioned on the almost sure event $E_1 = \{\Delta_p^k \to 0\}$. All that follows is conditioned on the event E_1. Assume that with nonzero probability, there exists a random variable $\mathcal{E}' > 0$ such that

$$\Psi_k^h \geq \mathcal{E}', \quad \text{for all } k \in \mathbb{N}. \tag{100}$$

Let $\{x_{\text{inf}}^k\}_{k\in\mathbb{N}}$, $\{s^k\}_{k\in\mathbb{N}}$, $\{\delta_p^k\}_{k\in\mathbb{N}}$ and $\epsilon' > 0$ be realizations of $\{X_{\text{inf}}^k\}_{k\in\mathbb{N}}$, $\{S^k\}_{k\in\mathbb{N}}$, $\{\Delta_p^k\}_{k\in\mathbb{N}}$ and $\mathcal{E}'$, respectively for which (100) holds. Let $\hat{z}$ be the same parameter of Algorithm 1 satisfying $\delta_p^k \leq \tau^{-\hat{z}}$ for all $k \geq 0$. Since $\delta_p^k \to 0$ because of the conditioning on E_1, there exists $k_0 \in \mathbb{N}$ such that

$$\delta_p^k < \lambda := \min\left\{\frac{\epsilon'}{m\varepsilon(\gamma+2)}, \tau^{1-\hat{z}}\right\}, \quad \text{for all } k \geq k_0. \tag{101}$$

Consequently and since $\tau < 1$, the random variable R_k with realizations $r_k := -\log_\tau\left(\frac{\delta_p^k}{\lambda}\right)$ satisfies $r_k < 0$ for all $k \geq k_0$. The main idea of the proof is to show that such realizations occur only with probability zero, thus leading to a contradiction. Let first show that $\{R_k\}_{k\in\mathbb{N}}$ is a submartingale. Let $k \geq k_0$ be an iteration for which the events I_k and J_k both occur, which happens with probability of at least $\alpha\beta > 1/2$. Then, it follows from the definition of the event I_k (see Definition 8) that

$$\begin{aligned} h(x_{\text{inf}}^k) &\leq u_0^k(x_{\text{inf}}^k) \leq \sum_{j=1}^m \max\left\{c_{j,0}^k(x_{\text{inf}}^k), 0\right\} + m\varepsilon(\delta_p^k)^2 = h_0^k(x_{\text{inf}}^k) + m\varepsilon(\delta_p^k)^2, \quad (102)\\ \text{and} \quad h(x_{\text{inf}}^k + s^k) &\geq \ell_s^k(x_{\text{inf}}^k + s^k) \geq h_s^k(x_{\text{inf}}^k + s^k) - m\varepsilon(\delta_p^k)^2. \quad (103)\end{aligned}$$

$$\begin{aligned}\text{Hence,} \quad h_s^k(x_{\text{inf}}^k + s^k) - h_0^k(x_{\text{inf}}^k) &= [h(x_{\text{inf}}^k + s^k) - h(x_{\text{inf}}^k)] + [h(x_{\text{inf}}^k) - h_0^k(x_{\text{inf}}^k)]\\ &\quad + [h_s^k(x_{\text{inf}}^k + s^k) - h(x_{\text{inf}}^k + s^k)] \quad (104)\\ &\leq 2m\varepsilon(\delta_p^k)^2 - \epsilon'\delta_p^k \leq 2m\varepsilon(\delta_p^k)^2 - m\varepsilon(\gamma+2)(\delta_p^k)^2 = -\gamma m\varepsilon(\delta_p^k)^2\end{aligned}$$

where the first inequality in (104) follows from (100), (102) and (103) while the last one follows from (101). Consequently, the iteration k of Algorithm 1 can not be unsuccessful. Thus, the frame size parameter is updated according to $\delta_p^{k+1} = \tau^{-1}\delta_p^k$ since $\delta_p^k < \tau^{1-\hat{z}}$. Hence, $r_{k+1} = r_k + 1$.

