Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

QUANTILE-BOOSTED STOCHASTIC APPROXIMATION

Jinyang Jiang[1,2], Bernd Heidergott[3], and Yijie Peng[1,2]

[1]Guanghua School of Management, Peking University, Beijing, CHINA
[2]Xiangjiang Laboratory, Changsha, CHINA
[3]Vrije Universiteit Amsterdam, Amsterdam, THE NETHERLANDS

ABSTRACT

Stochastic approximation (SA) offers a recursive framework for tracking the quantiles of a parameterized system's output distribution using observed samples. In this paper, we employ SA-based quantile trackers to approximate the gradient of an objective function and integrate them into a unified SA scheme for finding stationary points. The proposed gradient estimation framework accommodates both finite-difference and score-function methods. Our method allows for dynamically adjusting the number of trackers within a single optimization run. This adaptability enables more efficient and accurate approximation of the true objective gradient. The resulting single time-scale estimator is also applicable to stationary performance measures. Numerical experiments confirm the effectiveness and robustness of the proposed approach.

1 INTRODUCTION

Stochastic approximation (SA) is a standard technique for simulation/observation-driven optimization of stochastic models; see, e.g., Vázquez-Abad and Heidergott (2025), Kushner and Yin (2003), Borkar (2008). In general, SA is most commonly used to solve expectation minimization problems of the form $\min_\theta \mathbb{E}_\theta[Y]$ via gradient descent. The standard SA algorithm takes the form $\theta_{k+1} = \theta_k + \varepsilon_k G_k$, where θ_k is the parameter estimate at the k-th state of the process, ε_k is the gain size sequence, and G_k represents the update from θ_k, scaled by ε_k to the next subsequent value θ_{k+1}. Typically, G_k is taken to be a proxy/estimator for $d\mathbb{E}_{\theta_k}[Y]/d\theta$. A rich literature on unbiased gradient estimation exists, which is applicable in case we have explicit knowledge on the dependency of Y on θ. For example, when $Y = h(V, X(\theta))$, where $X(\theta)$ represents the θ-dependent random input, V denotes the θ independent input, and $h(\cdot)$ is some performance mapping known in closed form, the infinitesimal perturbation analysis (IPA) gradient estimator

$$X'(\theta)\frac{\partial}{\partial x}h(V,x)\big|_{x=X(\theta)}$$

is available; see, e.g., Ho and Cao (1991), Glasserman (1991), Cassandras and Lafortune (2008), Kroese et al. (2013). While IPA is known to have good performance as a gradient proxy in *white-box model* optimization, it requires explicit knowledge of $h(\cdot)$ and path-wise differentiability of $h(\cdot)$ and $X(\theta)$. When only *partial model* information is available, specifically, the distribution of $X(\theta)$ is known but only Y is available to us through observation, the score function (SF) estimator can be used

$$h(V,X(\theta))\frac{d}{d\theta}\ln\varphi_\theta(x)\big|_{x=X(\theta)},$$

where $\varphi_\theta(\cdot)$ is the differentiable density of $X(\theta)$; see, e.g., Rubinstein and Shapiro (1993), Rubinstein and Kroese (2016), Cassandras and Lafortune (2008), Rubinstein and Melamed (1998), Kroese et al. (2013). In the most general case, we have no information on the explicit dependence of Y on θ, i.e., facing a *black-box model*, one has to resort to brute force approximation via finite difference (FD), e.g., Simultaneous Perturbation Stochastic Approximation (SPSA); see Spall (2005), Bhatnagar et al. (2013), and use

$$\frac{1}{\Delta}\left(\mathbb{E}_{\theta+\Delta}[Y] - \mathbb{E}_\theta[Y]\right),$$

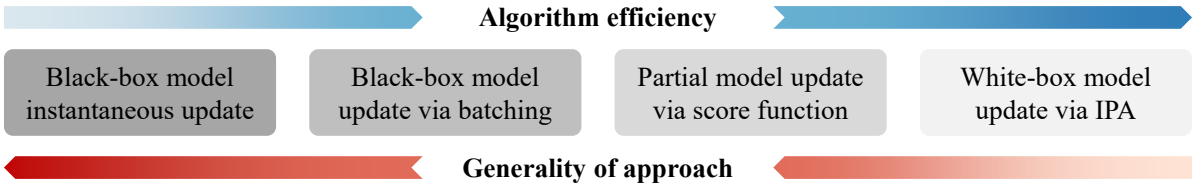


Figure 1: Trade-off between model knowledge and algorithm efficiency.

where Δ is the perturbation size. A representative black-box example could be when Y is the market feedback of operating a complex production system with some controllable input parameter θ.

In this paper, we propose a novel approach for optimization under black-box settings. Our method builds upon the availability of an SA-inspired algorithm for quantile tracking, as introduced in Section 2. We propose to apply the quantile tracking while running the optimization algorithm with negligible computation overhead and no extra simulation cost. The quantile trackers enable us to exploit historical data to numerically approximate $\mathbb{E}_\theta[Y]$ using a quantile-based reconstruction of the distribution functions. The discretization error is roughly related to the number of quantile trackers. Using quantile tracking in combination with SA is a well-established approach in the analysis of distortion risk measures (Jiang et al. 2024). The conventional approaches, such as optimization under the nested simulation framework (Li and Peng 2024), use a fixed number of quantile trackers throughout the entire experiment, which leads to a hard trade-off between computation burden and accuracy. The convergence analysis often focuses on the reduction of estimation error across multiple experimental runs, as the number of outer-layer samples increases, rather than on the vanishing discretization error within a single run (Gordy and Juneja 2010; Feng and Song 2024). In contrast, our method allows for dynamically adjusting the number of trackers within a single optimization run. This adaptability enables a more efficient and accurate approximation of the true objective gradient. From a numerical perspective, our approach partially transforms the standard Monte Carlo integration of $\mathbb{E}_\theta[Y]$ into a quantile-based numerical integration, using tracked quantiles as a discretization of the underlying distribution functions. When we integrate this quantile-descretization approach into gradient-descent optimization, it becomes comparable to the finite-difference method, and shows the potential of variance reduction with the score function.

