Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

SIMULATION OPTIMIZATION BY REUSING PAST REPLICATIONS: DON'T BE AFRAID OF DEPENDENCE

Tianyi Liu
Enlu Zhou

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30318, USA

2020 Winter Simulation Conference (WSC) | 978-1-7281-9499-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/WSC48552.2020.9384032

ABSTRACT

The main challenge of simulation optimization is the limited simulation budget because of the high computational cost of simulation experiments. One approach to overcome this challenge is to reuse simulation outputs from previous iterations in the current iteration of the optimization procedure. However, due to the dependence among iterations, simulation replications from different iterations are not independent, which leads to the lack of theoretical justification for the good empirical performance. In this paper, we fill this gap by theoretically studying the stochastic gradient descent method with reusing past simulation replications. We show that reusing past replications does not change the convergence of the algorithm, which implies the bias of the gradient estimator is asymptotically negligible. Moreover, we show that reusing past replications reduces the variance of gradient estimators conditioned on the history, which implies that the algorithm can use larger step size sequences to achieve faster convergence.

1 INTRODUCTION

As complex real-world systems are often modeled by stochastic simulation, simulation optimization is used to find the best design of such models (see reviews papers such as Fu 2002; Fu et al. 2005; Amaran et al. 2016; Fu and Henderson 2017). Specifically, we consider the following simulation optimization problem:

$$\min_{\theta \in \Theta} H(\theta) \tag{1}$$

where θ is the decision variable, Θ is the solution space, and H is a performance measure of interest. In particular, given an θ the value of H can only be estimated from simulation experiments. Then a typical simulation optimization method repeats two steps to find the best solution to (1): 1) running simulation replication(s) at the current candidate solution to get an estimate of the objective-relevant value, such as the objective value H or the gradient ∇H; 2) search Θ for the next candidate solution following some optimization strategy.

One major challenge in simulation optimization is that the simulation alone is computationally expensive in many practical applications. To capture necessary features of a real-world system, the simulation model usually has a huge size and complex logic, and thus often takes a long time to run even one replication. Therefore, the question of how to make full use of these valuable simulation replications is of great interest to the simulation optimization community.

To address this question, one simple yet efficient method is to reuse the simulation outputs from previous candidate solutions by the importance sampling technique. This idea can be traced back to Rubinstein and Shapiro (1990), Rubinstein and Shapiro (1993), and is recently studied by Feng and Staum (2015) and Feng and Staum (2017) in repeated experiment design under the name of green simulation. Eckman and Henderson (2018) studied this reusing technique in ranking and selection, and later Eckman

and Feng (2018) extended this method to simulation optimization. To better describe this method, we consider the performance measure to be an expectation, i.e., $H(\theta) = \mathbb{E}_{\xi \sim f(\cdot;\theta)}[h(\xi)]$, where h is a sample performance function,$f(\cdot;\theta)$ is the probability density function (pdf) of a family of parametric distributions defined on Ξ.. We solve the optimization problem by stochastic gradient descent (SGD) with the gradient $\nabla H(\theta) = \mathbb{E}_{\xi \sim f(\cdot;\theta)}[h(\xi)\nabla_\theta \ln f(\xi;\theta)]$ (see detailed calculation in Section 2). When the sample size B is small, the naive gradient estimator $\frac{1}{B}\sum_{i=1}^{B} h(\xi^i)\ln f(\xi^i;\theta)$, where $\xi^1,\ldots,\xi^B$ are independent and identically distributed (i.i.d.) samples drawn from $f(\xi;\theta)$, can have a large variance. Instead, we can reuse the simulation outputs from previous iterations by the principle of importance sampling to obtain a new weighed set of simulation outputs under the current solution. This leads to a new gradient estimator that uses more simulation replications than the naive gradient estimator. We will refer to this new estimator as the gradient estimator with past replications.

The potential issue of this method is the bias caused by the dependence between iterations. The importance sampling estimator is unbiased when each solution involved is independent. However, this is violated in an iterative method where the current solution is decided by the past candidate solutions. Therefore, the gradient estimator with past replications is biased, which makes the convergence analysis difficult and the algorithm behavior not well understood. However, we empirically observe that SGD with reusing past replications still converges and performs even better than the vanilla SGD. Eckman and Henderson (2018), Eckman and Feng (2018) pointed out this issue and suggested to be careful when applying the reusing technique, but there still lacks a theoretical justification of the good empirical performance.

In this paper, we theoretically study a mini-batch stochastic gradient descent with reusing past replications (RSGD for short). We apply the fixed state chain method proposed in Kushner and Yin (2003) to prove that the continuous-time interpolation of the solution trajectory of RSGD satisfies the same limit ordinary differential equation (ODE) as the vanilla stochastic gradient descent (VSGD). This directly implies that the bias of the gradient estimator with past replications in RSGD is asymptotically negligible and RSGD converges asymptotically. We further show that compared to the naive gradient estimator, the gradient estimator with past replications has a greatly-reduced variance conditioned on the past trajectory. These results are observed in our numerical experiments, which suggest under the same simulation budget and with the same step size sequence, RSGD achieves a smaller mean and standard deviation of the solution error. Moreover, we empirically find that RSGD can adopt a larger step size sequence than VSGD, and thus RSGD can converge faster than VSGD.

The idea of reusing simulation replications has also been used in other problems. For example, Dong et al. (2018) applied this idea to discover new metamodels that are unbiased and have low variance. Liu and Zhou (2019) developed an efficient online quantification method of input model uncertainty by reusing the simulation outputs from previous time stages. This idea has also found application in Reinforcement learning (e.g., Fernández and Veloso 2006) and logged bandit (e.g., Swaminathan and Joachims 2015).

2 PROBLEM AND ALGORITHM

Let $\Theta \subseteq \mathbb{R}^{d_\theta}$ denote the solution space. Given a solution $\theta \in \Theta$ we consider the system performance $H(\theta)$ defined as the following expectation: $H(\theta) = \mathbb{E}_\theta[h(\xi)]$, where $\xi \in \Xi \subseteq \mathbb{R}^{d_\xi}$ is a random vector that satisfies $\xi \sim f(\cdot;\theta)$, and $f(\cdot;\theta)$ is the probability density function (pdf) of a family of parametric distributions defined on Ξ. The function h is not available directly, but there is a simulation oracle that returns the value $h(\xi)$ given a $\xi \in \Xi$. To find the best solution, we need solve the following optimization problem:

$$\min_{\theta\in\Theta} H(\theta) = \mathbb{E}_\theta[h(\xi)] = \int_{x\in\Xi} h(x)f(x;\theta)dx,$$

where the expectation is with respect to $\xi \sim f(\cdot;\theta)$. Note that we consider the case where θ only appears in the density function instead of a more general case where $H(\theta) = \mathbb{E}_{\xi\sim f(\cdot;\theta)}[h(\xi,\theta)]$, as Fu (2015) pointed out that the general case can often be reformulated such that the sample performance function h does not

explicitly depend on θ. This idea first showed up as the push-out method in Rubinstein (1992).We will leave the analysis on the general case to our future research.