Let $\mathcal{F}_{k-1}^{I\cdot J} = \sigma(I_0, I_1, \ldots, I_{k-1}) \cap \sigma(J_0, J_1, \ldots, J_{k-1})$. For all other outcomes of I_k and J_k, which will occur with a total probability of at most $1 - \alpha\beta$, the inequality $\delta_p^{k+1} \geq \tau\delta_p^k$ always holds, thus implying that $r_{k+1} \geq r_k - 1$. Hence,

$$\begin{aligned} \mathbb{E}\left(\mathbb{1}_{I_k\cap J_k}(R_{k+1} - R_k)|\mathcal{F}_{k-1}^{I\cdot J}\right) &= \mathbb{P}\left(I_k \cap J_k|\mathcal{F}_{k-1}^{I\cdot J}\right) \geq \alpha\beta\\ \text{and} \quad \mathbb{E}\left(\mathbb{1}_{\overline{I_k\cap J_k}}(R_{k+1} - R_k)|\mathcal{F}_{k-1}^{I\cdot J}\right) &\geq -\mathbb{P}\left(\overline{I_k \cap J_k}|\mathcal{F}_{k-1}^{I\cdot J}\right) \geq \alpha\beta - 1.\end{aligned}$$

Thus, $\mathbb{E}\left(R_{k+1} - R_k|\mathcal{F}_{k-1}^{I\cdot J}\right) \geq 2\alpha\beta - 1 > 0$, implying that $\{R_k\}$ is a submartingale. The remainder of the proof is almost identical to that of the proof of the $\liminf$-type first-order result in [23].

Now, let construct a random walk W_k with realizations w_k on the same probability space as R_k, which will serve as a lower bound on R_k. Define W_k as in (15) by

$$W_k = \sum_{i=0}^k (2 \cdot \mathbb{1}_{I_i}\mathbb{1}_{J_i} - 1), \quad (105)$$

where the indicator random variables $\mathbb{1}_{I_i}$ and $\mathbb{1}_{J_i}$ are such that $\mathbb{1}_{I_i} = 1$ if I_i occurs, $\mathbb{1}_{I_i} = 0$ otherwise, and similarly, $\mathbb{1}_{J_i} = 1$ if J_i occurs while $\mathbb{1}_{J_i} = 0$ otherwise. Then following the proof of Theorem 1, it is easy to notice that $\{W_k\}$ is a $\mathcal{F}_{k-1}^{I\cdot J}$-submartingale (see also [23] for the same result), thus leading to the conclusion that $\left\{\limsup_{k\to+\infty} W_k = +\infty\right\}$ almost surely. Since by construction

$$r_k - r_{k_0} = -\log_\tau\left(\frac{\delta_p^k}{\delta_p^{k_0}}\right) = k - k_0 \geq w_k - w_{k_0},$$

then with probability one, R_k has to be positive infinitely often. Thus, the sequence of realizations r_k such that $r_k < 0$ for all $k \geq k_0$ occurs with probability zero. Consequently, the assumption that $\Psi_k^h \geq \mathcal{E}'$ holds for all $k \in \mathbb{N}$ with a positive probability is false, which implies that (38) holds. □

Proof of Theorem 4

Proof. The theorem is proved using ideas derived in [8, 11]. Define the events E_1 and E_2 by

$$E_1 = \left\{\omega \in \Omega : \Delta_p^k(\omega) \to 0\right\} \quad \text{and} \quad E_2 = \left\{\omega \in \Omega : \exists K'(\omega) \subset \mathbb{N} \text{ such that } \lim_{K'(\omega)} \Psi_k^h(\omega) \leq 0\right\}.$$

