Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

THE NUMBER OF RANDOM RESTARTS REQUIRED TO IDENTIFY ALL SOLUTIONS TO A NONLINEAR SYSTEM: APPLICATIONS TO GLOBAL STOCHASTIC OPTIMIZATION

Raghu Pasupathy

Department of Statistics
Purdue University
250 University St.
West Lafayette, IN 47907, USA

ABSTRACT

We consider the question of identifying the set of all solutions to a system of nonlinear equations, when the functions involved in the system can only be observed through a stochastic simulation. Such problems frequently arise as first order necessary conditions in global simulation optimization problems. A convenient method of "solving" such problems involves generating (using a fixed sample) a sample-path approximation of the functions involved, and then executing a convergent root-finding algorithm with several random restarts. The various solutions obtained thus are then gathered to form the estimator of the true set. We investigate the quality of the returned set in terms of the expected Hausdorff distance between the returned and true sets. Our message is that a certain simple logarithmic relationship between the sample size and the number of random restarts ensures maximal efficiency.

1 INTRODUCTION

In this paper we consider the question of identifying the "true set" π^* of all solutions to a system of *stochastic* nonlinear equations $g(x) = 0$, for $g : \mathcal{D} \subset \mathbb{R}^q \to \mathbb{R}^q$. We call the nonlinear systems we consider "stochastic" because we assume that the underlying vector function g can be observed only using a stochastic simulation. In other words, we do not know the function g but instead have a consistent estimator $G_m : \mathcal{D} \to \mathbb{R}^q$ of g. The subscript m refers to some measure of simulation effort (usually sample size) expended to obtain an estimate of g at the requested location. A typical example arises when g is an expectation of some random function G (i.e., $g(x) = \mathrm{E}[Y(x)]$) and its value at any location $x \in \mathcal{D}$ is obtained as a sample average of independent and identically distributed (iid) replicates of $Y(x)$ (i.e., $G_m(x) = m^{-1} \sum_{i=1}^{m} Y_i(x)$).

The question of identifying solutions (some or all) to a system of stochastic nonlinear equations has recently generated much attention. This is primarily owing to its flexibility — instead of having to analytically specify the functions involved, a user can simply "embed" these functions within a simulation of virtually any complexity. Some key application areas include equilibrium problems appearing within transport systems, problems in financial portfolio management, stocking problems in inventory management, and problems on bandwidth management in telecommunication networks. See (Pasupathy 2010) for a long list of references pertaining to several application areas.

In each of the above application contexts, depending on the prevailing needs, the user might seek *some solution* to the stochastic nonlinear system, or *all solutions* to the stochastic nonlinear system. The latter scenario, which is the broad focus of this paper, is typical of global stochastic optimization contexts, that is, contexts where the user seeks the optimal value (or location) of an estimable objective function, subject to a set of estimable constraints. In such cases, the need to identify all solutions of a stochastic nonlinear system arises either from specifying the first-order necessary conditions for local optimality, or simply due

to the feasible region being specified through equality constraints. In either case, the identification of all solutions to the specified nonlinear system is a necessary first step within the global optimization process.

While such global optimization contexts are important motivators, the need to identify all solutions to a stochastic nonlinear system also arises in its own right. For instance, the long-term equilibrium behavior of vehicular-traffic networks (Bierlaire and Crittin 2006; Sheffi 1985), telecommunication networks, and economic systems are frequently of interest for purely descriptive purposes. The equilibrium behavior in such contexts is routinely characterized through a stochastic fixed-point (vector) equation. The "manager" within such contexts might then be interested in all solutions to such a fixed-point equation, since the solutions describe the potential final states of the system that he commands.

How does one solve a stochastic nonlinear system of equations? While various solution strategies are available (Pasupathy 2010; Kushner and Yin 2003; Spall 2003; Pasupathy and Kim 2011), our interest in this paper is limited to a conceptually simple solution method called *sample average approximation* (SAA). SAA, in its basic form, involves generating a sample-path of the function g using an "appropriately chosen" sample size m, and then solving the resulting deterministic problem to desired precision using a chosen numerical procedure. Owing to its simplicity, SAA has recently found enormous expediency alongside stochastic approximation methods (Pasupathy 2010; Kushner and Yin 2003; Spall 2003; Pasupathy and Kim 2011) amongst researchers and practitioners. SAA and more generally sampling based simulation optimization methods have been investigated extensively towards establishing large-sample properties (Robinson 1996; Shapiro et al. 2009; Shashaani et al. 2018; Pasupathy et al. 2018; Pasupathy and Song 2019), small-sample theories on solution quality (Mak et al. 1999; Bayraksan and Morton 2007), and various implementable refinements (**?**; Pasupathy 2010; Pasupathy and Schmeiser 2009).

We emphasize that SAA, as described, implicitly solves only for *a solution* of the specified stochastic nonlinear system. For the current context of identifying *all solutions* to a system of stochastic nonlinear equations, the following simple modification of SAA, henceforth called the method of random restarts (MRR), is necessary: (i) generate a sample-path problem using an "appropriately chosen" sample size m; (ii) generate r random initial solutions according to a chosen sampling measure supported on $\mathcal{D}$; and (iii) execute a convergent root-finding algorithm from each of the r generated initial solutions, and gather the resulting solutions to form the returned set $\Pi^*_m(\mathbf{X_0}(r))$. Such "random restart" algorithms are very popular in the global optimization community (Pardalos and Romeijn 2002).

Our primary concern in this paper relates to efficiency issues within MRR. It seems intuitively clear that when implementing MRR as described above, the number of restarts r and the sample size m should be "in balance" in order to maximize any reasonable measure of efficiency. For a given total computing budget, "too many" restarts would translate to a small sample size, and a correspondingly large sampling variance, in the returned set $\Pi^*_m(\mathbf{X_0}(r))$. "Too few" restarts, by contrast, will result in a strong bias in $\Pi^*_m(\mathbf{X_0}(r))$ due to the potential non-identification of all solutions to the nonlinear system. This apparent tradeoff brings us to the central questions addressed in this paper.

- In MRR, can the quality of the returned set $\Pi^*_m(\mathbf{X_0}(r))$ be rigorously quantified, particularly as a function of the chosen number of random restarts r and the chosen sample size m?
- Does there exist an "optimal" trade-off between the number of restarts r and the sample size m that ensures maximal efficiency in some precise sense? Furthermore, can this trade-off be characterized in a manner that aids implementation?

