Non-asymptotic Analysis of Biased Stochastic Approximation Scheme

Belhal Karimi[1,2], Blazej Miasojedow[3], Eric Moulines[2] and Hoi-To Wai[4]

INRIA[1], École Polytechnique[2], University of Warsaw[3], Chinese University of Hong Kong[4]

belhal.karimi@polytechnique.edu, bmiasojedow@gmail.com, eric.moulines@polytechnique.edu, htwai@se.cuhk.edu.hk

Stochastic Approximation

- **Objective:** Find a *stationary point* of smooth Lyapunov function $V(\eta)$.
- SA scheme (Robbins and Monro, 1951) is a stochastic process:

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

where $\eta_n \in \mathcal{H} \subseteq \mathbb{R}^d$ is the nth 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

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

where $\mathcal{F}_n = \sigma(\eta_0, \{X_m\}_{m\le n})$. In this case, SA is better known as the SGD method.

Biased SA Scheme

- The **mean field is biased** $\Leftarrow$ gradient is sometimes difficult to compute...

We have $h(\eta) \neq \nabla V(\eta)$ and for some $c_0 \geq 0, c_1 > 0$,

$$c_0 + c_1 \langle \nabla V(\eta) \,|\, h(\eta) \rangle \geq \|h(\eta)\|^2, \ \forall \eta \in \mathcal{H}$$

- The **drift term** $\{H_{\eta_n}(X_{n+1})\}_{n\geq 1}$ is **not i.i.d.**. For example, in reinforcement learning, η_n controls the policy in a MDP & $H_{\eta_n}(X_{n+1})$ is computed from the MDP's state.

The random elements $\{X_n\}_{n\geq 1}$ form a **state-dependent Markov chain**:

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

where $P_{\eta_n} : \mathsf{X} \times \mathcal{X} \to \mathbb{R}_+$ is Markov kernel with a unique stationary distribution π_{η_n}.

- In the latter case, the mean field is given by $h(\eta) = \int H_\eta(x) \pi_\eta(\mathrm{d}x)$.
- **Stopping criterion**: fix any $n \geq 1$, we stop the SA at a random iteration N with

$$\mathbb{P}(N = \ell) = \Big(\sum_{k=0}^n \gamma_{k+1}\Big)^{-1} \gamma_{\ell+1}, \quad \text{with} \quad N \in \{1, \ldots, n\}.$$

Prior Work

- We focus on the **non-asymptotic convergence** analysis of SA scheme, where the relevant results are rare. Define:

$$e_{n+1} := H_{\eta_n}(X_{n+1}) - h(\eta_n) \tag{2}$$

Case 1: When $\{e_n\}_{n\geq 1}$ is Martingale difference — $\mathbb{E}[e_{n+1}|\mathcal{F}_n] = 0$

- *Asymptotic analysis*: (Robbins and Monro, 1951); *Non-asymptotic analysis*: (Ghadimi and Lan, 2013).

Case 2: When $\{e_n\}_{n\geq 1}$ is state-controlled Markov noise

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

- *Asymptotic analysis*: (Tadić and Doucet, 2017); *Non-asymptotic analysis*: (Sun et al., 2018), (Duchi et al., 2012), (Bhandari et al., 2018)

Analysis For Martingale Difference Noise (Case 1)

Assumption: $\mathbb{E}\,[e_{n+1}\,|\,\mathcal{F}_n] = 0$, $\mathbb{E}\,[\|e_{n+1}\|^2\,|\,\mathcal{F}_n] \leq \sigma_0^2 + \sigma_1^2 \|h(\eta_n)\|^2$. (e.g., 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* $V_{0,n} := \mathbb{E}[V(\eta_0) - V(\eta_{n+1})]$,
>
> $$\mathbb{E}[\|h(\eta_N)\|^2] \leq \frac{2c_1\big(V_{0,n} + \sigma_0^2 L \sum_{k=0}^n \gamma_{k+1}^2\big)}{\sum_{k=0}^n \gamma_{k+1}} + 2c_0\,,$$

Set $\gamma_k = (2c_1 L(1+\sigma_1^2)\sqrt{k})^{-1} \implies \mathbb{E}[\|h(\eta_N)\|^2] = \mathcal{O}(c_0 + \log n/\sqrt{n})$. *Remark*: if $h(\eta) = \nabla V(\eta)$ (with $c_0 = d_0 = 0$), it recovers (Ghadimi and Lan, 2013, Theorem 2.1).

Analysis For State-dependent Markov Noise (Case 2)

Assumptions: we need a few regularity conditions in this case,

1. There exists a Borel measurable function $\hat{H} : \mathcal{H} \times \mathsf{X} \to \mathcal{H}$,

$$\hat{H}_\eta(x) - P_\eta \hat{H}_\eta(x) = H_\eta(x) - h(\eta), \ \forall \eta \in \mathcal{H}, x \in \mathsf{X}.$$

$\implies$ existence of solution to the Poisson equation.

2. For all $\eta \in \mathcal{H}$ and $x \in \mathsf{X}$, $\|\hat{H}_\eta(x)\| \leq L_{PH}^{(0)}$, $\|P_\eta \hat{H}_\eta(x)\| \leq L_{PH}^{(0)}$, and

$$\sup_{x\in\mathsf{X}} \|P_\eta \hat{H}_\eta(x) - P_{\eta'} \hat{H}_{\eta'}(x)\| \leq L_{PH}^{(1)} \|\eta - \eta'\|, \ \forall (\eta, \eta') \in \mathcal{H}^2.$$

$\implies$ smoothness of $\hat{H}_\eta(x)$, satisfied if P_η, $H_\eta(X)$ are smooth w.r.t. η.

