Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

BUDGET-CONSTRAINED STOCHASTIC APPROXIMATION

Uday V. Shanbhag

Department of Indust. and Manuf. Engg.
Penn. State University
310 Leonhard Building
University Park, PA 16803, USA

Jose H. Blanchet

Department of IEOR and Department of Statistics
Columbia University
500 W 120th St, 3rd Floor
New York, NY 10027, USA.

Abstract

Traditional stochastic approximation (SA) schemes employ a single gradient or a fixed batch of noisy gradients in computing a new iterate. We consider SA schemes in which N_k samples are utilized at step k and the total simulation budget is M, where $\sum_{k=1}^{K} N_k \leq M$ and K denotes the terminal step. This paper makes the following contributions in the strongly convex regime: (I) We conduct an error analysis for constant batches ($N_k = N$) under constant and diminishing steplengths and prove linear convergence in terms of expected error in solution iterates based on prescribing N_k in terms of simulation and computational budgets; (II) we extend the linear convergence rates to the setting where N_k is increased at a prescribed rate dependent on simulation and computational budgets; (III) finally, when steplengths are constant, we obtain the optimal number of projection steps that minimizes the bound on the mean-squared error.

1 INTRODUCTION

First suggested by Robbins and Monro (1951) in the context of root finding problems, stochastic approximation schemes (Kushner and Yin 2003, Borkar 2008) have proven to be effective on a breadth of stochastic computational problems including convex optimization, variational inequality problems, and Markov decision processes. Stochastic approximation schemes closely resemble deterministic counterparts such as gradient descent, which under strong convexity assumptions, exhibit exponentially fast rates of convergence.

By introducing batch sizes, one may embed both stochastic approximations and gradient descent into a single parametric family of algorithms so that at one extreme (batch size equal to unity), we recover stochastic approximations, and at the other extreme (batch size equal to infinity), we obtain gradient descent. Given a fixed budget and various choices of steplengths, our goal is to study how one may choose batch sizes in order to minimize a certain bound on the rate of convergence. This bound, in turn, is built to recover the standard rates of convergence well known for stochastic approximation algorithms. In order to describe our contributions more precisely, let us discuss the stochastic approximation method in the context of stochastic convex optimization:

$$\min_{x \in X} \quad \mathbb{E}[f(x, \xi(\omega))], \tag{1}$$

where $X \subseteq \mathbb{R}^n$, $\xi : \Omega \to \mathbb{R}^d$, $f : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ and $(\Omega, \mathscr{F}, \mathbb{P})$ denotes the associated probability space and $\mathbb{E}[\bullet]$ denotes the expectation with respect to $\mathbb{P}[\bullet]$. We shall assume that X is a compact and convex set and $f(x)$ is a convex function in x where $f(x) := \mathbb{E}[f(x, \xi(\omega))]$. In fact, we shall focus on problems wherein the function $f(x)$ is continuously differentiable and strongly convex. Recall that strong convexity implies that there exists an $\eta > 0$ such that $(\nabla_x f(x) - \nabla_x f(y))^T (x - y) \geq \eta \|x - y\|^2$. Moreover, we shall assume that $f(x)$ has Lipschitz continuous gradients, that is, $\|\nabla_x f(x) - \nabla_x f(y)\| \leq L\|x - y\|$ for some $L > 0$. We

978-1-4673-9743-8/15/$31.00 ©2015 IEEE

use $\|\bullet\|$ to denote the Euclidian norm. Note that since X is compact, we have that there exists $D > 0$ such that $\max_{x\in X} \|x-x^*\|^2 \leq D$, where x^* is an optimal solution of (1).

Vanilla implementations of stochastic approximation schemes have comprised of the following update rule: Given an $x_1 \in X$, an SA scheme is based on the following update rule:

$$x_{k+1} := \Pi_X\left(x_k - \gamma_k(\nabla_x f(x_k) + w_k)\right), \quad k \geq 1$$

where $w_k := \nabla_x f(x_k;\omega_k) - \nabla_x f(x_k)$ and $\nabla_x f(x,\xi(\omega))$ is referred to as $\nabla_x f(x,\omega)$. If $\{\gamma_k\}$ is a square-summable but non-summable sequence, then $\{x_k\} \to x^*$. We shall write $\mathscr{F}_k \triangleq \{x_1,\omega_1,\ldots,\omega_k\}$ and assume that $\mathbb{E}[\|w_k\|^2 \mid \mathscr{F}_k] \leq \nu^2$ for some $\nu \in (0,\infty)$. Under strong convexity, it is known that when $x \in \text{int}(X)$, $\mathbb{E}[f(x_k;\omega_k) - f^*] = \mathscr{O}(1/k)$. Meanwhile, it is well known that iterating the sequence $x_{k+1} := \Pi_X\left(x_k - \nabla_x f(x_k)\right)$, which corresponds to gradient descent, we have that $f(x_k) - f(x^*) = \mathscr{O}\left(\rho^k\right)$ for some $\rho \in (0,1)$. (As a side note, when both strong convexity and differentiability of the function are weakened, $\mathbb{E}[f(x_k;\omega_k) - f^*] = \mathscr{O}(1/\sqrt{k})$, shown to be unimprovable by Nemirovski and Yudin (1983).)

Scheme	Sample size: N_k	Steplength: γ_k	q_k	β_K	Rate for $K \leq \bar{K}$	Optimal K
Const. N, Const. γ	$N_k := N = \lceil \beta_K q^{-K} \rceil$	$\gamma_k := \gamma$	$q := (1-2\eta\gamma+\gamma^2L^2)$	$\left(\frac{M}{K}-1\right)q^K$	Linear	K^* solves (4)
Const. N, Dim. γ	$N_k := N = \lceil \beta_K q_K^{-K} \rceil$	$\gamma_k := \theta/k$	$q_k := (1-2\eta\gamma_k+\gamma_k^2L^2)$	$\left(\frac{M}{K}-1\right)q_K^K$	Linear	-
Inc. N, Const. γ	$N_k := \lceil \frac{\beta_K}{q^k} \rceil$	$\gamma_k := \gamma$	$q := (1-2\eta\gamma+\gamma^2L^2)$	$\frac{(M-K)}{\sum_{k=1}^K q^{-k}}$	Linear	K^* solves (6)
Inc. N, Dim. γ	$N_k := \lceil \frac{\beta_K}{\prod_{j=1}^k q_j} \rceil$	$\gamma_k := \theta/k$	$q_k := (1-2\eta\gamma_k+\gamma_k^2L^2)$	$\frac{(M-K)}{\sum_{k=1}^K \frac{1}{\prod_{j=1}^k q_j}}$	Linear	-