Then E_1 and E_2 are almost sure due to Corollary 1 and (38) respectively. Let $\omega \in E_1 \cap E_2$ be an arbitrary outcome and note that the event $E_1 \cap E_2$ is also almost sure as countable intersection of almost sure events. Then $\lim_{K'(\omega)} \Delta_p^k(\omega) = 0$. It follows from the compactness hypothesis of Assumption 2 that there exists $K(\omega) \subseteq K'(\omega)$ for which the subsequence $\{X_{\text{inf}}^k(\omega)\}_{k\in K(\omega)}$ converges to a limit $\hat{X}_{\text{inf}}(\omega)$. Specifically, $\hat{X}_{\text{inf}}(\omega)$ is a refined point for the refining subsequence $\{X_{\text{inf}}^k(\omega)\}_{k\in K(\omega)}$. Let $v \in T_{\mathcal{X}}^H(\hat{X}_{\text{inf}}(\omega))$ be a refining direction for $\hat{X}_{\text{inf}}(\omega)$. Denote by V the random vector with realizations v, i.e., $v = V(\omega)$, and let $\hat{x}_{\text{inf}} = \hat{X}_{\text{inf}}(\omega)$, $x_{\text{inf}}^k = X_{\text{inf}}^k(\omega)$, $\delta_p^k = \Delta_p^k(\omega)$, $\delta_m^k = \Delta_m^k(\omega)$, $\psi_k^h = \Psi_k^h(\omega)$ and $\mathcal{K} = K(\omega)$. Since v is a refining direction, then there exists $\mathcal{L} \subseteq \mathcal{K}$ and polling directions $d^k \in \mathbb{D}_p^k(x_{\text{inf}}^k)$ such that $v = \lim\limits_{k\in\mathcal{L}} \frac{d^k}{\|d^k\|_\infty}$. For each $k \in \mathcal{L}$, define

$$t_k = \delta_m^k \left\|d^k\right\|_\infty \to 0, \qquad\qquad y^k = x_{\text{inf}}^k + t_k\left(\frac{d^k}{\|d^k\|_\infty} - v\right) \to \hat{x}_{\text{inf}},$$
$$a_k = \frac{h(y^k + t_k v) - h(x_{\text{inf}}^k)}{t_k} \quad \text{and} \quad b_k = \frac{h(x_{\text{inf}}^k) - h(y^k)}{t_k},$$

where the fact that $t_k \to 0$ follows from Definition 4, specifically the inequality $\delta_m^k\left\|d^k\right\|_\infty \leq \delta_p^k b$. Since h is λ^h–locally Lipschitz, then

$$|a_k| \leq \frac{\lambda^h}{t_k}\left\|(y^k + t_k v) - x_{\text{inf}}^k\right\|_\infty = \lambda^h \quad \text{and} \quad |b_k| \leq \frac{\lambda^h}{t_k}\left\|x_{\text{inf}}^k - y^k\right\|_\infty = \lambda^h\left\|\frac{d^k}{\|d^k\|_\infty} - v\right\|_\infty \to 0,$$

which shows that Lemma 2 applies for both subsequences $\{a_k\}_{k\in\mathcal{L}}$ and $\{b_k\}_{k\in\mathcal{L}}$. Moreover, combining the inequality $\lim_{\mathcal{L}} \psi_k^h \leq 0$ and Assumption 6 (the fact that $\delta_p^k\left\|d^k\right\|_\infty \geq d_{\min} > 0$), yields

$$\lim_{k\in\mathcal{L}}\left(\frac{-\psi_k^h}{\delta_p^k\|d^k\|_\infty}\right) = \lim_{k\in\mathcal{L}} \frac{h(x_{\text{inf}}^k + \delta_m^k d^k) - h(x_{\text{inf}}^k)}{t_k} \geq -d_{\min}^{-1} \lim_{k\in\mathcal{L}} \psi_k^h \geq 0. \tag{106}$$

Thus, by adding and subtracting $h(x_{\text{inf}}^k)$ to the numerator of the definition of the Clarke derivative, and using the fact that $x_{\text{inf}}^k + \delta_m^k d^k \in \mathcal{X}$ for sufficiently large $k \in \mathcal{L}$ since v is a hypertangent direction,

$$\begin{aligned} h^\circ(\hat{x}_{\text{inf}}; v) \quad &\geq \quad \limsup_{k\in\mathcal{L}} \frac{h(y^k + t_k v) - h(x_{\text{inf}}^k) + h(x_{\text{inf}}^k) - h(y^k)}{t_k} = \limsup_{k\in\mathcal{L}}(a_k + b_k) \\ &= \quad \limsup_{k\in\mathcal{L}} a_k + \lim_{k\in\mathcal{L}} b_k = \limsup_{k\in\mathcal{L}} \frac{h(x_{\text{inf}}^k + \delta_m^k d^k) - h(x_{\text{inf}}^k)}{t_k} \geq 0, \end{aligned}$$

where the last inequality follows from (106). Now, notice that it has been showed that every outcome ω arbitrarily chosen in $E_1 \cap E_2$, belongs to the event