As we shall see, the answer to first question is in the affirmative and we establish two results that together connect the quality of the provided estimator G_m, and the number of restarts r, with the quality of the returned set $\Pi_m(\mathbf{X_0}(r))$. This lays the foundation for an answer to the second question, where we show that there indeed needs to be a balance between r and m in order to ensure maximal efficiency. More importantly, we demonstrate that such balance is characterized very simply, through a certain logarithmic relationship between r and m.

1.1 Contributions

The problem of identifying all solutions to a system of stochastic nonlinear equations holds great significance due to its applicability within global stochastic optimization settings, and in general within all stochastic fixed-point settings. While the broad solution strategy (modify SAA with random restarts) to tackle such problems seems clear, crucial specific questions surrounding "optimal" parameter choice remain largely unresolved. This paper, with a view toward aiding efficient implementation, addresses some of these questions. The following are specific contributions.

(i) We rigorously establish the effect of sample size (Proposition 2), and the number of restarts (Proposition 3), on the rate at which the expected Hausdorff distance between the returned set and the true set tends to zero in MRR. We show that the said rate's dependence on sample size is canonical, i.e., it is the same as the rate at which the provided simulation estimator G_m tends to its limit g. By contrast, the dependence on the number of random restarts is relatively weak, and intimately connected to the nature of sampling employed within MRR.

(ii) Exploiting the rate established in (i), we demonstrate (Proposition 5) that if (and only if) the sample size and the number of random restarts are chosen to satisfy a simply specified logarithmic relationship, the resulting behavior of the returned set is optimal in a certain precise sense. Furthermore, we show that at such optimality, there is no loss in efficiency due to algorithmic choices. The implications for implementation are easily evident.

(iii) A minor contribution is clarifying the stipulations and the nature of convergence of MRR (Proposition 1). We demonstrate that there are essentially no surprises, i.e., convergence happens in a form and under conditions that one might intuitively expect.

1.2 Organization

The rest of the paper is organized as follows. In Section 2, we introduce some preliminaries including a formal listing of MRR for clarity, notation and terminology used throughout the paper, and the nature of basic convergence within MRR. Section 3 contains the main results of this note, i.e., results on convergence rates and the optimal trade-off between sample size and the number of random restarts. We provide concluding remarks in Section 4.

2 PRELIMINARIES

We remind the reader of the general problem context: identify *all* solutions of the vector equation $g(x) = 0$, where the function $g : \mathcal{D} \to \mathbf{R}^q$ is consistently estimated using $G_m : \mathcal{D} \to \mathbf{R}^q$ for all $x \in \mathcal{D}$. Also, for clarity, we list the algorithmic framework of MRR.

Method of Random Restarts (MRR):

1. Generate a sample-path function $G_m(x)$ using the sample size m.
2. Randomly generate r initial solutions $X_0^1, X_0^2, \ldots, X_0^r$ independently, and each governed by the measure $\mu_{X_0}(\cdot)$ supported on $\mathcal{D}$.
3. Solve (r times) the sample-path root-finding problem $G_m(x) = 0$ using the initial solutions $X_0^1, X_0^2, \ldots, X_0^r$. Record the solutions obtained $X_1^*, X_2^*, \ldots, X_r^*$.
4. Deliver the returned set $\Pi_m^*(\mathbf{X_0}(r)) = \{X_1^*, X_2^*, \ldots, X_r^*\}$ to the user.

2.1 Notation and Terminology

The following notation is used throughout the paper: (i) the generic initial guess for any random restart is denoted by the random variable X_0; (ii) $X_0^1, X_0^2, \ldots, X_0^r$ are r iid copies of the random variable X_0, and $\mathbf{X_0}(r) = (X_0^1, X_0^2, \ldots, X_0^r)$; (iii) $X_n \xrightarrow{p} X$ means that the sequence of random variables $\{X_n\}$ converges to

the random variable X in probability; (iv) $X_n \to X$ wp1 means that the sequence of random variables $\{X_n\}$ converges to the random variable X with probability one; (v) for two sets A,B, the difference set $\Delta(A,B) = (A \cap B^c) \cup (B \cap A^c)$; (vi) the ε-ball around $x \in \mathcal{D}$ will be denoted as $V(x,\varepsilon) = \{y \in \mathcal{D} : \|y-x\| \leq \varepsilon\}$; (vii) for a set A, the diameter $\text{diam}(A) = \sup_{x,y \in A}\{\|x-y\|\}$; (viii) the Lebesgue measure of a measurable set A is denoted $\mathcal{L}(A)$; (ix) the distance of the set $A \subset \mathbb{R}^q$ from the set $B \subset \mathbb{R}^q$ is $\text{dist}(A,B) = \sup_{x \in A}\{\inf_{y \in B} \|x-y\|\}$; (x) the Hausdorff distance $\mathbb{H}(A,B)$ between the sets A and B is $\mathbb{H}(A,B) = \max(\text{dist}(A,B), \text{dist}(B,A))$; (xi) throughout the paper we interchangeably write $\mu_{X_0}(A)$ and $\Pr\{X_0 \in A\}$ to refer to the probability assigned by the sampling measure $\mu_{X_0}(\cdot)$ to the set $A \subset \mathcal{D}$.

We will write $\pi^*(x)$ and $\Pi^*_m(x)$ to denote the solution that will be attained by the root-finding algorithm when executed (hypothetically) on the functions g and G_m respectively, with the initial guess $x \in \mathcal{D}$. Accordingly, the functions g and G_m, along with the root-finding algorithm in use, partition the space $\mathcal{D}$ into *attraction regions*. Specifically, suppose the true set $\pi^* = \{x^*_1, x^*_2, \ldots, x^*_k\}$. Then the attraction regions $B^1, B^2, \ldots, B^k$ corresponding to the function g are $B^i = \{x \in \mathcal{D} : \pi^*(x) = x^*_i\}$. We assume that the algorithm is such that the attraction regions partition $\mathcal{D}$, i.e., $B^i \bigcap B^j = \emptyset$ for $i \neq j$, and $\bigcup_{i=1}^k B^i = \mathcal{D}$. The attraction regions $B^1_m, B^2_m, \ldots,$ corresponding to G_m are defined in similar fashion.

For regularity, we assume that the attraction regions corresponding to the functions g and G_m contain their respective roots in their interior. Formally, for each $x^*_i \in \pi^*$, there exists $\lambda > 0$ such that $V(x^*_i, \lambda) \subset B^i$. A similar condition is assumed to hold for each $X^*_m \in \Pi^*_m$. Accordingly, we define the the minimal-size (denoted msize$(\cdot)$) of an attraction region as the radius of the "largest ball" that can be fitted around the root, while remaining within the attraction region. So, the minimal-size of the attraction region B^i is

$$\text{msize}(B^i) = \sup\{\lambda : V(x^*_i, \lambda) \subset B^i\}.$$

We will frequently talk about the convergence of double sequences, particularly with respect to the number of random restarts r and the sample size m. In such contexts, whenever we write $\lim_{m,r \to \infty} a(m,r)$, we are referring to the double limit of the sequence $\{a(m,r)\}$, as opposed to its iterative limits. (We have found little reason to investigate the latter limits in the current context.) For more on double and iterative limits, and their connection, see Royden (1988).