Table 1: Budget constrained stochastic approximation schemes (Simulation budget $= M$)

Research question: We consider a generalization where at step k, N_k samples of the gradient are obtained. As a consequence, given a randomly generated $x_1 \in X$, the sequence $\{x_{k+1}\}$ is given by the following update rule:

$$x_{k+1} := \Pi_X\left(x_k - \gamma_k \frac{\sum_{j=1}^{N_k} \nabla_x f(x_k,\omega_{j,k})}{N_k}\right), \quad k \geq 1. \qquad (\text{SA}_k)$$

One may immediately note that when $N_k := 1$ for all k, this reduces to the standard SA scheme, and if $N_k = \infty$, and $\gamma_k = \gamma$, the scheme reduces to gradient descent. In the context of this scheme, we consider the following question. **Given a simulation budget of M samples and a computational budget of K projection steps, how should N_k be selected as a function of M, K, and problem parameters so as to attain bounds on the rate of convergence that resemble the exponential rate of convergence seen in deterministic schemes such as gradient descent?** To this end, this paper makes the following contributions:

(I) We consider a constant sample-size batch ($N_k = N$) SA scheme under constant ($\gamma_k = \gamma$) and diminishing ($\gamma_k = \theta/k$) steplength regimes. Specifically, given simulation and computational budgets, we prove linear convergence in a mean-squared sense (see Proposition 2 and Theorem 3 below) when N is selected in accordance with M, K, and problem parameters;

(II) Next, we extend the two prior SA schemes to allow for an increasing sequence of sample sizes N_k and provide a scheme for updating the sample size in terms of simulation and computational budgets that allows for recovering the linear rate of convergence when steplength sequences are either constant or diminishing (see Proposition 5 and Theorem 6 below);

(III) Finally, when steplengths are constant, we resolve the question of the optimal number of projection steps that minimize the bound on the mean-squared error (see Proposition 4 and Lemma 7 below).

When the projection operation in (SA_k) is on relatively complex set and the simulation budget is M, standard SA schemes take $K = M$ projection steps, each requiring the solution of a (possibly challenging) convex program. We observe that in our preliminary numerics suggest that accurate solutions are available when $K << M$. Furthermore, in constant steplength regimes, an optimally chosen K (denoted by K^*) is much smaller than M.

We summarize our main findings in the Table 1 and conclude with a brief discussion of the relation between our findings and prior research on the analysis of stochastic approximation schemes related to the choice of steplengths and sample sizes for gradient estimates. The bottom line in this discussion is that our contribution differs from previous work in that we consider a fixed budget M and combine different choices of steplengths and batch sizes. Additional results in the literature are discussed next:

Steplength choices: The choice of the steplengths, γ_k, can prove devastating from the standpoint of algorithm performance, motivating diverse efforts to design schemes where steplengths are chosen in a sensible fashion. Nemirovski et al. (2009) presented a constant steplength scheme where the choice is contingent on the termination length and achieves the optimal rate. While these choices require an a priori specification of the termination length, such schemes naturally provide approximate solutions at best. Asymptotically exact schemes have been developed by Yousefian et al. (2012) in which the steplength sequence is selected in accordance with problem parameters (such as Lipschitz constant, convexity constant etc.) and are seen to display the optimal rate. Adaptively chosen steplengths have also been considered with a study of the associated rates (See (Cicek, Broadie, and Zeevi 2011) and references therein).

Constant sample size SA schemes: Constant sample size SA schemes are generally referred to as mini-batch SA schemes and there has been significant analysis of the associated error bounds (cf. (Ghadimi, Lan, and Zhang 2014)). However, much of this work has assumed that the size of the batch is taken as a parameter. In contrast, the present work additionally focuses on determinining the optimal batch size so as to recover fast rates of convergence while accommodating a budget constraint on the number of samples.

Variable sample size SA schemes: Amongst related prior work is that in which sample-averages of gradients are utilized within the gradient method with changing sample sizes (cf. (Friedlander and Schmidt 2012, Byrd, Chin, Nocedal, and Wu 2012, Pasupathy, Glynn, Ghosh, and Hashemi 2014)). In particular, Friedlander and Schmidt investigate how rates of deterministic methods can be achieved through increasing sample sizes. Our work differs in that we assume that the sampling budget is constrained and precise schemes for updating the sampling size are provided so as to recover linear rates of convergence.

The remainder of the paper is organized as follows. In Section 2, we consider the constant sample size regime where $N_k = N$ while in Section 3, we investigate the increasing sample size regime. The paper concludes after providing a set of numerics in Section 4.

2 STOCHASTIC APPROXIMATION WITH CONSTANT SAMPLE SIZES

In this section, we consider the setting in which a fixed sample size is utilized at every step. Our original research question then reduces to what the size of this sample should be given that a fixed simulation budget is available. In fact, based on this sample size, the overall number of computational steps can be defined. We begin with a simple bound on the conditional second moment of the error.

Lemma 1 Consider the variable sample SA scheme denoted by (SA_N) and suppose $f(x)$ is a strongly convex function with convexity constant η. Furthermore suppose f is continuously differentiable in x with Lipschitz continuous gradients with constant L. Then for every nonnegative k, the following holds

$$\mathbb{E}[\|x_{k+1} - x^*\|^2 \mid \mathscr{F}_k] \leq (1 - 2\eta\gamma_k + \gamma_k^2 L^2)\|(x_k - x^*)\|^2 + \frac{\gamma_k^2 \nu^2}{N_k}.$$

Proof. We begin by noting that x_k can be expressed as follows:

$$x_{k+1} = \Pi_X\left(x_k - \gamma_k \frac{\sum_{j=1}^{N_k} \nabla_x f(x_k,\omega_{k,j})}{N_k}\right) = \Pi_X\left(x_k - \gamma_k \frac{\sum_{j=1}^{N_k}(\nabla_x f(x_k)+w_{k,j})}{N_k}\right)$$
$$= \Pi_X\left(x_k - \gamma_k\left(\nabla_x f(x_k) + \frac{\sum_{j=1}^{N_k} w_{k,j}}{N_k}\right)\right) = \Pi_X\left(x_k - \gamma_k\left(\nabla_x f(x_k) + \bar{w}_{k,N_k}\right)\right),$$

where $\bar{w}_{k,N_k} := \frac{\sum_{j=1}^{N_k}\nabla f(x_k,\omega_{k,j})}{N_k}$. By leveraging the non-expansivity of the Euclidean projector, we may express $\|x_{k+1}-x^*\|^2$ as follows:

$$\begin{aligned}\|x_{k+1}-x^*\|^2 &= \|\Pi_X(x_k-\gamma_k(\nabla_x f(x_k)+\bar{w}_{k,N_k}) - \Pi_X(x^*-\gamma_k\nabla_x f(x^*))\|^2\\ &\leq \|(x_k-\gamma_k(\nabla_x f(x_k)+\bar{w}_{k,N_k}) - (x^*-\gamma_k\nabla_x f(x^*))\|^2\\ &= (1-2\eta\gamma_k+\gamma_k^2L^2)\|(x_k-x^*)\|^2\\ &\quad -2\gamma_k\bar{w}_{k,N_k}^T((x_k-x^*)-\gamma_k(\nabla_x f(x_k)-\nabla_x f(x^*))+\gamma_k^2\|\bar{w}_{k,N_k}\|^2.\end{aligned}$$