This quantile-informed expectation approximation enables a finite-difference gradient estimator, which is detailed in Section 3, without requiring explicit knowledge of the relationship between θ and Y. Our algorithm allows updates with only a single observation of Y per iteration, making it data-efficient and suitable for online settings. We first present a basic version of the quantile-boosted finite-difference approximation algorithm for convergence analysis and then extend it to a more stable variant using mini-batches of observations. Furthermore, in Section 3.3, we generalize the method to partial-model settings, integrating quantile trackers to SF-based gradient estimators. Our quantile tracking framework fundamentally reshapes the algorithmic design of SA for non-risk problems. The relationships between our proposed method and existing gradient estimation strategies, as well as their targeted scenarios, are summarized in Figure 1. The paper concludes with numerical experiments on financial investment and queueing examples. The experimental results are better than or comparable to the corresponding baselines, demonstrating the efficiency of our proposed approach. While we use the expectation objective as a working example in this paper, our adaptive quantile-based reconstruction technique introduces minimal degradation in optimization performance and will be extended to more complex performance measures, including ones that are difficult to handle using standard Monte Carlo estimation, such as distortion risk measures, in the full version.

2 STEADY-STATE QUANTILE TRACKING

2.1 Quantile Tracking of Limiting Distribution

Our approach is based on the recursive quantile tracking introduced by Spall (2005). Consider an independent and identically distributed (i.i.d.) sequence $\{Y_k\}$ with cumulative distribution function (CDF) $F(\cdot)$. Given an

$\alpha \in (0,1)$, the α-quantile q^α is defined by $q^\alpha = \inf\{y : F(y) \geq \alpha\} = F^{-1}(\alpha)$, assuming $F(\cdot)$ is continuous and strictly increasing. Finding q^α can be formulated as an optimization problem as follows:

$$\min_{q \in \mathbb{R}} (\alpha - F(q))^2. \tag{1}$$

Observe that the function $G(q) = \alpha - F(q)$ provides a descent direction for problem (1). The α-quantile can then be tracked by the SA algorithm:

$$q^\alpha_{k+1} = q^\alpha_k + \varepsilon_k (\alpha - \mathbf{1}\{Y_k \leq q^\alpha_k\}), \tag{2}$$

where $\mathbf{1}\{Y_k \leq q\}$ provides an unbiased stochastic estimator of $F(q)$, for an appropriate choice of gain size ε_k. Since the variance of the stochastic updates $\alpha - \mathbf{1}\{Y_k \leq q^\alpha_k\}$ is bounded, the algorithm (2) converges, under mild conditions, a.s. to the true quantile.

Theorem 1 Let $\{Y_k\}$ be an i.i.d. sequence with continuous and strictly increasing CDF $F(\cdot)$. Then, the SA algorithm (2) converges a.s. to q^α for $\alpha \in (0,1)$, provided that $\sum_k \varepsilon_k = \infty$ and $\sum_k \varepsilon_k^2 < \infty$:

$$\lim_{k \to \infty} q^\alpha_k = q^\alpha \quad \text{a.s.}$$

Theorem 1 is also compatible with a dynamic generating mechanism of Y. We assume that Y is an observable one-dimensional output variable defined on some underlying θ dependent process $\xi(\theta)$, e.g., $Y_k = h(\xi_k(\theta_k))$, where $\xi_k(\theta_k)$ could represent the current state of a production system operating at parameter θ_k, which controls the speed of one of the machines, and $h(\cdot)$ denotes the evaluation procedure of the system. The process $\xi(\theta)$ captures the inherent randomness of the system. For our analysis, we assume that $\{\xi_k(\theta_k)\}$ is a Markov chain, or more formally,

$$\mathbb{P}_{\theta_{k+1}}(B,x) = \mathbb{P}_{\theta_{k+1}}(\xi_{k+1} \in B \,|\, \xi_k(\theta_k) = x) = \mathbb{P}_{\theta_{k+1}}(\xi_{k+1} \in B \,|\, \xi_i(\theta_i) = x_i; 1 \leq i \leq k),$$

for all $k \geq 1$, $\{x_i\}$ representing a sample path of the underlying system, and all measurable sets B. This model covers the case where θ_k is adapted while running the system (i.e., without starting the system anew after each update of θ_k). Hence, we can let the system run and update θ after each observation of Y. More explicitly, the value of the quantile tracker q^α_{k+1} depends on the sample path information $Y^{\theta_{1:k}}_{1:k} = (Y_k(\theta_k), Y_{k-1}(\theta_{k-1}), \ldots, Y_1(\theta_1))$, i.e.,

$$q^\alpha_{k+1} = q^\alpha_{k+1}(Y^{\theta_{1:k}}_{1:k}) = q^\alpha_{k+1}(Y_k(\theta_k), Y_{k-1}(\theta_{k-1}), \ldots, Y_1(\theta_1)).$$

We make the following technical assumptions for further analysis:

Assumption 1 Let $\Theta \subset \mathbb{R}$ be a compact set, and consider a parameterized family of Markov chains $\{\xi_k(\theta)\}$, $\theta \in \Theta$, with state space Ξ, each admitting a unique stationary measure $\pi_\theta(\cdot)$. For all $\theta \in \Theta$, the process $\xi_k(\theta)$ is ergodic with respect to $h(\cdot)$, i.e.

$$\int_\Xi h(x)\pi_\theta(dx) = \lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} h(\xi_i(\theta)) \quad \text{a.s.}$$

We define $Y_k(\theta) = h(\xi_k(\theta))$, for $k \geq 1$. The transition probabilities $\mathbb{P}_\theta(\cdot, \xi)$ are continuous in (θ, ξ) on $\mathbb{R} \times \Xi$. The stationary measure of the fixed-θ process $\{\xi(\theta)\}$, denoted by π_θ, is unique, and the set $\{\pi_\theta(\cdot), \theta \in \Theta\}$ of stationary measures of the fixed-θ processes is tight on Θ.

Assumption 2 For all $k \geq 1$, the CDF of Y_k at θ, $F_{k,\theta}(y) = \mathbb{P}_\theta(Y_k \leq y)$, has a connected support within $[a,b] \subset \mathbb{R}$ and continuous with respect to y for all $\theta \in \Theta$. Moreover, the limiting distribution of Y_k at θ is given by

$$F_\theta(y) = \int_\Xi 1\{h(x) \leq y\}\pi_\theta(dx),$$

for y in $[a,b]$, and is continuous with respect to y for all $\theta \in \Theta$.

We now extend the i.i.d. limit put forward in Theorem 1 to the case of tracking the quantile of a limiting distribution. As this result holds for all θ fixed, we suppress θ in the notation for simplicity.

Theorem 2 If Assumptions 1 and 2 hold, then the SA algorithm (2) converges a.s. to q^α for $\alpha \in (0,1)$, provided that $\sum_k \varepsilon_k = \infty$ and $\sum_k \varepsilon_k^2 < \infty$:

$$\lim_{k\to\infty} q_k^\alpha(Y_{1:k-1}) = q^\alpha \quad \text{a.s.},$$

where q^α denotes the α-quantile of the limiting distribution of Y_k.