$$\begin{aligned} E_3 := \Big\{\omega \in \Omega : \exists K(\omega) \subseteq \mathbb{N} \text{ and } \exists \hat{X}_{\text{inf}}(\omega) = \lim_{k\in K(\omega)} X_{\text{inf}}^k(\omega), \hat{X}_{\text{inf}}(\omega) \in \mathcal{X}, \text{ such that} \\ \forall V(\omega) \in T_{\mathcal{X}}^H(\hat{X}_{\text{inf}}(\omega)),\ h^\circ(\hat{X}_{\text{inf}}(\omega); V(\omega)) \geq 0\Big\}, \end{aligned}$$

thus implying that $E_1 \cap E_2 \subseteq E_3$. Then the proof is complete by noticing that $\mathbb{P}\left(E_1 \cap E_2\right) = 1$. □

Proof of Lemma 3

Proof. The proof is almost identical to those of Lemma 1 and a similar result in [11]. Hence, full details are not provided here again. Unless otherwise stated, all the sequences, events and constants considered are defined as in the proof of Lemma 1. The result is proved by contradiction and all that follows is conditioned on the almost sure event $E_1 \cap \{T < +\infty\}$. Assume that with nonzero probability there exists a random variable $\mathcal{E}'' > 0$ such that

$$\Psi_k^{f,T} \geq \mathcal{E}'', \quad \text{for all } k \geq 0. \tag{107}$$

Let $\{x_{\text{feas}}^{k\vee t}\}_{k\in\mathbb{N}}$, $\{s^k\}_{k\in\mathbb{N}}$, $\{\delta_p^k\}_{k\in\mathbb{N}}$ and $\epsilon'' > 0$ be realizations of $\{X_{\text{feas}}^{k\vee T}\}_{k\in\mathbb{N}}$, $\{S^k\}_{k\in\mathbb{N}}$, $\{\Delta_p^k\}_{k\in\mathbb{N}}$ and $\mathcal{E}''$, respectively for which (107) holds. Let $\bar{k}_0 \in \mathbb{N}^*$ be such that

$$\delta_p^k < \lambda := \min\left\{\frac{\epsilon''}{\varepsilon(\gamma+2)}, \tau^{1-\hat{z}}\right\} \quad \text{for all } k \geq \bar{k}_0. \tag{108}$$

The key element of the proof is to show that an iteration $k \geq k_0 := \max\{\bar{k}_0, t\}$ for which the events I_k and J_k both occur can not be unsuccessful, thus leading to the fact that $\{R_k\}$ is a submartingale.

It follows from (107) and (108) that

$$f(x_{\text{feas}}^k + s^k) - f(x_{\text{feas}}^k) \leq -\epsilon''\delta_p^k \leq -(\gamma+2)\varepsilon(\delta_p^k)^2, \quad \text{for all } k \geq k_0.$$

Since J_k occurs,
$$\begin{aligned} f_s^k(x_{\text{feas}}^k + s^k) - f_0^k(x_{\text{feas}}^k) &= [f(x_{\text{feas}}^k + s^k) - f(x_{\text{feas}}^k)] + [f(x_{\text{feas}}^k) - f_0^k(x_{\text{feas}}^k)] \\ &\quad + [f_s^k(x_{\text{feas}}^k + s^k) - f(x_{\text{feas}}^k + s^k)] \\ &\leq -(\gamma+2)\varepsilon(\delta_p^k)^2 + 2\varepsilon(\delta_p^k)^2 = -\gamma\varepsilon(\delta_p^k)^2, \end{aligned}$$

which implies that the iteration $k \geq k_0$ of Algorithm 1 can not be unsuccessful. □