Taking conditional expectations on $\mathscr{F}_k$, we obtain the following inequality:

$$\mathbb{E}[\|x_{k+1}-x^*\|^2 \mid \mathscr{F}_k] \leq (1-2\eta\gamma_k+\gamma_k^2L^2)\|(x_k-x^*)\|^2 + \gamma_k^2\mathbb{E}[\|\bar{w}_{k,N_k}\|^2 \mid \mathscr{F}_k],$$

where $\mathbb{E}[\bar{w}_{k,N_k} \mid \mathscr{F}_k] = 0$. Furthermore, since we have that $\mathbb{E}[\|w_k\|^2 \mid \mathscr{F}_k] \leq \nu^2$, it follows that $\mathbb{E}[\|w_{k,N_k}\|^2 \mid \mathscr{F}_k] \leq \frac{\nu^2}{N_k}$. The result follows. ∎

Our first result assumes that a fixed number of samples, namely N, are employed at each iteration implying that the no more than $K \triangleq \lfloor \frac{M}{N} \rfloor$ steps are taken where

$$N := \begin{cases}\lceil \beta_K q^{-K}\rceil, & \text{if } \gamma_k := \gamma\\ \lceil \beta_K q_K^{-K}\rceil, & \text{if } \gamma_k := \theta/k\end{cases} \text{ and } \beta_K := \begin{cases}\left(\frac{M}{K}-1\right)q^K, & \text{if } \gamma_k = \gamma\\ \left(\frac{M}{K}-1\right)q_K^K. & \text{if } \gamma_k = \theta/k\end{cases} \tag{2}$$

Note that the choice of β_K is based on the specification $K\beta_K q^{-K} = M-K$. We now analyze the error bounds at the Kth iteration under constant and diminishing steplength regimes.

Proposition 2 Consider the scheme where N is defined as per (2). Then the following hold:

(i) Suppose $\gamma_k := \gamma$ for all k and $q \triangleq (1-2\eta\gamma+\gamma^2L^2)$. Then the mean-squared error may be bounded as follows:

$$\mathbb{E}[\|x_{K+1}-x^*\|^2] \leq \left(D + \frac{\min(K,(1-q)^{-1})\gamma^2\nu^2}{\beta_K}\right)q^K, \qquad \forall k \geq 1.$$

(ii) Suppose $\gamma_k := \theta/k$ for all k and $q_k \triangleq (1-2\eta\gamma_k+\gamma_k^2L^2)$. Then the mean-squared error may be bounded as follows:

$$\mathbb{E}[\|x_{K+1}-x^*\|^2] \leq \left(D + \frac{\pi^2\theta^2\nu^2}{6\beta_K}\right)q_K^K, \qquad \forall k \geq 1.$$

Proof. We begin by considering the general case in which the steplength is denoted by γ_k. Then after K gradient steps, the mean-squared error is given by the following:

$$\mathbb{E}[\|x_{K+1}-x^*\|^2 \mid \mathscr{F}_K] \leq (1-2\eta\gamma_K+\gamma_K^2L^2)\|(x_K-x^*)\|^2 + \frac{\gamma_K^2\nu^2}{N}.$$

Taking unconditional expectations, we obtain the following:

$$\mathbb{E}[\|x_{K+1}-x^*\|^2] \leq q_K \mathbb{E}[\|(x_K - x^*)\|^2] + \frac{\gamma_K^2 \nu^2}{N},$$

where $q_K := (1-2\eta\gamma_K + \gamma_K^2 L^2)$. Consequently, we have the following:

$$\begin{aligned}\mathbb{E}[\|x_{K+1}-x^*\|^2] &\leq q_K q_{K-1}\mathbb{E}[\|(x_{K-1}-x^*)\|^2] + q_K \frac{\gamma_{K-1}^2\nu^2}{N} + \frac{\gamma_K^2\nu^2}{N} \\ &\leq D\prod_{i=1}^{K} q_i + \prod_{i=2}^{K} q_i \frac{\gamma_1^2\nu^2}{N} + \cdots + q_K q_{K-1}\frac{\gamma_{K-2}^2\nu^2}{N} + q_K\frac{\gamma_{K-1}^2\nu^2}{N} + \frac{\gamma_K^2\nu^2}{N}. \qquad (3)\end{aligned}$$

(i) Suppose $\gamma_k := \gamma$ for all k, implying that $q_k = q$ for all k. Therefore, by recalling that $q < 1$ for sufficiently small γ, we may further bound (3) by recalling that $\sum_{j=0}^{K-1} q^j \leq \min((1-q)^{-1}, K)$. It follows that

$$\text{RHS of (3)} = q^K D + \sum_{j=0}^{K-1} q^j \frac{\gamma^2\nu^2}{N} \leq q^K D + \frac{\min(K,(1-q)^{-1})\gamma^2\nu^2}{N}.$$

By choice, $N = \lceil \beta_K q^{-K} \rceil \geq \beta_K q^{-K}$, allow for deriving the following bound:

$$q^K D + \frac{\min(K,(1-q)^{-1})\gamma^2\nu^2}{N} \leq q^K D + \frac{\min(K,(1-q)^{-1})\gamma^2\nu^2 q^K}{\beta_K}.$$

In effect, we have that the following holds for $k \geq 1$:

$$\mathbb{E}[\|x_{K+1}-x^*\|^2] \leq q^K\left(D + \frac{\min(K,(1-q)^{-1})\gamma^2\nu^2}{\beta_K}\right).$$