Finally, we re-emphasize that the true set π^* is the set of solutions to the vector equation $g(x) = 0$, the sample-path set Π^*_m is the set of solutions to the vector equation $G_m(x) = 0$, and the returned set $\Pi^*_m(\mathbf{X_0}(r))$ is the set of solutions that will be returned to the user upon execution of MRR with sample-size m and restarts $\mathbf{X_0}(r) = (X^1_0, X^2_0, \ldots, X^r_0)$.

2.2 Convergence

We begin by clarifying the conditions under which the returned set $\Pi^*_m(\mathbf{X_0}(r))$ in MRR converges to the true set π^*.

Proposition 1 Consider the following conditions.

- A_1. The set of zeros Π^*_m of the function G_m is non-empty for m large enough, wp1.
- A_2. $G_m(x) \to g(x)$ uniformly (in x) as $m \to \infty$ wp1.
- A_3. Constants $t_1, t_2 > 0$ are such that $\|g(x)\| \geq t_1 \mathbb{H}(\{x\}, \pi^*)$, and $\|G_m(x)\| \geq t_2 \mathbb{H}(\{x\}, \Pi^*_m)$ wp1, for all $x \in \mathcal{D}$.
- A_4. The minimal-size (see Section 2.1) of the attraction regions $B^1_m, B^2_m, \ldots$ corresponding to the function G_m are such that $\liminf_{m \to \infty} \inf_j \{\text{msize}(B^j_m)\} > 0$ wp1.
- A_5. Let $R \subset \mathcal{D}$ be a bounded region. If $\mathcal{A}$ is a class of sets in R such that $\inf_{A \in \mathcal{A}}\{\mathcal{L}(A)\} > 0$, then the sampling measure $\mu_{X_0}(\cdot)$ is such that $\inf_{A \in \mathcal{A}}\{\mu_{X_0}(A)\} > 0$.

The following assertions are true.

(i) If Assumptions A_1, A_2, A_3 hold, then $\mathbb{H}(\Pi^*_m, \pi^*) \to 0$ wp1 as $m \to \infty$.

(ii) If Assumptions $A_1, A_2 \ldots, A_5$ hold, then $\lim_{m,r\to\infty} \mathrm{E}[\mathbb{H}^q(\Pi^*_m(\mathbf{X_0}(r)), \pi^*)] = 0$ for $q > 0$.

Proof. Consider $X^*_m \in \Pi^*_m$ (which exists for large enough m by Assumption A_1), and denote $d(X^*_m) = \mathrm{dist}(X^*_m, \pi^*)$. Since π^* is finite, we know that $\mathrm{dist}(X^*_m, \pi^*) = \|X^*_m - x^*\|$ for some $x \in \pi^*$. Applying Assumption A_3 we then have $\|g(X^*_m) - g(x^*)\| = \|g(X^*_m)\| \geq t_1 d(X^*_m)$. However, we also know by Assumption A_2 that for given $\varepsilon > 0$, $\|G_m(X^*_m) - g(X^*_m)\| = \|g(X^*_m)\| \leq \varepsilon$ for large enough m wp1. We thus see that for given $\varepsilon > 0$, $d(X^*_m) \leq \varepsilon/t_1$ for large enough m wp1. This implies that $\lim_{m\to\infty} \mathrm{dist}(\Pi^*_m, \pi^*) = 0$ wp1. Likewise, by analogously considering $x^*_i \in \pi^*$ and using the slope condition on G_m, we have that $\lim_{m\to\infty} \mathrm{dist}(\pi^*, \Pi^*_m) = 0$ wp1. Since both $\mathrm{dist}(\Pi^*_m, \pi^*)$ and $\mathrm{dist}(\pi^*, \Pi^*_m) \to 0$ wp1, the first assertion of the proposition holds.

Towards helping prove the second assertion, we now note three facts.

(a) From assertion (i) and since π^* is finite, the set Π^*_m (and the random variables $\mathbb{H}(\Pi^*_m, \pi^*)$ $\mathbb{H}(\Pi^*_m(\mathbf{X_0}(r)), \pi^*)$)) are uniformly bounded for large enough m wp1.

(b) From (a) and Assumption A_4, there exists $l < \infty$ such that for large enough m, the cardinality $\mathrm{card}(\Pi^*_m) < l$ wp1.

(c) Using Assumption A_4, let $\nu > 0$ be such that for large enough m, $\mathrm{msize}(B^i_m) > \nu$ wp1 for any attraction region B^i_m corresponding to G_m. Also, since Π^*_m is uniformly bounded for large enough m, there exists a bounded rectangular region R such that $\bigcup_{x\in\Pi^*_m} V(x, \nu) \subset R$ for large enough m wp1. Now apply Assumption A_5 by setting $\mathcal{A} = \{A : \mathcal{L}(A) \geq \nu\}$ to see that $\rho = \liminf_{m\to\infty} \inf_i \Pr\{X_0 \in B^i_m\} > 0$.

Now we are ready to prove assertion (ii) of the proposition. We write

$$\begin{aligned}\mathrm{E}[\mathbb{H}^q(\Pi^*_m(\mathbf{X_0}(r)), \pi^*)] = \mathrm{E}[\mathbb{H}^q(\Pi^*_m(\mathbf{X_0}(r)), \pi^*) | \Pi^*_m(\mathbf{X_0}(r)) \neq \Pi^*_m] \Pr\{\Pi^*_m(\mathbf{X_0}(r)) \neq \Pi^*_m\} + \\ \mathrm{E}[\mathbb{H}^q(\Pi^*_m(\mathbf{X_0}(r)), \pi^*) | \Pi^*_m(\mathbf{X_0}(r)) = \Pi^*_m] \Pr\{\Pi^*_m(\mathbf{X_0}(r)) = \Pi^*_m\}. \quad (1)\end{aligned}$$

Consider the second term $\Pr\{\Pi^*_m(\mathbf{X_0}(r)) \neq \Pi^*_m\}$ on the right-hand side of (1). From Boole's inequality (Ross 1998, pp. 66) and using (b), (c) in preceding paragraph, we have for large enough m,

$$\Pr\{\Pi^*_m(\mathbf{X_0}(r)) \neq \Pi^*_m\} \leq \sum_i \Pr\{\cap^r_{j=1} X^j_0(r) \notin B^i_m\} \leq l(1-\rho)^r \to 0 \text{ wp1 as } r \to \infty. \quad (2)$$

