Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

A STOCHASTIC COMPOSITIONAL GRADIENT METHOD USING MARKOV SAMPLES

Mengdi Wang

Department of Operations Research and Financial Engineering
Princeton University
Sherrerd Hall, Charlton Street
Princeton, NJ 08540 USA

Ji Liu

Department of Computer Science
Computer Studies Building
Rochester University
Rochester, NY 14627 USA

ABSTRACT

Consider the convex optimization problem $\min_x f(g(x))$ where both f and g are unknown but can be estimated through sampling. We consider the stochastic compositional gradient descent method (SCGD) that updates based on random function and subgradient evaluations, which are generated by a conditional sampling oracle. We focus on the case where samples are corrupted with Markov noise. Under certain diminishing stepsize assumptions, we prove that the iterate of SCGD converges almost surely to an optimal solution if such a solution exists. Under specific constant stepsize assumptions, we obtain finite-sample error bounds for the averaged iterates of the algorithm. We illustrate an application to online value evaluation in dynamic programming.

1 INTRODUCTION

Consider the convex optimization problem

$$\min_{x\in X}\{F(x) = f(g(x))\}. \tag{1}$$

where $F:\Re^n \mapsto \Re$ is the objective function, $f:\Re^m \mapsto \Re$ is referred as the *outer function*, $g:\Re^n \mapsto \Re^m$ is referred as the *inner function*, and X is a convex and compact set in $\Re^n$. We assume throughout that the objective function F is convex and has at least one minimal solution in X, whereas f and g can be nonconvex. We also assume that f and g have Lipschitz continuous gradients.

We are interested in the simulation setting where values and gradients of $f(\cdot)$ and $g(\cdot)$ can only be queried from a sampling oracle and their samples are subject to Markov noise. A related stochastic optimization problem is

$$\min_{x\in X}\mathbf{E}\Big[f_w\left(\mathbf{E}\big[g_v(x)\mid w\big]\right)\Big], \tag{2}$$

where f_w and g_v are random realizations of the unknown deterministic functions f and g, respectively. Another related problem is the sample average approximation to problem (2), given by

$$\min_{x\in X} \frac{1}{N}\sum_{i=1}^{N} f_{w_i}\left(\frac{1}{N}\sum_{j=1}^{N} g_{v_{ij}}(x)\right),$$

in which the expectation is replaced by empirical means over some data set

$$\left\{\{w_i, v_{ij}\}_{j=1}^{N}\right\}_{i=1}^{N}.$$

Stochastic composition problems (1)-(2) find wide application in data analytics and operations research. Examples of applications include statistical learning, dynamic programming, estimation of large deviation rate and risk-averse optimization; see (Wang, Fang, and Liu 2016) Sections 4 and 5. Note that the composition of two stochastic functions is also related to three-stage stochastic programming, for which at least $\mathcal{O}(1/\varepsilon^4)$ sample paths are needed to reach an ε-optimal solution; see (Shapiro, Dentcheva, and Ruszczyński 2014) Section 5.8.

For the case where unbiased samples are available, (Wang, Fang, and Liu 2016) has proposed and analyzed a class of *stochastic compositional gradient/subgradient descent methods* (SCGD). The SCGD involves two iterations of different time scales, one for estimating x^* by a stochastic gradient-like iteration, the other for maintaining a running estimate of $g(x^*)$. Almost sure convergence and rate of convergence have been obtained, assuming that all samples are independent and identically distributed.

In this paper, we will focus on the application of SCGD to Markov sampling oracles. We emphasize that SCGD is a simulation-driven method. It finds the optimal solution to problem (1) even if one can only simulate the outer and inner functions. Markov noise is very common in Monte Carlo simulation and simulation of dynamic systems. An example of problem (1) with Markov simulator is the online value evaluation problem which arises from dynamic programming; see Section 5 for more details.

However, Markov noise makes the analysis much more complicated, because samples per iteration now become severely biased. Our analysis involves breaking down the iterate sequence into an infinite number of segments with increasing lengths. In this way, we use the Markov property to control the overall bias associated with iterates within the same segment. This leads to an almost sure convergence result as well as several error bounds and sample complexity results. To the authors' best knowledge, this is the first convergence and rate of convergence result for stochastic composition optimization under a Markov simulator.

Similar to several sources on convergence analysis of stochastic algorithms (see e.g., (Bertsekas and Tsitsiklis 1989), (Kushner and Yin 2003), (Borkar 2008)), we use a supermartingale convergence argument towards a specially constructed sequence. The idea of stochastic gradient is also related to the class of incremental methods, which are developed for minimizing the sum of a large number of component functions. These methods update incrementally by making use of one component at a time, through a gradient-type or proximal-type iteration; see for example, (Nedić and Bertsekas 2001), (Bertsekas 2011), (Nedić 2011), (Wang and Bertsekas 2016) and (Wang, Chen, Liu, and Gu 2015). The SCGD method considered in this paper also applies to the incremental problem $\min_x f\left(\sum_{i=1}^{M} g_i(x)\right)$ where component functions g_i are sampled according to a Markov chain. The idea of using two timescales existed in literature of stochastic approximation; see for examples (Borkar 1997), (Bhatnagar and Borkar 1998), (Konda and Tsitsikilis 2004). It is also related to the quasi-gradient methods that have been extensively studied by (Ermoliev 1976). However, there has been little analysis on the convergence rate and sample complexity, especially for the Markov case.

Notation For $x \in \Re^n$, we denote by x' its transpose, and by $\|x\|$ its Euclidean norm (i.e., $\|x\| = \sqrt{x'x}$). The abbreviation "$\xrightarrow{a.s.}$" means "converges almost surely to," while the abbreviation "i.i.d." means "independent identically distributed." For a function $f(x)$, we denote by $\nabla f(x)$ its gradient at x if f is

differentiable. For two sequences $\{y_k\}$ and $\{z_k\}$, we write $y_k = \mathcal{O}(z_k)$ if there exists a uniform constant $c > 0$ such that $\|y_k\| \leq c\|z_k\|$ for each k with probability 1. For simplicity, we will use the $\mathcal{O}(\cdot)$ notation frequently to avoid defining too many constants. It does not affect the convergence and asymptotic rate of convergence results.