(ii) Suppose $\gamma_k := \theta/k$ for all $k \geq 1$ and θ is sufficiently small. Then $q_1 := q(\theta_1) < 1$ and since $q(\gamma)$ is a decreasing function in γ, $q_k \leq q_{k-1} \leq \cdots \leq q_2 \leq q_1 \leq 1$, facilitating the bound below:

$$\begin{aligned}\text{RHS of (3)} &= D\prod_{i=1}^{K} q_i + \prod_{i=2}^{K} q_i \frac{\theta^2\nu^2}{N} + \cdots + q_K q_{K-1}\frac{\theta^2\nu^2}{N(K-2)^2} + q_K\frac{\theta^2\nu^2}{N(K-1)^2} + \frac{\theta^2\nu^2}{NK^2} \\ &= q_K^K D + \frac{\theta^2\nu^2}{N}\sum_{j=0}^{K-1}\frac{\prod_{i=1}^{j} q_{K+1-i}}{(K-j)^2} \leq q_K^K D + \frac{\theta^2\nu^2}{N}\left(\sum_{j=0}^{K-1}\frac{1}{(K-j)^2}\right).\end{aligned}$$

By recalling that $\sum_{j=0}^{K-1} 1/(K-j)^2 = \sum_{j=1}^{K-1} 1/j^2 \leq \sum_{j=1}^{\infty} 1/j^2 = \frac{\pi^2}{6}$, we can further bound this expression as follows:

$$q_K^K D + \frac{\theta^2\nu^2}{N}\left(\sum_{j=0}^{K-1}\frac{1}{(K-j)^2}\right) \leq q_K^K D + \frac{\pi^2\theta^2\nu^2}{6N}.$$

Finally, by utilizing the requirement that $N = \lceil \beta_K q_K^{-K} \rceil \geq \beta_K q_K^{-K}$, we may derive the following bound:

$$\mathbb{E}[\|x_{K+1}-x^*\|^2] \leq q_K^K\left(D + \frac{\pi^2\theta^2\nu^2}{6\beta_K}\right), \qquad \forall k \geq 1.$$

∎

Remark: While a cursory look might suggest a geometric rate of convergence, a closer study suggests that this is not the case given that β_K and q_K are both dependent on K. However, if $K \leq \bar{K}$, a consequence of imposing a computational budget, we note that a linear rate of convergence may be obtained.

Theorem 3 (**Linear convergence rate under finite computational budget**) Suppose $\bar{K}$ projection steps are available.

(i) Let $\gamma_k = \gamma$ for all k. Then the following holds:

$$\mathbb{E}[\|x_{k+1} - x^*\|^2] \leq q^k \left(D + \frac{\min\left(\bar{K}, (1-q)^{-1}\right)\gamma^2 \nu^2}{\beta_{\bar{K}}} \right), \qquad \text{for all } k \leq \bar{K}.$$

(ii) Let $\gamma_k = \theta/k$ for all k. Then the following holds:

$$\mathbb{E}[\|x_{k+1} - x^*\|^2] \leq q_{\bar{K}}^k \left(D + \frac{\pi^2 \theta^2 \nu^2}{6\beta_{\bar{K}}} \right), \qquad \text{for all } k \leq \bar{K}.$$

Proof. (i) Suppose $K \leq \bar{K}$. Then by leveraging (2), we have that

$$\beta_K = \left(\frac{M}{K} - 1\right) q^K \geq \left(\frac{M}{\bar{K}} - 1\right) q^K \geq \left(\frac{M}{\bar{K}} - 1\right) q^{\bar{K}} \triangleq \beta_{\bar{K}}.$$

It follows that when computational budget is bounded and steplengths are fixed, we have that for $K \leq \bar{K}$:

$$\mathbb{E}[\|x^{K+1} - x^*\|^2] \leq q^K \left(D + \frac{\min(\bar{K}, (1-q)^{-1})\gamma^2 \nu^2}{\beta_{\bar{K}}} \right).$$

(ii) Consider the case where $\gamma_k = \theta/k$. When computational budget is bounded and steplengths are diminshing, by the decreasing nature of β_K in k, we have that for $K \leq \bar{K}$:

$$\mathbb{E}[\|x_{K+1} - x^*\|^2] \leq q_K^K \left(D + \frac{\pi^2 \theta^2 \nu^2}{6\beta_K} \right) \leq q_{\bar{K}}^K \left(D + \frac{\pi^2 \theta^2 \nu^2}{6\beta_{\bar{K}}} \right), \qquad \forall K \leq \bar{K}.$$

∎

Remark: This result is instructive in that one can derive an error bound that reduces to that observed in strongly convex deterministic regimes when the uncertainty disappears. Specifically, if $\nu = 0$, we notice that $\|x^{k+1} - x^*\|^2 \leq q^K \|x_0 - x^*\|^2$, in the constant steplength regime while in the diminishing steplength regime, we obtain $\|x^{k+1} - x^*\|^2 \leq \Pi_{j=1}^k q_i \|x_0 - x^*\|^2 \leq q_k^k \|x_0 - x^*\|^2$.

We conclude this section with an investigation of the optimal number of computational steps or K (or equivalently the optimal N) in the constant steplength regime, obtained by minimizing the upper bound.

Proposition 4 (**Optimal choice of** K) Consider the SA scheme in which $\gamma_k = \gamma$ and $N_k = N$ for all k. Then the following hold:

(i) The error bound $h(K)$ is convex in K for all K where

$$h(K) \triangleq q^K D + \frac{1}{\left(\frac{M}{K} - 1\right)} \frac{\gamma^2 \nu^2}{1-q}.$$

(ii) Suppose there exists a γ such that

$$\log(1/q)(1-q)D > \gamma^2 \nu^2 / M.$$

Then there is a unique root $K^* \in (0, M)$ to the equation

$$\ln(1/q)(1-q)Dq^K = \frac{\gamma^2 \nu^2 M}{(M-K)^2}, \tag{4}$$

and $h(K^*) \leq h(K)$ for all $K \in [0, M]$.

Proof. **(i)**: Consider $h(K)$ which is defined as follows:

$$\mathbb{E}[\|x_{K+1}-x^*\|^2] \leq q^K\left(D+\frac{\gamma^2\nu^2}{\beta_K}\frac{1}{1-q}\right) = q^K D + \frac{1}{\left(\frac{M}{K}-1\right)}\frac{\gamma^2\nu^2}{1-q} \triangleq h(K).$$

The first and second derivatives of h can then be derived.

$$h'(K) = Dq^K\ln(q) + \frac{1}{\left(\frac{M}{K}-1\right)^2}\frac{M}{K^2}\frac{\gamma^2\nu^2}{1-q} = Dq^K\ln(q) + \frac{M}{(M-K)^2}\frac{\gamma^2\nu^2}{1-q}$$

$$h''(K) = Dq^K(\ln(q))^2 + \frac{2M}{(M-K)^3}\frac{\gamma^2\nu^2}{1-q} > 0, \qquad \text{for all } K.$$

It follows that $h(K)$ is convex in K.