Furthermore, since $\mathbb{H}(\Pi^*_m(\mathbf{X_0}(r)), \pi^*)$ is uniformly bounded for large enough m wp1, (2) implies

$$\lim_{m,r\to\infty} \mathrm{E}[\mathbb{H}^q(\Pi^*_m(\mathbf{X_0}(r)), \pi^*) | \Pi^*_m(\mathbf{X_0}(r)) \neq \Pi^*_m] \Pr\{\Pi^*_m(\mathbf{X_0}(r)) \neq \Pi^*_m\} = 0 \text{ wp1}. \quad (3)$$

Next consider the third term $\mathrm{E}[\mathbb{H}^q(\Pi^*_m(\mathbf{X_0}(r)), \pi^*) | \Pi^*_m(\mathbf{X_0}(r)) = \Pi^*_m] = \mathrm{E}[\mathbb{H}^q(\Pi^*_m, \pi^*)]$ appearing on the right-hand side of (1). Since $\mathbb{H}(\Pi^*_m, \pi^*)$ is uniformly bounded for large enough m wp1, and since $\mathbb{H}(\Pi^*_m, \pi^*) \to 0$ wp1, Lebesgue's dominated convergence theorem (Royden 1988, pp. 268) gives $\lim_{m\to\infty} \mathrm{E}[\mathbb{H}^q(\Pi^*_m, \pi^*)] = 0$. We thus have

$$\lim_{m,r\to\infty} \mathrm{E}[\mathbb{H}^q(\Pi^*_m(\mathbf{X_0}(r)), \pi^*) | \Pi^*_m(\mathbf{X_0}(r)) = \Pi^*_m] \Pr\{\Pi^*_m(\mathbf{X_0}(r)) = \Pi^*_m\} = 0 \text{ wp1}. \quad (4)$$

The statements in (3) and (4) imply that $\lim_{m,r\to\infty} \mathrm{E}[\mathbb{H}^q(\Pi^*_m, \pi^*)] = 0$ for $q > 0$. □

Assumptions A_1, A_2 are fairly standard in the context of proving convergence in solution space (Pflug 2004; Shapiro et al. 2009). Condition A_3 stipulates minimum slopes on the function g and its estimator G_m, and is meant to ensure that "nearness" in the function space translates to "nearness" in the domain. Assumptions A_4 is imposed to ensure that roots with vanishing attraction regions do not persist. This is arguably reasonable since no algorithm without specific prior information will be able to systematically locate such roots. Assumption A_5 is very mild and relates the Lebesgue measure of sets in $\mathcal{D}$ to the sampling measure that is in use within the algorithm. Most widely known continuous measures satisfy this assumption.

Since the assertion (ii) of Proposition 1 implies convergence in mean square, $\mathbb{H}(\Pi^*_m(\mathbf{X_0}(r)),\pi^*) \xrightarrow{p} 0$ (in probability) is also guaranteed. Nothing can be said, however, about the almost sure convergence of $\mathbb{H}(\Pi^*_m(\mathbf{X_0}(r)),\pi^*)$ to π^* without further assumptions on the nature of the realizations of the random variable X_0.

3 CONVERGENCE RATES AND OPTIMAL PARAMETER CHOICE

In Section 2.2, we demonstrated that under certain conditions, the moments of the Hausdorff distance between the returned set $\Pi^*_m(\mathbf{X_0}(r))$ and the true set π^* tends to zero. In this section, we delve a little further to establish the rate at which such convergence happens. These results on the rate of convergence will then set the stage for inference on the optimal choices of the number of random restarts r, and sample size m.

For separating the effects of sample size and the number of restarts, we write

$$\begin{aligned}\mathrm{E}[\mathbb{H}(\Pi^*_m(\mathbf{X_0}(r)),\pi^*)] &= \mathrm{E}[\mathbb{H}(\Pi^*_m(\mathbf{X_0}(r)),\pi^*)|\Pi^*_m(\mathbf{X_0}(r))=\Pi^*_m]\mathrm{Pr}\{\Pi^*_m(\mathbf{X_0}(r))=\Pi^*_m\}+\\ &\quad \mathrm{E}[\mathbb{H}(\Pi^*_m(\mathbf{X_0}(r)),\pi^*)|\Pi^*_m(\mathbf{X_0}(r))\neq\Pi^*_m]\mathrm{Pr}\{\Pi^*_m(\mathbf{X_0}(r))\neq\Pi^*_m\}. \qquad (5)\end{aligned}$$

For ease of exposition, we will individually analyze the terms appearing in (5). We first show that under certain conditions, the first multiplier of the first term in (5) simply inherits the convergence rate of the simulation estimator G_m. Specifically, Proposition 2 asserts that, if G_m converges to g according to the rate function $\beta^{-1}(m)$, then $\mathrm{E}[\mathbb{H}(\Pi^*_m(\mathbf{X_0}(r)),\pi^*)|\Pi^*_m(\mathbf{X_0}(r))=\Pi^*_m]$ converges to zero according to $\beta^{-1}(m)$ as well. Since $\mathrm{E}[\mathbb{H}(\Pi^*_m(\mathbf{X_0}(r)),\pi^*)|\Pi^*_m(\mathbf{X_0}(r))=\Pi^*_m]=\mathrm{E}[\mathbb{H}(\Pi^*_m,\pi^*)]$, we have chosen to state Proposition 2 without reference to the number of random restarts r.

Proposition 2 Let conditions A_1, A_2 of Proposition 1 hold. In addition, assume the following.

- A_3'. Constants $t_1, t_1', t_2, t_2' > 0$ satisfy $t_1\mathbb{H}(\{x\},\pi^*) \le \|g(x)\| \le t_1'\mathbb{H}(\{x\},\pi^*)$, and $t_2\mathbb{H}(\{x\},\Pi^*_m) \le \|G_m(x)\| \le t_2'\mathbb{H}(\{x\},\Pi^*_m)$ wp1, for all $x \in \mathcal{D}$.
- A_6. The random variable $\liminf_{m\to\infty}\beta(m)\sup_{z\in\pi^*}\|G_m(z)-g(z)\|$ does not degenerate to zero, i.e., $\mathrm{Pr}\{\liminf_{m\to\infty}\beta(m)\sup_{z\in\pi^*}\|G_m(z)-g(z)\|>s\}>0$ for some $s>0$.
- A_7. $\varepsilon(c):[0,\infty)\to(0,1]$ is a non-increasing function such that the tail probability function $\mathrm{Pr}\{\beta(m)\sup_{x\in\mathcal{D}}\|G_m(x)-g(x)\|\ge c\} \le \varepsilon(c)$ for large-enough m, and $\int_0^\infty \varepsilon(c^{1/2})\,dc<\infty$.