Stochastic gradient descent (SGD), which can be tracked back to Robbins and Monro (1951), has been widely used to solve the simulation optimization problem above. SGD takes the following iteration:

$$\theta_{n+1} = \text{Proj}_\Theta\left(\theta_n - \alpha_n \widetilde{\nabla H}(\theta_n)\right),$$

where $\alpha_n > 0$ is the step size, $\widetilde{\nabla H}(\theta_n)$ is a gradient estimator, and $\text{Proj}_\Theta(\theta)$ is a projection operator that projects the iterate of θ to Θ. The gradient estimator $\widetilde{\nabla H}(\theta_n)$ can be derived as follows. Under some regularity conditions, the gradient of H can be rewritten as

$$\begin{aligned}\nabla H(\theta) = \nabla_\theta \int_{x\in\Xi} h(x) f(x;\theta) dx &= \int_{x\in\Xi} h(x) \nabla_\theta f(x;\theta) dx = \int_{x\in\Xi} h(x) \frac{\nabla_\theta f(x;\theta)}{f(x;\theta)} f(x;\theta) dx \\ &= \int_{x\in\Xi} h(x) \nabla_\theta \ln f(x;\theta) f(x;\theta) dx = \mathbb{E}_\theta[h(\xi)\nabla_\theta \ln f(\xi;\theta)]. \end{aligned} \tag{2}$$

Hence, a mini-batch vanilla gradient estimator is

$$\widetilde{\nabla H}(\theta_n) = \frac{1}{B}\sum_{i=1}^{B} G(\xi_n^i, \theta_n) = \frac{1}{B}\sum_{i=1}^{B} h(\xi_n^i)\nabla_\theta \ln f(\xi_n^i;\theta_n),$$

where $B \geq 1$ is the batch size, $\{\xi_n^1, \ldots, \xi_n^B\} \overset{\text{i.i.d}}{\sim} f(\cdot;\theta_n)$, and $G(\xi,\theta) = h(\xi)\nabla_\theta \ln f(\xi;\theta)$. One can easily verify $\frac{1}{B}\sum_{i=1}^{B} G(\xi_n^i,\theta_n)$ is an unbiased estimator of the gradient. However, $\widetilde{\nabla H}(\theta_n)$ may have a large variance when B is small, which can happen when the simulation replication is expensive and the total budget is limited. In the following we refer to SGD with the vanilla gradient estimator as Vanilla SGD (VSGD).

To overcome the issue of limited simulation budget, we reuse the past simulation outputs based on the principle of importance sampling, i.e., changing the underlying probability measures under previous iterations to the one under the current iteration. Specifically, for two different $\theta_1, \theta_2 \in \Theta$, $\nabla H(\theta_1)$ can be rewritten as follows:

$$\begin{aligned}\nabla_\theta \int_{x\in\Xi} h(x) f(x;\theta_1) dx &= \int_{x\in\Xi} h(x)\nabla_\theta \ln f(x;\theta_1) f(x;\theta_1) dx \\ &= \mathbb{E}_{\theta_2}\left[h(\xi)\nabla_\theta \ln f(\xi;\theta_1)\frac{f(\xi;\theta_1)}{f(\xi;\theta_2)}\right] = \mathbb{E}_{\theta_2}\left[\omega(\xi,\theta_1|\theta_2) G(\xi,\theta_1)\right], \end{aligned} \tag{3}$$

where $\omega(\xi,\theta_1|\theta_2) = f(\xi;\theta_1)/f(\xi;\theta_2)$ is the likelihood ratio. Following (3), we can estimate the gradient at θ_n using not only the new simulation outputs obtained at the solution θ_n, but also simulation outputs from the past $K-1$ iterations, where $K \geq 1$ is a parameter to choose, i.e.,

$$\widehat{\nabla H}(\theta_n) = \frac{1}{KB}\sum_{m=n-K+1}^{n}\sum_{i=1}^{B} \omega(\xi_m^i, \theta_n|\theta_m) G(\xi_m^i,\theta_n), \tag{4}$$

where $\{\xi_m^i, i = 1,\ldots,B\} \overset{\text{i.i.d}}{\sim} f(\cdot;\theta_m)$ for $m = n-K+1, ..., n$. We refer to (4) as the gradient estimator with past replications at θ. Then the updating of SGD with reusing past replications (RSGD) is:

$$\theta_{n+1} = \text{Proj}_\Theta\left(\theta_n - \alpha_n \frac{1}{KB}\sum_{m=n-K+1}^{n}\sum_{i=1}^{B} \omega(\xi_m^i,\theta_n|\theta_m) G(\xi_m^i,\theta_n)\right). \tag{5}$$

The RSGD algorithm is described in details below.

- At $n = 0$, we choose an initial solution $\theta_0 \in \Theta$. Draw B i.i.d. samples $\{\xi_0^i\} \overset{\text{i.i.d}}{\sim} f(\cdot;\theta_0)$, and record their pdf values $\{f(\xi_0^i;\theta_0)\}$ in the memory. Then run simulation and obtain simulation outputs $\{h(\xi_0^i)\}$.
- At iteration $n+1$, carry out the following steps.
  1. Optimization: Update θ_{n+1} according (5).
  2. Simulation: Draw B i.i.d. samples $\{\xi_{n+1}^i\} \overset{\text{i.i.d}}{\sim} f(\cdot;\theta_{n+1})$, and record their pdf values $\{f(\xi_{n+1}^i;\theta_{n+1})\}$. Run simulation and obtain the simulation output $\{h(\xi_{n+1}^i)\}$.
  3. Updating the memory: Discard $\xi_{n-K+1}^i, f(\xi_{n-K+1}^i,\theta_{n-K+1})$, and save $\xi_n^i, f(\xi_n^i,\theta_n)$, $i = 1,...,B$.
- Output θ_n when some stopping criteria is satisfied.

Ideally, since we reuse past simulation outputs, RSGD should perform better than VSGD under limited simulation budget (i.e., small B). However, reusing past replications is not a free lunch. The dependence between iterations makes the gradient estimator with past replications $\widehat{\nabla H}(\theta_n)$ biased. Specifically, $\xi_{n-1}^i|\theta_{n-1}$ and $\xi_{n-1}^i|(\theta_{n-1},\theta_n)$ are not identically distributed since θ_n can provide additional information. That is, $\xi_{n-1}^i|\theta_{n-1}^i$ and θ_n are not independent. As a direct result, we have the following inequality.

$$\mathbb{E}\left[G(\xi_m^i,\theta_n)\frac{f(\xi_m^i;\theta_n)}{f(\xi_m^i;\theta_m)}\right] = \mathbb{E}\left[\mathbb{E}\left[G(\xi_m^i,\theta_n)\frac{f(\xi_m^i;\theta_n)}{f(\xi_m^i;\theta_m)}\middle|\theta_n\right]\right] \neq \mathbb{E}_{\theta_n}\left[h(\xi)\right].$$

This bias issue has been pointed out in Eckman and Henderson (2018), Eckman and Feng (2018). Despite of the bias, we empirically observe that reusing past replications works pretty well under many different circumstances. Thus, we conjecture that the bias in the gradient estimator with past replications is asymptotically negligible. In the next section, we theoretically prove this conjecture by establishing the convergence result of RSGD.