2 SIMULATION ORACLE, ALGORITHM, AND ASSUMPTIONS

Suppose that we have access to a **Conditional Sampling Oracle** (*CSO*) such that:

- Given some $y \in \Re^m$, the *CSO* returns a noisy gradient $\nabla f(y) + w \in \Re^m$.
- Given some $x \in X$ and conditioned on w, the *CSO* returns $g(x) + v \in \Re^n$ and $\nabla g(x) + \tilde{v} \in \Re^{n\times m}$.

We focus on the case where the sample errors $v, \tilde{v}, w$ are Markov random variables with a zero-mean invariant distribution.

For solution of the stochastic program (1), we use the *stochastic compositional gradient* (SCGD) method proposed by (Wang, Fang, and Liu 2016), taking the form of Algorithm 1. Note that $x_k \in \Re^n$, $y_k \in \Re^m$, $v_k, w_k \in \Re^m$, $\tilde{v}_k \in \Re^{n\times m}$, ∇g is the $n \times m$ matrix with each column being a gradient of the corresponding entry of g, $\{\alpha_k\}$, $\{\beta_k\}$ are sequences of positive scalars in $(0,1)$, and Π_X denotes the orthogonal projection onto X with respect to the Euclidean norm $\|\cdot\|$.

Algorithm 1 Stochastic Compositional Gradient Descent

Input: $x_0 \in \Re^n$, $y_0 \in \Re^m$, *CSO*, number of queries T, positive stepsizes $\{\alpha_k\}, \{\beta_k\} \subset (0,1)$.

1: **for** $k = 0, 1, \ldots, T$ **do**
2: Query *CSO* for the sample values of g at x_k, obtaining $g(x_k) + v_k$ and $\widetilde{\nabla} g(x_k) + \tilde{v}_k$.
3: Update

$$y_{k+1} = (1 - \beta_k) y_k + \beta_k (g(x_k) + v_k). \tag{3}$$

4: Query *CSO* for the sample gradient of f at y_{k+1}, obtaining $\nabla f(y_{k+1}) + w_k$.
5: Update

$$x_{k+1} = \Pi_X \big\{ x_k - \alpha_k (\nabla g(x_k) + \tilde{v}_k)(\nabla f(y_{k+1}) + w_k) \big\}. \tag{4}$$

6: **end for**

Output: The averaged iterate $\frac{1}{T}\sum_{k=1}^{T} x_k$.

In the case where f is a linear function and $w \equiv 0$, Algorithm 1 is equivalent to the classical stochastic gradient/approximation method. In the case where f is nonlinear, the auxiliary variable y_k plays the important role of "tracking" the unknown value $g(x_k)$, so that the iteration for x_k behaves like a gradient descent update.

Let us denote by $\mathscr{F}_k$ the collection of random variables

$$\{x_0, \ldots, x_k, y_0, \ldots, y_k, (w_0, v_0, \tilde{v}_0), \ldots, (w_{k-1}, v_{k-1}, \tilde{v}_{k-1})\}.$$

For simplicity of analysis, we make the following assumptions regarding the structure of problem (1) and the Markov property of the random process $\{(w_k, \tilde{v}_k, v_k)\}$.

Assumption 1 (Boundedness and Continuity)

(i) The constraint set X is compact and convex.
(ii) The function g has bounded and Lipschitz continuous gradient over X.
(iii) The function f has bounded and Lipschitz continuous gradient over the set $Y = \{g(x) \mid x \in X\}$.

Note that Assumption 1 requires the existence of Lipschitz constants without specifying their values. These unspecified constants should play a role in the convergence rate of stochastic algorithms. However, the focus of the current paper is the sample complexity, i.e., how the convergence rate relates to the number of oracle queries. For simplicity of analysis, we will omit these constants in the big O notation. For the same reason, we assume boundedness of the constraint set as well as the gradients.

Assumption 2 (Markov Noise in the Conditional Simulation Oracle)

(i) The random variables $\{(w_0, v_0, \tilde{v}_0), (w_1, v_1, \tilde{v}_1), (w_2, v_2, \tilde{v}_2), \ldots\}$ are uniformly bounded with probability 1.

(ii) There exists a scalar $\rho \in (0,1)$ such that with probability 1,

$$\left|\mathbf{E}\left[(w_t, v_t, \tilde{v}_t) \mid \mathscr{F}_s\right]\right| \leq \mathscr{O}(\rho^{t-s+1}), \qquad \left|\mathbf{E}\left[\tilde{v}_t w_t \mid \mathscr{F}_s\right]\right| \leq \mathscr{O}(\rho^{t-s+1}), \qquad \forall\ 0 < s < t.$$

Assumption 2(ii) is critical to our analysis. The first part $\left|\mathbf{E}\left[(w_t, v_t, \tilde{v}_t) \mid \mathscr{F}_s\right]\right| \leq \mathscr{O}(\rho^{t-s+1})$ requires that the additive sampling error $(w_t, v_t, \tilde{v}_t)$ becomes asymptotically zero-mean, whose conditional bias decreases to zero at a geometric speed. This is a typical property of Markov chain sampling oracle. The second part $\left|\mathbf{E}\left[\tilde{v}_t w_t \mid \mathscr{F}_s\right]\right| \leq \mathscr{O}(\rho^{t-s+1})$ requires the sample errors of the outer and inner functions be asymptotically uncorrelated. This is satisfied when the inner samples are simulated conditioned on the outer sample while being corrupted with Markov noise.

In contrast to the prior work (Wang, Fang, and Liu 2016) which assumes independent samples, Assumption 2 allows the samples to be corrupted with Markov noise instead of i.i.d. zero-mean noise. For simplicity of analysis, we assume that sample errors are uniformly bounded instead of having bounded second moments.

Under Assumptions 1 and 2, we will show in Theorem 1 that the SCGD algorithm converges almost surely to an optimal solution, in the case where certain diminishing stepsizes are used. In Theorem 2 and Theorem 3, we will provide upper bounds on the sub-optimality of the averaged iterates of SCGD, in the case where constant stepsizes are used.

3 MAIN RESULTS

In this section, we give the main results on the convergence and rate of convergence for SCGD under Markov noise. Our first result states that the SCGD algorithm converges to an optimal solution with probability 1, as long as the two stepsize sequences diminish to zero at favorable rates.