Then for all $\alpha \in [0,2]$,

$$0 < \liminf_{m\to\infty}\mathrm{E}[\beta^\alpha(m)\mathbb{H}^\alpha(\Pi^*_m,\pi^*)] \le \limsup_{m\to\infty}\mathrm{E}[\beta^\alpha(m)\mathbb{H}^\alpha(\Pi^*_m,\pi^*)] < \infty. \qquad (6)$$

Proof. We prove the assertion of the proposition in two parts, corresponding to the first and last inequalities in (6). (The second inequality in (6) holds trivially.) Write

$$\begin{aligned}A_m = \sup_{z\in\Pi^*_m}\{\|G_m(z)-g(z)\|\} = \sup_{z\in\Pi^*_m}\{\|g(z)\|\} &\le \sup_{z\in\Pi^*_m}\{t_1'\,\mathrm{dist}(\{z\},\pi^*)\} = t_1'B_m;\\ C_m = \sup_{z\in\pi^*}\{\|G_m(z)-g(z)\|\} = \sup_{z\in\pi^*}\{\|G_m(z)\|\} &\le \sup_{z\in\pi^*}\{t_2'\,\mathrm{dist}(\{z\},\Pi^*_m)\} = t_2'D_m;\end{aligned}$$

and hence

$$\mathrm{Max}(t_1',t_2')\mathrm{E}[\mathrm{Max}(B_m,D_m)] \ge \mathrm{E}[\mathrm{Max}(A_m,C_m)]. \qquad (7)$$

However, we see that

$$\liminf_{m\to\infty}\mathrm{E}[\beta^\alpha(m)(\mathrm{Max}(A_m,C_m))^\alpha] \ge \liminf_{m\to\infty}\mathrm{E}[\beta^\alpha(m)C_m^\alpha] \ge \mathrm{E}[\liminf_{m\to\infty}\beta^\alpha(m)C_m^\alpha] > 0, \qquad (8)$$

where the second inequality in (8) follows from Fatou's lemma (Serfling 1980, pp. 92) and the third inequality in (8) is by assumption A_6. From (7) and (8), and noticing that $\mathrm{E}[\mathbb{H}^{\alpha}(\Pi_m^*,\pi^*)] = \mathrm{E}[(\mathrm{Max}(B_m,D_m))^{\alpha}]$, we see that the first inequality in (6) has to hold. Next write

$$\begin{aligned} A_m = \sup_{z\in\Pi_m^*}\{\|G_m(z)-g(z)\|\} = \sup_{z\in\Pi_m^*}\{\|g(z)\|\} &\geq \sup_{z\in\Pi_m^*}\{t_1\,\mathrm{dist}(\{z\},\pi^*)\} = t_1 B_m \\ C_m = \sup_{z\in\pi^*}\{\|G_m(z)-g(z)\|\} = \sup_{z\in\pi^*}\{\|G_m(z)\|\} &\geq \sup_{z\in\pi^*}\{t_2\,\mathrm{dist}(\{z\},\Pi_m^*)\} = t_2 D_m; \end{aligned}$$

and hence

$$\mathrm{E}[\mathrm{Max}(A_m,C_m)] \geq \mathrm{Min}(t_1,t_2)\mathrm{E}[\mathrm{Max}(B_m,D_m)]. \tag{9}$$

From Assumption A_7, particularly the stipulation on the tail $\varepsilon(c)$, we have for large enough m,

$$\mathrm{E}[(\beta(m)\mathrm{Max}(A_m,C_m))^{\alpha}] = \int_0^{\infty}\Pr\{(\beta(\mathrm{m})\sup_{\mathrm{x}\in\mathcal{D}}\|\mathrm{G_m(x)-g(x)}\|)^{\alpha} > \mathrm{c}\}\,\mathrm{dc} \leq \int \varepsilon(\mathrm{c}^{1/\alpha})\,\mathrm{dc} < \infty. \tag{10}$$

Now use (9) and (10) to conclude that last inequality in (6) holds as well. □

The assertion of Proposition 2 — the distance between the sample-path set Π_m^* and the true set π^* converges to zero at the same rate as the estimator G_m converges to the limit function g — should not come as a surprise. The main ingredient driving this assertion is Assumption A_3'. By assuming that G_m and g are neither "too flat" nor "too steep," A_3 guarantees that nearness in function values translates into nearness in the domain. (Assumption A_3' replaces the milder assumption Assumption A_3 of Proposition 1 where we only needed the condition that G_m and g are not too flat.) Assumption A_6 identifies $\beta^{-1}(m)$ as the exact rate at which G_m tends to g, as opposed to just saying $G_m = O_p(\beta(m))$. If the function G_m is a sample mean (as is often the case), under very mild conditions (Serfling 1980) the function $\beta(m) = \sqrt{m}$. This gives the canonical case where the squared distance between the sample-path set and the true set is of the order of the reciprocal of the sample size. Finally, Assumption A_7 stipulates that the tails of the random variables $\beta(m)\sup_{x\in\mathcal{D}}\|G_m(x)-g(x)\|$ fall to zero fast enough. This is a standard assumption used to justify the exchange of limits and integrals.

We now turn our attention back to the expression in (5). We have just shown through Proposition 2 that the first term appearing in the right-hand side of (5) converges to zero at the same rate as G_m converges to g, i.e., according to the function $\beta^{-1}(m)$. We now establish, through Proposition 3, the behavior of the other three terms appearing in (5). Specifically, assertion (i) in Proposition 3 demonstrates that the third term appearing on the right-hand side of (5) remains bounded away from 0 and ∞ as $m,r\to\infty$. Similarly, assertion (ii) in Proposition 3 demonstrates that the fourth term appearing on the right-hand side of (5) tends to zero exponentially in the number of random restarts r. Propositions 2 and 3 thus collectively (and completely) describe the convergence rate behavior of the moments of $\mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)),\pi^*)$.

Proposition 3 Consider the following assumptions.

