Stochastic Approximation Schemes with Decision Dependent Data

Hoi-To Wai

Department of Systems Engineering & Engineering Management,
The Chinese University of Hong Kong (CUHK), Hong Kong

April 28, 2023
Seminar @ Universidad Rey Juan Carlos de Madrid

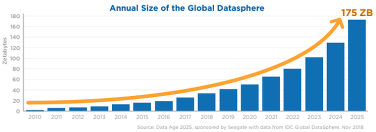
Motivation

We are living in a highly dynamical world ...

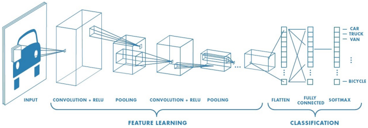
- Huge amount of data generated in **real time** and may be reacting to decision.

  → **dynamical** data → challenging for ML.

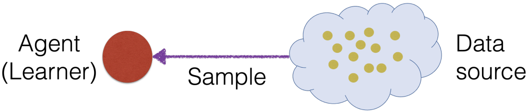
Figure 1 - Annual Size of the Global Datasphere



Do we have to modify our ML algorithms? Any theoretical insights?



Stochastic Approximation (SA) Scheme



- Many engineering/ML problems can be reduced to:

$$\text{find} \quad \boldsymbol{\theta}^\star \quad \text{such that} \quad h(\boldsymbol{\theta}^\star) = \mathbf{0}$$

Examples: to solve $\min_{\boldsymbol{\theta}} V(\boldsymbol{\theta})$, take $h(\cdot) := \nabla V(\cdot)$.

- When $h(\boldsymbol{\theta})$ is unknown, SA[1] (commonly used as **SGD**) is a remedy:

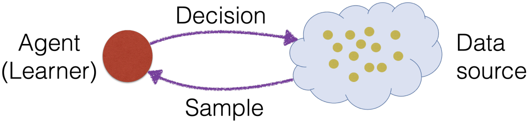
$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_{t+1} H(\boldsymbol{\theta}_t; X_{t+1}), \quad \text{with} \quad H(\boldsymbol{\theta}_t; X_{t+1}) \approx h(\boldsymbol{\theta}_t)$$

where X_{t+1} is the **data** drawn $\leftarrow$ common assumption: **i.i.d. data**.

- **Fact**: (+ appropriate step size) $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ such that $h(\bar{\boldsymbol{\theta}}) = 0$.

[1] [Robbins and Monro, 1951] H. Robbins, S. Monro. A stochastic approximation method. The Annals of Mathematical Statistics, 1951.

SA with Decision-Dependent Data



- What if data $\color{red}{X_{t+1}}$ is not i.i.d., and depends on $\boldsymbol{\theta}_t$?

$$\underline{\text{SA}}: \quad \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_{t+1} H(\boldsymbol{\theta}_t; \color{red}{X_{t+1}}).$$

- Data source may be '**organic**' and react to the agent's decision. (e.g., when they represent real user)

SA with Decision-Dependent Data

- What if data X_{t+1} is not i.i.d., and depends on $\boldsymbol{\theta}_t$?

$$\underline{\text{SA}}: \quad \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_{t+1} H(\boldsymbol{\theta}_t; X_{t+1}).$$

- *Example 1*: in reinforcement learning (RL),

$$X_t = (S_t, A_t) \text{ -- state/action}, \quad \boldsymbol{\theta}_t \text{ -- policy}.$$

  A policy describes conditional probability for selecting A_t.

- **Online policy gradient** [Baxter and Bartlett, 2001] –

$$\cdots \to A_t \to S_t \underbrace{\longrightarrow}_{\text{SA step}} \boldsymbol{\theta}_t \underbrace{\longrightarrow}_{\text{Use Policy}} \underbrace{A_{t+1} \to S_{t+1}}_{\text{Calc. Reward}} \to \cdots$$

SA with Decision-Dependent Data

- What if data X_{t+1} is not i.i.d., and depends on $\boldsymbol{\theta}_t$?

$$\underline{\text{SA}}: \quad \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_{t+1} H(\boldsymbol{\theta}_t; X_{t+1}).$$

- *Example 2*: (Performative Prediction) data may react to your decision,

  $\boldsymbol{\theta}_t$ – classifier/prediction model, $\quad X_{t+1} \sim \mathcal{D}(\boldsymbol{\theta}_t)$ – observed outcome

  such as in loan application, spam email classification, etc.
- **Greedy Deployment** [Perdomo et al., 2020]:

$$\cdots \to X_t \underbrace{\longrightarrow}_{\text{SA step}} \boldsymbol{\theta}_t \qquad \underbrace{\longrightarrow}_{\text{Adopt/deploy the decision}} \qquad X_{t+1} \to \cdots$$

SA with Decision-Dependent Data

- What if data X_{t+1} is not i.i.d., and depends on $\boldsymbol{\theta}_t$?

$$\underline{\text{SA}}: \quad \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_{t+1} H(\boldsymbol{\theta}_t; X_{t+1}).$$

- *Example 3*: in Actor-Critic algorithm,

  X_t – intermediate policy evaluation (critic), $\boldsymbol{\theta}_t$ – policy (actor)

- **Actor-critic Algorithm** [Konda and Tsitsiklis, 1999]:

$$\cdots \to X_t \underbrace{\longrightarrow}_{\text{SA Step / 'Actor' update}} \boldsymbol{\theta}_t \underbrace{\longrightarrow}_{\text{'Critic' update}} X_{t+1} \to \cdots$$

  requires two timescales update to 'stabilize' the critic.