Theorem 1 (Almost Sure Convergence) Let Assumptions 1-2 hold. Let the stepsizes $\{\alpha_k\}$ and $\{\beta_k\}$ be decreasing positive scalars such that

$$\sum_{k=0}^{\infty} \min\{\alpha_k, \beta_k\} = \infty, \qquad \sum_{k=0}^{\infty} \left(\alpha_k^2 + \beta_k^2 + \frac{\alpha_k^2}{\beta_k}\right) < \infty.$$

In addition, let there be a sequence of increasing integers $\{N_0, N_1, \ldots\}$ such that

$$\sum_{t=0}^{\infty} \left(\alpha_{N_t} + \beta_{N_t} + \left(\sum_{k=N_t}^{N_{t+1}-1} (\alpha_k + \beta_k)\right)^2\right) < \infty.$$

Then the SCGD Algorithm 1 generates a sequence $\{(x_k, y_k)\}$ such that $\|y_{k+1} - g(x_k)\| \xrightarrow{a.s.} 0$ and x_k converges almost surely to an optimal solution of problem (1).

In Theorem 1, a key requirement is $\sum_{k=0}^{\infty} \left(\frac{\alpha_k^2}{\beta_k}\right) < \infty$, which implies that $\alpha_k \ll \beta_k$ for k sufficiently large. In other words, the algorithm produces consistent estimates of the optimal solution only if the x-step updates conservatively while the auxillary y-step updates aggressively fast.

For an example, the stepsizes $\alpha_k = k^{-1}$ and $\beta_k = k^{-3/4}$ satisfy all the conditions required by Theorem 1 if we choose $N_t = t^{5/3}$. The main proof idea is to construct a special sequence of merit functions based on the subsequence $\{X_{N_t}\}_{t=1}^{\infty}$. Since N_t increases superlinearly fast, the overall sample error incurred between the N_t-th and N_{t+1}-th iterates become less and less biased as $t \to \infty$. The formal proof is deferred to Section 4.

Our second result concerns the expected optimality error when constant stepsizes are used. The analysis follows from that of Theorem 1.

Theorem 2 (Constant Stepsize Error Bound) Let Assumptions 1-2 hold, and let the stepsizes be constant scalars

$$\alpha_k = \alpha, \qquad \beta_k = \beta, \qquad k = 1, \ldots, T,$$

where $\alpha, \beta \in (0,1)$. Then the averaged iterate generated by Algorithm 1 using T oracle queries satisfies

$$\mathbf{E}\left[F\left(\frac{1}{T}\sum_{t=0}^{T} x_t\right) - F^*\right] \leq \mathscr{O}\left(\frac{1}{T\alpha} + \alpha + \frac{\beta^2}{\alpha} + \frac{\alpha}{\beta} + \sqrt{\frac{\rho}{1-\rho}}\frac{(\alpha+\beta)^{3/2}}{\alpha}\right).$$

Suppose that the total number of oracle queries T is known in advance. Theorem 2 allows us to pick values of the constant stepsizes α, β in order to optimize the error bound. This lead to the following sample-error complexity result.

Theorem 3 (Sample-Error Complexity) Let Assumptions 1-2 hold, and let the stepsizes be constant scalars

$$\alpha_k = \frac{1}{T^{5/6}}, \qquad \beta_k = \frac{1}{T^{2/3}}, \qquad k = 1, \ldots, T.$$

Then the averaged iterate generated by Algorithm 1 using T oracle queries satisfies

$$\mathbf{E}\left[F\left(\frac{1}{T}\sum_{t=0}^{T} x_t\right) - F^*\right] \leq \mathscr{O}\left(\left(1 + \sqrt{\frac{\rho}{1-\rho}}\right) \cdot \frac{1}{T^{1/6}}\right).$$

In Theorem 3, the stepsizes are chosen in a way such that the asymptotic error bound is minimized. Note that such stepsizes do not satisfy the conditions required by Theorem 1 for almost sure convergence. To the authors' best knowledge, one cannot get the best sampler-error complexity and almost sure convergence *simultaneously*. This illustrates a tradeoff between pathwise convergence and minimal expected error.

When the noises are independent ($\rho = 0$), the simulation oracle generates unbiased sample at every query. In this case, the error bound can be improved to $\mathscr{O}(1/T^{1/4})$ with $\alpha = 1/T^{3/4}$ and $\beta = 1/\sqrt{T}$, which matches the error bound obtained in the earlier work (Wang, Fang, and Liu 2016).

When the noises are Markov ($\rho > 0$), the error-sample complexity deterioriates from $\mathscr{O}(1/T^{1/4})$ to $\mathscr{O}\left(\left(1 + \sqrt{\frac{\rho}{1-\rho}}\right) \cdot \frac{1}{T^{1/6}}\right)$. We conjecture that the result of Theorem 3 can be further improved to achieve $\mathscr{O}(1/T^{1/4})$. This remains an open question for future research.

4 PROOF OF CONVERGENCE

In this section, we develop the almost sure convergence and convergence rate results step by step. The key to the Markov-noise analysis which differs from the classical analysis is to divide the sequence of iterates $\{x_k\}_{k=0}^{\infty}$ into a sequence of increasingly long segments

$$\left\{(x_{N_0}, \ldots, x_{N_1-1}), \ldots, (x_{N_t}, \ldots, x_{N_{t+1}-1}), \ldots\right\},$$

where the segment length $N_{t+1} - N_t$ increases to infinity as $t \to \infty$. As long as the segments are properly constructed, we can use the Markov property of the *CSO* to show that the overall error incurred within the t-th segment $(x_{N_t}, \ldots, x_{N_{t+1}-1})$ is increasingly close to be zero-mean, as $t \to \infty$. This allows us to apply a coupled supermartingale analysis to a specifically constructed merit function and prove the convergence.

4.1 Preliminaries

The main idea of the convergence analysis is that the two sequences $\{x_k - x^*\}$ and $\{g(x_k) - y_{k+1}\}$ are coupled in their asymptotic behaviors and they converge together to zero. Our proof will use the following coupled supermartingale convergence lemma by Robbins and Siegmund (Robbins and Siegmund 1971) .