- A_8. There exists a labeling $B_m^1, B_m^2, \ldots,$ of the attraction regions corresponding to the function G_m such that they satisfy $\mathcal{L}(\Delta(B_m^i, B^i)) \to 0$ wp1 as $m\to\infty$, for $i = 1,2,\ldots,k$.
- A_9. The sampling measure $\mu_{X_0}(\cdot)$ is continuous in the Lebesgue measure, i.e., if $A\subset\mathcal{D}$ and $\{A_i\}$ is a sequence of sets in $\mathcal{D}$ satisfying $\mathcal{L}(\Delta(A_i,A))\to 0$, then $\mu_{X_0}(A_i)\to\mu_{X_0}(A)$.

Let Assumptions $A_1, A_2, \ldots, A_9$ hold, with the more stringent A_3' (Proposition 2) substituted for A_3. Then the following assertions are true.

(i) Denoting $C(r,m) = \mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)),\pi^*)|\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m^*$,

$$0 < \liminf_{m,r\to\infty}\mathrm{E}[C(r,m)] \leq \limsup_{m,r\to\infty}\mathrm{E}[C(r,m)] < \infty.$$

(ii) Denote the "smallest attainment probability" $\rho_* = \min_{x\in\pi^*} \Pr\{\pi^*(X_0) = x\}$, we have for all $\delta > 0$ (assuming $\delta < 1-\rho_*$),

$$\liminf_{m,r\to\infty} \frac{\Pr\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m^*\}}{(1-(\rho_*+\delta))^r} = \infty; \quad \limsup_{m,r\to\infty} \frac{\Pr\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m^*\}}{(1-(\rho_*-\delta))^r} = 0.$$

Proof. To prove the last inequality in assertion (i), recall that π^* is finite, and the set Π_m^* is uniformly bounded for large enough m. Hence, there exists $\kappa_1 < \infty$ such that for large-enough m and all r, we can write $C(m,r) \leq \text{diam}(\pi^* \cup \Pi_m^*) \leq \text{diam}(\pi^*) + \text{diam}(\Pi_m^*) < \kappa_1$. It thus follows that $\limsup_{m,r\to\infty} \mathrm{E}[C(r,m)] < \infty$.

For proving the first inequality in assertion (i), we see from Proposition 1 and Assumption A_4 that there have to be exactly k attraction regions (and k elements in Π_m^*) corresponding to G_m for large enough m wp1. Therefore, for given small-enough $\varepsilon > 0$ and large-enough m, there exists $\kappa_2 > 0$ such that $C(m,r) \geq \inf_{x\in\Pi_m^*} \text{dist}(\pi^*, \Pi_m^* \backslash \{x\}) \geq \inf_{i\neq j}\{\|x_i^* - x_j^*\| - \varepsilon\} > \kappa_2$, after recalling that $\pi^* = \{x_1^*, x_2^*, \ldots, x_k^*\}$. Apply Fatou's lemma (Royden 1988, pp. 86) to get $\liminf_{m,r\to\infty} \mathrm{E}[C(r,m)] \geq \mathrm{E}[\liminf_{m,r\to\infty} C(r,m)] > 0$.

For proving assertion (ii), denote $\Pr\{X_0 \in B_m^i | G_m\} = \rho_{i,m}$. Then, we note that

$$\begin{aligned} &\Pr\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m^* | G_m\} \\ &= \sum_{i_1} \rho_{i_1,m}^r + \sum_{i_1\neq i_2} (\rho_{i_1,m} + \rho_{i_2,m})^r + \cdots + \sum_{i_1\neq i_2 \neq \cdots \neq i_{k-1}} (\rho_{i_1,m} + \rho_{i_2,m} + \cdots + \rho_{i_{k-1},m})^r, \end{aligned} \tag{11}$$

where each index i_j satisfies $i_j \in \{1,2,\ldots,k\}$ if m is large enough. Due to Assumption A_8 and Assumption A_9, we see that $\rho_{*m} = \text{Min}\{\rho_{i,m} : i = 1,2,\ldots,k\} < \rho_* + \delta$ (and hence $1-\rho_{*m} > 1-\rho_*-\delta$) for large enough m wp1, and for any fixed $\delta > 0$. Since the largest summand in the last term appearing in (11) can be expressed as simply $(1-\rho_{*m})^r$, we see that

$$\frac{\Pr\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m | G_m\}}{(1-(\rho^*+\delta))^r} > \frac{(1-\rho_{*m})^r}{(1-(\rho^*+\delta))^r} \to \infty \text{ as } m,r \to \infty, wp1.$$

Now noting that $\Pr\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m\} = \int \Pr\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m | G_m\}\, \nu(dG_m)$ and using Fatou's lemma (Royden 1988, pp. 86), we conclude that the first statement in assertion (ii) should hold.

The second statement in assertion (ii) follows in similar fashion. Note that $1-\rho_{*m} < 1-(\rho_*-\delta)$ for large enough m wp1, implying that $\lim_{m,r\to\infty}(1-\rho_{*m})^r/(1-(\rho_*-\delta))^r = 0$. Since $(1-\rho_{*m})^r$ is the largest summand in (11), we have $\lim_{m,r\to\infty} s^r/(1-(\rho_*-\delta))^r = 0$ for any summand s^r in (11). Moreover, since there are $\binom{k}{1} + \binom{k}{2} + \cdots + \binom{k}{k-1} = 2^k - 2 < \infty$ summands in (11) for large enough m, we see that $\limsup_{m,r\to\infty}(1-(\rho_*-\delta))^{-r}\Pr\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m^* | G_m\} = 0$. Finally note that $\Pr\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m\} = \int \Pr\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m | G_m\}\, \nu(dG_m)$ and use Lebesgue's dominated convergence theorem (Royden 1988, pp. 268) to conclude that the second statement in (ii) holds as well. □

We see assumption A_8 as a certain type of continuity condition on the algorithm. It stipulates that as G_m tends to its limit g, the behavior of the root-finding algorithm on G_m should tend to its (hypothetical) behavior on g. Assumption A_9 is a weak continuity assumption on the choice of the sampling measure $\mu_{X_0}(\cdot)$. This assumption is again satisfied by virtually all well-known continuous measures.

Propositions 2 and 3 are crucial in that they give clues into how the number of restarts r should relate to the sample size m, if $\mathrm{E}[\mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)), \pi^*)]$ is to converge to zero rapidly. For example, from the convergence rates $\beta^{-1}(m)$ and $(1-(\rho_* \pm \delta))^r$ that appear within Propositions 2 and 3, it seems that the slower of the two sequences $\beta^{-1}(m)$ and $(1-\rho_*)^r$ will determine the speed at which $\mathrm{E}[\mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)), \pi^*)]$ tends to zero, where ρ_* measures the lowest "attainment probability" amongst all solutions. Loosely speaking, it thus makes intuitive sense to choose the number of restarts r such that $\beta^{-1}(m)$ and $(1-\rho_*)^r$ converge to zero at the same rate. In what follows, Proposition 5 clarifies this intuitive notion. We first state a simple result (without proof) that will be helpful in proving Proposition 5.