Proof. Under Assumptions 1 and 2, we have the SA Markovian noise model in Vázquez-Abad and Heidergott (2025) or the Markov state-dependent noise model in Chapter 6 of Kushner and Yin (2003). The analysis follows these references. By ergodicity of ξ_k (for fixed θ), it holds w.p.1 for every $q \in \mathbb{R}$:

$$G(q) = \lim_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}[\alpha - \mathbf{1}\{Y_k \le q\}] = \lim_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}[\alpha - \mathbf{1}\{h(\xi_k) \le q\}] = \lim_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} (\alpha - \mathbf{1}\{h(\xi_k) \le q\}).$$

Since $G(q)$ is a descent direction for the quantile tracking problem in (1) with $F(\cdot)$ being the limiting distribution of Y_k, and the variance of the instantaneous update has bounded variance, it follows from Theorem 5.1 in Vázquez-Abad and Heidergott (2025) that (2) converges a.s. to the α-quantile of the limiting distribution. □

2.2 Quantile Tracking for Distribution Approximation

In this section, we develop a numerical approximation of

$$\mathbb{E}_\theta[Y] = \int_\Xi h(x)\pi_\theta(dx) = \int_a^b y F_\theta(dy) = \int_0^1 F_\theta^{-1}(z)dz$$

based on quantile values. As in Section 2.1, we omit θ in the formulas, since the analysis in this section is conducted for a fixed θ. We split the unit interval into M segments with $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_M = 1$ and denote the approximation of $F(\cdot)$ as $F_M(\cdot)$ given by the linear interpolation of quantile points $\{q^{\alpha_m}\}$, i.e.,

$$F_M(y) = \frac{\alpha_m - \alpha_{m-1}}{q^{\alpha_m} - q^{\alpha_{m-1}}}(y - q^{\alpha_{m-1}}) + \alpha_{m-1}, \quad \text{for } y \in [q^{\alpha_{m-1}}, q^{\alpha_m}],\ 1 \le m \le M, \tag{3}$$

for $y \in [a,b]$, where $q^{\alpha_0} = a$ and $q^{\alpha_M} = b$. Note that $F_M(y)$ tends to $F(y)$ for all y as $M \to \infty$. Denote the inverse of $F_M(\cdot)$ by $F_M^{-1}(\cdot)$, and note that $F_M^{-1}(z)$ tends to $F^{-1}(z)$ for all z as $M \to \infty$.

Provided that Assumptions 1 and 2 hold, Theorem 2 establishes a.s. convergence of the quantile trackers. We now write $F_{M,k}^{-1}(\cdot)$ for inverse approximation of F_M in (3) built on the quantiles $q_k^{\alpha_m}(Y_{1:k-1})$, $0 \le m \le M$, learned from k iterations of the quantile tracking in (2). It is worth noting that the algorithm (2) does not guarantee that the values of the quantile trackers $\{q_k^{\alpha_m}\}_{m=0}^M$ remain ordered in accordance with their respective indices during training. However, their limiting values $\{q^{\alpha_m}\}_{m=0}^M$ are in the order of the quantile indices. Therefore, although we do not explicitly order the actual quantile tracker, we can utilize the sorted values of the current quantile trackers when evaluating $F_{M,k}^{-1}(\cdot)$. An important observation is that a single observation of a Y_k allows for *simultaneously updating all M quantile trackers*. Applying (2) simultaneously to learn the $M-1$ undetermined quantiles, it follows from Theorem 2 that

$$\lim_{k\to\infty} \int_0^1 F_{M,k}^{-1}(z)dz = \int_0^1 F_M^{-1}(z)dz.$$

Hence, we further have

$$\lim_{M\to\infty}\lim_{k\to\infty} \int_0^1 F_{M,k}^{-1}(z)dz = \int_0^1 F^{-1}(z)dz = \mathbb{E}[Y].$$

The above suggests increasing M while running the quantile tracking algorithm for learning the steady-state performance characteristic $\mathbb{E}[Y]$.

When M changes over the course of the iterations, we denote the corresponding α-values with an iteration-dependent index. Specifically, we write $\alpha_0(M(k)) < \alpha_1(M(k)) < \ldots < \alpha_{M(k)}(M(k)) = 1$. At certain iterations k, the number of quantile trackers $M(k)$ may increase to $M(k+1)$, which requires the addition of a new evaluation point in $[0,1]$ and a corresponding quantile value. To accomplish this, we identify the largest interval in the current set of α-values and insert a new quantile tracker via linear interpolation. For notational convenience, we define the *quantile-tracking refinement operation* as:

$$M', m', \alpha', q' = \text{Refine}(M, \alpha, q), \tag{4}$$

where $M' = M+1$, and m is the index of the largest interval in the partition of $[0,1]$ induced by the current α-values, i.e., $m' = \min\{\arg\max_{1\le m\le M}(\alpha_m - \alpha_{m-1})\}$; we insert the midpoint of the m-th interval as a new quantile point:

$$\alpha' = \big(\alpha_{0:m'-1},\ \frac{1}{2}(\alpha_{m'-1} + \alpha_{m'}),\ \alpha_{m':M}\big),$$

and define the updated quantile values via linear interpolation:

$$q' = \big(q^{\alpha_{0:m'-1}},\ \frac{1}{2}(q^{\alpha_{m'-1}} + q^{\alpha_{m'}}),\ q^{\alpha_{m':M}}\big).$$

We define a set of refinement times $\mathscr{K} := \{k_n\}_{n\ge 1}$ with $k_n \in \mathbb{N}$ and $k_n \le k_{n+1}$, such that a refinement step is performed after $k_{n+1} - k_n$ SA steps, simultaneously with the SA update. Under this scheme, the number of quantile trackers at iteration k satisfies

$$M(k) = M(k-1) + 1\{k \in \mathscr{K}\},$$

for $M(0) > 0$ given. This setting also includes the case when no refinement is conducted by letting $k_1 = \infty$. We summarize the behavior of this refinement procedure in the following theorem.