3 CONVERGENCE ANALYSIS

In this section, we theoretically study the convergence behavior of RSGD. We first characterize the limit ODE of the solution trajectory of RSGD. Our main result shows that RSGD and VSGD share the same limit ODE, and the bias introduced by the dependence between iterations become asymptotically negligible. Thus, RSGD has the same asymptotic convergence as VSGD. We further show that RSGD reduces variance of each iterate compared to VSGD conditioned on the history, which implies that RSGD has more stable convergence behavior and can take larger step sizes.

3.1 Asymptotic Convergence by the ODE Method

We study the algorithmic behavior of RSGD by the ODE method. The main idea is that SGD can be viewed as a noisy discretization of an ODE, and under certain conditions, asymptotically the noise effects in SGD average out so that the SGD iterates converge to the solution trajectory of an ODE. By comparing such ODEs of RSGD and VSGD, we can reveal the convergence of both methods and thus the influence of bias caused by reusing past replications. Before proceeding to our analysis, we first make some assumptions.

Assumption 1.

- The step size sequence $\{\alpha_n\}_n$ satisfies $\sum \alpha_n = \infty$, $\sum \alpha_n^2 < \infty$.
- Compactness and Convexity: Ξ and Θ are nonempty compact sets in $\mathbb{R}^{d_\xi}$ and $\mathbb{R}^{d_\theta}$, respctively. Moreover, Θ is convex.
- Continuous System Performance: $h(x)$ is a continuous function and thus, $|h(\xi)|$ is bounded by some constant $C > 0$.

- Uniform Lipschitz Continuity: for any $\xi \in \Xi$, $f(\xi,\cdot)$ is continuously differentiable and there exists a constant $L > 0$, such that

$$\|\nabla_\theta f(\xi,\theta_1) - \nabla_\theta f(\xi,\theta_2)\|\| \le L\|\theta_1 - \theta_2\|, \forall \xi \in \Xi, \theta_1, \theta_2 \in \Theta.$$

- Uniform Lower Bound: there exists a constant $\varepsilon > 0$, such that $f(\xi,\theta) > \varepsilon,\ \forall \xi \in \Xi, \theta \in \Theta$.

We remark that Assumption 1 can be easily verified in practice. The assumption on the step size sequence is standard for the convergence of SGD type algorithms. Under the compactness assumption, the last three assumptions are satisfied by any continuous function h and density function $f(\xi;\theta)$ that is jointly continuous in ξ and θ and bounded away from zero, such as truncated exponential family of distributions. The convexity of the parameter space guarantees the uniqueness of the projection in the SGD iteration. Moreover, Assumption 1 guarantees the interchange of integration and differentiation in the derivation of gradient estimators (2) and (3). We then rewrite the update (5) with an explicit projection term z_n as follows

$$\theta_{n+1} = \text{Proj}_\Theta\left(\theta_n - \alpha_n \widehat{\nabla H}(\theta_n)\right) = \theta_n - \alpha_n \widehat{\nabla H}(\theta_n) - \alpha_n z_n. \tag{6}$$

The ODE method considers the asymptotic behavior of the continuous-time interpolation of the solution trajectory $\{\theta_n\}$. To facilitate the analysis, we explicitly define the interpolated time scale and processes. A natural time scale is defined in terms of the step size sequence. Define $t_0 = 0$ and $t_n = \sum_{i=0}^{n-1} \alpha_i$. For $t \ge 0$, let $m(t)$ denote the unique value of n such that $t_n \le t < t_{n+1}$. For $t < 0$, set $m(t) = 0$. We then define the continuous-time interpolation $\theta^0(\cdot)$ on $(-\infty,\infty)$ by $\theta^0(t) = \theta_0$ for $t \le 0$ and for $t > 0$, $\theta^0(t) = \theta_n$, for $t_n \le t < t_{n+1}$. For notational simplicity, we further define the shift process $\theta^n(\cdot)$ by shifting the process $\theta^0(\cdot)$ by t_n, i.e.,

$$\theta^n(t) = \theta^0(t_n + t), \quad \text{for } t \in (-\infty,\infty).$$

Moreover, we define the interpolation of the projection term z_n by $Z^0(t) = 0$ for $t \le 0$ and $Z^0(t) = \sum_{i=0}^{m(t)-1} \alpha_i z_i$, for $t > 0$.

As we have mentioned before, due to the lack of independence between iterations, the gradient estimators at different iterations are correlated. However, we can dig out some important Markov properties here to help us show that this bias is asymptotically negligible and thus RSGD converges asymptotically. We have the next theorem to characterize the limiting behavior of the solution trajectory.

Theorem 1 Under Assumption 1, for a fixed $K > 0$, the solution trajectory of RSGD, $\{\theta_n\}$, converges almost surely to some limit set of the following ODE in Θ.

$$\dot{\theta}(t) = \nabla_\theta H(\theta(t)) + Z(\theta(t)), \tag{7}$$

where Z is the minimum force needed to keep the solution $\theta(t)$ in Θ.

Proof. We first rewrite the update (6) as follows

$$\theta_{n+1} = \theta_n - \{\alpha_n \nabla H(\theta_n) + \alpha_n \delta M_n + \alpha_n \gamma_n + \alpha_n z_n\},$$

where δM_n is the noise caused by the simulation estimation error at the current iterate, and γ_n is the bias introduced by reusing past replications. Following the main idea of the ODE method, we will show that the continuous time interpolations of both noise δM_n and bias γ_n do not change asymptotically (see Lemma 2, Corollary 2), and thus the limit ODE is determined by the gradient ∇H and the projection.

For any $s > 0$, denote $\boldsymbol{\xi_s} = (\xi_s^1, ..., \xi_s^B)$ and $\boldsymbol{f_s} = \left(f(\xi_s^1,\theta), ..., f(\xi_s^B,\theta)\right)$. Specifically, let $\{\mathscr{F}_n\}_n$ be a nondecreasing σ-algebras such that $\{\theta_m, \boldsymbol{d_m}, m \le n\}$ is measurable with respect to $\mathscr{F}_n$, where $\boldsymbol{d_m}$ is a random vector that represents the 'effective memory' to compute the gradient, i.e., $\boldsymbol{d_m} = (\boldsymbol{\xi}_{m-K+1}, \boldsymbol{f}_{m-K+1}, , ..., \boldsymbol{\xi}_{m-1}, \boldsymbol{f}_{m-1})$. We have the following two observations regarding $\mathscr{F}_n$ and $\boldsymbol{d_n}$.