Proposition 4 Let $\{a(m)\}$ and $\{b(m)\}$ be positive valued sequences satisfying $a(m) \to \infty, b(m) \to \infty$ as $m \to \infty$, and let the limit $\lim_{m\to\infty} a(m)/b(m)$ exist (allowing for ∞). Let q be a constant satisfying $0 < q < 1$. Then the following hold.

(i) If $\lim_{m\to\infty} a(m)/b(m) < 1$, then $\liminf_{m\to\infty} q^{a(m)}/q^{b(m)} = \infty$.
(ii) If $\lim_{m\to\infty} a(m)/b(m) > 1$, then $\limsup_{m\to\infty} q^{a(m)}/q^{b(m)} = 0$.

We are now ready for Proposition 5. We use the notation $r(m)$ instead of r in the statement of Proposition 5 to make explicit the dependence of the number of restarts r on the sample size m.

Proposition 5 Let the assumptions $A_1, \ldots, A_9$ introduced in Propositions 1 through 3 hold, with A_3' substituted for A_3. Assume the limit $\lim_{m\to\infty} r(m)/\log\beta(m) = l$ exists, after allowing $l = \infty$. Then the following assertions hold for any $\delta > 0$ (assuming $\delta < 1 - \rho_*$).

(i) If $l \in [0, -1/\log(1-(\rho_*+\delta))]$, then $\liminf_{m\to\infty} \beta(m)\mathrm{E}[\mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)), \pi^*)] = \infty$.
(ii) If $l \in [-1/\log(1-(\rho_*-\delta)), \infty]$, then

$$0 < \liminf_{m\to\infty} \beta(m)\mathrm{E}[\mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)), \pi^*)] \le \limsup_{m\to\infty} \beta(m)\mathrm{E}[\mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)), \pi^*)] < \infty.$$

Proof. By dividing (5) through by $\beta^{-1}(m)$ we have

$$\begin{aligned}\frac{\mathrm{E}[\mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)), \pi^*)]}{\beta^{-1}(m)} &= \frac{\mathrm{E}[\mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)), \pi^*)|\Pi_m^*(\mathbf{X_0}(r)) = \Pi_m^*]}{\beta^{-1}(m)}\mathrm{Pr}\{\Pi_m^*(\mathbf{X_0}(r)) = \Pi_m^*\} + \\ &\quad \mathrm{E}[\mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)), \pi^*)|\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m^*]\frac{\mathrm{Pr}\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m^*\}}{\beta^{-1}(m)}. \qquad (12)\end{aligned}$$

We will now individually analyze the behavior of the four terms on the right-hand side of (12) to deduce the assertions of the proposition. The first term $t_1(m) = \beta(m)\mathrm{E}[\mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)), \pi^*)|\Pi_m^*(\mathbf{X_0}(r)) = \Pi_m^*]$ appearing on the right-hand side of (12) is independent of r and satisfies

$$0 < \liminf_{m\to\infty} t_1(m) \le \limsup_{m\to\infty} t_1(m) < \infty, \qquad (13)$$

by Proposition 2. From (2) in proof of Proposition 1, the second term $t_2(m,r) = \mathrm{Pr}\{\Pi_m^*(\mathbf{X_0}(r)) = \Pi_m^*\}$ appearing on the right-hand side of (12) satisfies

$$\lim_{m,r\to\infty} t_2(m,r) = 1. \qquad (14)$$

The third term $t_3(m,r) = \mathrm{E}[\mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)), \pi^*)|\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m^*]$ appearing on the right-hand side of (12) satisfies

$$0 < \liminf_{m,r\to\infty} t_3(m,r) \le \limsup_{m,r\to\infty} t_3(m,r) < \infty, \qquad (15)$$

by assertion (i) of Proposition 3.

To understand the behavior of the last term $t_4(m,r) = \beta(m)\mathrm{Pr}\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m^*\}$, first apply assertion (i) of Proposition 4 with $a(m) = r$, $b(m) = \log\beta(m)/\log 1/q$, and $q = 1-(\rho_*+\delta)$. We get $\liminf_{m\to\infty}(1-(\rho_*+\delta))^r/\beta^{-1}(m) = \infty$. This in combination with assertion (ii) of Proposition 3 implies that the last term $t_4(m,r)$ appearing on the right-hand side of (12) satisfies

$$\liminf_{m,r\to\infty} t_4(m,r) = \liminf_{m,r\to\infty} \frac{(1-(\rho_*+\delta))^r}{\beta^{-1}(m)}\frac{\mathrm{Pr}\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m^*\}}{(1-(\rho_*+\delta))^r} = \infty. \qquad (16)$$

Apply (13), (14), (15), (16) to (12), and conclude that the first assertion of the proposition holds.

To prove the second assertion, we adopt a similar approach to deduce the behavior of $t_4(m,r)$. Apply (ii) of Proposition 4 with $a(m) = r$, $b(m) = \log\beta(m)/\log 1/q$, and $q = 1-(\rho_* - \delta)$. We get $\limsup_{m\to\infty}(1-(\rho_*-\delta))^r/\beta^{-1}(m) = 0$. This in combination with assertion (ii) of Proposition 3 implies that $t_4(m,r)$ satisfies

$$\limsup_{m,r\to\infty} t_4(m,r) = \limsup_{m,r\to\infty} \frac{(1-(\rho_*-\delta))^r}{\beta^{-1}(m)} \frac{\Pr\{\Pi_m^*(\mathbf{X_0}(r)) \neq \Pi_m^*\}}{(1-(\rho_*-\delta))^r} = 0. \tag{17}$$

Apply (13), (14), (15), (17) to (12), and conclude that the second assertion of the proposition holds.