Theorem 3 Under the conditions put forward in Theorem 2, it holds for any refinement sequence that

$$\mathbb{E}[Y] = \lim_{k\to\infty} \int_0^1 F^{-1}_{M(k),k}(z)dz.$$

Proof. We give here a sketch of the proof. When a new quantile tracker is introduced, the interpolated quantile value is biased with respect to the true value based on the observation history. For $k \in \mathscr{K}$, suppose that a new quantile tracker is inserted via linear interpolation at the m'-th interval of $[0,1]$. The error introduced by this refinement step has two components. The first is the deterministic interpolation error, arising from the fact that $q^{\frac{\alpha_{m'-1}+\alpha_{m'}}{2}} \neq \frac{1}{2}(q^{\alpha_{m'-1}} + q^{\alpha_{m'}})$. This error shrinks as $M(k)$ increases, since more quantile points are placed in $[a,b]$ and the largest interval is refined at each iteration. The second component is the estimation error in the quantile values used for interpolation, given by $\left|q^{\alpha_{m'-1}} - q_k^{\alpha_{m'-1}}(Y_{1:k-1})\right| + \left|q^{\alpha_{m'}} - q_k^{\alpha_{m'}}(Y_{1:k-1})\right|$, which decreases after the corresponding quantile levels are included in the tracker set. Suppose that, during the update process, all quantiles corresponding to the current α-levels have been initialized for tracking at iteration k. Then, the initial estimation error diminishes as k increases, for both existing and newly added quantile points by interpolation. As the algorithm proceeds, the quantile tracking procedure learn each q^α from the incoming data $Y_{k:k+\Delta k}$ at the same rate, as $\Delta k \to \infty$. This shows that (up to the changing labeling of the quantile trackers), as k tends to infinity, q^α are correctly approximated in the limit. By Theorem 2, the quantiles are correctly tracked for the current α values. As we increase the number of α-evaluation points as k increases and as these new quantiles are initialized close to the unbiased true value, the claimed result follows. □

It should be noted that $\int_0^1 F^{-1}_{M(k),k}(z)dz$ is a numerical integral approximation of $\mathbb{E}[Y]$ that is easily computed for a given vector of α values and the corresponding quantile values.

3 OPTIMIZATION ALGORITHMS

In this section, we develop a framework for recursive optimization based on distribution approximations constructed via quantile tracking. We first present a theoretical algorithm that serves as the basis for convergence analysis, and then introduce two practical algorithms in the subsequent subsections.

3.1 Black-Box Gradient Algorithm

For ease of reference, we introduce the following assumption:

Assumption 3 $\mathbb{E}_\theta[Y]$ is twice continuously differentiable for $\theta \in \Theta$.

We consider the following optimization problem: $\min_{\theta\in\Theta} \mathbb{E}_\theta[Y]$. To solve this problem iteratively, we define the gradient approximation at iteration k as

$$g(\theta,\Delta,k) := \frac{1}{\Delta}\left(\int_0^1 F^{-1}_{M(k),k,\theta}(z)dz - \int_0^1 F^{-1}_{M(k),k,\theta-\Delta}(z)dz\right), \tag{5}$$

for $\Delta \neq 0$. In words, the function $g(\theta,\Delta,k)$ approximates the gradient of the objective via a finite difference scheme, where each term is evaluated using numerical integration of the quantile function corresponding to the approximate distribution $F_{M(k),k,\theta}(\cdot)$. We introduce our general SA algorithm for finding a local minimum of $\mathbb{E}_\theta[Y]$. This algorithm combines the quantile-tracking update from (2) with the gradient approximation in (5). The overall algorithm is given by two recursions running at the same time-scale:

$$q^{\alpha_m}_{k+1} = q^{\alpha_m}_k + \varepsilon_k(\alpha_m - 1\{Y(\theta_k) \leq q^{\alpha_m}_k\}), \tag{6}$$

$$\theta_{k+1} = \theta_k - \varepsilon_k g(\theta_k,\Delta_k,k), \tag{7}$$

for $k \geq 0$. We have the following theorem for capturing the behavior of this algorithm.

Theorem 4 Let Assumptions 1, 2 and 3 hold, and let $\varepsilon_k = 1/(k+1)$, $\Delta_k = k^{-p}$, and $M(k) = \lceil k^q \rceil$, for $p,q > 0$ and all k. If the second-order derivative of $\mathbb{E}_\theta[Y]$ with respect to θ is bounded on Θ, and $q > p > 0$, then the SA algorithm (7) converges a.s. to a stationary point of $\mathbb{E}_\theta[Y]$.

Proof. For the proof, we first show that

$$\left|\int_0^1 F^{-1}_{M(k),k,\theta}(z)dz - \int_0^1 F^{-1}_{\theta}(z)dz\right| = \mathscr{O}\left(\frac{1}{M(k)}\right). \tag{8}$$

To see this note that

$$\left|\int_{\alpha^m}^{\alpha^{m+1}} F^{-1}_{M(k),k,\theta}(z)dz - \int_{\alpha^m}^{\alpha^{m+1}} F^{-1}_{\theta}(z)dz\right| \leq (\alpha_{m+1}-\alpha_m)\left(\max(q^{\alpha_{m+1}}_k, q^{\alpha_{m+1}}) - \min(q^{\alpha_m}_k, q^{\alpha_m})\right).$$

As we add new α values by splitting the largest interval in the present α vector, for $M(k)$ large, we have $(\alpha_{m+1}-\alpha_m)$ to be of order $1/M(k)$. By assumption, the $M(k)$ quantiles are in $[a,b]$. Since the CDF is assumed to be continuous, the quantiles q^{α_m} can have no accumulation point in $[a,b]$. This shows that $(q^{\alpha_{m+1}} - q^{\alpha_m})$ is of order $1/M(k)$. Since $q^{\alpha_m}_k(Y_{1:k-1})$ approximates q^{α_m} as $k \to \infty$, it follows that $\max(q^{\alpha_{m+1}}_k, q^{\alpha_{m+1}}) - \min(q^{\alpha_m}_k, q^{\alpha_m})$ is of order $1/M(k)$, which shows (8). By further computation, we have

$$\begin{aligned}
&\left|\frac{1}{\Delta_k}\left(\int_0^1 F^{-1}_{M(k),k,\theta}(z)dz - \int_0^1 F^{-1}_{M(k),k,\theta-\Delta_k}(z)dz\right) - \frac{d}{d\theta}\int_0^1 F^{-1}_{\theta}(z)dz\right| \\
\leq\ &\left|\frac{1}{\Delta_k}\left(\int_0^1 F^{-1}_{M(k),k,\theta}(z)dz - \int_0^1 F^{-1}_{\theta}(z)dz + \int_0^1 F^{-1}_{\theta-\Delta_k}(z)dz - \int_0^1 F^{-1}_{M(k),k,\theta-\Delta_k}(z)dz)\right)\right| \\
&+ \left|\frac{1}{\Delta_k}\left(\int_0^1 F^{-1}_{\theta}(z)dz - \int_0^1 F^{-1}_{\theta-\Delta_k}(z)dz\right) - \frac{d}{d\theta}\int_0^1 F^{-1}_{\theta}(z)dz\right| \ \leq\ \mathscr{O}\left(\frac{1}{|\Delta_k|}\right)\mathscr{O}\left(\frac{1}{M(k)}\right) + \mathscr{O}(\Delta_k),
\end{aligned}$$

