Non-asymptotic Analysis of Biased Stochastic Approximation Scheme

32^{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:

$$\boldsymbol{\eta}_{n+1} = \boldsymbol{\eta}_n - \gamma_{n+1} H_{\boldsymbol{\eta}_n}(X_{n+1}), \quad n \in \mathbb{N}$$

  where $\boldsymbol{\eta}_n \in \mathcal{H} \subseteq \mathbb{R}^d$ is the nth state, $\gamma_n > 0$ is the step size.
- The *drift term* $H_{\boldsymbol{\eta}_n}(X_{n+1})$ depends on an **i.i.d. random element** X_{n+1} and the mean-field satisfies

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

  where $\mathcal{F}_n$ is the filtration generated by $\{\boldsymbol{\eta}_0, \{X_m\}_{m \leq n}\}$.
- 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:

$$\boldsymbol{\eta}_{n+1} = \boldsymbol{\eta}_n - \gamma_{n+1} H_{\boldsymbol{\eta}_n}(X_{n+1}), \quad n \in \mathbb{N}. \tag{1}$$

- The **mean field** $h(\boldsymbol{\eta}) \neq \nabla V(\boldsymbol{\eta})$

  $\Longrightarrow$ 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**

  $\Longrightarrow$ relevant to *SGD with non-iid noise* and *policy gradient*. E.g., $\boldsymbol{\eta}_n$ controls the policy in a Markov decision process, and the gradient estimate $H_{\boldsymbol{\eta}_n}(x)$ is computed from the intermediate reward.

Biased SA Scheme

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

$$\boldsymbol{\eta}_{n+1} = \boldsymbol{\eta}_n - \gamma_{n+1} H_{\boldsymbol{\eta}_n}(X_{n+1}), \quad n \in \mathbb{N}. \tag{1}$$

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

$$c_0 + c_1 \left\langle \nabla V(\boldsymbol{\eta}) \,|\, h(\boldsymbol{\eta}) \right\rangle \geq \|h(\boldsymbol{\eta})\|^2$$

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

$$\mathbb{E}[H_{\boldsymbol{\eta}_n}(X_{n+1}) | \mathcal{F}_n] = P_{\boldsymbol{\eta}_n} H_{\boldsymbol{\eta}_n}(X_n) = \int H_{\boldsymbol{\eta}_n}(x) P_{\boldsymbol{\eta}_n}(X_n, \mathrm{d}x),$$

  where $P_{\boldsymbol{\eta}_n} : \mathsf{X} \times \mathcal{X} \to \mathbb{R}_+$ is Markov kernel with a unique stationary distribution $\pi_{\boldsymbol{\eta}_n}$, and the mean field $h(\boldsymbol{\eta}) = \int H_{\boldsymbol{\eta}}(x) \pi_{\boldsymbol{\eta}}(\mathrm{d}x)$.

Prior Work & Biased SA Scheme

Consider two cases for the noise sequence

$$\boldsymbol{e}_{n+1} = H_{\boldsymbol{\eta}_n}(X_{n+1}) - h(\boldsymbol{\eta}_n)$$

Case 1: When $\{\boldsymbol{e}_n\}_{n\geq 1}$ is Martingale difference —

$$\mathbb{E}\big[\boldsymbol{e}_{n+1}|\mathcal{F}_n\big] = 0 \ \text{ and other conditions...}$$

- *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 $\{\boldsymbol{e}_n\}_{n\geq 1}$ is state-controlled Markov noise —

$$\mathbb{E}\big[\boldsymbol{e}_{n+1}|\mathcal{F}_n\big] = P_{\boldsymbol{\eta}_n} H_{\boldsymbol{\eta}_n}(X_n) - h(\boldsymbol{\eta}_n) \neq 0 \ \text{ and other conditions....}$$

- *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

$$\mathbb{E}[\|h(\boldsymbol{\eta}_N)\|^2] = \mathcal{O}\Big(c_0 + \frac{\log n}{\sqrt{n}}\Big)$$

  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) $\{\boldsymbol{e}_n\}_{n\geq 1}$ is a Martingale difference sequence such that $\mathbb{E}\left[\boldsymbol{e}_{n+1} \,|\, \mathcal{F}_n\right] = \mathbf{0}$, $\mathbb{E}\left[\|\boldsymbol{e}_{n+1}\|^2 \,\middle|\, \mathcal{F}_n\right] \leq \sigma_0^2 + \sigma_1^2 \|h(\boldsymbol{\eta}_n)\|^2$ for any $n \in \mathbb{N}$.