- Since $\mathscr{F}_n$ provides all the information required to achieve θ_n, we have $\theta_n \in \mathscr{F}_n$. Moreover, conditioned on $\mathscr{F}_n$, the expectation of the gradient estimator with past replications takes the following form.

$$\begin{aligned}\mathbb{E}[\widehat{\nabla H}(\theta_n)\big|\mathscr{F}_n] &= \mathbb{E}\left[\frac{1}{KB}\sum_{m=n-K+1}^{n}\sum_{i=1}^{B}\omega(\xi_m^i,\theta_n|\theta_m)G(\xi_m^i,\theta_n)\Big|\mathscr{F}_n\right]\\ &= \frac{1}{KB}\sum_{m=n-K+1}^{n-1}\sum_{i=1}^{B}\omega(\xi_m^i,\theta_n|\theta_m)G(\xi_m^i,\theta_n)+\frac{1}{K}\nabla H(\theta_n).\end{aligned}$$

  We further denote the noise $\delta M_n = \widehat{\nabla H}(\theta_n) - \mathbb{E}[\widehat{\nabla H}(\theta_n)\big|\mathscr{F}_n] = \frac{1}{KB}\sum_{i=1}^{B}\left(G(\xi_n^i,\theta_n)-\nabla H(\theta_n)\right)$.
- Let $\boldsymbol{P}(\boldsymbol{d_n},\cdot\,|\theta_n)$ be the transition probability from $\boldsymbol{d_n}$ given the current iterate θ_n. Then $\boldsymbol{d_n}$ satisfies the following Markov property,

$$\mathbb{P}(\boldsymbol{d}_{n+1}\in\cdot|\boldsymbol{d_m},\theta_m,m\le n) = \boldsymbol{P}(\boldsymbol{d_n},\cdot\,|\theta_n), \tag{8}$$

In fact, one can observe this Markov property if we explicitly compare $\boldsymbol{d}_{n+1}$ and $\boldsymbol{d}_n$ as follows:

$$\begin{aligned}\boldsymbol{d}_n &= \left(\boldsymbol{\xi}_{n-K+1},\boldsymbol{f}_{n-K+1},\boldsymbol{\xi}_{n-K+1},\boldsymbol{f}_{n-K+1},...,\boldsymbol{\xi}_{n-1},\boldsymbol{f}_{n-1}\right),\\ \boldsymbol{d}_{n+1} &= \left(\boldsymbol{\xi}_{n-k+2},\boldsymbol{f}_{n-k+2},...,\boldsymbol{\xi}_{n-1},\boldsymbol{f}_{n-1},\ \boldsymbol{\xi}_n,\ \boldsymbol{f}_n\right).\end{aligned}$$

Clearly, given $\boldsymbol{d}_n$, the component of $\boldsymbol{d}_{n+1}$ that remains unknown are $\boldsymbol{\xi_n}$ and $\boldsymbol{f}_n$, which are random variables only depend on θ_n. This justifies (8).

Given the Markov property above, we can apply the fixed-state chain method, Theorem 6.6.1 in Kushner and Yin (2003), to show the convergence of RSGD. Specifically, for a fixed state θ, the transition probability $\boldsymbol{P}(\cdot,\cdot|\theta)$ defines a Markov chain denoted as $\{\boldsymbol{d}_n(\theta)\}_{n\ge0}$. We expect that the probability law of this chain for a given θ is close to the probability law of the true $\{\boldsymbol{d}_n\}_{n\ge0}$ if θ_n varies slowly around θ. Hence, the limit ODE obtained in terms of this fixed-state chain approximates that of the original iterates. This method has also been used by Liu et al. (2018) to prove the convergence of momentum stochastic gradient descent algorithm.

For notational simplicity, we denote

$$\widehat{\nabla H}(\theta,\boldsymbol{d}_m) = \frac{1}{BK}\sum_{j=m-K+1}^{m}\sum_{i=1}^{B}\frac{f(\xi_j^i,\theta)}{f(\xi_j^i,\theta_j)}G(\xi_j^i,\theta),$$

where $\boldsymbol{d}_m = \left(\boldsymbol{\xi}_{m-K+1},\boldsymbol{f}_{m-K+1},\boldsymbol{\xi}_{m-K+1},\boldsymbol{f}_{m-K+1},...,\boldsymbol{\xi}_{m-1},\boldsymbol{f}_{m-1}\right)$, and $\{\xi_m^i, i=1,..,.B\}\overset{\text{i.i.d}}{\sim} f(\cdot;\theta_m)$. Thus, $\widehat{\nabla H}(\theta_n,\boldsymbol{d}_n) = \widehat{\nabla H}(\theta_n)$. Define a function $v_n(\theta,\boldsymbol{d_n})$ as follows:

$$v_n(\theta,\boldsymbol{d_n}) = \sum_{i=0}^{\infty}\alpha_{n+i}\mathbb{E}\left[\widehat{\nabla H}(\theta,\boldsymbol{d}_{n+i}(\theta))-\nabla H(\theta)\big|\boldsymbol{d_n}(\theta)=\boldsymbol{d_n},\theta\right].$$

where $\boldsymbol{d}_{n+i}(\theta)\big|\boldsymbol{d_n}(\theta)=\boldsymbol{d_n}$ is the i-th state of the Markov chain defined by the transition probability $\boldsymbol{P}(\cdot,\cdot|\theta)$ and initialized at $\boldsymbol{d_n}$. The term $v_n(\theta,\boldsymbol{d_n})$ actually represents the accumulated bias in the fixed-state chain with fixed state θ. The next lemma shows that the bias in the fixed-state chain with fixed state θ_n vanishes.

Lemma 1. $\lim_{n\to\infty} v_n(\theta_n,\boldsymbol{d_n}) = 0, a.s.$.

Please refer to Appendix A.1 for the detailed proof of Lemma 1. We then consider a perturbed iteration $\widetilde{\theta}_n = \theta_n + v_n(\theta_n,\boldsymbol{d_n})$. By Lemma 1, we know $\widetilde{\theta}_n$ asymptotically equals to θ_n. Thus, instead of the original iterations, we study the perturbed iterations. By simple manipulation, we obtain the update of the perturbed iterations as follows.

$$\widetilde{\theta}_{n+1} = \widetilde{\theta}_n - \alpha_n\left[\nabla H(\theta_n)+\delta M_n+z_n\right]-\beta_n-\delta N_n, \tag{9}$$

where α_n and the martingale difference δN_n are given by

$$\beta_n = v_{n+1}(\theta_{n+1}, \boldsymbol{d}_{n+1}) - v_{n+1}(\theta_n, \boldsymbol{d}_{n+1}), \quad \delta N_n = v_{n+1}(\theta_n, \boldsymbol{d}_{n+1}) - \mathbb{E}[v_{n+1}(\theta_n, \boldsymbol{d}_{n+1}) | \boldsymbol{d_n}(\theta) = \boldsymbol{d_n}, \theta_n].$$

We next show that the three terms β_n, δN_n, and δM_n vanish as $n \to \infty$ because their continuous-time interpolations almost do not change as $t \to \infty$, which is formally defined as follows.

Definition 1 (Zero Asymptotic Rate of Change) We call a stochastic process $X(t)$ has zero asymptotic rate of change almost surely if for any $T > 0$, $\mathbb{P}\left(\lim_n \sup_{j \geq n} \max_{0 \leq t \leq T} |X(jT+t) - X(jT)| = 0\right) = 1$.

We then have the following lemma.

Lemma 2. Define the continuous-time interpolations of β_n, δN_n and δM_n as follows.

$$B(t) = \sum_{i=0}^{m(t)-1} \beta_i, \quad N(t) = \sum_{i=0}^{m(t)-1} \delta N_i, \quad M(t) = \sum_{i=0}^{m(t)-1} \alpha_i \delta M_i.$$

Then $B(t)$, $N(t)$ and $M(t)$ have zero asymptotic rate of change almost surely.