SA with Decision-Dependent Data: Challenges

Key Q: *When will SA with decision-dependent data converge to some meaningful point (e.g., $h(\bar{\boldsymbol{\theta}}) = 0$)? How fast?*

Challenges —

- The drift term is *biased*, i.e., $\mathbb{E}\big[H(\boldsymbol{\theta}_t; X_{t+1})|\mathcal{F}_t\big] \neq h(\boldsymbol{\theta}_t)$.
- If X_{t+1} is *(too) sensitive* to $\boldsymbol{\theta}_t$, the SA process may not converge.

This Talk —

- Recent results on convergence to **stationary or stable** solution with decision dependent data SA.
- Applications to online policy gradient, performative prediction, two-timescale SA, etc.

Goal of This Talk

- **General Convergence for Biased SA**[2]
  - Highlight on techniques for analyzing with non-i.i.d. samples.
- **Applications of Biased SA**
  - Design and analysis of Online Policy Gradient.
  - Modeling stochastic algorithm for performative prediction via SA[3] (relies on an extension of general convergence to **strongly convex** functions).
- **Extension to Coupled SA for Bilevel Optimization**[4]
  - Bi-level optimization where *lower level* gives decision-dependent data.

[2] [Karimi et al., 2019] B. Karimi B. Miasojedow, É. Moulines, **H.-T.**, "Non-asymptotic Analysis of Biased Stochastic Approximation Scheme", in COLT 2019.

[3] [Li and Wai, 2022] **Q. Li**, **H.-T.**, "State Dependent Performative Prediction with Stochastic Approximation", in AISTATS 2022.

[4] [Hong et al., 2023] M. Hong, **H.-T.**, Z. Wang, Z. Yang, "A two-timescale framework for bilevel optimization: Complexity analysis & application to actor-critic", SIOPT, 2023.

Roadmap

1. General Convergence of (Biased) SA

2. Applications of Biased SA

3. Extension: Two-timescale SA

4. Conclusions and Perspectives

Roadmap

Biased SA Scheme

Consider:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_{t+1} H(\boldsymbol{\theta}_t; X_{t+1}). \tag{1}$$

- $\{X_t\}_{t\geq 1}$ is not i.i.d. and form a **decision-dependent Markov chain**:

$$\mathbb{E}[H(\boldsymbol{\theta}_t; X_{t+1}) | \mathcal{F}_t] = P_{\boldsymbol{\theta}_t} H(\boldsymbol{\theta}_t; X_t) = \int H(\boldsymbol{\theta}_t; x) P_{\boldsymbol{\theta}_t}(X_t, \mathrm{d}x),$$

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

This Part: *We analyze the SA scheme with the mean field $h(\boldsymbol{\theta}) = \nabla V(\boldsymbol{\theta})$ (= SGD) for tackling:*

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^d} \; V(\boldsymbol{\theta}). \tag{2}$$

(In the paper) Analysis when $h(\boldsymbol{\theta}) \neq \nabla V(\boldsymbol{\theta})$ is available $\Rightarrow$ the 'beyond' gradient setting.

Biased SA Scheme

Consider:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_{t+1} H(\boldsymbol{\theta}_t; X_{t+1}). \tag{1}$$

Prior Works —

- *Asymptotic Analysis*: studied with $h(\boldsymbol{\theta}) = \nabla V(\boldsymbol{\theta})$ in [Kushner and Yin, 2003], similar biased SA setting in [Tadić and Doucet, 2017].
- *Non-asymptotic Analysis*:
  - Analysis for (projected) SA with convex optimization [Atchadé et al., 2017, Doan, 2022].
  - Sun et al. [2018] and Duchi et al. [2012] assumed $h(\boldsymbol{\theta}) = \nabla V(\boldsymbol{\theta})$ & **decision-independent** Markov chain.
  - Bhandari et al. [2018] studied a similar setting but focuses on linear SA with convex Lyapunov function.
  - Recent works [Chen et al., 2020, Mou et al., 2020, Durmus et al., 2021a,b, 2022, Chen et al., 2022b] provided tight bounds for linear SA.

Assumptions

- For simplicity, we assume that $h(\boldsymbol{\theta}) = \nabla V(\boldsymbol{\theta})$ (the paper presents a more general case for **non-gradient** algorithms).

(A1) The Lyapunov function V is L-smooth,

$$\|\nabla V(\boldsymbol{\theta}) - \nabla V(\boldsymbol{\theta}')\| \leq L\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|, \ \forall \ \boldsymbol{\theta}, \boldsymbol{\theta}'.$$

- Requires smooth Lyapunov function, $V(\boldsymbol{\theta})$ is possibly *non-convex*.