Lemma 1 (Supermartingale Convergence (Robbins and Siegmund 1971)) Let $\{z_k\}$, $\{u_k\}$, $\{a_k\}$ and $\{b_k\}$ be sequences of nonnegative random variables so that

$$\mathbf{E}[z_{k+1} \mid \mathscr{G}_k] \leq (1+a_k)z_k - u_k + b_k, \qquad \text{for all } \ k \geq 0 \ \ w.p.1,$$

where $\mathscr{G}_k$ denotes the collection $z_0, \ldots, z_k, u_0, \ldots, u_k, a_0, \ldots, a_k, b_0, \ldots, b_k$. Also, let $\sum_{k=0}^{\infty}(a_k + b_k) < \infty$ with probability 1. Then z_k converges almost surely to a random variable and $\sum_{k=0}^{\infty} u_k < \infty$ with probability 1.

Lemma 2 (Basic Facts) Let Assumptions 1-2 hold. Then with probability 1 for any $t, k > 0$,

(a) $\|x_{k+1} - x_k\| \leq \mathscr{O}(\alpha_k), \quad \|x_{k+t} - x_k\| \leq \sum_{i=k}^{k+t-1} \mathscr{O}(\alpha_i)$.
(b) $\|y_{k+1} - y_k\| \leq \mathscr{O}(\beta_k), \quad \|y_{k+t} - y_k\| \leq \sum_{i=k}^{k+t-1} \mathscr{O}(\beta_i)$.
(c) For all $y_1, y_2 \in Y$, $\|g(y_1) - g(y_2)\| \leq \mathscr{O}(\|y_1 - y_2\|)$.

Proof. The proof directly follows from Assumptions 1-2 and triangle/matrix norm inequalities. ■

Let us analyze the iteration for x_k. We will show that it behaves in a way similar to a gradient descent step.

Lemma 3 (Contraction of the x-Iteration) Let Assumptions 1-2 hold, let x^* be an arbitrary optimal solution of problem (1), and let $F^* = F(x^*)$. Then with probability 1, for any $k > 0$

$$\begin{aligned}\mathbf{E}\left[\|x_{k+1} - x^*\|^2 \mid \mathscr{F}_k\right] \leq \|x_k - x^*\|^2 - 2\alpha_k(F(x_k) - F^*) + \mathscr{O}\left(\alpha_k^2 + \frac{\alpha_k^2}{\beta_k}\right) \\ + \alpha_k \mathbf{E}[L_1(x_k, \tilde{v}_k, w_k) \mid \mathscr{F}_k] + \beta_k \mathbf{E}\left[\|g(x_k) - y_{k+1}\|^2 \mid \mathscr{F}_k\right],\end{aligned} \tag{5}$$

where L_1 is a function given by $L_1(x_k, \tilde{v}_k, w_k) = -2(x_k - x^*)'\tilde{v}_k(\nabla f(g(x_k)) + w_k) - 2(x_k - x^*)'\nabla g(x_k) w_k$.

Proof. By using the definition of x_k [cf. Eq. (4)], the nonexpansiveness of Π_X, and the fact $x^* \in X$, we have

$$\begin{aligned}\|x_{k+1} - x^*\|^2 &\leq \|x_k - x^* - \alpha_k(\nabla g(x_k) + \tilde{v}_k)(\nabla f(y_{k+1}) + w_k)\|^2 \\ &= \|x_k - x^*\|^2 + \alpha_k^2\|(\nabla g(x_k) + \tilde{v}_k)(\nabla f(y_{k+1}) + w_k)\|^2 - 2\alpha_k(x_k - x^*)'(\nabla g(x_k) + \tilde{v}_k)(\nabla f(y_{k+1}) + w_k) \\ &= \|x_k - x^*\|^2 + \alpha_k^2\|(\nabla g(x_k) + \tilde{v}_k)(\nabla f(y_{k+1}) + w_k)\|^2 - 2\alpha_k(x_k - x^*)'\nabla g(x_k)\nabla f(g(x_k)) \\ &\quad - 2\alpha_k(x_k - x^*)'\tilde{v}_k(\nabla f(g(x_k)) + w_k) - 2\alpha_k(x_k - x^*)'\nabla g(x_k) w_k \\ &\quad + 2\alpha_k(x_k - x^*)'(\nabla g(x_k) + \tilde{v}_k)(\nabla f(g(x_k)) - \nabla f(y_{k+1})) \\ &= \|x_k - x^*\|^2 + \alpha_k^2\|(\nabla g(x_k) + \tilde{v}_k)(\nabla f(y_{k+1}) + w_k)\|^2 - 2\alpha_k(x_k - x^*)'\nabla g(x_k)\nabla f(g(x_k)) \\ &\quad + \alpha_k L_1(x_k, \tilde{v}_k, w_k) + u_k,\end{aligned} \tag{6}$$

where we define u_k to be

$$u_k = 2\alpha_k(x_k - x^*)'(\nabla g(x_k) + \tilde{v}_k)(\nabla f(g(x_k)) - \nabla f(y_{k+1})).$$

By using matrix norm inequalities, the Lipschitz continuity of ∇f, and the boundedness of x_k, $\nabla g(x_k)$, and $\tilde{v}_k$, we obtain

$$\begin{aligned} u_k &\leq 2\alpha_k \|x_k - x^*\| \|\nabla g(x_k) + \tilde{v}_k\| \|\nabla f(g(x_k)) - \nabla f(y_{k+1})\| \\ &\leq \mathscr{O}(\alpha_k) \|g(x_k) - y_{k+1}\| \\ &\leq \beta_k \|g(x_k) - y_{k+1}\|^2 + \mathscr{O}\left(\frac{\alpha_k^2}{\beta_k}\right). \end{aligned}$$

By the convexity of $F = f \circ g$ and the chain rule of differential, we obtain $\nabla g(x_k) \nabla f(g(x_k)) = \nabla F(x_k)$, therefore

$$(x_k - x^*)' \nabla g(x_k) \nabla f(g(x_k)) \geq F(x_k) - F^*.$$

Taking expectation on both sides of Eq. (6) and applying the preceding inequalities, we obtain

$$\begin{aligned} \mathbf{E}\left[\|x_{k+1} - x^*\|^2 \mid \mathscr{F}_k\right] \leq \|x_k - x^*\|^2 &+ \mathscr{O}\left(\alpha_k^2 + \frac{\alpha_k^2}{\beta_k}\right) - 2\alpha_k \left(F(x_k) - F^*\right) \\ &+ \alpha_k \mathbf{E}\left[L_1(x_k, \tilde{v}_k, w_k) \mid \mathscr{F}_k\right] \\ &+ \beta_k \mathbf{E}\left[\|g(x_k) - y_{k+1}\|^2 \mid \mathscr{F}_k\right]. \end{aligned}$$

By using Assumptions 1 and 2, we can verify that $L_1(x, \tilde{v}, w)$ is Lipschitz continuous in $x \in X$ and is bilinear in $(\tilde{v}, w)$. Thus we have completed the proof. ■

Let us we analyze the behavior of y_k, which is obtained by averaging past samples $\{g(x_t) + v_t\}_{t=0}^k$ as an approximation to the unknown quantity $g(x_k)$. The next lemma shows that the approximation error $\|y_{k+1} - g(x_k)\|$ decreases "on average" according to a supermartingle-type inequality.