Please refer to Appendix A.2 for detailed proof of Lemma 2. The results on $B(t)$ and $N(t)$ imply zero asymptotic rate of change of the continuous time interpolation of γ_t (see Corollary 2). Thus, when n is large, $\theta_{n+1} \approx \theta_n + \alpha_n [\nabla H(\theta_n) + z_n]$, which is exactly the discretization of ODE (7). In fact, Lemmas 1 and 2 are Conditions A 6.6.1 and A 6.6.2 for Theorem 6.6.1 in Kushner and Yin (2003). One can verify other regularity conditions in Theorem 6.6.1 hold. We then apply Theorem 6.6.1 and prove the result. □

Note that ODE (7) does not depend on the number of reused iterations K. Therefore, VSGD satisfies the same limit ODE. Moreover, the following corollary shows that the bias in the gradient estimator with past replications has zero asymptotic rate of change. Thus, it is asymptotically negligible and does not affect the asymptotic convergence of RSGD.

Corollary 2 Recall that the bias $\gamma_n = \widehat{\nabla H}(\theta_n) - \nabla H(\theta_n) - \delta M_n$. Then the continuous-time interpolation $\Gamma(t) = \sum_{i=0}^{m(t)-1} \alpha_i \gamma_i$ has zero asymptotic rate of change almost surely.

Proof. Note that according to (9), we have

$$\theta_{n+1} = \theta_n - \{\alpha_n [\nabla H(\theta_n) + \delta M_n + z_n] + \beta_n + \delta N_n + v_n(\theta_n, \boldsymbol{d}_n) - v_{n+1}(\theta_{n+1}, \boldsymbol{d}_{n+1})\}.$$

Thus, $\alpha_n \gamma_n = \beta_n + \delta N_n + v_n(\theta_n, \boldsymbol{d}_n) - v_{n+1}(\theta_{n+1}, \boldsymbol{d}_{n+1})$. Then $\Gamma(t)$ can be rewritten as follows.

$$\Gamma(t) = \sum_{i=0}^{m(t)-1} \beta_i + \delta N_i + v_i(\theta_i, \boldsymbol{d}_i) - v_{i+1}(\theta_{i+1}, \boldsymbol{d}_{i+1}) = B(t) + N(t) + v_0(\theta_0, \boldsymbol{d}_0) - v_{m(t)}(\theta_{m(t)}, \boldsymbol{d}_{m(t)}).$$

Thus,

$$\begin{aligned} |\Gamma(jT+t) - \Gamma(jT)| \leq & |B(jT+t) - B(jT)| + |M(jT+t) - M(jT)| \\ & + |v_{m(jT)}(\theta_{m(jT)}, \boldsymbol{d}_{m(jT)})| + |v_{m(jT+t)}(\theta_{m(jT+t)}, \boldsymbol{d}_{m(jT+t)})|. \end{aligned}$$

Then the result follows directly from Lemma 1 and 2. □

Remark 1. Theorem 1 shows that RSGD converges to the limit set of the ODE (7) in Θ. We need further conditions to determine the stability of limit points. However, this depends on the problem to be solved but not the algorithm. Thus, we do not include the analysis on the limit set of ODEs in this paper. Moreover, the ODE method only studies the mean behavior of RSGD and does not justify the advantage of RSGD due to the variance reduction effect of reusing past replications. We will show this effect in the next section.

3.2 Reduction of Conditional Variance

The main idea behind RSGD is to reuse more simulation outputs obtained in the previous iterations to get a better gradient estimator with smaller variance. We show that reusing past replications reduces the variance of each iterate conditioned on the history. For notational simplicity, we denote the update of VSGD and RSGD as θ_n^{VSGD} and θ_n^{RSGD}, respectively. Correspondingly, the filtrations are denoted as

$$\mathscr{F}_n^{\text{VSGD}} = \sigma\{\theta_m^{\text{VSGD}}, m \le n\}, \quad \mathscr{F}_n^{\text{RSGD}} = \sigma\{\theta_m^{\text{RSGD}}, \boldsymbol{d_m}, m \le n\}.$$

Then we have the following theorem to characterize the conditional variance reduction:

Theorem 3 For any vector $v \in \mathbb{R}^{d_\theta}$, let $v^{(i)}$ denote the i-th dimension of v for $i \le d_\theta$. When K, the number of reused iterations, satisfies the following condition

$$K \ge \max_{i \le d_\theta} \sqrt{\max_{\theta\in\Theta} \mathbb{E}_\theta \left[\left(G(\xi,\theta)^{(i)} - \nabla H(\theta)^{(i)}\right)^2\right] \Big/ \min_{\theta\in\Theta} \mathbb{E}_\theta \left[\left(G(\xi,\theta)^{(i)} - \nabla H(\theta)^{(i)}\right)^2\right]}, \tag{10}$$

we have for any $n > 0$, $i \le d_\theta$, $\text{Var}[\theta_{n+1}^{(i),\text{RSGD}} \big| \mathscr{F}_n^{\text{RSGD}}] \le \text{Var}[\theta_{n+1}^{(i),\text{VSGD}} \big| \mathscr{F}_n^{\text{VSGD}}]$, almost surely.

Proof. To prove this theorem, we first provide an upper bound for $\text{Var}[\theta_{n+1}^{(i),\text{RSGD}} \big| \mathscr{F}_n^{\text{RSGD}}]$ and a lower bound for $\text{Var}[\theta_{n+1}^{(i),\text{VSGD}} \big| \mathscr{F}_n^{\text{VSGD}}]$. Comparing these two bounds, we will prove the result.

$$\begin{aligned}
\text{Var}[\theta_{n+1}^{(i),\text{RSGD}} \big| \mathscr{F}_n^{\text{RSGD}}] &= \mathbb{E}\left[(\theta_{n+1}^{(i),\text{RSGD}} - \mathbb{E}[\theta_{n+1}^{(i),\text{RSGD}} \big| \mathscr{F}_n^{\text{RSGD}}])^2 \big| \mathscr{F}_n^{\text{RSGD}}\right] \\
&\le \frac{\alpha_n^2}{K^2 B} \max_{\theta\in\Theta} \mathbb{E}_\theta \left[\left(G(\xi,\theta)^{(i)} - \nabla H(\theta)^{(i)}\right)^2\right].
\end{aligned}$$

Following similar lines, we have the lower bound for $\text{Var}[\theta_{n+1}^{\text{VSGD}} \big| \mathscr{F}_n^{\text{VSGD}}]$.

$$\text{Var}[\theta_{n+1}^{(\text{i}),\text{VSGD}} \big| \mathscr{F}_n^{\text{VSGD}}] \ge \frac{\alpha_n^2}{B} \min_{\theta\in\Theta} \mathbb{E}_\theta \left[\left(G(\xi,\theta)^{(i)} - \nabla H(\theta)^{(i)}\right)^2\right].$$

If $K \ge \sqrt{\frac{\max_{\theta\in\Theta} \mathbb{E}_\theta \left[\left(G(\xi,\theta)^{(i)} - \nabla H(\theta)^{(i)}\right)^2\right]}{\min_{\theta\in\Theta} \mathbb{E}_\theta \left[\left(G(\xi,\theta)^{(i)} - \nabla H(\theta)^{(i)}\right)^2\right]}}$, then $\text{Var}[\theta_{n+1}^{(i),\text{RSGD}} \big| \mathscr{F}_n^{\text{RSGD}}] \le \text{Var}[\theta_{n+1}^{(i),\text{VSGD}} \big| \mathscr{F}_n^{\text{VSGD}}]$. Then take K satisfying (10), we prove the result. □