(A2) It holds that $\sup_{\boldsymbol{\theta} \in \mathbb{R}^d, x \in \mathsf{X}} \|H(\boldsymbol{\theta}; x) - h(\boldsymbol{\theta})\| \leq \sigma$.

Remark: **(A2)** requires noise is *uniformly bounded* for all $x \in \mathsf{X}$.

Assumptions (on the Markov Chain)

(A3) There exists a bounded measurable function $\hat{H} : \mathbb{R}^d \times \mathsf{X} \to \mathbb{R}^d$ s.t.

$$\hat{H}_{\boldsymbol{\theta}}(x) - P_{\boldsymbol{\theta}} \hat{H}_{\boldsymbol{\theta}}(x) = H(\boldsymbol{\theta}; x) - h(\boldsymbol{\theta}),\ \forall\ \boldsymbol{\theta} \in \mathbb{R}^d, x \in \mathsf{X},$$

and $\quad \sup_{x \in \mathsf{X}} \| P_{\boldsymbol{\theta}} \hat{H}_{\boldsymbol{\theta}}(x) - P_{\boldsymbol{\theta}'} \hat{H}_{\boldsymbol{\theta}'}(x) \| \leq L_{PH}^{(1)} \| \boldsymbol{\theta} - \boldsymbol{\theta}' \|,\ \forall\ (\boldsymbol{\theta}, \boldsymbol{\theta}').$

- $\hat{H}_{\boldsymbol{\theta}}(.)$ exists if MC $P_{\boldsymbol{\theta}}$ is uniformly geometric ergodic [Douc et al., 2018].
- **Consequence**: implies the *error decomposition*

$$H(\boldsymbol{\theta}_t; X_{t+1}) - h(\boldsymbol{\theta}_t) = \hat{H}_{\boldsymbol{\theta}_t}(X_{t+1}) - P_{\boldsymbol{\theta}_t} \hat{H}_{\boldsymbol{\theta}_t}(X_{t+1})$$
$$= \underbrace{\hat{H}_{\boldsymbol{\theta}_t}(X_{t+1}) - P_{\boldsymbol{\theta}_t} \hat{H}_{\boldsymbol{\theta}_t}(X_t)}_{\text{Martingale with conditional 0-mean}} + P_{\boldsymbol{\theta}_t} \hat{H}_{\boldsymbol{\theta}_t}(X_t) - P_{\boldsymbol{\theta}_t} \hat{H}_{\boldsymbol{\theta}_t}(X_{t+1})$$

Main Results

Theorem

Let A1–A3 hold. Suppose that the step sizes satisfy

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

for $a, a' > 0$ and all $t \ge 0$, then

$$\mathbb{E}[\|\nabla V(\boldsymbol{\theta}_T)\|^2] \le \frac{2\big(V_{0,t} + C_{0,t} + (\sigma^2 L + C_\gamma)\sum_{k=0}^{t} \gamma_{k+1}^2\big)}{\sum_{k=0}^{t} \gamma_{k+1}},$$

where C_h, C_γ, $C_{0,t}$, $V_{0,t}$ are $\mathcal{O}(1)$ constants.

- *Stopping Criterion*: fix any $t \ge 1$ and $T \in \{0, \ldots, t\}$ is a discrete random variable with (see [Ghadimi and Lan, 2013])

$$\mathbb{P}(T = \ell) = \big(\textstyle\sum_{t=0}^{n} \gamma_{t+1}\big)^{-1} \gamma_{\ell+1}.$$

- If $\gamma_t = (2L(1 + C_h)\sqrt{t})^{-1}$, then $\mathbb{E}[\|\nabla V(\boldsymbol{\theta}_T)\|^2] = \mathcal{O}(\log t/\sqrt{t})$.

Proof: From L-smoothness of V, set $H(\boldsymbol{\theta}_t; X_{t+1}) = h(\boldsymbol{\theta}_t) + \boldsymbol{e}_{t+1}$, we have

$$\underbrace{V(\boldsymbol{\theta}_{k+1}) - V(\boldsymbol{\theta}_k)}_{\text{telescoping sum} \to \text{repeated terms are cancelled}} \leq$$

$$\underbrace{-\gamma_{k+1}\langle \nabla V(\boldsymbol{\theta}_k) \,|\, h(\boldsymbol{\theta}_k)\rangle + \frac{\gamma_{k+1}^2 L}{2}\|h(\boldsymbol{\theta}_k) + \boldsymbol{e}_{k+1}\|^2}_{\approx -\mathcal{O}(\gamma_{k+1}\|h(\boldsymbol{\theta}_k)\|^2) \leftarrow \text{controlled by biasedness}} \underbrace{-\gamma_{k+1}\langle \nabla V(\boldsymbol{\theta}_k) \,|\, \boldsymbol{e}_{k+1}\rangle}_{\text{good if summable!}}$$