where the second part follows from our assumption that the second order derivative of $\int_0^1 F_\theta^{-1}(z)dz$ is bounded. To satisfy the bias condition in stochastic approximation, we let $q > p > 0$ and have $|d\mathbb{E}_\theta[Y]/d\theta - g(\theta_k,\Delta_k,k)| \leq \beta_k$, where we conclude from the above arguments that β_k is of order $k^{(p-q)}$. Hence, $\varepsilon_k\beta_k = \mathcal{O}(k^{p-q-1})$, which shows that $\sum_k \varepsilon_k\beta_k < \infty$ for $q > p$. This together with the fact that due the bounded support assumption the second moment of $g(\theta_k,\Delta_k,k)$ is bounded, establishes convergence of θ_k towards a stationary point of $\mathbb{E}_\theta[Y]$, see Theorem 5.1 in Vázquez-Abad and Heidergott (2025) or the Markov state-dependent noise model in Chapter 6 of Kushner and Yin (2003). □

Algorithm (7) allows to link the perturbation size Δ_k to the iteration counter, by letting $\Delta_k = \min(a(k), \theta_k - \theta_{k-1})$, for some sequence $a(k) = k^{-p}$ slowly tending to zero (i.e., for p small). As the gradient updates are scaled by ε_k, the minimum is quickly attained by $\theta_k - \theta_{k-1}$ and we may use

$$\theta_{k+1} = \theta_k - \varepsilon_k g(\theta_k, \theta_k - \theta_{k-1}, k), \tag{9}$$

for $k \geq 0$. Algorithm (9) utilize the past evaluation of $\int_0^1 F^{-1}_{M(k-1),k-1;\theta_{k-1}}(z)dz$ as proxy for $\int_0^1 F^{-1}_{M(k),k;\theta_{k-\Delta_k}}(z)dz$, so that an update of (9) is obtained from a single observation of Y_k at θ_k as follows:

$$\theta_{k+1} = \theta_k - \frac{\varepsilon_k}{\theta_k - \theta_{k-1}} \left(\sum_{m=1}^{M} q_k^{\alpha_m}(Y_{1:k-1})(\alpha_m - \alpha_{m-1}) - \sum_{m=1}^{M} q_{k-1}^{\alpha_m}(Y_{1:k-2})(\alpha_m - \alpha_{m-1}) \right), \tag{10}$$

for $k \geq 2$, where the numerical value of the second sum is known from the previous iteration at $k-1$. Inserting the recursive relation for the quantile tracker in (6) into (10) yields the simplified recursion:

$$\theta_{k+1} = \theta_k - \frac{\varepsilon_k}{\theta_k - \theta_{k-1}} \left(\sum_{m=1}^{M} \varepsilon_k \big(\alpha_m - 1\{Y(\theta_k) \leq q_k^{\alpha_m}\}\big)(\alpha_m - \alpha_{m-1}) \right).$$

3.2 Gradient Approximation via Batching

In this section, we discuss a practical adaptation of the general quantile-boosted algorithm that incorporates batching for estimating $dF_\theta/d\theta$ via an FD approach. For ease of presentation, we focus on the case where Y_k is an i.i.d. sequence for fixed θ. The extension of the algorithm to more general settings is deferred to the full-length version of this paper. By computation, we can obtain

$$\begin{aligned} G^{\mathrm{FD}}(\theta) &= \int_0^1 \frac{dF_\theta^{-1}(z)}{d\theta}dz = -\int_0^1 \frac{dF_\theta(q)}{d\theta}\Big|_{q=F_\theta^{-1}(z)} dF_\theta^{-1}(z) = \lim_{\Delta\to 0} -\int_0^1 \frac{F_{\theta+\Delta}(q) - F_\theta(q)}{\Delta}\Big|_{q=F_\theta^{-1}(z)} dF_\theta^{-1}(z) \\ &= \lim_{\Delta\to 0} -\int_0^1 \frac{1}{\Delta}(F_{\theta+\Delta}(q) - z)\Big|_{q=F_\theta^{-1}(z)} dF_\theta^{-1}(z), \end{aligned} \tag{11}$$

where the interchange of integration and differentiation is justified if $dF_\theta^{-1}/d\theta$ is continuous on $[a,b]\times\Theta$. The resulting algorithm consists of two main components: the first component tracks the quantiles via the standard quantile tracker (2), and the second component carries out the learning of θ making use of the discretized version of (11), where $Y(l,\theta)$ for $1 \leq l \leq L$ are i.i.d. samples of Y under θ. The algorithm now updates the quantile trackers according to (6) and the parameter as follows:

$$\theta_{k+1} = \theta_k - \varepsilon_k \sum_{m=1}^{M} \left(\frac{1}{L}\sum_{l=1}^{L} \frac{1}{\Delta_k}\big(1\{Y(l,\theta_k+\Delta_k) \leq q_k^{\alpha_m}\} - \alpha_m\big) \right) \big(q_k^{\alpha_m} - q_k^{\alpha_{m-1}}\big), \tag{12}$$

for $k \geq 0$. To ensure the stability of the FD approximation of the CDF derivative via indicator functions, the batch size L should be chosen sufficiently large. The numerical examples in Section 4 will show the performance of this algorithm.

3.3 Partially Available Models

We now consider the case where $Y(\theta) = h(V, X(\theta))$, with $X(\theta)$ being a random variable with a known distribution and density $\varphi_\theta(\cdot)$. This setting allows us to apply the SF method to estimate $dF_\theta/d\theta$, which, under suitable smoothness conditions, leads to the following expression:

$$\begin{aligned} G^{\mathrm{SF}}(\theta) &= \int_0^1 \frac{d}{d\theta} F_\theta^{-1}(z)dz = -\int_0^1 \frac{d}{d\theta} F_\theta(q)\big|_{q=F_\theta^{-1}(z)} dF_\theta^{-1}(z) \\ &= -\int_0^1 \mathbb{E}\left[1\{h(V,X(\theta)) \le F_\theta^{-1}(z)\}\frac{d}{d\theta}\ln\varphi_{\theta_k}(X(\theta))\right] dF_\theta^{-1}(z). \end{aligned}$$

For clarity of exposition, we again focus on the case where $\{Y_k\}$ is an i.i.d. sequence for fixed θ. We write $Y_k(X_k(\theta_k))$ to emphasize the fact that the θ-dependence of the k-th sample of Y is given through the k-th sample of X at θ_k. For the SF-based algorithm, the quantile trackers are updated as in (6), and the parameter update reads

$$\theta_{k+1} = \theta_k - \varepsilon_k \sum_{m=1}^{M} 1\{Y_k(X_k(\theta_k)) \le q_k^{\alpha_m}\}\left(\frac{d}{d\theta}\ln\varphi_{\theta_k}(X_k(\theta_k))\right)\left(q_k^{\alpha_m} - q_k^{\alpha_{m-1}}\right), \tag{13}$$

for $k \ge 0$. The performance of this algorithm is shown in Section 4. The pseudo code of our algorithms proposed in this section can be summarized as Algorithm 1.