Theorem 3 suggests that reusing past replications is guaranteed to reduce the variance conditioned on the past history if the number of reused iterations is large enough. Therefore, RSGD has a much smoother trajectory than VSGD, which will be shown in our numerical experiment in Section 4. Moreover, the result also indicates the larger K is, the smaller the conditional variance will be. From this point of view, at iteration n, we should choose K as large as possible, i.e., $K = n$. However, this can possibly lead to a large total variance $\text{Var}[\theta_n^{(i),\text{RSGD}}]$. In fact, by the law of total variance, $\text{Var}[\theta_{n+1}^{(i),\text{RSGD}}] = \mathbb{E}[\text{Var}[\theta_{n+1}^{(i),\text{RSGD}} \big| \mathscr{F}_n^{\text{RSGD}}]] + \text{Var}[\mathbb{E}[\theta_{n+1}^{(i),\text{RSGD}} \big| \mathscr{F}_n^{\text{RSGD}}]]$. The second term is the variance of $\frac{1}{KB}\sum_{m=n-K+1}^{n-1} \sum_{i=1}^{B} \omega(\xi_m^i, \theta_n^{(i),\text{RSGD}} | \theta_m^{\text{RSGD}}) G(\xi_m^i, \theta_n^{(i),\text{RSGD}})$, which increases as $|\theta_n^{(i),\text{RSGD}} - \theta_{n-K+1}^{(i),\text{RSGD}}|$ becomes large. This problem is crucial when we have an unbounded parameter space. In practice, the parameter K should be carefully chosen.

Remark 2. Note that our theoretical result considers the conditional variance instead of the total variance. Since RSGD and VSGD have different past trajectories, the total variance of these two algorithms can behave differently from the conditional variance. However, our numerical experiments find that RSGD can greatly reduce the total variance as well. Please refer to Section 4 for more details.

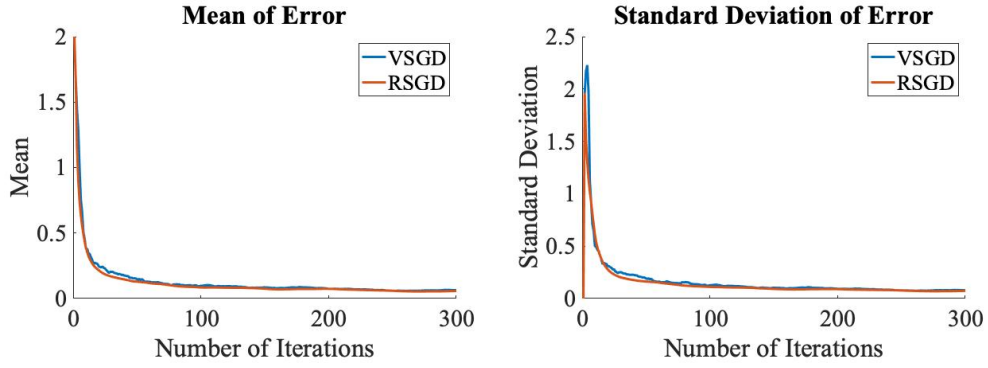

Figure 1: Mean and standard deviation of $|\theta_n - \theta^*|$ of 100 runs of VSGD and RSGD with $\alpha_n = \frac{1}{n}, K = 30$.

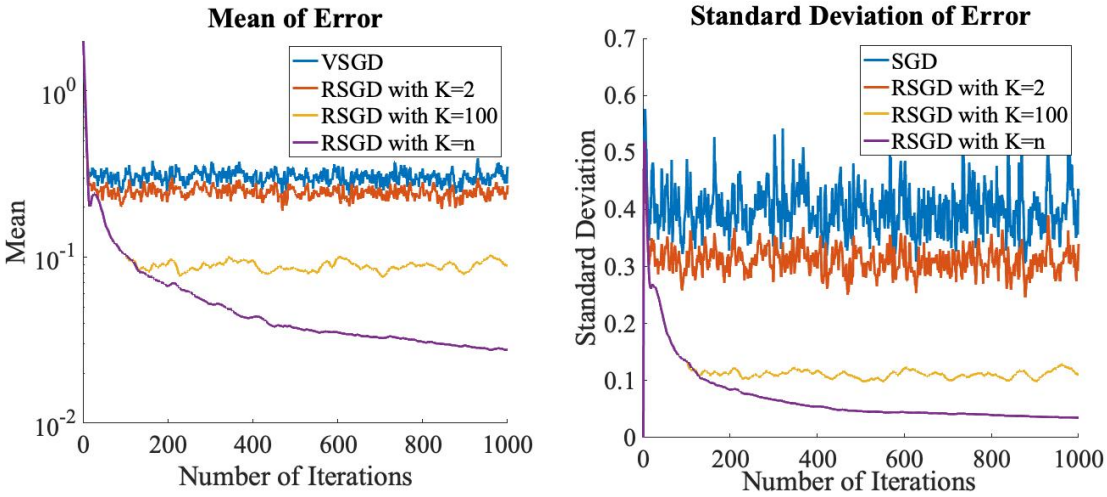

Figure 2: Mean and standard deviation of $|\theta_n - \theta^*|$ of 100 runs of VSGD and RSGD with $\alpha_n = 0.1, K = 2, 100, n$.

Remark 3. Note that when the total number of iterations is small, (10) may not be satisfied. In fact, (10) is a loose bound and only a sufficient condition for variance reduction. In practice, we do not need a very large K to obtain variance reduction. Our numerical experiment in the next section shows that even RSGD with $K = 2$ performs better than VSGD.

4 NUMERICAL EXPERIMENTS

We use a simple quadratic problem to demonstrate the performance improvement of RSGD over VSGD. We take the problem setting from Eckman and Feng (2018). Specifically, we consider the performance measure $H(\theta) = \mathbb{E}_\theta[\xi^2]$, where $\xi \sim f(\cdot;\theta)$ and $f(\cdot;\theta)$ is the pdf of a normal distribution with mean θ and fixed variance $\sigma^2 = 1$, i.e., $f(\xi;\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\xi-\theta)^2}{2}\right)$. Then $H(\theta)$ is the second moment of $\mathcal{N}(\theta, 1)$, and is minimized at $\theta^* = 0$. The function $G(\xi, \theta)$ has the explicit form $G(\xi, \theta) = \xi^2(\xi - \theta)$. We run both VSGD and RSGD to solve this problem. Both algorithms are initialized at $\theta_0 = -2$.

• Experiment 1: In this experiment, we observe that with a decreasing step size sequence, RSGD converges to the global minimizer and has similar performance as VSGD asymptotically. More specifically, we run both algorithms 100 times each, with a step size sequence $\{\alpha_n = \frac{1}{n}\}$, under a limit budget case with $B = 3$. For RSGD, we choose $K = 30$. The mean and standard deviation of the solution error $|\theta_n - \theta^*|$ over $N = 300$ iterations are shown in Figure 1. As can be seen clearly, both algorithms converge, and when

Table 1: Mean and standard deviation of $|\theta_N - \theta^*|$ of 100 runs of VSGD and RSGD with $K = 2, 100, n$.