Idea — there exists $\hat{H}_{\boldsymbol{\theta}}(\cdot)$ such that $\boldsymbol{e}_{k+1} = \hat{H}_{\boldsymbol{\theta}_k}(X_{k+1}) - P_{\boldsymbol{\theta}_k}\hat{H}_{\boldsymbol{\theta}_k}(X_k)$ (Poisson equation), consequently,

$$\sum_{k=0}^n \gamma_{k+1}\left\langle \nabla V(\boldsymbol{\theta}_k) \,|\, \hat{H}_{\boldsymbol{\theta}_k}(X_{k+1}) - P_{\boldsymbol{\theta}_k}\hat{H}_{\boldsymbol{\theta}_k}(X_{k+1})\right\rangle \equiv A_1 + A_2 + A_3 + A_4 + A_5$$

Martingale $\rightarrow$ $A_1 = \sum_{k=1}^n \gamma_{k+1}\left\langle \nabla V(\boldsymbol{\theta}_k) \,|\, \hat{H}_{\boldsymbol{\theta}_k}(X_{k+1}) - P_{\boldsymbol{\theta}_k}\hat{H}_{\boldsymbol{\theta}_k}(X_k)\right\rangle$

Smoothness $\rightarrow$ $A_2 = \sum_{k=1}^n \gamma_{k+1}\left\langle \nabla V(\boldsymbol{\theta}_k) \,|\, P_{\boldsymbol{\theta}_k}\hat{H}_{\boldsymbol{\theta}_k}(X_k) - P_{\boldsymbol{\theta}_{k-1}}\hat{H}_{\boldsymbol{\theta}_{k-1}}(X_k)\right\rangle$

Smoothness $\rightarrow$ $A_3 = \sum_{k=1}^n \gamma_{k+1}\left\langle \nabla V(\boldsymbol{\theta}_k) - \nabla V(\boldsymbol{\theta}_{k-1}) \,|\, P_{\boldsymbol{\theta}_{k-1}}\hat{H}_{\boldsymbol{\theta}_{k-1}}(X_k)\right\rangle$

Step size $\rightarrow$ $A_4 = \sum_{k=1}^n (\gamma_{k+1} - \gamma_k)\left\langle \nabla V(\boldsymbol{\theta}_k) \,|\, P_{\boldsymbol{\theta}_{k-1}}\hat{H}_{\boldsymbol{\theta}_{k-1}}(X_k)\right\rangle$

Finite number $\rightarrow$ $A_5 = \gamma_1\left\langle \nabla V(\boldsymbol{\theta}_0) \,|\, \hat{H}_{\boldsymbol{\theta}_0}(X_1)\right\rangle - \gamma_{n+1}\left\langle \nabla V(\boldsymbol{\theta}_n) \,|\, P_{\boldsymbol{\theta}_n}\hat{H}_{\boldsymbol{\theta}_n}(X_{n+1})\right\rangle$

Roadmap

1. General Convergence of (Biased) SA

2. Applications of Biased SA

3. Extension: Two-timescale SA

4. Conclusions and Perspectives

Policy Optimization: Average Reward Maximization

- **Goal:** Find a policy $\boldsymbol{\theta}$ (that governs the conditional prob. of taking action a' when in state s') to maximize an average reward:

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

  where $\upsilon_{\boldsymbol{\theta}}(s,a)$ is invariant distribution under $\boldsymbol{\theta}$ $\leftarrow$ difficult!.

- What is the gradient of $J(\boldsymbol{\theta})$ *w.r.t.* $\boldsymbol{\theta}$?

$$\nabla J(\boldsymbol{\theta}) = \lim_{T \to \infty} \mathbb{E}_{\boldsymbol{\theta}} \big[\mathsf{R}(S_T, A_T) \sum_{i=0}^{T-1} \nabla \log \Pi_{\boldsymbol{\theta}}(A_{T-i}; S_{T-i}) \big].$$

- Use a *biased* estimate of $\nabla J(\boldsymbol{\theta})$. Let $\lambda \in [0,1)$, consider the approximation [Baxter and Bartlett, 2001]:

$$\lim_{T \to \infty} \widehat{\nabla}_T J(\boldsymbol{\theta}) := \lim_{T \to \infty} \mathsf{R}(S_T, A_T) \sum_{i=0}^{T-1} \lambda^i \, \nabla \log \Pi_{\boldsymbol{\theta}}(A_{T-i}; S_{T-i}).$$

This Part: Design and analyze online policy gradient method via designing a *Markov chain* that converges to the limit.