(ii) An unconstrained minimizer K^* is given by the solution to $h'(K)=0$, which is equivalent to (4). We can see that $K^* \in (0,M)$ is unique because of two reasons: first the left hand side of (4) is decreasing in K, and the right hand side increases up to infinity as $K \nearrow M$; second, the left hand side equals $\log(1/q)(1-q)D$, at $K=0$, which, by assumption, is strictly larger than $\gamma^2\nu^2/M$, which in turn is the value of the righ hand side at $K=0$. ∎

Remark: The reader may observe that an optimal selection of K^* is always feasible and can be made independently of γ if the budget M is large enough. We note that the optimal choice of K lies in the interior of $[0,M]$. In fact, this choice of K^* contrasts with the standard value of $K=M$ which is a result of using a single sample at every k. Furthermore, given K^*, one may then compute a β_{K^*} and an N^* based on (2).

3 STOCHASTIC APPROXIMATION WITH INCREASING SAMPLE SIZES

In the prior section, we considered a setting where the sample size N_k was fixed for every step at N. In this section, we consider an alternate approach in which the sample size is raised at every step through a prescribed update rule, with the overall goal of obtaining linear convergence rates over a finite horizon.

$$N_k := \begin{cases} \lceil \frac{\beta_K}{q^k} \rceil, & \text{if } \gamma_k := \gamma \\ \lceil \frac{\beta_K}{\pi_{j=1}^k q_j} \rceil, & \text{if } \gamma_k := \theta/k \end{cases} \text{ and } \beta_K := \begin{cases} \frac{(M-K)}{\sum_{k=1}^K q^{-k}}, & \text{if } \gamma_k = \gamma \\ \frac{(M-K)}{\sum_{k=1}^K \frac{1}{\Pi_{j=1}^k q_j}}. & \text{if } \gamma_k = \theta/k \end{cases} \tag{5}$$

Proposition 5 Consider the scheme where N_k is defined as per (5). Then the following hold:

(i) Suppose $\gamma_k := \gamma$ for all k. Then the mean-squared error may be bounded as follows:

$$\mathbb{E}[\|x_{K+1}-x^*\|^2] \leq q^K\left(D+\frac{\gamma^2\nu^2 K}{\beta_K}\right).$$

(ii) Suppose $\gamma_k := \theta/k$ for all k. Then the mean-squared error may be bounded as follows:

$$\mathbb{E}[\|x_{K+1}-x^*\|^2] \leq q_K^K\left(D+\frac{\theta^2\nu^2\pi^2}{6\beta_K}\right).$$

Proof. After K gradient steps, the mean-squared error is given by the following:

$$\mathbb{E}[\|x_{K+1}-x^*\|^2 \mid \mathscr{F}_K] \leq (1-2\eta\gamma_K+\gamma_K^2L^2)\|(x_K-x^*)\|^2 + \frac{\gamma_K^2\nu^2}{N_K}.$$

Taking unconditional expectations, we obtain the following:

$$\mathbb{E}[\|x_{K+1}-x^*\|^2] \leq q_K \mathbb{E}[\|(x_K - x^*)\|^2] + \frac{\gamma_K^2 \nu^2}{N_K},$$

where $q_K := (1-2\eta\gamma_K + \gamma_K^2 L^2)$. Consequently, we have the following:

$$\begin{aligned}\mathbb{E}[\|x_{K+1}-x^*\|^2] &\leq q_K q_{K-1}\mathbb{E}[\|(x_{K-1}-x^*)\|^2] + q_K \frac{\gamma_{K-1}^2\nu^2}{N_{K-1}} + \frac{\gamma_K^2\nu^2}{N_K} \\ &\leq D\prod_{i=1}^K q_i + \left(\prod_{i=2}^K q_i\right)\frac{\gamma_1^2\nu^2}{N_1} + \cdots + q_K q_{K-1}\frac{\gamma_{K-2}^2\nu^2}{N_{K-2}} + q_K\frac{\gamma_{K-1}^2\nu^2}{N_{K-1}} + \frac{\gamma_K^2\nu^2}{N_K}.\end{aligned}$$

(i) Suppose $\gamma_k := \gamma$ for all k, implying that $q_k = q$ for all k. Therefore, we may further bound the above expression as follows:

$$\begin{aligned}&D\prod_{i=1}^K q_i + \prod_{i=2}^K q_i\frac{\gamma_1^2\nu^2}{N_1} + \cdots + q_K q_{K-1}\frac{\gamma_{K-2}^2\nu^2}{N_{K-2}} + q_K\frac{\gamma_{K-1}^2\nu^2}{N_{K-1}} + \frac{\gamma_K^2\nu^2}{N_K} \\ &\leq Dq^K + q^K\frac{\gamma^2\nu^2}{\beta_K} + \cdots + q^K\frac{\gamma^2\nu^2}{\beta_K} + q^K\frac{\gamma^2\nu^2}{\beta_K} + q^K\frac{\gamma^2\nu^2}{\beta_K} = Dq^K + q^K\frac{K\gamma^2\nu^2}{\beta_K},\end{aligned}$$

where $\beta_K = \frac{M-K}{\sum_{k=1}^K \frac{1}{q^k}}$, implying that

$$\mathbb{E}[\|x_{K+1}-x^*\|^2] \leq q^K\left(D + \frac{\gamma^2\nu^2 K}{\beta_K}\right).$$

(ii) Suppose $\gamma_k := \theta/k$ for all k and by recalling that q is a decreasing function in γ, we may further bound the above expression as follows:

$$\begin{aligned}&D\prod_{i=1}^K q_i + \prod_{i=2}^K q_i\frac{\gamma_1^2\nu^2}{N_1} + \cdots + q_K q_{K-1}\frac{\gamma_{K-2}^2\nu^2}{N_{K-2}} + q_K\frac{\gamma_{K-1}^2\nu^2}{N_{K-1}} + \frac{\gamma_K^2\nu^2}{N_K} \\ &= D\prod_{i=1}^K q_i + \prod_{i=2}^K q_i\frac{\gamma_1^2\nu^2}{\lceil\beta_K q_1^{-1}\rceil} + \cdots + q_K q_{K-1}\frac{\gamma_{K-2}^2\nu^2}{\lceil\beta_K\prod_{i=0}^{K-2}q_i^{-1}\rceil} + q_K\frac{\gamma_{K-1}^2\nu^2}{\lceil\beta_K\prod_{i=0}^{K-1}q_i^{-1}\rceil} + \frac{\gamma_K^2\nu^2}{\lceil\beta_K\prod_{i=0}^{K}q_i^{-1}\rceil} \\ &= D\prod_{i=1}^K q_i + \prod_{i=2}^K q_i\frac{\gamma_1^2\nu^2}{\lceil\beta_K q_1^{-1}\rceil} + \cdots + q_K q_{K-1}\frac{\gamma_{K-2}^2\nu^2}{\lceil\beta_K\prod_{i=0}^{K-2}q_i^{-1}\rceil} + q_K\frac{\gamma_{K-1}^2\nu^2}{\lceil\beta_K\prod_{i=0}^{K-1}q_i^{-1}\rceil} + \frac{\gamma_K^2\nu^2}{\lceil\beta_K\prod_{i=0}^{K}q_i^{-1}\rceil} \\ &\leq D\left(\prod_{i=1}^K q_i\right) + \left(\prod_{i=1}^K q_i\right)\sum_{j=1}^{K-1}\frac{\theta^2\nu^2}{\beta_K j^2} = \left(\prod_{i=1}^K q_i\right)\left(D + \theta^2\nu^2\sum_{j=1}^{K-1}\frac{1}{\beta_K j^2}\right),\end{aligned}$$

where the inequality follows from noting that $\lceil\beta_K\prod_{i=0}^{K-j}q_i^{-1}\rceil \geq \beta_K\prod_{i=0}^{K-j}q_i^{-1}$. By recalling that $\sum_{j=1}^{K-1}1/j^2 \leq \sum_{j=1}^{\infty}1/j^2 = \frac{\pi^2}{6}$, we can further bound this expression as follows:

$$\left(\prod_{i=1}^K q_i\right)\left(D + \theta^2\nu^2\sum_{j=1}^{K-1}\frac{1}{\beta_K j^2}\right) \leq \left(\prod_{i=1}^K q_i\right)\left(D + \frac{\theta^2\nu^2\pi^2}{6\beta_K}\right) \implies \mathbb{E}[\|x_{K+1}-x^*\|^2] \leq q_K^K\left(D + \frac{\theta^2\nu^2\pi^2}{6\beta_K}\right).$$

∎

We now examine the convergence rate when constrained by a finite computational budget.

Theorem 6 (Linear convergence rate under finite computational budget) Suppose $\bar{K}$ projection steps are available.

(i) Let $\gamma_k = \gamma$ for all k. Then the following holds:

$$\mathbb{E}[\|x_{k+1} - x^*\|^2] \leq q^k \left(D + \frac{\gamma^2 \nu^2 \bar{K}}{\beta_{\bar{K}}} \right), \qquad \forall k \leq \bar{K}.$$

(ii) Let $\gamma_k = \theta/k$ for all k. Then the following holds:

$$\mathbb{E}[\|x_{k+1} - x^*\|^2] \leq q_{\bar{K}}^k \left(D + \frac{\theta^2 \nu^2 \pi^2}{6\beta_{\bar{K}}} \right), \qquad \forall k \leq \bar{K}.$$

Proof. **(i):** Consider the error given by $q^K \left(D + \frac{\gamma^2 \nu^2 K}{\beta_K} \right)$. If $K \leq \bar{K}$, then we have that

$$\beta_K = \frac{M}{\sum_{k=1}^{K} \frac{1}{q^k}} \geq \frac{M}{\sum_{k=1}^{\bar{K}} \frac{1}{q^k}} \triangleq \beta_{\bar{K}}.$$

It follows that when computational budget is bounded and steplengths are fixed, we have that for $K \leq \bar{K}$:

$$\mathbb{E}[\|x_{k+1} - x^*\|^2] \leq q^k \left(D + \frac{\gamma^2 \nu^2 k}{\beta_K} \right) \leq q^k \left(D + \frac{\gamma^2 \nu^2 \bar{K}}{\beta_{\bar{K}}} \right), \qquad \forall k \leq \bar{K}.$$

(ii): When computational budget is bounded and steplengths are diminshing, we have that for $K \leq \bar{K}$:

$$\mathbb{E}[\|x_{K+1} - x^*\|^2] \leq q_K^K \left(D + \frac{\theta^2 \nu^2 \pi^2}{6\beta_K} \right) \leq q_{\bar{K}}^K \left(D + \frac{\theta^2 \nu^2 \pi^2}{6\beta_{\bar{K}}} \right), \qquad \forall K \leq \bar{K}.$$

∎

We conclude this section with a discussion of how to optimally choose K when steplength sequences are constant and sample sizes are increasing. We discuss the convexity of a bound on the mean-squared error which allows for deriving an optimal choice of K. Rather than minimize the bound derived in Theorem 6 (i), we minimize an upper bound based on the following observation: Since $\beta_K = \frac{M-K}{\sum_{j=1}^{K} \frac{1}{q^j}}$, it follows that $\beta_K \geq \frac{M-K}{Kq^{-K}}$ from noting that $1/q^j \geq 1/q^K$, leading to the following bound:

$$Dq^K + \frac{\gamma^2 \nu^2 K q^K}{\beta_K} \leq Dq^K + \frac{\gamma^2 \nu^2 K}{(\frac{M}{K} - 1)}.$$

Lemma 7 Consider the SA scheme in which N_k is increasing as per the prescribed rule and $\gamma_k = \gamma$ for all k. Then the function is $h(K)$ convex in K where $h(K) = Dq^K + \frac{\gamma^2 \nu^2}{\frac{(M-K)}{K^2}}$. Furthermore, the optimal choice $K^* \in (0, M)$ is the unique solution to the equation

$$Dq^K \ln(1/q) = \frac{\gamma^2 \nu^2 K}{(M-K)^2} \left((2M - K) \right). \tag{6}$$