Lemma 4 (Contraction of the y-Iteration) Let Assumptions 1-2 hold. Then with probability 1, for all k

$$\begin{aligned} \mathbf{E}\left[\|y_{k+1} - g(x_k)\|^2 \mid \mathscr{F}_k\right] \leq (1-\beta_k)\|y_k - g(x_{k-1})\|^2 &+ \mathscr{O}\left(\frac{\alpha_k^2}{\beta_k} + \alpha_k^2 + \beta_k^2\right) \\ &+ \beta_k \mathbf{E}\left[L_2\Big((x_{k-1}, y_k), v_k\Big) \mid \mathscr{F}_k\right]. \end{aligned} \tag{7}$$

where L_2 is a function given by $L_2\Big((x,y),v\Big) = (y - g(x))'v$.

Proof. By using the definition of y_k [cf. Eq. (3)], we have

$$y_{k+1} - g(x_k) + e_k = (1-\beta_k)(y_k - g(x_{k-1})) + \beta_k v_k, \tag{8}$$

where we define $e_k = (1-\beta_k)(g(x_k) - g(x_{k-1}))$. By using Assumption 1 we obtain

$$\|e_k\| \leq \mathscr{O}\left(\|x_k - x_{k-1}\|\right) \leq \mathscr{O}\left(\alpha_k\right). \tag{9}$$

Taking squared norm expectation on both sides of Eq. (8) and using Assumptions 1-2, we have

$$\begin{aligned} &\mathbf{E}\left[\|y_{k+1} - g(x_k) + e_k\|^2 \mid \mathscr{F}_k\right] \\ &= (1-\beta_k)^2 \|y_k - g(x_{k-1})\|^2 + \beta_k^2 \mathbf{E}\left[\|v_k\|^2 \mid \mathscr{F}_k\right] + 2(1-\beta_k)\beta_k \mathbf{E}\left[(y_k - g(x_{k-1}))' v_k \mid \mathscr{F}_k\right] \\ &\leq (1-\beta_k)^2 \|y_k - g(x_{k-1})\|^2 + \mathscr{O}(\beta_k^2) + 2(1-\beta_k)\beta_k \mathbf{E}\left[(y_k - g(x_{k-1}))' v_k \mid \mathscr{F}_k\right]. \end{aligned} \tag{10}$$

By using the basic inequality $\|a+b\|^2 \leq (1+\varepsilon)\|a\|^2 + (1+1/\varepsilon)\|b\|^2$ for any $\varepsilon > 0$, we have

$$\|y_{k+1} - g(x_k)\|^2 \leq (1+\beta_k)\|y_{k+1} - g(x_k) + e_k\|^2 + (1+1/\beta_k)\|e_k\|^2.$$

Taking expectation on both sides and applying Eqs. (9)-(10), we obtain that

$$\mathbf{E}\left[\|y_{k+1}-g(x_k)\|^2 \mid \mathscr{F}_k\right] \leq (1-\beta_k)\|y_k-g(x_{k-1})\|^2 + \mathscr{O}\left(\frac{a_k^2}{\beta_k}+\alpha_k^2+\beta_k^2\right) + 2\beta_k \mathbf{E}\left[(y_k-g(x_{k-1}))'v_k \mid \mathscr{F}_k\right],$$

for all k with probability 1. By letting $L_2\Big((x,y),v\Big) = 2(y-g(x))'v$, we obtain Eq. (7). It can be easily seen that L_2 is Lipschitz continuous in (x,y) and linear in v. ■

4.2 Almost Sure Convergence

Next we develop our first main result. Its proof idea is to combine the results for the x-iteration and y-iteration (Lemmas 3 and 4) and to make use of the Markov property (Assumption 2(ii)).

Proof of Theorem 1. Define the random variable

$$J_k = \|x_k - x^*\|^2 + 2\|y_k - g(x_{k-1})\|^2.$$

We multiply Eq. (7) with 2 and take its sum with Eq. (5), and obtain

$$\begin{aligned}\mathbf{E}[J_{k+1} \mid \mathscr{F}_k] \leq J_k - 2\alpha_k(F(x_k)-F^*) - \beta_k \mathbf{E}\left[\|y_{k+1}-g(x_k)\|^2 \mid \mathscr{F}_k\right] + \varepsilon_k \\ + \alpha_k \mathbf{E}[L_1(x_k,\tilde{v}_k,w_k) \mid \mathscr{F}_k] + 2\beta_k \mathbf{E}[L_2((x_{k-1},y_k),v_k) \mid \mathscr{F}_k].\end{aligned} \tag{11}$$

where we define ε_k to be the deterministic sequence given by

$$\varepsilon_k = \mathscr{O}\left(\alpha_k^2+\beta_k^2+\frac{\alpha_k^2}{\beta_k}\right),$$

and L_1, L_2 are functions defined in Lemmas 3 and 4 respectively.