	VSGD	RSGD ($K = 2$)	RSGD ($K = 100$)	RSGD ($K = n$)
Mean	0.3488	0.2703	0.0890	0.0278
Standard Deviation	0.4251	0.3393	0.1106	0.0351

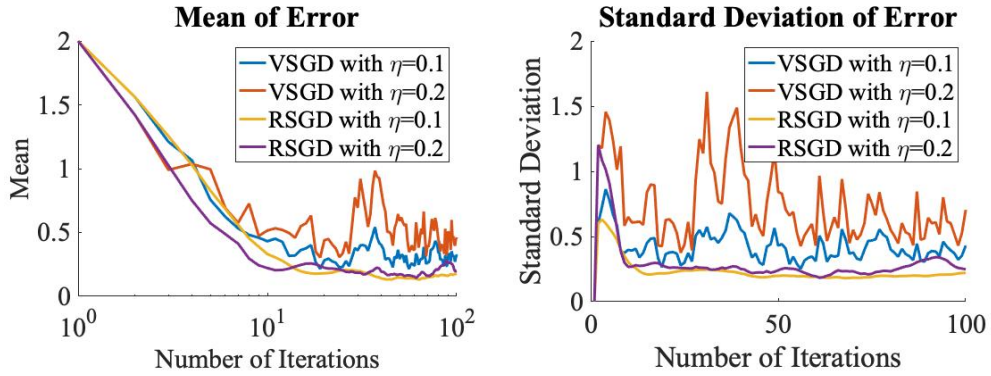

Figure 3: Mean and standard deviation of $|\theta_n - \theta^*|$ of 100 runs of VSGD and RSGD with $\alpha = 0.1, 0.2$.

number of iterations is large, RSGD and VSGD have almost the same performance. This verifies our conclusion in Theorem 1 that RSGD and VSGD have the same limit ODE. When the number of iterations is small, RSGD performs slightly better than SGD. This suggests that RSGD has advantage over VSGD when step size is large, which can be observed in the next experiment.

• Experiment 2: In this experiment, we compare RSGD and VSGD with a fixed step size $\alpha_n = \alpha = 0.1$ to demonstrate the advantage of RSGD. We consider a limit budget case with $B = 3$. For RSGD, we choose $K = 2, 100$ and n (i.e., reuse all the previous simulation outputs). The mean and standard deviation of the solution error over $N = 1000$ iterations are shown in Figure 2 and Table 1. As we can see clearly, even reusing the simulation outputs from only the last iteration can significantly decrease the mean and standard deviation of the solution error. Moreover, the larger K is, the better result we obtain. When reusing all the past simulation outputs, i.e., $K = n$, the solution error of RSGD is less than one tenth of that of VSGD. Another important observation is that RSGD has a much smoother trajectory, which indicates reusing past replications can significantly decrease the variance of iterates and improves the stability of the algorithm.

• Experiment 3: When using VSGD, we usually need to choose a sufficiently small step size to control the variance, which however, may lead to slow convergence. Since RSGD can reduce the variance, we can choose a large step size to achieve fast convergence and good accuracy at the same time. To verify this, we run VSGD and RSGD with various fixed step sizes $\alpha = 0.1, 0.2$, over $N = 100$ iterations. The batch size is chosen as $B = 3$, and the number of reused iterations for RSGD is set as $K = 20$. The mean and standard deviation of the solution error are shown in Figure 3. We observe that RSGD with $\alpha = 0.2$ converges fastest and has smaller variance than VSGD. VSGD with $\alpha = 0.2$, however, has the largest mean and variance. This is partly because n VSGD with $\alpha = 0.2$ even diverges in some runs. This implies that RSGD can take larger step sizes to achieve faster convergence and smaller variance than VSGD.

5 CONCLUSION

In this paper, we study the convergence of stochastic gradient descent with reusing past simulation replications (RSGD). We show that RSGD shares the same limit mean ODE with VSGD, and the bias in the RSGD gradient estimator caused by the dependence between iterations is asymptotically negligible. Moreover, the RSGD has a reduced conditional variance of iterates compared to VSGD and has a much smoother solution trajectory. As a direct result, RSGD can take large step sizes, which lead to faster convergence than VSGD. All the theoretical results are verified in our numerical experiments.

ACKNOWLEDGMENTS

The authors are grateful to the support by the National Science Foundation under Grant CAREER CMMI-1453934, and Air Force Office of Scientific Research under Grant FA9550-19-1-0283.

A APPENDICES

A.1 Proof of Lemma 1

Proof. For notational simplicity, we denote $\mathbb{E}_n[\cdot] = \mathbb{E}[\ \cdot\ |\boldsymbol{d_n}(\theta) = \boldsymbol{d_n}, \theta]$. Note that for $\forall \theta \in \Theta$, $\boldsymbol{d}_i(\theta) = \left(\boldsymbol{\xi}_{i-K+1}, \boldsymbol{f}_{i-K+1}, ..., \boldsymbol{\xi}_{i-1}, \boldsymbol{f}_{i-1}\right)$. Thus, $\boldsymbol{d}_i(\theta)$ is independent of $\boldsymbol{d}_n(\theta)$, when $i \geq n+K-1$. Then we can rewrite the conditional expectation as follows.

$$\mathbb{E}_n\left[\widehat{\nabla H}(\theta, \boldsymbol{d_i}(\theta)) - \nabla H(\theta)\right] = \frac{1}{BK}\sum_{s=i-K+1}^{i-1}\sum_{j=1}^{B}\mathbb{E}\left[h(\xi_s^j)\nabla_\theta \ln f(\xi_s^j;\theta)\frac{f(\xi_s^j;\theta)}{f(\xi_s^j;\theta)} - \nabla H(\theta)\big|\theta\right] = 0.$$

Thus, we have $v_n(\theta_n, \boldsymbol{d_n}) = \sum_{i=n}^{n+K-1}\alpha_i\mathbb{E}_n\left[\widehat{\nabla H}(\theta_n, \boldsymbol{d_i}(\theta_n)) - \nabla H(\theta_n)\right]$. Moreover, under Assumption 1, there exists some constant C_1 such that $\left|\sum_{i=n}^{n+K-1}\alpha_i\mathbb{E}_n\left[\widehat{\nabla H}(\theta, \boldsymbol{d_i}(\theta)) - \nabla H(\theta)\right]\right| \leq C_1\sum_{i=n}^{n+K-1}\alpha_i \to 0$, as t goes to infinity and the step size α_i diminishes. Then we prove the lemma. □

A.2 Proof of Lemma 2

Proof. Recall that for $m \leq n+K-2$, we have the following equation.

$$\mathbb{E}_n\left[\widehat{\nabla H}(\theta, \boldsymbol{d_m}(\theta)) - \nabla H(\theta)\right] = \frac{1}{BK}\sum_{s=m-K+1}^{n-1}\sum_{i=1}^{B}h(\xi_s^i)\nabla_\theta \ln f(\xi_s^i;\theta)\frac{f(\xi_s^i;\theta)}{f(\xi_s^i;\theta_s)} - \nabla H(\theta).$$

This can help us further simplify the formulation of $v_n(\theta, \boldsymbol{d_n})$ as follows.