Algorithm 1 Quantile-Boosted Stochastic Approximation

1: **Input:** Initial parameter θ_0; parameter space Θ; quantile levels $\alpha = \{\alpha_m\}_m$; initial quantile trackers $q_0 = \{q_0^{\alpha_m}\}_m$; initial number of trackers $M(0)$; and refinement schedule $\mathscr{K} = \{k_n\}_n$.
2: **for** $k = 0$ to $K-1$ **do**
3: Generate a sample $Y(\theta_k)$ under parameter θ_k (and additional $L-1$ samples if using recursion (12));
4: Update quantile trackers $q_k \to q_{k+1}$ using recursion (6) with $Y(\theta_k)$;
5: Update parameter $\theta_k \to \theta_{k+1}$ using either recursion (12) or (13) with projection onto Θ;
6: **if** $k \in \mathscr{K}$ **then** Refine quantile levels and trackers using operation (4) **end if**.
7: **end for**
8: **Output:** Final parameter θ_K.

4 NUMERICAL EXPERIMENTS

We present two numerical examples. The first example stems from portfolio optimization, and we illustrate our algorithm for various settings discussed above. The portfolio example is based on i.i.d. observations, and, in a second example, we illustrate our algorithm for optimizing steady-state characteristics with a simple queueing example. We evaluate four algorithms on both scenarios: FD and SF represent baseline methods that estimate gradients directly from observations, using vanilla finite differences and the score function method, respectively. FDQT and SFQT leverage the same gradient estimation technique, but are built upon our Quantile Tracking (QT) framework. The additional computational cost from quantile tracking in our method is negligible, as it only requires a linear update of a tracker vector and is independent of the problem complexity. In all comparative experiments, we employ an identical step-size schedule for updating parameters in both our proposed algorithms and their counterparts, i.e., $\varepsilon_k = aK_0^\alpha/(k+K_0)^\alpha$, where K_0 is 10% of the maximum number of iterations allowed, as suggested by Spall (2005). In particular, our proposed SA algorithms utilize a shared step-size decay factor for both parameter update and quantile tracking, differing only in initial values, which is consistent with single-timescale SA requirements. The quantile tracking step size is set to $\varepsilon_k' = bK_0^\alpha/(k+K_0)^\alpha = (b/a)\varepsilon_k$. The perturbation magnitude for the black-box algorithms follows $\Delta_k = cK_0^\beta/(k+K_0)^\beta$. The choices of hyperparameters (a,b,c,α,β) are

specified in the respective experiments. Our standard algorithms consider scalar parameters, while the extension of our black-box approach to vector-valued parameters can be obtained by SPSA.

4.1 Financial Example

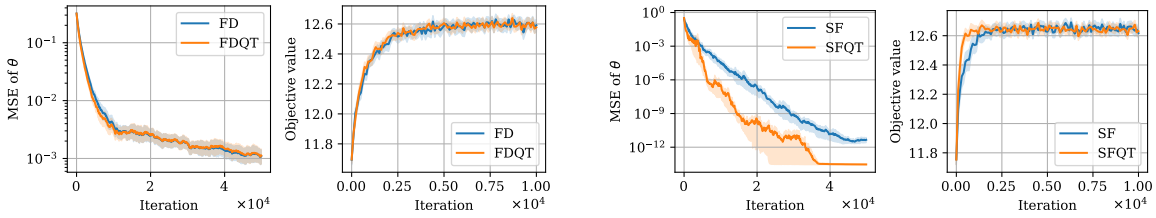
Consider an investor who allocates an initial capital, normalized to 1, across n companies. Each company's payoff depends on a predefined threshold. Let X_i denote the market value of company i at the end of the investment period. If X_i exceeds a predefined threshold x_i, the investor receives a return $Y_i \sim U[0, X_i]$, meaning Y_i is uniformly distributed between zero and the company's final market value. Conversely, if $X_i < x_i$, no return is realized. Market values are modeled to exhibit extremal dependence among companies, following the framework suggested by Bassamboo et al. (2008). Specifically, the market value X_i of company i is given by

$$X_i = \frac{\rho V + \sqrt{1-\rho^2}\eta_i}{\max(W,1)}, \quad i = 1, \ldots, n,$$

where η_i represents firm-specific noise, V is a common standard normally distributed market factor, and W is a non-negative random variable capturing systemic market shocks. The parameter $\rho \in [0,1]$ controls the strength of systemic dependence. In our example, W is modeled as an exponential distribution with a rate parameter λ that depends on a tunable parameter θ_1, reflecting the investor's belief about overall market risk: $\lambda(\theta_1) = \frac{1}{0.3}(2(1-\theta_1)\theta_1 + 0.5)$. Each η_i is drawn from a normal distribution with mean zero and variance i, except for η_1, whose mean is set to $\mu_1 = 2(1-\theta_2)\theta_2$. Each sample of Z represents the return over a single trading day. The investor is interested in the cumulative performance over T consecutive days and aims to determine the optimal investment strategy by maximizing the risk-adjusted return (i.e., the Sharpe ratio). This leads to the following optimization problem:

$$\max_{(\theta_1,\theta_2)\in[0,1]^2} \mathbb{E}\left[\frac{1}{\sigma}\sum_{t=1}^{T} Z_t\right],$$

where σ denotes the sample standard deviation of the i.i.d. returns $\{Z_t\}_{t=1}^{T}$.



(a) Black-box setting.

(b) Partially available setting.

Figure 2: Mean square error (MSE) of θ and the objective value during optimization. The curves represent the mean over 50 independent runs, and the shaded regions indicate the 95% confidence intervals.

We provide numerical results for the black-box SA with batching and the partial-model SA algorithm with SF. These methods correspond respectively to recursions (6) and (12), and recursions (6) and (13). The optimal solution, known from numerical simulations, is located at $(\theta_1^*, \theta_2^*) = (0.5, 0.5)$, providing a benchmark for evaluating algorithmic performance. The hyperparameters of FD methods are $(a, b, c, \alpha, \beta) = (10^{-4}, 10^{-2}, 10^{-1}, 0.99, 0.25)$ with batch size 20, while those of SF methods are $(a, b, \alpha) = (10^{-4}, 2 \times 10^{-1}, 0.51)$ without batching. The number of quantile trackers is fixed to 64 and 8, respectively. As shown in Figure 2, the numerical results demonstrate that quantile tracking effectively captures distributional information and achieves optimization efficiency comparable to direct gradient-based methods across

different algorithmic frameworks. In Figure 2(b), our algorithm converges even faster than only applying the SF method with observed samples.