$\Longrightarrow$ 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* $V_{0,n} := \mathbb{E}[V(\boldsymbol{\eta}_0) - V(\boldsymbol{\eta}_{n+1})]$,

$$\mathbb{E}[\|h(\boldsymbol{\eta}_N)\|^2] \leq \frac{2c_1\left(V_{0,n} + \sigma_0^2 L \sum_{k=0}^n \gamma_{k+1}^2\right)}{\sum_{k=0}^n \gamma_{k+1}} + 2c_0 \ ,$$

If we set $\gamma_k = (2c_1 L(1+\sigma_1^2)\sqrt{k})^{-1}$, then the SA scheme (1) finds an $\mathcal{O}(c_0 + \log n/\sqrt{n})$ quasi-stationary point within n iterations.

$\Longrightarrow$ if $h(\boldsymbol{\eta}) = \nabla V(\boldsymbol{\eta})$ it recovers *(Ghadimi and Lan, 2013, Theorem 2.1)*.

Case 2: State-dependent Markov Noise

In this case, $\{\boldsymbol{e}_n\}_{n\geq 1}$ is not a Martingale sequence. Instead,

$$\mathbb{E}\big[\boldsymbol{e}_{n+1}|\mathcal{F}_n\big] = P_{\boldsymbol{\eta}_n} H_{\boldsymbol{\eta}_n}(X_n) - h(\boldsymbol{\eta}_n) \neq 0.$$

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

Theorem 2

Suppose that the step sizes satisfy

$$\gamma_{n+1} \leq \gamma_n, \;\; \gamma_n \leq a\gamma_{n+1}, \;\; \gamma_n - \gamma_{n+1} \leq a'\gamma_n^2, \;\; \gamma_1 \leq 0.5\big(c_1(L + C_h)\big)^{-1} \;,$$

for $a, a' > 0$ and all $n \geq 0$. Let $V_{0,n} := \mathbb{E}[V(\boldsymbol{\eta}_0) - V(\boldsymbol{\eta}_{n+1})]$,

$$\mathbb{E}[h(\boldsymbol{\eta}_N)\|^2] \leq \frac{2c_1\big(V_{0,n} + C_{0,n} + \big(\sigma^2 L + C_\gamma\big)\sum_{k=0}^{n}\gamma_{k+1}^2\big)}{\sum_{k=0}^{n}\gamma_{k+1}} + 2c_0 \;,$$

- If $\gamma_k = (2c_1 L(1 + C_h)\sqrt{k})^{-1}$, then $\mathbb{E}[h(\boldsymbol{\eta}_N)\|^2] = \mathcal{O}(c_0 + \log n/\sqrt{n})$ 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**: $\boldsymbol{\theta} = (\{\omega_m\}_{m=1}^{M-1}, \{\mu_m\}_{m=1}^{M})$ and

$$g(y;\boldsymbol{\theta}) \propto \left(1-\sum_{m=1}^{M-1}\omega_m\right)\exp\left(-\frac{(y-\mu_M)^2}{2}\right)+\sum_{m=1}^{M-1}\omega_m\exp\left(-\frac{(y-\mu_m)^2}{2}\right),$$

- Data $\{Y_n\}_{n\geq 1}$ arrives in a streaming fashion, the ro-EM method (modified from (Cappé and Moulines, 2009)) does:

$$\begin{aligned} \text{E-step:} \quad & \hat{\mathbf{s}}_{n+1} = \hat{\mathbf{s}}_n + \gamma_{n+1}\left\{\overline{\mathbf{s}}(Y_{n+1};\hat{\boldsymbol{\theta}}_n) - \hat{\mathbf{s}}_n\right\}, \\ \text{M-step:} \quad & \hat{\boldsymbol{\theta}}_{n+1} = \overline{\boldsymbol{\theta}}(\hat{\mathbf{s}}_{n+1}). \end{aligned}$$