Proof of Theorem 5

Proof. The proof results from Corollary 2 by observing that for all outcome ω in the almost sure event

$$E_4 := \left\{\omega \in \Omega : \forall K(\omega) \subseteq \mathbb{N}, \lim_{k\in K(\omega)} \left|H_0^k(X_{\text{feas}}^{k\vee T})(\omega) - h(X_{\text{feas}}^{k\vee T}(\omega))\right| = 0\right\} \cap \{T < +\infty\},$$

$$\lim_{k\in K(\omega)} \left|H_0^k(X_{\text{feas}}^{k\vee T})(\omega) - h(X_{\text{feas}}^{k\vee T}(\omega))\right| = \lim_{k\in K(\omega)} h(X_{\text{feas}}^{k\vee T}(\omega)) = h(\hat{X}_{\text{feas}}(\omega)) = 0,$$

where the penultimate equality follows from the continuity of h in $\mathcal{X}$. This means that

$$\mathbb{P}\left(h(\hat{X}_{\text{feas}}) = 0\right) = \mathbb{P}\left(\hat{X}_{\text{feas}} \in \mathcal{D}\right) = 1.$$

□

Proof of Theorem 6

Proof. First, notice that the fact that $\mathbb{P}\left(\hat{X}_{\text{feas}} \in \mathcal{D}\right) = 1$ follows from Theorem 5. Then the proof easily follows from that of Theorem 4, by replacing h by f, $\hat{x}_{\text{inf}} = \hat{X}_{\text{inf}}(\omega)$ by $\hat{x}_{\text{feas}} = \hat{X}_{\text{feas}}(\omega)$, $x^k_{\text{inf}} = X^k_{\text{inf}}(\omega)$ by $x^{k\vee t}_{\text{feas}} = X^{k\vee T}_{\text{feas}}(\omega)$, $\psi^h_k = \Psi^h_k(\omega)$ by $\psi^{f,t}_k = \Psi^{f,T}_k(\omega)$ with $t = T(\omega)$ and $T^H_{\mathcal{X}}(\cdot)$ by $T^H_{\mathcal{D}}(\cdot)$, for ω fixed and arbitrarily chosen in the almost sure event $E_1 \cap E_5 \cap \{T < +\infty\}$, where

$$E_5 = \Big\{\omega \in \Omega : \exists K(\omega) \subseteq \mathbb{N} \text{ such that } \hat{X}_{\text{feas}}(\omega) = \lim_{k\in K(\omega)} X^{k\vee T}_{\text{feas}}(\omega),\ \hat{X}_{\text{feas}}(\omega) \in \mathcal{D},$$
$$\lim_{k\in K(\omega)} \Psi^{f,T}_k(\omega) \leq 0 \text{ and } \lim_{k\in K(\omega)} H^k_0(X^{k\vee T}_{\text{feas}})(\omega) = 0\Big\}. \tag{109}$$

□

References

[1] M.A. Abramson, C. Audet, J.E. Dennis, Jr., and S. Le Digabel. OrthoMADS: A Deterministic MADS Instance with Orthogonal Directions. *SIAM Journal on Optimization*, 20(2):948–966, 2009.

[2] S. Alarie, C. Audet, P.-Y. Bouchet, and S. Le Digabel. Optimization of noisy blackboxes with adaptive precision. Technical Report G-2019-84, Les cahiers du GERAD, 2019.

[3] E.J. Anderson and M.C. Ferris. A Direct Search Algorithm for Optimization with Noisy Function Evaluations. *SIAM Journal on Optimization*, 11(3):837–857, 2001.

[4] E. Angün, J. Kleijnen, D. den Hertog, and G. Gürkan. Response surface methodology with stochastic constraints for expensive simulation. *Journal of the operational research society*, 60(6):735–746, 2009.

[5] L. Armijo. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16(1):1–3, 1966.

[6] C. Audet. A survey on direct search methods for blackbox optimization and their applications. In P.M. Pardalos and T.M. Rassias, editors, *Mathematics without boundaries: Surveys in interdisciplinary research*, chapter 2, pages 31–56. Springer, 2014.

[7] C. Audet and J.E. Dennis, Jr. Mesh Adaptive Direct Search Algorithms for Constrained Optimization. *SIAM Journal on Optimization*, 17(1):188–217, 2006.

[8] C. Audet and J.E. Dennis, Jr. A Progressive Barrier for Derivative-Free Nonlinear Programming. *SIAM Journal on Optimization*, 20(1):445–472, 2009.

[9] C. Audet, J.E. Dennis, Jr., and S. Le Digabel. Parallel Space Decomposition of the Mesh Adaptive Direct Search Algorithm. *SIAM Journal on Optimization*, 19(3):1150–1170, 2008.