Proof. Since $\beta_K = \frac{M-K}{\sum_{j=1}^{K} \frac{1}{q^j}}$ for $j < K$, it follows that $\beta_K \geq \frac{M-K}{Kq^{-K}}$ from noting that $1/q^j \geq 1/q^K$. Consequently $Dq^K + \frac{\gamma^2 \nu^2 K q^K}{\beta_K} \leq Dq^K + \frac{\gamma^2 \nu^2}{\frac{(M-K)}{K^2}}$ and

$$
\begin{aligned}
h'(K) &= Dq^K \ln(q) - \frac{\gamma^2 \nu^2}{\frac{(M-K)^2}{K^4}} \left(-\frac{2M}{K^3} + \frac{1}{K^2} \right) = Dq^K \ln(q) + \frac{\gamma^2 \nu^2}{(\frac{M}{K} - 1)^2} \left((2\frac{M}{K} - 1) \right) \\
h''(K) &= Dq^K \ln^2(q) + \frac{M}{K^2} 2 \frac{\gamma^2 \nu^2}{(\frac{M}{K} - 1)^3} \left((2\frac{M}{K} - 1) \right) + \frac{\gamma^2 \nu^2}{(\frac{M}{K} - 1)^2} \left((-\frac{M}{K^2}) \right) \\
&= Dq^K \ln^2(q) + \frac{M\gamma^2 \nu^2}{(\frac{M}{K} - 1)^2 K^2} \left(\frac{4(M/K) - 2}{((M/K) - 1))} - 1 \right) \\
&= Dq^K \ln^2(q) + \frac{M\gamma^2 \nu^2}{(\frac{M}{K} - 1)^2 K^2} \left(\frac{3(M/K) - 1}{((M/K) - 1))} \right) \geq 0, \text{ when } K \leq M.
\end{aligned}
$$

We select K^* satisfying the equation $h'(K) = 0$, which is equivalent to (6). The fact that there is a unique $K^* \in (0, M)$ satisfying (6) follows from the fact that the left hand side of (6) is decreasing in $[0, M]$ with initial value equal to $D \ln(1/q) > 0$ at $K = 0$; meanwhile, the right hand side of (6), whose nonnegative derivative has been evaluated in the analysis of $h''(K)$, increases up to infinity as $K \nearrow M$. ∎

4 NUMERICAL RESULTS

We provide some preliminary numerical investigations on a stochastic quadratic program defined as follows:

$$
\min_{x \in X} \quad \mathbb{E}\left[\frac{1}{2} x^T Q(\xi) x - d^T x\right],
$$

where $X \triangleq [0, 10]^n$. We assume that $\mathbb{E}[Q(\xi)] = 2I + R^T R$ where R is a randomly generated $n-$dimensional square matrix (using a uniform distribution over $[0, 1]$) and I denotes the identity matrix. Furthermore $d = -2e$ where e denotes the column of ones in $n-$space. The optimal solution lies in the interior of the set and is given by $Q^{-1} d$ where $Q = \mathbb{E}[Q(\xi)]$. We examine the performance of four SA schemes when $M = 1e6$.

Performance of SA schemes: In Table 2, we consider constant sample size SA schemes with constant and diminishing steplength sequences (with $\theta = 1$). It can be seen that while deterministic projected gradient schemes require 20 projection steps to reach an error of 2.3e-5, the constant sample-size schemes provide solutions with an error of the order 1e-2 in the same number of steps. Minimizing the theoretical error bound in the constant steplength regime leads to a $K^* = 48$ which corresponds well with what is seen from the profile of the thoeretical error. When one compares the diminishing steplength scheme, it can be seen that there is a significant improvement in terms of empirical performance. This is not surprising since the diminishing nature of steplength helps mute the stochastic error. It is also observed that the theoretical error bound tends to have a minimzer around $K^* = 40$ and then slowly edges upward, a characteristic that is supported by the empirical behavior. When one examines increasing sample-size schemes with

#	$\|x_K^{\text{det}} - x^*\|$	k^{det}	$\|x - x^*\|$	k^{const}	Theor. bound
10	2.288e-05	20	1.471e-02	12	9.536e-01
20	2.288e-05	20	9.058e-03	22	2.421e-01
30	2.288e-05	20	1.189e-02	33	6.331e-02
40	2.288e-05	20	1.319e-02	42	2.398e-02
50	2.288e-05	20	1.424e-02	52	2.079e-02
60	2.288e-05	20	1.607e-02	62	2.238e-02
70	2.288e-05	20	1.757e-02	72	2.415e-02
80	2.288e-05	20	1.888e-02	82	2.582e-02
48	2.288e-05	20	1.647e-02	51	2.066e-02

#	$\|x_K^{\text{det}} - x^*\|$	k^{det}	$\|x - x^*\|$	k^{const}	Theor. bound
10	2.288e-05	20	6.862e-03	11	9.535e-01
20	2.288e-05	20	6.240e-03	21	6.577e-01
30	2.288e-05	20	5.708e-03	32	5.961e-01
40	2.288e-05	20	6.027e-03	41	5.703e-01
50	2.288e-05	20	5.978e-03	51	5.562e-01
60	2.288e-05	20	5.973e-03	61	5.473e-01
70	2.288e-05	20	5.976e-03	71	5.413e-01
80	2.288e-05	20	5.922e-03	81	5.369e-01
90	2.288e-05	20	5.849e-03	92	5.335e-01

Table 2: Constant sample size SA schemes with constant and diminishing steplengths

constant and diminishing steplength sequences, the results tend to be somewhat better. For instance, with

SA schemes with constant steplength and increasing sample sizes, it is seen that the obtained accuracy is of the order of 1e-3 and is nearly an order of magnitude better than the constant sample size counterpart. The diminishing steplength counterpart provides similar results but with somewhat poorer theoretical error bounds. We should note that the theoretical error bound employed is slightly weaker than that proved in theory and we expect the bounds to be tighter in practice.

#	$\|x_K^{det}-x^*\|$	k^{det}	$\|x-x^*\|$	k^{const}	Theor. bound
10	2.288e-05	20	1.411e-02	12	9.536e-01
20	2.288e-05	20	3.385e-03	22	2.421e-01
30	2.288e-05	20	3.385e-03	32	6.331e-02
40	2.288e-05	20	3.385e-03	42	2.398e-02
50	2.288e-05	20	3.389e-03	52	2.079e-02
60	2.288e-05	20	3.392e-03	62	2.238e-02
70	2.288e-05	20	3.389e-03	72	2.415e-02
80	2.288e-05	20	3.389e-03	82	2.582e-02
38	2.288e-05	20	3.384e-03	40	2.711e-02

#	$\|x_K^{det}-x^*\|$	k^{det}	$\|x-x^*\|$	k^{const}	Theor. bound
10	2.288e-05	20	4.264e-01	4	9.535e-01
20	2.288e-05	20	3.920e-03	11	6.577e-01
30	2.288e-05	20	3.410e-03	18	5.961e-01
40	2.288e-05	20	3.676e-03	24	5.703e-01
50	2.288e-05	20	3.739e-03	30	5.562e-01
60	2.288e-05	20	3.790e-03	36	5.473e-01
70	2.288e-05	20	3.824e-03	42	5.413e-01
80	2.288e-05	20	3.874e-03	48	5.369e-01
90	2.288e-05	20	3.770e-03	54	5.335e-01

Table 3: Increasing sample size SA schemes with constant and diminishing steplengths

Optimal choices of K and N: In the prior subsections, we have examined the question of minimizing the theoretical bound in K to ascertain the optimal number of projection steps (and therefore the optimal sample size). In Figure 1, we plot the theoretical and empirical error for constant steplength SA schemes with constant and increasing sample sizes. From this figure, we see that the theoretical error bound is minimized in both regimes and this behavuor is matched to some degree by the empirical error.