4.2 Queueing Example

Consider a first-come, first-served M/M/1 queue. Specifically, the arrival times $\{A_n : n \in \mathbb{N}\}$ and the service times $\{S_n(\theta) : n \in \mathbb{N}\}$ are independent and identically distributed exponential random variables with rates λ and $1/\theta$, respectively. The sequence of sojourn times $\{X_n\}$ in this queue forms a Markov chain, governed by Lindley's recursion:

$$X_{n+1}(\theta) = \max(0, X_n(\theta) - A_{n+1}) + S_{n+1}(\theta), \quad n \geq 0,$$

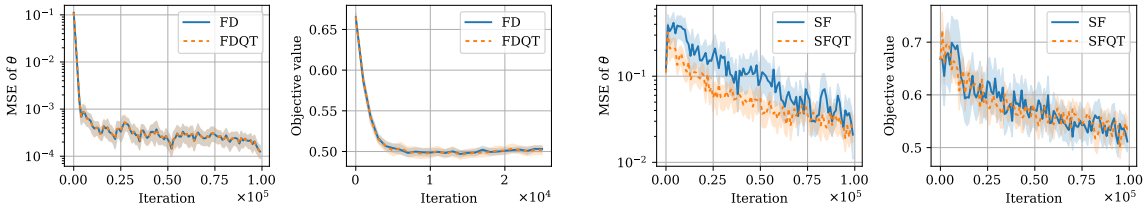
with the initial condition $X_0 = 0$, representing an initially empty system. The system is stable provided that $\theta \in \Theta = (0, 1/\lambda)$. Under this stability condition, the sojourn times converge in distribution, and the stationary sojourn time, denoted by $X_\infty(\theta)$, is a well-defined random variable. Let $\hat{F}_\theta(\cdot)$ denote the CDF of $X_\infty(\theta)$. By the ergodic theorem, for any continuous and bounded function $h(\cdot)$, it holds almost surely that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} h(X_k(\theta)) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[h(X_k(\theta))] = \mathbb{E}[h(X_\infty(\theta))] = \int_{a_h}^{b_h} h(x)\, \hat{F}_\theta(dx),$$

for $\theta \in \Theta$, where $a_h \leq h(x) \leq b_h$ for all $x \geq 0$. For our experiments, let $Y_k = h(X_k(\theta))$, where $h(x) = 1 - e^{-x}$ for $x \geq 0$, representing the scaled sojourn times with $a_h = 0$ and $b_h = 1$. This transformation emphasizes smaller sojourn times, assigning greater utility to reducing short delays compared to equivalent reductions in longer delays. Assume that the server's energy consumption is given by $1/(4\theta)$, and consider the following optimization problem:

$$\min_{\theta\in\Theta} \mathbb{E}\left[1 - e^{-X_\infty(\theta)}\right] + \frac{1}{4\theta},$$

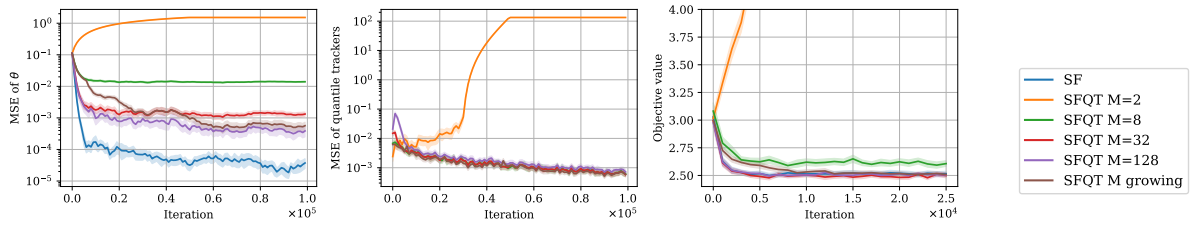
where the optimum is $\theta^* = 2/3$ for $\lambda = 1/2$. It is straightforward to verify that Assumptions 1-3 are satisfied for this model. When simulating the queue, the steady state sojourn time is approximated by the sojourn time of the 100th customer after each iteration. In this example, the system is not reset between iterations during the execution of the algorithm, and the optimization is performed in an online manner, where decisions are updated continuously based on real-time feedback from the evolving system state. The FD methods adopt hyperparameters $(a, b, c, \alpha, \beta) = (10^{-3}, 10^{-2}, 10^{-1}, 0.99, 0.25)$ with batch size 10, while those of SF methods are $(a, b, \alpha) = (10^{-2}, 10^{-2}, 0.99)$ without batching. The number of quantile trackers is fixed to 128 for both paradigms. As illustrated in Figure 3, gradient computation using a distribution function reconstructed via the quantile trackers introduces negligible impact on the optimization performance. Our method even demonstrates slightly better efficiency under the partially available setting.



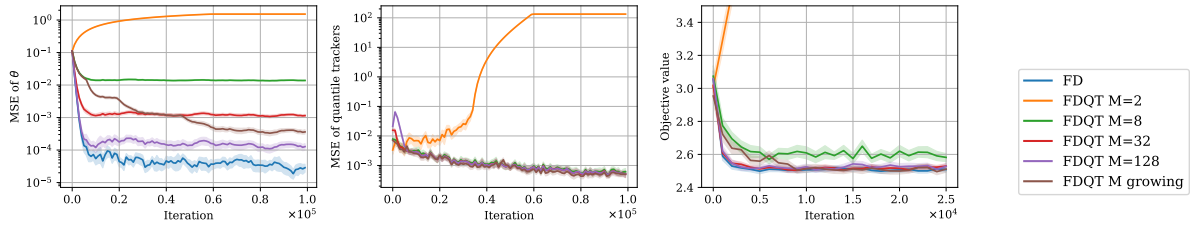
(a) Black-box setting.

(b) Partially available setting.

Figure 3: MSE of θ and the objective value in the bounded example. The curves represent the mean over 50 independent runs, and the shaded regions indicate the 95% confidence intervals.



(a) Black-box setting.



(b) Partially available setting.

Figure 4: MSE of θ and quantile trackers as well as the objective value in the unbounded example. The solid lines represent the mean over 50 independent runs, and the shaded regions indicate the 95% confidence intervals.