[10] C. Audet, J.E. Dennis, Jr., and S. Le Digabel. *Parallel Space Decomposition of the Mesh Adaptive Direct Search algorithm*, volume 5 of *The GERAD newsletters (Eleven articles published in leading journals)*, page 3. 2008.

[11] C. Audet, K. J. Dzahini, M. Kokkolaras, and S. Le Digabel. StoMADS: Stochastic blackbox optimization using probabilistic estimates. Technical Report G-2019-30, Les cahiers du GERAD, 2019.

[12] C. Audet and W. Hare. *Derivative-Free and Blackbox Optimization*. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, Cham, Switzerland, 2017.

[13] C. Audet, A. Ihaddadene, S. Le Digabel, and C. Tribes. Robust optimization of noisy blackbox problems using the Mesh Adaptive Direct Search algorithm. *Optimization Letters*, 12(4):675–689, 2018.

[14] C. Audet, S. Le Digabel, and C. Tribes. The Mesh Adaptive Direct Search Algorithm for Granular and Discrete Variables. *SIAM Journal on Optimization*, 29(2):1164–1189, 2019.

[15] F. Augustin and Y.M. Marzouk. NOWPAC: A provably convergent derivative-free nonlinear optimizer with path-augmented constraints. *arXiv*, 2014.

[16] F. Augustin and Y.M. Marzouk. A trust-region method for derivative-free nonlinear constrained stochastic optimization. *arXiv*, 2017.

[17] A.S. Bandeira, K. Scheinberg, and L.N. Vicente. Convergence of trust-region methods based on probabilistic models. *SIAM Journal on Optimization*, 24(3):1238–1264, 2014.

[18] R.R. Barton and J.S. Ivey, Jr. Nelder-Mead simplex modifications for simulation optimization. *Management Science*, 42(7):954–973, 1996.

[19] D. Bertsimas, O. Nohadani, and K. M. Teo. Nonconvex robust optimization for problems with constraints. *INFORMS Journal on Computing*, 22(1):44–58, 2010.

[20] R.N. Bhattacharya and E.C. Waymire. *A basic course in probability theory*, volume 69. Springer, 2007.

[21] J. Blanchet, C. Cartis, M. Menickelly, and K. Scheinberg. Convergence Rate Analysis of a Stochastic Trust-Region Method via Supermartingales. *INFORMS Journal on Optimization*, 2019.

[22] K.H. Chang. Stochastic nelder-mead simplex method - a new globally convergent direct search method for simulation optimization. *European Journal of Operational Research*, 220(3):684–694, 2012.

[23] R. Chen, M. Menickelly, and K. Scheinberg. Stochastic optimization using a trust-region method and random models. *Mathematical Programming*, 169(2):447–487, 2018.

[24] X. Chen and N. Wang. Optimization of short-time gasoline blending scheduling problem with a DNA based hybrid genetic algorithm. *Chemical Engineering and Processing: Process Intensification*, 49(10):1076–1083, 2010.

[25] F.H. Clarke. *Optimization and Nonsmooth Analysis*. John Wiley & Sons, New York, 1983. Reissued in 1990 by SIAM Publications, Philadelphia, as Vol. 5 in the series Classics in Applied Mathematics.

[26] A.R. Conn, K. Scheinberg, and L.N. Vicente. *Introduction to Derivative-Free Optimization*. MOS-SIAM Series on Optimization. SIAM, Philadelphia, 2009.

[27] F. E. Curtis and K. Scheinberg. Adaptive Stochastic Optimization. *arXiv*, 2020.

[28] F.E. Curtis, K. Scheinberg, and R. Shi. A Stochastic Trust Region Algorithm Based on Careful Step Normalization. *arXiv*, 2017.

[29] M. A. Diniz-Ehrhardt, D. G. Ferreira, and S. A. Santos. A pattern search and implicit filtering algorithm for solving linearly constrained minimization problems with noisy objective functions. *Optimization Methods and Software*, 34(4):827–852, 2019.

[30] M. A. Diniz-Ehrhardt, D. G. Ferreira, and S. A. Santos. Applying the pattern search implicit filtering algorithm for solving a noisy problem of parameter identification. *Computational Optimization and Applications*, pages 1–32, 2020.

