Non-asymptotic Analysis of Biased Stochastic Approximation Scheme

32\textsuperscript{nd} annual Conference On Learning Theory, COLT 2019

Belhal Karimi, Blazej Miasojedow, Eric Moulines, Hoi-To Wai

June 18, 2019
Stochastic Approximation (SA) Scheme

- Consider a smooth Lyapunov function $V : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ (possibly non-convex) that we wish to find its stationary point.

- SA scheme (Robbins and Monro, 1951) is a stochastic process:

$$\eta_{n+1} = \eta_n - \gamma_{n+1} H_{\eta_n}(X_{n+1}), \quad n \in \mathbb{N}$$

where $\eta_n \in \mathcal{H} \subseteq \mathbb{R}^d$ is the $n$th state, $\gamma_n > 0$ is the step size.

- The *drift term* $H_{\eta_n}(X_{n+1})$ depends on an i.i.d. random element $X_{n+1}$ and the mean-field satisfies

$$h(\eta_n) = \mathbb{E}[H_{\eta_n}(X_{n+1})|\mathcal{F}_n] = \nabla V(\eta_n),$$

where $\mathcal{F}_n$ is the filtration generated by $\{\eta_0, \{X_m\}_{m \leq n}\}$.

- In this case, the SA scheme is better known as the SGD method.
Biased SA Scheme

In this work, we relax a few restrictions of the classical SA. Consider:

$$\eta_{n+1} = \eta_n - \gamma_{n+1} H_{\eta_n}(X_{n+1}), \quad n \in \mathbb{N}. \quad (1)$$

- The mean field $h(\eta) \neq \nabla V(\eta)$
  $$\implies$$ relevant to non-gradient method where the gradient is hard to compute, e.g., online EM.

- $\{X_n\}_{n \geq 1}$ is not i.i.d. and form a state-dependent Markov chain
  $$\implies$$ relevant to SGD with non-iid noise and policy gradient. E.g., $\eta_n$ controls the policy in a Markov decision process, and the gradient estimate $H_{\eta_n}(x)$ is computed from the intermediate reward.
Biased SA Scheme

In this work, we relax a few restrictions of the classical SA. Consider:

\[ \eta_{n+1} = \eta_n - \gamma_{n+1} H_{\eta_n}(X_{n+1}), \quad n \in \mathbb{N}. \tag{1} \]

- The **mean field** \( h(\eta) \neq \nabla V(\eta) \) but satisfies for some \( c_0 \geq 0, c_1 > 0, \)

\[ c_0 + c_1 \langle \nabla V(\eta) | h(\eta) \rangle \geq \| h(\eta) \|^2 \]

- \( \{X_n\}_{n \geq 1} \) is not i.i.d. and form a **state-dependent Markov chain**:

\[ \mathbb{E}[H_{\eta_n}(X_{n+1}) | \mathcal{F}_n] = P_{\eta_n} H_{\eta_n}(X_n) = \int H_{\eta_n}(x) P_{\eta_n}(X_n, dx), \]

where \( P_{\eta_n} : X \times \mathcal{X} \to \mathbb{R}_+ \) is Markov kernel with a unique stationary distribution \( \pi_{\eta_n} \), and the mean field \( h(\eta) = \int H_\eta(x) \pi_\eta(dx) \).
Prior Work & Biased SA Scheme

Consider two cases for the noise sequence

\[ e_{n+1} = H_{\eta_n}(X_{n+1}) - h(\eta_n) \]

**Case 1: When \( \{e_n\}_{n \geq 1} \) is Martingale difference —**

\[ \mathbb{E}[e_{n+1}|\mathcal{F}_n] = 0 \quad \text{and other conditions...} \]

- Asymptotic (Robbins and Monro, 1951), (Benveniste et al., 1990), (Borkar, 2009);
  Non-asymptotic (Moulines and Bach, 2011) (Dalal et al., 2018), (Ghadimi and Lan, 2013).

**Case 2: When \( \{e_n\}_{n \geq 1} \) is state-controlled Markov noise —**

\[ \mathbb{E}[e_{n+1}|\mathcal{F}_n] = P_{\eta_n} H_{\eta_n}(X_n) - h(\eta_n) \neq 0 \quad \text{and other conditions...} \]

- Asymptotic (Kushner and Yin, 2003), (Tadić and Doucet, 2017);
  Non-asymptotic (Sun et al., 2018), (Bhandari et al., 2018)
Our Contributions

- First *non-asymptotic analysis* of biased SA scheme under the relaxed settings for *non-convex* Lyapunov function.