We take an *arbitrary* sequence of increasing integers $\{N_t\}$. Let t,k be positive integers such that $N_t \leq k < N_{t+1}$. Recall that $L_1(x_k,\tilde{v}_k,w_k) = -2(x_k-x^*)'\tilde{v}_k(\nabla f(g(x_k))+w_k) - 2(x_k-x^*)'\nabla g(x_k)w_k$. By using the Lipschitz continuity and bilinearity of $L_1(x,\tilde{v},w)$ in x and $(w,\tilde{v})$ respectively, and by using Assumption 2(ii), we obtain

$$\begin{aligned}\mathbf{E}[L_1(x_k,\tilde{v}_k,w_k) \mid \mathscr{F}_{N_t}] &= \mathbf{E}[L_1(x_{N_t},\tilde{v}_k,w_k) \mid \mathscr{F}_{N_t}] + \mathbf{E}[(L_1(x_k,\tilde{v}_k,w_k) - L_1(x_{N_t},\tilde{v}_k,w_k) \mid \mathscr{F}_{N_t}] \\ &\leq \mathscr{O}\left(|\mathbf{E}[w_k \mid \mathscr{F}_{N_t}]| + |\mathbf{E}[\tilde{v}_k \mid \mathscr{F}_{N_t}]| + |\mathbf{E}[\tilde{v}_k w_k \mid \mathscr{F}_{N_t}]|\right) + \mathscr{O}\left(\mathbf{E}[\|x_k - x_{N_t}\| \mid \mathscr{F}_{N_t}]\right) \\ &\leq \mathscr{O}\left(\rho^{k-N_t+1} + \sum_{i=N_t}^{k} \alpha_i\right).\end{aligned}$$

Similarly, we also have

$$\begin{aligned}\mathbf{E}[L_2(x_{k-1},y_k),v_k \mid \mathscr{F}_{N_t}] &= \mathbf{E}[L_2(x_{N_t-1},y_{N_t}),v_k) \mid \mathscr{F}_{N_t}] + \mathbf{E}[L_2(x_{k-1},y_k),v_k) - L_2(x_{N_t-1},y_{N_t}),v_k) \mid \mathscr{F}_{N_t}] \\ &\leq \mathscr{O}\left(|\mathbf{E}[v_k \mid \mathscr{F}_{N_t}]|\right) + \mathscr{O}\left(\mathbf{E}[\|x_{N_t-1}-x_{k-1}\| + \|y_k - y_{N_t}\| \mid \mathscr{F}_{N_t}]\right) \\ &\leq \mathscr{O}\left(\rho^{k-N_t+1} + \sum_{i=N_t}^{k}(\alpha_i+\beta_i)\right).\end{aligned}$$

We also have

$$-\mathbf{E}[F(x_k)-F^* \mid \mathscr{F}_{N_t}] \leq -(F(x_{N_t})-F^*) + \mathscr{O}\left(\mathbf{E}[\|x_k - x_{N_t}\| \mid \mathscr{F}_{N_t}]\right) \leq -(F(x_{N_t})-F^*) + \sum_{i=N_t}^{k} \mathscr{O}(\alpha_i),$$

and

$$-\mathbf{E}\left[\|y_k - g(x_{k-1})\|^2 \mid \mathscr{F}_{N_t}\right] \le -\|y_{N_t} - g(x_{N_t-1})\|^2 + \mathscr{O}\left(\mathbf{E}\left[\|x_{N_t-1} - x_{k-1}\| + \|y_k - y_{N_t}\| \mid \mathscr{F}_{N_t}\right]\right)$$
$$\le -\|y_{N_t} - g(x_{N_t-1})\|^2 + \sum_{i=N_t}^{k} \mathscr{O}(\alpha_i + \beta_i).$$

Taking expectation on both sides of (13) conditioned on $\mathscr{F}_{N_t}$ and applying the preceding relations, we obtain

$$\mathbf{E}\left[J_{k+1} \mid \mathscr{F}_{N_t}\right] \le \mathbf{E}\left[J_k \mid \mathscr{F}_{N_t}\right] + \varepsilon_k - 2\alpha_k\left(F(x_k) - F^*\right) - \beta_k \mathbf{E}\left[\|y_{k+1} - g(x_k)\|^2 \mid \mathscr{F}_{N_t}\right]$$
$$+ \mathscr{O}\left(\rho^{k-N_t+1}(\alpha_k + \beta_k) + (\alpha_k + \beta_k)\sum_{i=N_t}^{N_{t+1}-1}(\alpha_i + \beta_i)\right). \tag{12}$$

Applying the preceding relations inductively for $k = N_t, \ldots, N_{t+1} - 1$ and using the facts $\sum_{k=N_t}^{N_{t+1}-1} \rho^{k-N_t+1} \le \frac{\rho}{1-\rho}$, $\alpha_k \downarrow 0$, $\beta_k \downarrow 0$ we obtain

$$\mathbf{E}\left[J_{N_{t+1}} \mid \mathscr{F}_{N_t}\right] \le J_{N_t} + \sum_{k=N_t}^{N_{t+1}-1} \varepsilon_k - 2\sum_{k=N_t}^{N_{t+1}-1} \alpha_k\left(F(x_{N_t}) - F^*\right) - \sum_{k=N_t}^{N_{t+1}-1} \beta_k \mathbf{E}\left[\|y_{N_t+1} - g(x_{N_t})\|^2 \mid \mathscr{F}_{N_t}\right]$$
$$+ \mathscr{O}\left(\frac{\rho}{1-\rho}(\alpha_{N_t} + \beta_{N_t}) + \left(\sum_{k=N_t}^{N_{t+1}-1}(\alpha_k + \beta_k)\right)^2\right). \tag{13}$$

We note that the stepsize assumptions imply that $\sum_{t=0}^{\infty}\sum_{k=N_t}^{N_{t+1}-1} \varepsilon_k \le \sum_{k=0}^{\infty} \mathscr{O}(\alpha_k^2 + \beta_k^2 + \alpha_k^2/\beta_k) < \infty$, and that $\sum_{t=0}^{\infty}\left(\alpha_{N_t} + \beta_{N_t} + \left(\sum_{k=N_t}^{N_{t+1}-1}(\alpha_k + \beta_k)\right)^2\right) < \infty$. These together with the fact $F(x_k) - F^* \ge 0$ suggest that the supermartingale convergence Lemma 1 applies to Eq. (13).