[31] E.D. Dolan and J.J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213, 2002.

[32] R. Durrett. *Probability: theory and examples*. Cambridge university press, 2010.

[33] K. J. Dzahini. Expected complexity analysis of stochastic direct-search. Technical Report G-2020-18, Les cahiers du GERAD, 2020.

[34] S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang. Direct search based on probabilistic feasible descent for bound and linearly constrained problems. *Computational Optimization and Applications*, 72(3):525–559, 2019.

[35] W. Hock and K. Schittkowski. *Test Examples for Nonlinear Programming Codes*, volume 187 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Berlin, Germany, 1981.

[36] J. Jahn. *Introduction to the Theory of Nonlinear Optimization*. Springer, Berlin, 1994.

[37] S. Kitayama, M. Arakawa, and K. Yamazaki. Sequential approximate optimization using radial basis function network for engineering optimization. *Optimization and Engineering*, 12(4):535–557, 2011.

[38] K. J. Klassen and R. Yoogalingam. Improving performance in outpatient appointment services with a simulation optimization approach. *Production and Operations Management*, 18(4):447–458, 2009.

[39] T. Lacksonen. Empirical comparison of search algorithms for discrete event simulation. *Computers & Industrial Engineering*, 40(1-2):133–148, 2001.

[40] J. Larson and S.C. Billups. Stochastic derivative-free optimization using a trust region framework. *Computational Optimization and Applications*, 64(3):619–645, 2016.

[41] S. Le Digabel and S.M. Wild. A Taxonomy of Constraints in Simulation-Based Optimization. Technical Report G-2015-57, Les cahiers du GERAD, 2015.

[42] B. Letham, B. Karrer, G. Ottoni, and E. Bakshy. Constrained Bayesian optimization with noisy experiments. *Bayesian Analysis*, 14(2):495–519, 2019.

[43] L. Lukšan and J. Vlček. Test problems for nonsmooth unconstrained and linearly constrained optimization. Technical Report V-798, ICS AS CR, 2000.

[44] E. Mezura-Montes and C.A. Coello. Useful Infeasible Solutions in Engineering Optimization with Evolutionary Algorithms. In *Proceedings of the 4th Mexican International Conference on Advances in Artificial Intelligence*, MICAI'05, pages 652–662, Berlin, Heidelberg, 2005. Springer-Verlag.

[45] J. Mockus. *Bayesian approach to global optimization: theory and applications*, volume 37 of *Mathematics and Its Applications*. Springer Science & Business Media, 2012.

[46] J.J. Moré and S.M. Wild. Benchmarking derivative-free optimization algorithms. *SIAM Journal on Optimization*, 20(1):172–191, 2009.

[47] J.A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

[48] C. Paquette and K. Scheinberg. A Stochastic Line Search Method with Expected Complexity Analysis. *SIAM Journal on Optimization*, 30(1):349–376, 2020.

[49] H. Robbins and S. Monro. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951.

[50] R.T. Rockafellar. Generalized directional derivatives and subgradients of nonconvex functions. *Canad. J. Math.*, 32(2):257–280, 1980.

[51] J.F. Rodríguez, J.E. Renaud, and L.T. Watson. Trust Region Augmented Lagrangian Methods for Sequential Response Surface Approximation and Optimization. *Journal of Mechanical Design*, 120(1):58–66, 1998.

[52] S. Shashaani, F.S. Hashemi, and R. Pasupathy. ASTRO-DF: A class of adaptive sampling trust-region algorithms for derivative-free stochastic optimization. *SIAM Journal on Optimization*, 28(4):3145–3176, 2018.

[53] J. Tao and N. Wang. DNA Double Helix Based Hybrid GA for the Gasoline Blending Recipe Optimization Problem. *Chemical Engineering and Technology*, 31(3):440–451, 2008.

[54] Z. Wang and M. Ierapetritou. Constrained optimization of black-box stochastic systems using a novel feasibility enhanced Kriging-based method. *Computers & Chemical Engineering*, 118:210–223, 2018.

[55] J. Zhao and N. Wang. A bio-inspired algorithm based on membrane computing and its application to gasoline blending scheduling. *Computers and Chemical Engineering*, 35(2):272–283, 2011.