Although our theoretical development requires $h(\cdot)$ to be bounded, our approaches also work well for the unbounded case, and we show the output of the SA for $h(x) = x$, i.e., we consider $\min_{\theta\in\Theta} \mathbb{E}[X_\infty(\theta)] + 1/\theta$, where the solution is $\theta^* = 2/3$ for $\lambda = 1/2$. The hyperparameters of FD methods are set to $(a,b,c,\alpha,\beta) = (10^{-4}, 2\times10^{-1}, 10^{-1}, 0.99, 0.25)$ with batch size 10, and those of SF methods run with $(a,b,\alpha) = (10^{-4}, 2\times10^{-1}, 0.9)$ with batch size 10. We evaluate our algorithms with a fixed number of quantile trackers M, ranging from 2 to 128, and further investigate the case where M increases linearly from 8 to 128 during training. Since the steady-state sojourn time in an M/M/1 queue explicitly follows an exponential distribution with rate $1/\theta - \lambda$, this provides us with an explicit reference for evaluating the accuracy of the quantile trackers. As shown in Figure 4, the algorithms based on quantile trackers all function properly for all values of M except when $M = 2$, where the poor approximation of the distribution leads to a failure of the algorithm. The MSE curves of θ indicate that as M increases, the quantile tracker provides a more accurate reconstruction of the distribution, resulting in more precise learning of the objective function. Consequently, the performance of the quantile tracker-based algorithm increasingly approaches that of the baseline, in line with the earlier theoretical discussion that a larger M yields a better approximation of the true solution. The MSE curves of the quantile trackers show that the errors across all tracked quantiles are independent and converge at a similar rate. Furthermore, in the experiment with gradually increasing M, we observe that the MSE curve of θ sequentially intersects the curves corresponding to fixed M, eventually converging toward the curve associated with the final value of M. This demonstrates that the algorithm effectively realizes the idea of gradually increasing M during the optimization process, without the need to restart the experiment each time M is adjusted. The quantile MSE curves also indicate that the error introduced by adding new quantile trackers is well controlled. As a result, it significantly reduces the computational overhead in the early and middle stages.

5 CONCLUSIONS

We introduce an SA algorithm based on adaptive quantile tracking of the underlying distribution functions. The method is grounded in a convergence-guaranteed theoretical framework and compatible with both FD and SF methods. Numerical examples show that learning the entire distribution alongside optimization leads

to an efficient algorithm applicable to data streaming applications in non white-box models. Future research includes extensions to the vector-valued case and to performance characteristics that cannot straightforwardly be expressed as expected value, such as distortion risk measures.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 72325007, 72250065, and 72022001, Xiangjiang Laboratory Key Project under Grant 23XJ02004, and the Science and Technology Innovation Program of Hunan Province under Grant 2024RC7003.

REFERENCES

Bassamboo, A., S. Juneja, and A. Zeevi. 2008. "Portfolio Credit Risk with Extremal Dependence: Asymptotic Analysis and Efficient Simulation". *Operations Research* 56(3):593–606.

Bhatnagar, S., H. Prasad, and L. Prashanth. 2013. *Stochastic Recursive Algorithms for Optimization: Simultaneous Perturbation Methods*. New-York: Springer.

Borkar, S. 2008. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press.

Cassandras, C., and L. Lafortune. 2008. *Introduction to Discrete Event Systems (2nd Ed.)*. New York: Springer.

Feng, B. M., and E. Song. 2024. "Efficient Nested Simulation Experiment Design via the Likelihood Ratio Method". *INFORMS Journal on Computing*.

Glasserman, P. 1991. *Gradient Estimation via Perturbation Analysis*. Boston: Kluwer Academic Publishers.

Gordy, M. B., and S. Juneja. 2010. "Nested Simulation in Portfolio Risk Measurement". *Management Science* 56(10):1833–1848.

Ho, Y., and X.-R. Cao. 1991. *Perturbation Analysis of Discrete Event Systems*. Kluwer Academic: Boston.

Jiang, J., B. Heidergott, J. Hu, and Y. Peng. 2024. "Distortion Risk Measure-Based Deep Reinforcement Learning". In *2024 Winter Simulation Conference (WSC)*, 2595–2606. IEEE.

Kroese, D., T. Taimre, and Z. Botev. 2013. *Handbook of Monte Carlo Methods*, Volume 706. John Wiley & Sons.

Kushner, H., and G. Yin. 2003. *Stochastic Approximation and Recursive Algorithms*. Springer, New York.

Li, Z., and Y. Peng. 2024. "Eliminating Ratio Bias for Gradient-based Simulated Parameter Estimation". *arXiv preprint arXiv:2411.12995*.

Rubinstein, R., and D. Kroese. 2016. *Simulation and the Monte Carlo method*, Volume 10. John Wiley & Sons.

Rubinstein, R., and B. Melamed. 1998. *Moderns Simulation and Modelling*. New York: Wiley.

Rubinstein, R., and A. Shapiro. 1993. *Discrete Event Systems: Sensitivity Analysis and Optimization by the Score Function Method*. Chichester: Wiley.

Spall, J. C. 2005. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley & Sons.

Vázquez-Abad, F., and B. Heidergott. 2025. *Optimization and Learning via Stochastic Gradient Search*. Princeton University Press, to appear.

AUTHOR BIOGRAPHIES

JINYANG JIANG is a PhD candidate in the Department of Management Science and Information Systems in Guanghua School of Management at Peking University, Beijing, China. He received the BS degree in information and computational sciences from School of Mathematics and Statistics, Wuhan University, and the double BS degree in computer science and technology from School of Computer Science, Wuhan University, in 2021. His research interests include machine learning and simulation optimization. His email address is jinyang.jiang@stu.pku.edu.cn.

BERND HEIDERGOTT is the professor of Stochastic Optimization at the Department of Operations Analytics at the Vrije Universiteit Amsterdam, the Netherlands. He received his PhD degree from the University of Hamburg, Germany, in 1996, and held postdoc positions at various universities before joining the Vrije Universiteit. Bernd is research fellow of the Tinbergen Institute and board member of the Amsterdam Business Research Institute. His research interests are optimization and control of discrete event systems, perturbation analysis, Markov chains, max-plus algebra, and social networks. His email address is b.f.heidergott@vu.nl.

YIJIE PENG is an associate professor in Guanghua School of Management at Peking University. His research interests include stochastic modeling and analysis, simulation optimization, machine learning, data analytics, and healthcare. He is a member of INFORMS and IEEE, and serves as an Associate Editor of the Asia-Pacific Journal of Operational Research and the Conference Editorial Board of the IEEE Control Systems Society. His email address is pengyijie@pku.edu.cn.