By applying Lemma 1 to Eq. (13), we obtain that J_{N_t} converges almost surely to a random variables as $t \to \infty$, and *w.p.*1,

$$\sum_{t=0}^{\infty}\sum_{k=N_t}^{N_{t+1}-1}\left(\alpha_k\left(F(x_{N_t}) - F^*\right) + \beta_k \mathbf{E}\left[\|y_{N_t+1} - g(x_{N_t})\|^2 \mid \mathscr{F}_{N_t}\right]\right) < \infty.$$

This together with the stepsize assumption $\sum_{k=0}^{\infty} \min\{\alpha_k, \beta_k\} = \infty$ further implies that *w.p.*1,

$$\liminf_{t\to\infty}\left(F(x_{N_t}) - F^* + \mathbf{E}\left[\|y_{N_t+1} - g(x_{N_t})\|^2 \mid \mathscr{F}_{N_t}\right]\right) = 0.$$

Note that the sequence $\{x_{N_t}\}$ is bounded with probability 1. Consider an *arbitrary sample trajectory* of $\{(x_{N_t}, y_{N_t})\}$ such that the corresponding J_{N_t} converges. By the continuity of F, the sequence $\{(x_{N_t}, \mathbf{E}\left[\|y_{N_t+1} - g(x_{N_t})\|^2 \mid \mathscr{F}_{N_t}\right])\}$ must have a limit point $(\bar{x}, 0)$ with $\bar{x}$ being an optimal solution, i.e., $F(\bar{x}) = F^*$. Also we have $\|y_{N_t} - g(x_{N_t-1})\| \le \mathbf{E}\left[\|y_{N_t+1} - g(x_{N_t})\| \mid \mathscr{F}_{N_t}\right] + \mathscr{O}(\alpha_k + \beta_k) \to 0$. Since the choice of x^* is arbitrary, we take $x^* = \bar{x}$. On this sample trajectory, we have shown that $J_{N_t} = \|x_{N_t} - \bar{x}\|^2 + 2\|y_{N_t} - g(x_{N_t-1})\|^2 \to 0$ and $x_{N_t} \to \bar{x}$. Therefore x_{N_t} converges almost surely to a random point in the set of optimal solutions of problem (1) and $\|y_{N_t} - g(x_{N_t-1})\| \xrightarrow{a.s.} 0$, as $t \to \infty$.

Finally, since $\sum_{t=0}^{\infty}\left(\sum_{k=N_t}^{N_{t+1}-1}(\alpha_k + \beta_k)\right)^2 < \infty$, we have $\sum_{k=N_t}^{N_{t+1}-1} \alpha_k \le \sum_{k=N_t}^{N_{t+1}-1}(\alpha_k + \beta_k) \to 0$. We have, as $k \to \infty$,

$$\|x_k - x_{N_{t^*(k)}}\| \le \sum_{k=N_t}^{N_{t+1}-1} \mathscr{O}(\alpha_k) \to 0, \qquad \text{where } t^* = \max_t\{N_t \le k\}.$$

Therefore x_k converges almost surely to an optimal solution. Similarly $\|y_{k+1} - g(x_k)\| \xrightarrow{a.s.} 0$. ■

4.3 Constant Stepsize Error Bounds

Following the line of analysis of Theorem 1, we continue to prove the rate of convergence results.

Proof of Theorem 2 and Theorem 3. For any increasing integers $\{N_t\}$, we take expectation on both sides of Eq. (12) and apply it inductively, yielding

$$\begin{aligned}\mathbf{E}\left[J_{N_{t+1}}\right] \leq \mathbf{E}\left[J_{N_t}\right] - 2\sum_{k=N_t}^{N_{t+1}-1} \alpha_k \mathbf{E}\left[F(x_k) - F^*\right] \\ + \sum_{k=N_t}^{N_{t+1}-1} \mathscr{O}\left(\alpha_k^2 + \beta_k^2 + \frac{\alpha_k^2}{\beta_k}\right) + \mathscr{O}\left(\rho \frac{\alpha_{N_t} + \beta_{N_t}}{1-\rho} + \left(\sum_{k=N_t}^{N_{t+1}-1} (\alpha_k + \beta_k)\right)^2\right).\end{aligned} \tag{14}$$

We choose $N_t = Mt$ for all t, where M is an arbitrary positive integer. Applying Eq. (14) repeatedly with $\alpha_k = \alpha$ and $\beta_k = \beta$ yields

$$\begin{aligned}2\sum_{t=0}^{k} \mathbf{E}\left[F(x_t) - F^*\right] &\leq 2\sum_{t=0}^{\lceil k/M \rceil M} \mathbf{E}\left[F(x_t) - F^*\right] \\ &\leq \frac{\mathbf{E}\left[J_0\right]}{\alpha} + \mathscr{O}\left(\lceil k/M \rceil M\right)\left(\alpha + \frac{\beta^2}{\alpha} + \frac{\alpha}{\beta}\right) \\ &\quad + \frac{\mathscr{O}\left(\lceil k/M \rceil M\right)}{\alpha M} \frac{\rho(\alpha+\beta)}{1-\rho} + \frac{\mathscr{O}\left(\lceil k/M \rceil M\right)}{\alpha M} M^2(\alpha^2 + \beta^2).\end{aligned}$$

Minimizing the preceding upper bound over M for k sufficiently large, we take $M = \left(\frac{\rho(\alpha+\beta)}{(1-\rho)(\alpha^2+\beta^2)}\right)^{1/2}$ and obtain

$$2\sum_{t=0}^{k} \mathbf{E}\left[F(x_t) - F^*\right] \leq \frac{\mathbf{E}\left[J_0\right]}{\alpha} + \mathscr{O}(k)\left(\alpha + \frac{\beta^2}{\alpha} + \frac{\alpha}{\beta}\right) + \mathscr{O}(k)\sqrt{\frac{\rho}{1-\rho}}(\alpha+\beta)^{3/2}/\alpha.$$

By the convexity of F, we have $\mathbf{E}\left[F\left(\frac{1}{k}\sum_{t=0}^{k} x_t\right) - F^*\right] \leq \frac{1}{k}\sum_{t=1}^{k} \mathbf{E}\left[F(x_t) - F^*\right]$ and complete the proof of Theorem 2. By applying the specific stepsizes $\alpha = \frac{1}{T^{5/6}}, \beta = \frac{1}{T^{2/3}}$, the results of Theorem 3 follow immediately. ■

5 APPLICATION IN DYNAMIC PROGRAMMING: ONLINE BELLMAN ERROR MINIMIZATION

Optimization involving compositions of expected-value functions is very common. As an example of using stochastic optimization to solve dynamic programming, let us consider the Markov decision problem (MDP) with states $i = 1, \ldots, n$. Finding an optimal policy for the MDP can be equivalently casted into solving the fixed-point *Bellman equation*, i.e., finding J such that

$$J = \max_{a \in A}\{g_a + P_a J\}, \tag{15}$$

where $J \in \Re^n$ is the optimal value-per-state vector (also known as value function), $a \in A$ is an action, $g_a \in \Re^n$ is the transition reward-per-state vector given action a, $P_a \in \Re^{n \times n}$ is the matrix of transition probabilities given action a, and the maximization is elementwise. The optimal policy is the state to action mapping that achieves the elementwise maximization in the Bellman equation.