- For both cases, with $N$ being a r.v. drawn from $\{1, ..., n\}$, we show

$$\mathbb{E}[\| h(\eta_N) \|^2] = \mathcal{O}\left(c_0 + \frac{\log n}{\sqrt{n}}\right)$$

where $c_0$ is the *bias* of the mean field. If unbiased, then we find a stationary point.

- Analysis of two stochastic algorithms:
  - Online expectation maximization in *(Cappé and Moulines, 2009)*
  - Online policy gradient for infinite horizon reward maximization *(Baxter and Bartlett, 2001)*.

- We provide the first *non-asymptotic* rates for the above algorithms.
Case 1: Martingale Difference Noise

(A4) \( \{ e_n \}_{n \geq 1} \) is a Martingale difference sequence such that
\[
E [ e_{n+1} \mid \mathcal{F}_n ] = 0, \quad E [ \| e_{n+1} \|^2 \mid \mathcal{F}_n ] \leq \sigma_0^2 + \sigma_1^2 \| h(\eta_n) \|^2
\]
for any \( n \in \mathbb{N} \).

\( \implies \) can be satisfied when \( X_n \) is i.i.d. similar to the SGD setting.

Theorem 1

Let \( \gamma_{n+1} \leq (2c_1 L (1 + \sigma_1^2))^{-1} \) and \( V_{0,n} := E[V(\eta_0) - V(\eta_{n+1})] \),
\[
E[\| h(\eta_N) \|^2] \leq \frac{2c_1 (V_{0,n} + \sigma_0^2 L \sum_{k=0}^{n} \gamma_k^2)}{\sum_{k=0}^{n} \gamma_{k+1}} + 2c_0 \, ,
\]
If we set \( \gamma_k = (2c_1 L (1 + \sigma_1^2)\sqrt{k})^{-1} \), then the SA scheme (1) finds an
\( O(c_0 + \log n / \sqrt{n}) \) quasi-stationary point within \( n \) iterations.

\( \implies \) if \( h(\eta) = \nabla V(\eta) \) it recovers (Ghadimi and Lan, 2013, Theorem 2.1).
Case 2: State-dependent Markov Noise

In this case, \( \{e_n\}_{n \geq 1} \) is not a Martingale sequence. Instead,

\[
E[e_{n+1}|\mathcal{F}_n] = P_{\eta_n} H_{\eta_n}(X_n) - h(\eta_n) \neq 0.
\]

and \( P_{\eta}, H_{\eta}(X) \) are smooth w.r.t. \( \eta \) as well as the other conditions.

**Theorem 2**

*Suppose that the step sizes satisfy*

\[
\gamma_{n+1} \leq \gamma_n, \quad \gamma_n \leq a \gamma_{n+1}, \quad \gamma_n - \gamma_{n+1} \leq a' \gamma_n^2, \quad \gamma_1 \leq 0.5 \left(c_1 (L + C_h)\right)^{-1},
\]

*for \( a, a' > 0 \) and all \( n \geq 0 \). Let \( V_{0,n} := E[V(\eta_0) - V(\eta_{n+1})] \),

\[
E[h(\eta_N)||^2] \leq \frac{2c_1(V_{0,n} + C_{0,n} + (\sigma^2 L + C_\gamma) \sum_{k=0}^{n} \gamma_{k+1}^2)}{\sum_{k=0}^{n} \gamma_{k+1}} + 2c_0,
\]

*• If \( \gamma_k = (2c_1 L (1 + C_h) \sqrt{k})^{-1} \), then \( E[h(\eta_N)||^2] = \mathcal{O}(c_0 + \log n/\sqrt{n}) \) as in our case 1 with Martingale noise.*

*• Key idea to the proof is to use the Poisson equation [see Lemma 2], which is new to the SA analysis.*
Regularized Online EM (ro-EM)

- **GMM Fitting:** \( \theta = (\{\omega_m\}_{m=1}^{M-1}, \{\mu_m\}_{m=1}^M) \) and
  \[
g(y; \theta) \propto \left(1 - \sum_{m=1}^{M-1} \omega_m\right) \exp\left(-\frac{(y - \mu_M)^2}{2}\right) + \sum_{m=1}^{M-1} \omega_m \exp\left(-\frac{(y - \mu_m)^2}{2}\right),
\]

- Data \( \{Y_n\}_{n \geq 1} \) arrives in a streaming fashion, the ro-EM method (modified from (Cappé and Moulines, 2009)) does:
  