Online Policy Gradient (PG)

Online policy gradient [Baxter and Bartlett, 2001, Tadić and Doucet, 2017]:

$$G_{t+1} = \lambda G_t + \nabla \log \Pi_{\boldsymbol{\theta}_n}(A_{t+1}; S_{t+1}) \,, \tag{2a}$$

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \gamma_{t+1} G_{t+1} \, \mathrm{R}(S_{t+1}, A_{t+1}) \,. \tag{2b}$$

- Let the joint state be $X_t = (S_t, A_t, G_t) \in \mathsf{S} \times \mathsf{A} \times \mathbb{R}^d$. Eq. (2b) is SA with the drift term:

$$H(\boldsymbol{\theta}_t; X_{t+1}) = G_{t+1} \, \mathrm{R}(S_{t+1}, A_{t+1})$$

- $\{X_t\}_{t \geq 1}$ forms a Markov chain and

$$h(\boldsymbol{\theta}) = \lim_{T \to \infty} \mathbb{E}_{\tau_T \sim \Pi_{\boldsymbol{\theta}}, \; S_1 \sim \overline{\Pi}_{\boldsymbol{\theta}}} \big[\widehat{\nabla}_T J(\boldsymbol{\theta}) \big].$$

- (A1)–(A3) can be verified under (PG1) exponential family (or soft-max) policy, bounded reward, etc.

Convergence Analysis

Corollary

Under PG1. Set $\gamma_t = (2c_1 L(1 + C_h)\sqrt{t})^{-1}$. For any $t \in \mathbb{N}$, the algorithm (2) finds a policy $\boldsymbol{\theta}_T$ with

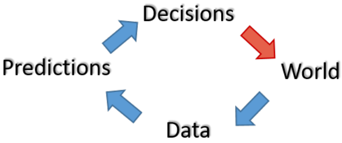
$$\mathbb{E}\big[\|\nabla J(\boldsymbol{\theta}_T)\|^2\big] = \mathcal{O}\Big((1-\lambda)^2\Gamma^2 + c(\lambda)\log t/\sqrt{t}\Big), \tag{3}$$

where $c(\lambda) = \mathcal{O}(\frac{1}{(1-\max\{\rho,\lambda\}}\big)^2)$ & expectation taken w.r.t. *T, (A_t, S_t).*

- *Variance-bias trade-off* with $\lambda \in (0,1)$: $\lambda \to 1$ reduces the bias, but increases the variance in static term as $c(\lambda) = \mathcal{O}((1-\lambda)^{-2})$.

Performative Prediction (PP)

- Predictions are used to support decisions and may **alter** the distribution of future observations.

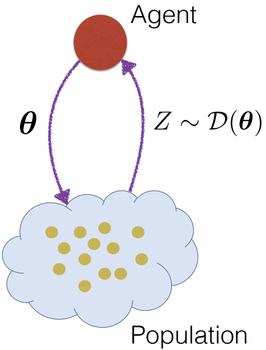


- **Empirical Risk Minimization (ERM)**: *static world* with i.i.d. data.
- But *decision* (classifier) can cause distribution shift in the *world*.
- **Example**: bank loan, spam filtering, healthcare systems, ...
- **Performative Prediction (PP)**[5]: capture the shift in decision-dependent data distribution

$$\underbrace{\min_{\boldsymbol{\theta}} \ \mathbb{E}_{Z\sim\mathcal{D}}\left[\ell(\boldsymbol{\theta};Z)\right]}_{\text{ERM}} \longrightarrow \underbrace{\min_{\boldsymbol{\theta}} \ \mathbb{E}_{Z\sim\mathcal{D}(\boldsymbol{\theta})}\left[\ell(\boldsymbol{\theta};Z)\right]}_{\text{PP}}.$$

[5] [Perdomo et al., 2020] J. Perdomo, T. Zrnic, C. Mendler-Dunner, M. Hardt. Performative prediction. ICML 2020.

Two Solution Concepts for PP



Performative Optimal Solution (PO):

$$\boldsymbol{\theta}^{\mathsf{PO}} \in \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \; \mathbb{E}_{Z \sim \mathcal{D}(\boldsymbol{\theta})} [\ell(\boldsymbol{\theta}; Z)] \,.$$

- Difficult $\because$ non-convexity, unknown $\mathcal{D}(\cdot)$, etc.
- Remedy: estimate $\nabla \mathcal{L}(\boldsymbol{\theta})$ [Izzo et al., 2021, Miller et al., 2021] $\rightarrow$ ✗ needs to know $\mathcal{D}(\cdot)$...

Performative Stable Solution (PS):

$$\boldsymbol{\theta}^{\mathsf{PS}} \in \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \; \mathbb{E}_{Z \sim \mathcal{D}(\boldsymbol{\theta}^{\mathsf{PS}})} [\ell(\boldsymbol{\theta}; Z)] \,.$$

- In general $\boldsymbol{\theta}^{\mathsf{PS}} \neq \boldsymbol{\theta}^{\mathsf{PO}}$. Fixed point of repeated risk minimization (RRM)

$$\boxed{\boldsymbol{\theta}^{+} \leftarrow \arg\min_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} \mathbb{E}_{Z \sim \mathcal{D}(\boldsymbol{\theta})} \big[\ell(\tilde{\boldsymbol{\theta}}; Z)\big] \,.}$$

- Algorithms based on RRM: [Perdomo et al., 2020, Mendler-Dünner et al., 2020] $\rightarrow$ ✓ no extra knowledge on $\mathcal{L}(\boldsymbol{\theta})$

Greedy Deployment Scheme

- Recall the **performative stable** solution:

$$\boldsymbol{\theta}_{PS} = \arg\min_{\boldsymbol{\theta}' \in \mathbb{R}^d} \mathbb{E}_{Z \sim \mathcal{D}(\boldsymbol{\theta}_{PS})}[\ell(\boldsymbol{\theta}'; Z)] \Leftrightarrow \mathbf{0} = \mathbb{E}_{Z \sim \mathcal{D}(\boldsymbol{\theta}_{PS})}[\nabla \ell(\boldsymbol{\theta}_{PS}; Z)].$$

(provided that $\boldsymbol{\theta}_{PS}$ exists).