- We can interpret **E-step** as an SA update (1) with drift term

$$H_{\hat{\mathbf{s}}_n}(Y_{n+1}) = \hat{\mathbf{s}}_n - \overline{\mathbf{s}}(Y_{n+1};\overline{\boldsymbol{\theta}}(\hat{\mathbf{s}}_n))\ ,$$

whose mean field is given by $h(\hat{\mathbf{s}}_n) = \hat{\mathbf{s}}_n - \mathbb{E}_\pi\left[\overline{\mathbf{s}}(Y_{n+1};\overline{\boldsymbol{\theta}}(\hat{\mathbf{s}}_n))\right]$

Convergence Analysis

Lyapunov function? We use the KL divergence

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

Corollary 1

Set $\gamma_k = (2c_1 L(1+\sigma_1^2)\sqrt{k})^{-1}$. The ro-EM method for GMM finds $\hat{\boldsymbol{s}}_N$ such that

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

The expectation is taken w.r.t. *N and the observation law π.*

- 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) $(\mathsf{S}, \mathsf{A}, \mathsf{R}, \mathsf{P})$:
  - S, A is the finite set of state/action.
  - $\mathsf{R} : \mathsf{S} \times \mathsf{A} \to [0, \mathsf{R}_{\max}]$ is a reward function; P is the transition model.
- A **policy** is parameterized by $\boldsymbol{\eta} \in \mathbb{R}^d$ as (e.g., soft-max):

$$\Pi_{\boldsymbol{\eta}}(a'; s') = \text{probability of taking action } a' \text{ in state } s'$$

- We update the policy $\boldsymbol{\eta}$ on-the-fly with an online policy gradient update (Baxter and Bartlett, 2001; Tadić and Doucet, 2017):

$$G_{n+1} = \lambda G_n + \nabla \log \Pi_{\boldsymbol{\eta}_n}(A_{n+1}; S_{n+1}) \ , \tag{2a}$$

$$\boldsymbol{\eta}_{n+1} = \boldsymbol{\eta}_n + \gamma_{n+1} G_{n+1} \mathsf{R}(S_{n+1}, A_{n+1}) \ , \tag{2b}$$

  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:

$$H_{\boldsymbol{\eta}_n}(X_{n+1}) = G_{n+1} \mathsf{R}(S_{n+1}, A_{n+1})$$

Convergence Analysis

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

$$J(\boldsymbol{\eta}) := \sum_{s\in\mathsf{S}, a\in\mathsf{A}} \upsilon_{\boldsymbol{\eta}}(s,a)\, \mathrm{R}(s,a) \, .$$

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

$$\mathbb{E}\big[\|\nabla J(\boldsymbol{\eta}_N)\|^2\big] = \mathcal{O}\Big((1-\lambda)^2\Gamma^2 + c(\lambda)\log n/\sqrt{n}\Big), \tag{3}$$

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 $\lambda \to 1$ 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

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

  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?

References

Baxter, J. and Bartlett, P. L. (2001). Infinite-horizon policy-gradient estimation. *Journal of Artificial Intelligence Research*, 15:319–350.

Benveniste, A., Priouret, P., and Métivier, M. (1990). *Adaptive Algorithms and Stochastic Approximation*.

Bhandari, J., Russo, D., and Singal, R. (2018). A finite time analysis of temporal difference learning with linear function approximation. In *Conference On Learning Theory*, pages 1691–1692.

Borkar, V. S. (2009). *Stochastic approximation: a dynamical systems viewpoint*, volume 48. Springer.

Cappé, O. and Moulines, E. (2009). On-line Expectation Maximization algorithm for latent data models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):593–613.

Dalal, G., Szorenyi, B., Thoppe, G., and Mannor, S. (2018). Finite sample analysis of two-timescale stochastic approximation with applications to reinforcement learning. In *Conference On Learning Theory*.

Ghadimi, S. and Lan, G. (2013). Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368.

Kushner, H. and Yin, G. G. (2003). *Stochastic approximation and recursive algorithms and applications*, volume 35. Springer Science & Business Media.

Moulines, E. and Bach, F. R. (2011). Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Advances in Neural Information Processing Systems*, pages 451–459.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407.

Sun, T., Sun, Y., and Yin, W. (2018). On Markov chain gradient descent. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 9918–9927. Curran Associates, Inc.

Tadić, V. B. and Doucet, A. (2017). Asymptotic bias of stochastic gradient search. *The Annals of Applied Probability*, 27(6):3255–3304.