3. It holds that $\sup_{\eta\in\mathcal{H}, x\in\mathsf{X}} \|H_\eta(x) - h(\eta)\| \leq \sigma$.

$\implies$ requires the noise is uniformly bounded for all $x \in \mathsf{X}$.

Example: assumptions 1 & 2 are satisfied if the Markov kernel P_{η_n} is geometrically ergodic + smooth, and the drift term is smooth w.r.t. η.

> **Theorem 2.** *Suppose that the step sizes are decreasing and* $\gamma_1 \leq 0.5(c_1(L + C_h))^{-1}$ *(+other conditions). Let* $V_{0,n} := \mathbb{E}[V(\eta_0) - V(\eta_{n+1})]$,
>
> $$\mathbb{E}[\|h(\eta_N)\|^2] \leq \frac{2c_1\big(V_{0,n} + C_{0,n} + (\sigma^2 L + C_\gamma) \sum_{k=0}^n \gamma_{k+1}^2\big)}{\sum_{k=0}^n \gamma_{k+1}} + 2c_0\,.$$

- Set $\gamma_k = (2c_1 L(1+C_h)\sqrt{k})^{-1} \implies \mathbb{E}[\|h(\eta_N)\|^2] = \mathcal{O}(c_0 + \log n/\sqrt{n})$ (same as Case 1).
- **Proof idea:** challenge is that e_{n+1} is not zero-mean $\implies$ bound the sum of $\mathbb{E}[\langle \nabla V(\eta_n) \,|\, e_{n+1}\rangle]$ w/ Poisson equation + a novel decomposition (cf. Lemma 2).

Regularized Online EM Algorithm

- **Special Case of GMM**: we fit the data $\{Y_n\}_{n\geq 1}, Y_n \sim \pi$ into the parametric model with $\theta = (\{\omega_m\}_{m=1}^{M-1}, \{\mu_m\}_{m=1}^M)$

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

- Data arrives in a streaming fashion, Cappé and Moulines (2009) does:

$$\text{E-step:} \quad \hat{s}_{n+1} = \hat{s}_n + \gamma_{n+1}\{\overline{s}(Y_{n+1}; \hat{\theta}_n) - \hat{s}_n\},$$
$$\text{M-step:} \quad \hat{\theta}_{n+1} = \overline{\theta}(\hat{s}_{n+1}).$$

- The **E-step** is a biased SA step on s with the drift term & mean field

$$H_{\hat{s}_n}(Y_{n+1}) = \hat{s}_n - \overline{s}(Y_{n+1}; \overline{\theta}(\hat{s}_n)), \quad h(\hat{s}_n) = \hat{s}_n - \mathbb{E}_\pi\big[\overline{s}(Y_{n+1}; \overline{\theta}(\hat{s}_n))\big]$$

Analysis of the ro-EM Algorithm (Application of Case 1)

Consider the KL divergence as a function of sufficient statistics s:

$$V(s) := \mathrm{KL}(\pi | g(\cdot; \overline{\theta}(s))) + \mathrm{R}(\overline{\theta}(s)) = \mathbb{E}_\pi\big[\log\big(\pi(Y)/g(Y; \overline{\theta}(s))\big)\big] + \mathrm{R}(\overline{\theta}(s)).$$

> **Corollary 1.** *Set* $\gamma_k = (2c_1 L(1+\sigma_1^2)\sqrt{k})^{-1}$. *Ro-EM method for GMM finds* $\hat{s}_N$ *such that*
>
> $$\mathbb{E}[\|\nabla V(\hat{s}_N)\|^2] = \mathcal{O}(\log n/\sqrt{n})$$
>
> *The expectation is taken w.r.t. N and the observation law π.*

- First explicit non-asymptotic rate given for online EM method.
- Consider a slightly modified/regularized M-step update for satisfaction of the technical conditions.

(Online) Policy Gradient Method

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

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

- Update η in an online fashion (Tadić and Doucet, 2017) using observed state-action pair:

$$G_{n+1} = \lambda G_n + \nabla \log \Pi_{\eta_n}(A_{n+1}; S_{n+1})\,,$$
$$\eta_{n+1} = \eta_n + \gamma_{n+1} G_{n+1} \mathrm{R}(S_{n+1}, A_{n+1})$$

where $\lambda \in (0,1)$ is a parameter for the variance-bias trade-off.

- The η-update is an biased SA step with the drift term:

$$H_{\eta_n}(X_{n+1}) = G_{n+1} \mathrm{R}(S_{n+1}, A_{n+1})$$

Analysis of Policy Gradient Method (Application of Case 2)

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

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

> **Corollary 2.** *Set* $\gamma_k = (2c_1 L(1+C_h)\sqrt{k})^{-1}$. *For any $n \in \mathbb{N}$, the policy gradient algorithm (3) finds a policy that*
>
> $$\mathbb{E}[\|\nabla J(\eta_N)\|^2] = \mathcal{O}\Big((1-\lambda)^2 \Gamma^2 + c(\lambda) \log n/\sqrt{n}\Big),$$
>
> *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 $\lambda \in (0,1)$.
- Setting $\lambda \to 1$ reduces the bias, but decreases the convergence speed.

Conclusion

- 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.

References

Jalaj Bhandari, Daniel Russo, and Raghav Singal. A finite time analysis of temporal difference learning with linear function approximation. In *Conference On Learning Theory*, pages 1691–1692, 2018.

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

John C Duchi, Alekh Agarwal, Mikael Johansson, and Michael I Jordan. Ergodic mirror descent. *SIAM Journal on Optimization*, 22(4):1549–1578, 2012.

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

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

Tao Sun, Yuejiao Sun, and Wotao Yin. On Markov chain gradient descent. In *Advances in Neural Information Processing Systems 31*, pages 9918–9927. Curran Associates, Inc., 2018.

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