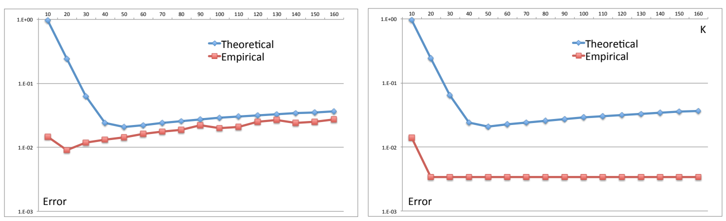

Figure 1: Theoretical vs Empirical error (const steplength and constant (l) and increasing (r) sample sizes

Linear convergence rates with finite computational budget: A key question that has motivated this work is whether one can develop methods that provide faster convergence rates which are valid in a finite (rather than non-asymptotic) regime. We note that by imposing a budget on the computational complexity, we may derive precisely such rates and these rates. Naturally these rates are weaker than the empirical behavior but Figure 2 demonstrates that the linear nature of these rates.

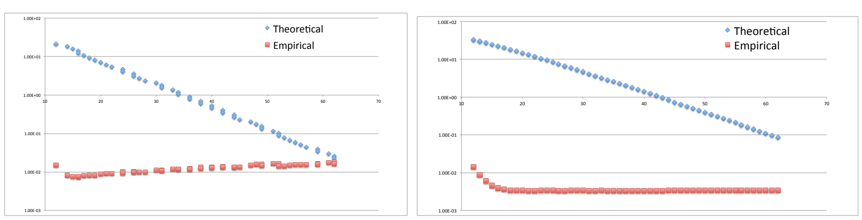

Figure 2: Linear convergence: (const steplength and constant (l) and increasing (r) sample sizes)

Computational benefits: A key concern in the implementation of standard stochastic approximation schemes is the effort associated with projection on convex sets which are not necessarily simple (such as box constraints, for instance). These projection problems lead to nonlinear convex programs and in traditional SA implementations with a budget of M samples, such problems are solved M times. In the proposed schemes, we observe that $K^* << M$ and the computational benefits are seen to be significant.

5 CONCLUDING REMARKS

We present a set of stochastic approximation schemes that can contend with bounds on the simulation budget under either constant or increasing sample sizes with either constant or diminishing steplength sequences. In particular, we show that by suitable allocation policies of the simulation budget, the schemes display linear convergence when the computational budget is restricted. Notably, the number of projection steps that minimize the theoretical error bound can be specified when steplengths are constant and sample sizes are either constant or increasing. Finally, preliminary numerics suggest that the schemes perform well in that they produce reasonably accurate solutions with significant computational benefit.

ACKNOWLEDGEMENTS

The first author is grateful for support of the NSF (CMMI (CAREER) 1246887).

REFERENCES

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

Byrd, R. H., G. M. Chin, J. Nocedal, and Y. Wu. 2012. "Sample size selection in optimization methods for machine learning". *Math. Program.* 134 (1): 127–155.

Cicek, D., M. Broadie, and A. Zeevi. 2011. "General Bounds and Finite-time Performance Improvement for the Kiefer-Wolfowitz Stochastic Approximation Algorithm". *Operations Research* 59:1211–1224.

Friedlander, M. P., and M. Schmidt. 2012. "Hybrid deterministic-stochastic methods for data fitting". *SIAM J. Scientific Computing* 34 (3): A1380A1405.

Ghadimi, S., G. Lan, and H. Zhang. 2014. "Mini-batch Stochastic Approximation Methods for Nonconvex Stochastic Composite Optimization". *Mathematical Programming*.

Kushner, H. J., and G. G. Yin. 2003. *Stochastic approximation and recursive algorithms and applications*. Second ed, Volume 35 of *Applications of Mathematics (New York)*. New York: Springer-Verlag. Stochastic Modelling and Applied Probability.

Nemirovski, A., A. Juditsky, G. Lan, and A. Shapiro. 2009. "Robust Stochastic Approximation Approach to Stochastic Programming". *SIAM Journal on Optimization* 19 (4): 1574–1609.

Nemirovski, A., and D. Yudin. 1983. *Problem Complexity and Method Efficiency in Optimization*. Wiley -Intersci. Ser. Discrete Math 15 John Wiley New York.

Pasupathy, R., P. Glynn, S. Ghosh, and F. Hashemi. 2014. "How much to sample in simulation-based stochastic recursions?". *Under review at SIAM Journal of Optimization*.

Robbins, H., and S. Monro. 1951. "A stochastic approximation method". *Ann. Math. Statistics* 22:400–407.

Yousefian, F., A. Nedić, and U. V. Shanbhag. 2012. "On Stochastic Gradient and Subgradient Methods with adaptive steplength sequences". *Automatica* 48 (1): 56–67.

AUTHOR BIOGRAPHIES

UDAY V. SHANBHAG has been an Associate Professor in Department of Industrial and Manufacturing Engineering at Penn. State University since 2012. From 2006, he was an assistant and subsequently a tenured associate professor at the department of Industrial and Enterprise Systems Engineering at the University of Illinois. Uday holds a Ph.D. in Management Science and Engineering from Stanford University. His research interests lies in stochastic and nonlinear optimization, nonsmooth analysis, and variational inequality problems. Dr. Shanbhag received the triennial A.W. Tucker Prize for his dissertation from the mathematical programming society (MPS) in 2006, the Computational Optimization and Applications (COAP) Best Paper Award in 2007 (with W. Murray), and the best theory paper award at the Winter Simulation Conference (2013) (with F. Yousefian and A. Nedić). His email address is udaybag@psu.edu.

JOSE H. BLANCHET is an Associate Professor in the Departments of IEOR and Statistics at Columbia University. Jose holds a Ph.D. in Management Science and Engineering from Stanford University. Prior to joining Columbia, he was a faculty member in the Department of Statistics at Harvard University. Jose is a recipient of the 2009 Best Publication Award given by the INFORMS Applied Probability Society and of the 2010 Erlang Prize. He also received a PECASE award given by NSF in 2010. He worked as an analyst in Protego Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability and Monte Carlo methods. His email is jose.blanchet@columbia.edu