Stochastic algorithm? **greedy deployment** scheme:

$$\begin{aligned} &\underline{\textit{Agent}}: && \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_{k+1} \nabla \ell(\boldsymbol{\theta}_t; Z_{t+1}) \leftarrow \text{SA scheme}, \\ &\underline{\textit{Population}}: && Z_{t+1} \sim \mathcal{D}(\boldsymbol{\theta}_t). \end{aligned}$$

- Mean field: $h(\boldsymbol{\theta}) = \mathbb{E}_{Z \sim \mathcal{D}(\boldsymbol{\theta})}[\nabla \ell(\boldsymbol{\theta}; Z)] \Rightarrow \boldsymbol{\theta}_{PS}$ is the SA's fixed point.
- **Fact**: $\ell(\cdot; Z)$ is strongly-convex + $\mathcal{D}(\boldsymbol{\theta})$ is 'insensitive' to $\boldsymbol{\theta}$, then[6]

$$\mathbb{E}[\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{PS}\|^2] = \mathcal{O}(1/t).$$

[6] [Mendler-Dünner et al., 2020] C. Mendler-Dunner, et al.. Stochastic optimization for performative prediction. NeurIPS 2020.

State-dependent Performative Prediction

- *Issue:* Greedy deployment in Mendler-Dünner et al. [2020]:

$$\begin{aligned} \underline{Agent}: \quad & \theta_{t+1} = \theta_k - \gamma_{t+1} \nabla \ell(\theta_t; Z_{t+1}), \\ \underline{Population}: \quad & Z_{t+1} \sim \mathcal{D}(\theta_t) \;\; \leftarrow \text{immediate adaptation} \end{aligned}$$

- Example: Loan applicants may take months to build up credit history to adapt to changes in classifier of bank.
- But both *agent* and *population* are possibly slow adapters[7]. ⇒ fully state dependent performative prediction.

This Part: *Greedy deployment comes naturally as SA with decision-dependent distribution. Is it stable? How to model it?*

[7] Brown et al. [2022] has similar setting but w/o sampling at learner.

State-dependent SA for PP

- **Idea**: models agents' adaptation via a *controlled Markov Chain*.
  - $\mathrm{P}_{\boldsymbol{\theta}} : \mathsf{Z} \times \mathcal{Z} \to \mathbb{R}_+$ = Markov kernel w/ stationary dist. $\mathcal{D}(\boldsymbol{\theta})$.

State-dependent Performative Prediction with SA

$$\underline{\textit{Population}}: \quad Z_{t+1} \sim \mathrm{P}_{\boldsymbol{\theta}_t}(Z_t, \cdot) \quad \text{(}\leftarrow \text{allows slow adaptation)}$$

$$\underline{\textit{Agent}}: \quad \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_{t+1} \nabla \ell(\boldsymbol{\theta}_t; Z_{t+1}) \text{ \& deploys } \boldsymbol{\theta}_{t+1}. \tag{4}$$

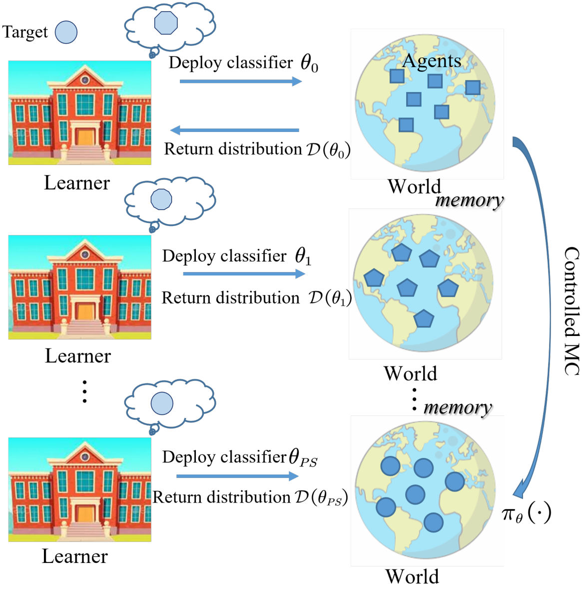
- **Example:** population runs SGD to adapt to $z \sim \mathcal{D}(\boldsymbol{\theta})$:

$$Z_{t+1} = Z_t + \alpha \nabla_z U(Z_t; \boldsymbol{\theta}_t, \zeta_{t+1}), \quad \leftarrow U = \text{utility fct.}$$