  **E-step:** \( \hat{s}_{n+1} = \hat{s}_n + \gamma_{n+1}\{s(Y_{n+1}; \hat{\theta}_n) - \hat{s}_n\} \),
  
  **M-step:** \( \hat{\theta}_{n+1} = \overline{\theta}(\hat{s}_{n+1}) \).

- We can interpret **E-step** as an SA update (1) with drift term
  
  \[
  H_{\hat{s}_n}(Y_{n+1}) = \hat{s}_n - \overline{s}(Y_{n+1}; \overline{\theta}(\hat{s}_n)),
  \]
  
  whose mean field is given by
  
  \[
  h(\hat{s}_n) = \hat{s}_n - \mathbb{E}_\pi[\overline{s}(Y_{n+1}; \overline{\theta}(\hat{s}_n))].
  \]
Lyapunov function? We use the KL divergence

\[ V(s) := \mathbb{E}_{\pi} \left[ \log \left( \pi(Y) / g(Y; \bar{\theta}(s)) \right) \right] + R(\bar{\theta}(s)). \]

**Corollary 1**

Set \( \gamma_k = (2c_1L(1 + \sigma_1^2)\sqrt{k})^{-1} \). The ro-EM method for GMM finds \( \hat{s}_N \) such that

\[ \mathbb{E}[\| \nabla V(\hat{s}_N) \|^2] = O(\log n / \sqrt{n}) \]

The expectation is taken w.r.t. \( N \) and the observation law \( \pi \).

- First *explicit non-asymptotic* rate given for online EM method.
- We consider a slightly modified/regularized M-step update to satisfy the technical convergence conditions.
Online Policy Gradient (PG)

- Consider a Markov Decision Process (MDP) \((S, A, R, P)\):
  - \(S, A\) is the finite set of state/action.
  - \(R : S \times A \to [0, R_{\text{max}}]\) is a reward function; \(P\) is the transition model.
- A **policy** is parameterized by \(\eta \in \mathbb{R}^d\) as (e.g., soft-max):
  \[
  \Pi_{\eta}(a'; s') = \text{probability of taking action } a' \text{ in state } s'
  \]
- We update the policy \(\eta\) on-the-fly with an online policy gradient update (Baxter and Bartlett, 2001; Tadić and Doucet, 2017):
  \[
  G_{n+1} = \lambda G_n + \nabla \log \Pi_{\eta_n}(A_{n+1}; S_{n+1}) , \tag{2a}
  \]
  \[
  \eta_{n+1} = \eta_n + \gamma_{n+1} G_{n+1} R(S_{n+1}, A_{n+1}) , \tag{2b}
  \]
  where \(\lambda \in (0, 1)\) is a parameter for the variance-bias trade-off.
- We can interpret (2b) as an SA step with the drift term:
  \[
  H_{\eta_n}(X_{n+1}) = G_{n+1} R(S_{n+1}, A_{n+1})
  \]
Convergence Analysis

Let \( \nu(\eta)(s, a) \) be the invariant distribution of \( \{(S_t, A_t)\}_{t \geq 1} \), we consider:

\[
J(\eta) := \sum_{s \in S, a \in A} \nu(\eta)(s, a) R(s, a).
\]

**Corollary 2**

Set \( \gamma_k = (2c_1 L(1 + C_h)\sqrt{k})^{-1} \). For any \( n \in \mathbb{N} \), the policy gradient algorithm (2) finds a policy that

\[
\mathbb{E}[\|\nabla J(\eta_N)\|^2] = \mathcal{O}\left( (1 - \lambda)^2 \Gamma^2 + c(\lambda) \log n / \sqrt{n} \right),
\]

(3)

where \( c(\lambda) = \mathcal{O}(\frac{1}{1 - \lambda}) \). Expectation is taken w.r.t. \( N \) and \( (A_n, S_n) \).

- It shows the first convergence rate for the online PG method.
- Our result shows the variance-bias trade-off with \( \lambda \in (0, 1) \).
- While setting \( \lambda \rightarrow 1 \) reduces the bias, but it decreases the convergence rate with \( c(\lambda) \).
Take-aways

- **Theorem 1 & 2** show the non-asymptotic convergence rate of biased SA scheme with smooth (possibly non-convex) Lyapunov function.
- With appropriate step size, in $n$ iterations the SA scheme finds

$$\mathbb{E}[\|h(\eta_N)\|^2] = \mathcal{O}(c_0 + \log n/\sqrt{n}),$$

where $c_0$ is the bias and $h(\cdot)$ is the mean field.
- Applications to online EM and online policy gradient with *rigorous* verification of the assumptions.
  - For *online EM*, we show the first non-asymptotic, global convergence rate.
  - For *online policy gradient*, we show the first non-asymptotic convergence rate under a dynamical setting.
Thank you! Questions?