$$v_n(\theta, \boldsymbol{d_n}) = \sum_{m=n}^{n+K-2}\alpha_m\frac{1}{BK}\sum_{s=m-K+1}^{n-1}\sum_{i=1}^{B}h(\xi_s^i)\nabla_\theta \ln f(\xi_s^i;\theta)\frac{f(\xi_s^i;\theta)}{f(\xi_s^i;\theta_s)} - \nabla H(\theta).$$

We then show that $v_n(\cdot, \boldsymbol{d})$ is Lipschitz continuous uniformly in n and $\boldsymbol{d}$. This can be directly implied by the uniform Lipschitz continuity of function $h(x)\nabla_\theta \ln f(x;\cdot)\frac{f(x;\cdot)}{f(x;\theta')}$ for $\forall x \in \Xi$ and $\theta' \in \Theta$ and the Lipschitz continuity of $\nabla H(\theta)$. For any $\theta_1, \theta_2 \in \Theta$, we have

$$\|h(x)(\nabla_\theta \ln f(x;\theta_1)\frac{f(x;\theta_1)}{f(x;\theta')} - \nabla_\theta \ln f(x;\theta_2)\frac{f(x;\theta_2)}{f(x;\theta')})\| = \left|\frac{h(x)}{f(x;\theta')}\right|\|\nabla_\theta f(x;\theta_1) - \nabla_\theta f(x;\theta_2)\| \leq \frac{C}{\varepsilon}L\|\theta_1 - \theta_2\|.$$

The last equality comes from the boundedness of h, boundedness away from zero of f and the uniform Lipschitz continuity of $\nabla_\theta f$. Thus, function $h(x)\nabla_\theta \ln f(x;\cdot)\frac{f(x;\cdot)}{f(x;\theta')}$ is Lipschitz continuity uniformly in x and θ'. On the other hand,

$$\|\nabla H(\theta_1) - \nabla H(\theta_2)\| = \|\int_{x\in\Xi} h(x)\nabla_\theta f(x;\theta_1) - h(x)\nabla_\theta f(x;\theta_2)dx\| \leq \int_{x\in\Xi}|h(x)|dxL\|\theta_1 - \theta_2\|.$$

Thus, ∇H is Lipschitz. Then we prove that $v_n(\cdot, \boldsymbol{d})$ is Lipschitz continuous uniformly in n and $\boldsymbol{d}$. One can then check that $\beta_t = O(\alpha_t^2)$, which implies the asymptotic rate of change of $B(t)$ is zero.

Moreover, because of the boundedness assumption, we have $\sup_n \mathbb{E}|\delta M_n|^2 < \infty$. For large enough j, we have $\sum_{i=m(j)}^{mj+1)-1}\alpha_i \leq 1$. Then, Burkholder's inequality (Theorem 6.3.10 in Stroock 2000) yields that for any $\mu > 0$, we have $\sum_j \mathbb{P}\left\{\max_{0\leq t\leq T}\left|\sum_{i=m(j)}^{mj+1)-1}\alpha_i\delta M_i\right| \geq \mu\right\} < \infty$. Together with $\sum_t \alpha_t^2 < \infty$, by Theorem 5.3.2 in Kushner and Yin (2003), $M(t)$ has zero asymptotic rate of change. Following similar lines, we can also show $N(t)$ has zero asymptotic rate of change. □

REFERENCES

Amaran, S., N. V. Sahinidis, B. Sharda, and S. J. Bury. 2016. "Simulation Optimization: A Review of Algorithms and Applications". *Annals of Operations Research* 240(1):351–380.

Dong, J., M. B. Feng, and B. L. Nelson. 2018. "Unbiased Metamodeling via Likelihood Ratios". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe et al., 1778–1789. Piscataway, New Jersy: Institute of Electrical and Electronics Engineers, Inc.

Eckman, D. J., and M. Feng. 2018. "Green Simulation Optimization Using Likelihood Ratio Estimators". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe et al., 2049–2060. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Eckman, D. J., and S. G. Henderson. 2018, July. "Reusing Search Data in Ranking and Selection: What Could Possibly Go Wrong?". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 28(3):1–15.

Feng, M., and J. Staum. 2015. "Green Simulation Designs for Repeated Experiments". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz et al., 403–413. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Feng, M., and J. Staum. 2017. "Green Simulation: Reusing the Output of Repeated Experiments". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 27(4):1–28.

Fernández, F., and M. Veloso. 2006. "Probabilistic Policy Reuse in A Reinforcement Learning Agent". In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, 720–727.

Fu, M. C. 2002. "Optimization for Simulation: Theory vs. Practice". *INFORMS Journal on Computing* 14(3):192–215.

Fu, M. C. 2015. "Stochastic Gradient Estimation". In *Handbook of Simulation Optimization*, 105–147. New York: Springer.

Fu, M. C., F. W. Glover, and J. April. 2005. "Simulation Optimization: A Review, New Developments, and Applications". In *Proceedings of the 2005 Winter Simulation Conference*, edited by M. E. Kuhl, et al., 13 pp.–. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Fu, M. C., and S. G. Henderson. 2017. "History of Seeking Better Solutions, AKA Simulation Optimization". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan et al., 131–157. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kushner, H., and G. G. Yin. 2003. *Stochastic Approximation and Recursive Algorithms and Applications*, Volume 35. New York: Springer.

Liu, T., S. Li, J. Shi, E. Zhou, and T. Zhao. 2018. "Towards Understanding Acceleration Tradeoff between Momentum and Asynchrony in Nonconvex Stochastic Optimization". In *Advances in Neural Information Processing Systems*, 3682–3692.

Liu, T., and E. Zhou. 2019. "Online Quantification of Input Model Uncertainty by Two-Layer Importance Sampling". *arXiv preprint arXiv:1912.11172*. accessed 17th August.

Robbins, H., and S. Monro. 1951. "A Stochastic Approximation Method". *The Annals of Mathematical Statistics*:400–407.

Rubinstein, R. Y. 1992. "Sensitivity Analysis of Discrete Event Systems by The "Push Out" Method". *Annals of Operations Research* 39(1):229–250.

Rubinstein, R. Y., and A. Shapiro. 1990. "Optimization of Static Simulation Models by The Score Function Method". *Mathematics and Computers in Simulation* 32:373–392.

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

Stroock, D. W. 2000. *Probability Theory, An Analytic View*. Cambridge: Cambridge University Press.

Swaminathan, A., and T. Joachims. 2015. "Batch Learning from Logged Bandit Feedback through Counterfactual Risk Minimization". *The Journal of Machine Learning Research* 16(1):1731–1755.

AUTHOR BIOGRAPHIES

Tianyi Liu is a Ph.D. student in the School of Industrial and Systems Engineering at Georgia Institute of Technology. His research interests include machine learning, stochastic optimization, and simulation optimization. His e-mail address is tliu341@gatech.edu.

Enlu Zhou is an Associate Professor in the School of Industrial and Systems Engineering at Georgia Institute of Technology. She received the B.S. degree with highest honors in electrical engineering from Zhejiang University, China, in 2004, and received the Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 2009. Her research interests include stochastic control and simulation optimization. Her email address is enlu.zhou@isye.gatech.edu.