Observation: Agent's updates (4) is biased SA with

$$H(\boldsymbol{\theta}_t; X_{t+1}) = \nabla \ell(\boldsymbol{\theta}_t; Z_{t+1})$$

Illustration



Assumptions

(PP1). We assume that $\ell(\boldsymbol{\theta}; Z)$ is μ-strongly convex, L-smooth, and the distribution $\mathcal{D}(\boldsymbol{\theta})$ satisfies $\epsilon-$sensitivity (W_1 denotes Wasserstein-1 distance)

$$W_1(\mathcal{D}(\boldsymbol{\theta}), \mathcal{D}(\boldsymbol{\theta}')) \leq \epsilon \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|, \ \forall \ \boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathbb{R}^d,$$

- PP1 is **sensitivity** *w.r.t.* $\mathcal{D}(\boldsymbol{\theta})$ to $\boldsymbol{\theta}$ [Mendler-Dünner et al., 2020].
- *Strategic* population with a linear utility function, $Z \sim \mathcal{D}(\boldsymbol{\theta})$ if

$$Z = Z_0 + \epsilon \, \boldsymbol{\theta}, \quad Z_0 \sim \mathcal{D}_0 \text{ - base/intrinsic distribution}$$

(PP2). σ-perturbation with sampled gradient

$$\sup_{z \in Z} \|\nabla \ell(\boldsymbol{\theta}; z) - \nabla f(\boldsymbol{\theta}; \boldsymbol{\theta}_{PS})\| \leq \sigma \left(1 + \|\boldsymbol{\theta} - \boldsymbol{\theta}_{PS}\|\right).$$

- PP2 allows $\nabla \ell(\boldsymbol{\theta}; z) = \mathcal{O}(1 + \|\boldsymbol{\theta} - \boldsymbol{\theta}_{PS}\|)$ - compatible with strongly convex loss.

Convergence of SA for Performative Prediction

Theorem

Under PP1–PP2, $\mathrm{P}_{\boldsymbol{\theta}}$ satisfies A3 and $\ell(\cdot; Z)$ is μ-strongly convex. Let $\epsilon < \frac{\mu}{L}$, non-increasing step sizes

$$\frac{\gamma_{t-1}}{\gamma_t} \leq 1 + \frac{\gamma_t(\mu - L\epsilon)}{4}, \quad \gamma_t \leq \min\left\{ \frac{\mu - L\epsilon}{2L^2}, \frac{\mu - L\epsilon}{2\mathcal{C}_2}, \frac{\min\{(\mu - L\epsilon)/3, 3\widehat{L}_P\}}{\mathcal{C}_3 + 3\widehat{L}_P(\mu - L\epsilon)}, \frac{1}{6\widehat{L}_P} \right\}.$$

For any $k \geq 1$, there exists $\mathbb{C}$ where it holds

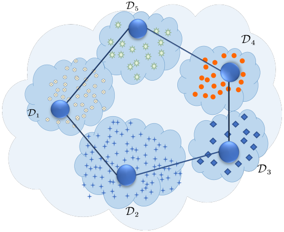
$$\mathbb{E}[\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{PS}\|^2] \leq \underbrace{\prod_{i=1}^{t}\left(1 - \gamma_i \frac{\mu - L\epsilon}{2}\right)\|\boldsymbol{\theta}_0 - \boldsymbol{\theta}_{PS}\|^2}_{\text{Transient}} + \underbrace{\mathbb{C}\,\gamma_t}_{\text{Fluctuation}} .$$

- Convergence needs $\epsilon < \mu/L$ (similar to [Mendler-Dünner et al., 2020]) + Step size constrained by mixing time of MC.
- Oscillation of stochastic gradient σ, mixing time of MC $\widehat{L}$ appear in fluctuation term $\mathbb{C}$.
- **Extensions**: training with NN / non-convex $\ell(\cdot)$ [Mehrnaz Mofakhami, 2023, Zhao, 2022], performative RL [Mandal et al., 2022].

Extension: Multi-agent Performative Prediction

- Decentralized learning uses **cooperative agents** to solve the learning problem[8]. *Example*: training from clinical data.

$$\min_{\boldsymbol{\theta}_i \in \mathbb{R}^d} \; \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{Z_i \sim \mathcal{D}_i(\boldsymbol{\theta}_i)}\left[\ell(\boldsymbol{\theta}_i; Z_i)\right]$$
$$\text{s.t.} \;\; \boldsymbol{\theta}_i = \boldsymbol{\theta}_j, \; \forall\, (i,j) \in E.$$

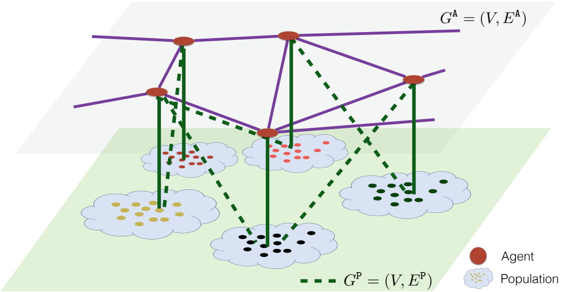


- *Benefits of consensus*: single agent case needs $\epsilon < \mu/L$ vs multi agent case: $\epsilon_{avg} < \mu/L$.
- Improved robustness to sensitive local distribution shifts
- The corresponding PS solution $\boldsymbol{\theta}_{PS}$ can be found using a slight modification to decentralized SGD.