Suppose that we are given a fixed policy $\pi : 1, \ldots, n \mapsto A$. Let the transition matrix and transitional cost vector be $P^\pi \in \Re^{n \times n}$ and $g^\pi \in \Re^n$, respectively. In order to improve the fixed policy via policy iteration,

one needs to calculate the value function associated with policy π, which is the solution to an $n \times n$ system of Bellman equations:

$$J = g^\pi + P^\pi J.$$

Meanwhile, the Bellman residual minimization approach is to solve the problem

$$\begin{aligned} &\min \|J - (g^\pi + P^\pi J)\|^2 \\ &\text{s.t.} \quad J \in S, \end{aligned} \tag{16}$$

where S is a convex constraint set (e.g., a linear subspace spanned by a small number of features). When $S = \Re^n$, the the set of optimal solutions of (16) coincides that of the Bellman equation. In *approximate dynamic programming*, the original high-dimensional problem may be solved by restricting J to some parametric family, which translates to a constraint S in the residual minimization problem.

We consider the simulation setting where P^π and g^π are not explicitly given. Instead, we are given a simulator of the MDP which generates random state transitions under the fixed policy π. This simulator produces a sample trajectories of state and reward pairs according to the unknown transition probabilities P^π:

$$\{(i_k, g_k), (i_{k+1}, g_{k+1}), \ldots\}.$$

Such a simulator is a special case of the conditional sample oracle described in Section 2 and satisfies our assumptions. We may rewrite the Bellman residual minimization (16) as

$$\min_{J \in S} \sum_{i=1}^{n} \left(J(i) - \mathbf{E}^\pi \left[g_{k+1} + J(i_{k+1}) \mid i_k = i\right]\right)^2,$$

which is a stochastic program that takes the form of problem (1). Therefore the proposed SCGD method is able to solve the Bellman minimization problem using only the simulation trajectory $\{(i_k, g_k)\}$, without knowing in advance the transition probabilities. The SCGD method can be applied to *online* policy evaluation, where the samples $\{(i_k, g_k)\}$ are generated from the actual stochastic system, rather than from a simulator which has to restart after every state transition in order to maintain independence. Customized algorithms and analysis for the dynamic programming application is a direction for future research.

6 SUMMARY

In this paper, we have considered a stochastic quasi-gradient scheme for minimizing the composition of a stochastic nonlinear function and an expected-value mapping. We have focused on the case where the function values and subgradient samples are perturbed with Markov noise. With certain diminishing step size assumptions, the proposed stochastic compositional gradient method converges almost surely to an optimal solution of the convex problem. We have derived finite-sample error bounds for the method, which show that the algorithm's performance indeed deteriorates when the noises are Markov instead of independent.

REFERENCES

Bertsekas, D. P. 2011. "Incremental Proximal Methods for Large Scale Convex Optimization". *Mathematical Programming, Series B* 129:163–195.

Bertsekas, D. P., and J. N. Tsitsiklis. 1989. *Parallel and Distributed Computation: Numerical Methods*. Belmont, M. A.: Athena Scientific.

Bhatnagar, S., and V. S. Borkar. 1998. "A Two Timescale Stochastic Approximation Scheme for Simulation-Based Parametric Optimization". *Probability in the Engineering and Informational Sciences* 12:519–531.

Borkar, V. S. 1997. "Stochastic Approximation with Two Time Scales". *Systems & Control Letters* 29:291–294.

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

Ermoliev, Y. M. 1976. *Methods of Stochastic Programming*. Nauka, Moscow.

Konda, R., and J. N. Tsitsikilis. 2004. "Convergence Rate of Linear Two-Time-Scale Stochastic Approximation". *Annals of Applied Probability* 14:796–819.

Kushner, H. J., and G. Yin. 2003. *Stochastic Approximation and Recursive Algorithms and Applications*. N. Y.: Springer.

Nedić, A. 2011. "Random Algorithms for Convex Minimization Problems". *Mathematical Programming, Ser. B* 129:225–253.

Nedić, A., and D. P. Bertsekas. 2001. "Incremental Subgradient Methods for Nondifferentiable Optimization". *SIAM Journal on Optimization* 12:109–138.

Robbins, H., and D. Siegmund. 1971. *A Convergence Theorem for Nonnegative Almost Supermartingales and Some Applications*, 233–257. New York: Academic Press.

Shapiro, A., D. Dentcheva, and A. Ruszczyński. 2014. *Lectures on Stochastic Programming: Modeling and Theory*, Volume 16. SIAM.

Wang, M., and D. P. Bertsekas. 2016. "Stochastic First-Order Methods with Random Constraint Projection". *SIAM Journal on Optimization* 26 (1): 681–717.

Wang, M., Y. Chen, J. Liu, and Y. Gu. 2015. "Random Multi-Constraint Projection: Stochastic Gradient Methods for Convex Optimization with Many Constraints". *arXiv preprint arXiv:1511.03760*.

Wang, M., E. X. Fang, and H. Liu. 2016. "Stochastic Compositional Gradient Descent: Algorithms for Minimizing Compositions of Expected-Value Functions". *Mathematical Programming Series A*.

AUTHOR BIOGRAPHIES

MENGDI WANG is an Assistant Professor of Operations Research and Financial Engineering at Princeton University. She received her PhD in Electrical Engineering and Computer Science from Massachusetts Institute of Technology. Her research interests lie in data-driven optimization, with an emphasis on designing stochastic algorithms with theoretical guarantees and applications in learning. Her email address is mengdiw@princeton.edu.

JI LIU is an Assistant Professor in Computer Science and Goergen Institute for Data Science at University of Rochester (UR). He received his Ph.D., Masters, and B.S. degrees from University of Wisconsin-Madison, Arizona State University, and University of Science and Technology of China respectively. His research interests cover a broad scope of machine learning, optimization, and their applications in other areas such as healthcare, bioinformatics, computer vision, and many other data analysis involved areas. His email address is ji.liu.uwisc@gmail.com.