□

Three implications are apparent from the assertions in Proposition 5. First, assertion (i) implies that selecting the number of restarts to increase at a rate slower than the rate at which $\log\beta(m)$ tends to infinity causes the expected Hausdorff distance (between the returned set $\Pi_m^*(\mathbf{X_0}(r))$ and the true set π^*) to converge to zero at a rate slower than $\beta^{-1}(m)$ — the rate at which G_m converges to g. Second, the maximal achievable convergence rate of the expected Hausdorff distance $\mathrm{E}[\mathbb{H}(\Pi_m^*(\mathbf{X_0}(r)),\pi^*)]$ is $\beta^{-1}(m)$. This rate is achieved, for instance, if $\log\beta(m) = o(r)$, i.e., the number of restarts is chosen to tend to infinity faster than $\log\beta(m)$. In fact, any choice of restarts such as $r = \log\beta^{-1}(m)/\log(1-u)$, where $u < \rho_*$, will achieve the maximal convergence rate. The implicit assertion here is that it is most efficient to choose r to be as close as possible but above $\log\beta^{-1}(m)/\log(1-\rho_*)$, since no additional benefit is incurred in the form of an increased convergence rate due to increasing the number of restarts any further. Third, it is apparent from the result that efficiency is critically dependent on the probability of sampling from within the "smallest attraction region." Specifically, it seems that efficiency gains are likely by maximizing $\rho_* = \mathrm{Min}_{x\in\pi^*}\{\pi^*(X_0) = x\}$ through our sampling strategy. In the extreme case, it is most efficient if μ_{X_0} is such that the probability of the algorithm evolving to any of the k zeros of g, when (hypothetically) executed on g directly, are equal.

4 SUMMARY AND CONCLUDING REMARKS

MRR is a simple adaptation of sample average approximation targeted at the question of identifying *all solutions* to a system of stochastic nonlinear equations, and with immediate relevance to global simulation optimization. The method involves generating an approximate deterministic problem with a fixed sample size, executing a convergent root-finding algorithm with several random restarts on the generated problem, and then gathering the obtained solutions to form the estimator of the set of true solutions to the system.

In this note, we argue with rigorous theoretical support that in order to guarantee maximal efficiency in MRR, the chosen number of random restarts should satisfy a certain simple logarithmic relationship with the chosen sample size. In addition to being maximally efficient, maintaining this relationship guarantees that efficiencies implicit in the simulation estimator are preserved, i.e., there is no loss in efficiency due to algorithmic choice. Since the recommended relationship between restarts and sample size is canonical, we intend its use as a broad guiding principle when devising specific implementations within simulation-based root finding and optimization contexts.

Three other remarks relating to future research are in order.

(i) The "optimal" relationship between restarts and sample size established in this paper, while useful as a guiding principle, still leaves constants unchosen. This is evident, for instance, from the fact that the recommended relationship is (seemingly) independent of dimension! Efficient implementation will crucially hinge on how these problem-specific constants are determined — an important area that is worthy of further investigation.

(ii) For convenience, we have assumed throughout the paper that the solution obtained by the root-finding algorithm during each random restart has "negligible error." While we believe this assumption is reasonable for the purposes of obtaining the broad relationship between restarts and sample size,

how much to solve each random execution of the root-finding algorithm is an important question. This problem has been addressed to a substantial extent for the context of finding *any solution* to a system of stochastic nonlinear equations (Pasupathy and Kim 2011; Pasupathy 2010) and we believe similar results will hold for the current context.

(iii) The quality of the estimator as obtained through MRR was shown to be inversely related to ρ_* — the attraction region having the smallest measure under the chosen sampling plan. It is thus true that convergence can be accelerated by making ρ_* as large as possible. In other words, our sampling measure μ_{X_0} should be tilted such that the likelihood of the algorithm evolving to any of the k zeros of g (during a hypothetical execution on g) should be equalized.

REFERENCES

Bayraksan, G., and D. P. Morton. 2007. "Assessing Solution Quality in Stochastic Programs". *Mathematical Programming Series B* 108:495–514.

Bierlaire, M., and F. Crittin. 2006. "Solving Noisy, Large-Scale Fixed-Point Problems and Systems of Nonlinear Equations". *Transportation Science* 40(1):44–63.

Kushner, H. J., and G. G. Yin. 2003. *Stochastic Approximation and Recursive Algorithms and Applications*. New York, NY: Springer-Verlag.

Mak, W. K., D. P. Morton, and R. K. Wood. 1999. "Monte Carlo Bounding Techniques for Determining Solution Quality in Stochastic Programs". *Operations Research Letters* 24:47–56.

Pardalos, P. M., and H. E. Romeijn. (Eds.) 2002. Volume 2 of *Handbook of Global Optimization*. Kluwer Academic Publishers.

Pasupathy, R. 2010. "On Choosing Parameters in Retrospective-Approximation Algorithms for Stochastic Root Finding and Simulation Optimization". *Operations Research* 58:889–901.

Pasupathy, R., P. W. Glynn, S. G. Ghosh, and F. S. Hashemi. 2018. "On Sampling Rates in Simulation-Based Recursions". *SIAM Journal on Optimization* 28(1):45–73.

Pasupathy, R., and S. Kim. 2011. "The Stochastic Root-Finding Problem: Overview, Solutions, and Open Questions". *ACM Transactions on Modeling and Computer Simulation* 21(3):1–23.

Pasupathy, R., and B. W. Schmeiser. 2009. "Retrospective-Approximation Algorithms for Multidimensional Stochastic Root-finding Problems". *ACM Transactions on Modeling and Computer Simulation* 19(2):5:1–5:36.

Pasupathy, R., and Y. Song. 2019. "An Adaptive Sequential SAA Algorithm for Solving Two-stage Stochastic Programs.". *Optimization Online*. http://www.optimization-online.org/DB_HTML/2019/02/7068.html, accessed October 28, 2019.

Pflug, G. C. 2004. "Stochastic Optimization and Statistical Inference". In *Stochastic Programming*, edited by A. Ruszczynski and A. Shapiro, Handbooks in Operations Research and Management Science, 427–482. Elsevier, New York, NY.

Robinson, S. 1996. "Analysis of Sample-path Optimization". *Mathematics of Operations Research* 21:513–528.

Ross, S. 1998. *A First Course in Probability*. New York, New York: Prentice Hall.

Royden, H. 1988. *Real Analysis*. New York, New York: Prentice Hall.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. New York, New York: John Wiley & Sons, Inc.

Shapiro, A., D. Dentcheva, and A. Ruszczynski. 2009. *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics.

Shashaani, S., F. S. Hashemi, and R. Pasupathy. 2018. "ASTRO-DF: A Class of Adaptive Sampling Trust-Region Algorithms for Derivative-Free Simulation Optimization.". *SIAM Journal on Optimization* 28(4):3145–3176.

Sheffi, Y. 1985. *Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods*. New York, New York: Prentice Hall.

Spall, J. C. 2003. *Introduction to Stochastic Search and Optimization*. Hoboken, New Jersey: John Wiley & Sons, Inc.

AUTHOR BIOGRAPHIES

RAGHU PASUPATHY is Associate Professor in the Department of Statistics at Purdue University. His research interests lie broadly in Monte Carlo methods with a focus on stochastic optimization. He is a member of INFORMS, IIE, and ASA. His email address is pasupath@purdue.edu.