[8][Li et al., 2022] **Q. Li, C.-Y. Yau, H.-T.**. Multi-agent performative prediction with greedy deployment and consensus seeking agents. In NeurIPS, 2022.

Extension: Performative Prediction Game

Setting[9]: *Multiplex network game* [Gómez-Gardenes et al., 2012] extension of PP with n agents, each agent i interacts with a local population $\mathcal{D}_i(\cdot)$.



- Agent network $\mathcal{G}^{\mathtt{A}}$: agent i decision depends on its neighbors.
- Population network $\mathcal{G}^{\mathtt{P}}$: $\mathcal{D}_i(\cdot)$ react to decisions $\boldsymbol{\theta}_i$ and neighbors.
- Networks affect the equilibrium solution $\Rightarrow$ graph learning.

[9] **X. Wang, C.-Y. Yau, H.-T.**. Network Effects on Performative Prediction Games. in ICML 2023 (preprint: available soon).

Roadmap

Bilevel Optimization

- Many problems can be described as **bilevel optimization**:

$$\min_{x \in X \subseteq \mathbb{R}^{d_1}} \quad \ell(x) := f(x, y^\star(x)) \qquad \text{s.t.} \quad y^\star(x) \in \arg\min_{y \in \mathbb{R}^{d_2}} \; g(x, y), \tag{Bi}$$

- **Upper-level** = *leader* / decision maker, **lower-level** = *follower*.
- Related to *mathematical program with equilibrium constraint (MPEC)* Luo et al. [1996], *stackelberg game* Stackelberg [1952].

- **Applications**: meta learning, policy optimization, etc..

This Part: f, g are *stochastic* – $f(x, y) = \mathbb{E}_{\xi \sim \mathcal{D}}[f(x, y, \xi)]$.
$\Rightarrow$ Consider tackling upper level by SA: samples $y^\star(x)$ are **decision-dependent**: there are more structure than previous part.

Motivation: Policy Optimization via Actor Critic

- Consider **tabular policy** given by $\pi : \mathsf{S} \times \mathsf{A} \to \mathbb{R}_+$ with $|\mathsf{S}|, |\mathsf{A}| < \infty$.
- Let ρ_0 be init. distribution, the γ-discounted reward[10] of π is:

$$\mathbb{E}_\pi[Q^\pi(S, A)] = \mathbb{E}_{S \sim \rho_0}\left[\langle Q^\pi(S, \cdot) \,|\, \pi(\cdot|S)\rangle\right],$$

$$\text{with} \quad Q^\pi(S, A) = \mathbb{E}_\pi\left[\textstyle\sum_{t \geq 0} \gamma^t R(S_t, A_t) | S_0 = S, A_0 = A\right]$$

- Note $Q^\pi(S, A)$ is γ-discounted reward (Q-function) given init. (S, A).
- With **fixed** π, $Q^\pi(S, A)$ can be evaluated by solving Bellman equation; or through linear approximation $Q^\pi(S, A) \approx \langle \theta^\star(\pi) \,|\, \phi(S, A)\rangle$.

A **Bilevel Optimization** problem:

$$\min_{\pi \in X \subseteq \mathbb{R}^{|S| \times |A|}} \ell(\pi) = -\langle Q_{\theta^\star(\pi)}, \pi \rangle_{\rho_0} \qquad \text{(Actor)}$$

$$\text{s.t.} \quad \theta^\star(\pi) \in \arg\min_{\theta \in \mathbb{R}^d} \; \tfrac{1}{2} \| Q_\theta - R - \gamma P^\pi Q_\theta \|^2_{\mu^\pi \otimes \pi}. \qquad \text{(Critic)}$$

[10] In Part II, we have considered average reward with paramterized policy.

Tackling the Bilevel Problem (Bi)

- Recall the bi-level optimization problem:

$$\min_{x \in X \subseteq \mathbb{R}^{d_1}} \ell(x) \iff \begin{array}{rl} \min\limits_{x \in X \subseteq \mathbb{R}^{d_1}} & \ell(x) := f(x, y^\star(x)) \\ \text{s.t.} & y^\star(x) \in \arg\min\limits_{y \in \mathbb{R}^{d_2}} g(x, y), \end{array}$$

- The gradient of $\ell(x)$ is:

$$\nabla_x \ell(x) = \nabla_x f(x, y^\star) - \nabla^2_{xy} g(x, y^\star)[\nabla^2_{yy} g(x, y^\star)]^{-1} \nabla_y f(x, y^\star)$$

Stationary Condition: (Bi) can be tackled by finding $(x^\star, y^\star)$ s.t.

$$F(x, y) = 0, \quad G(x, y) = \nabla_y g(x, y) = 0$$

where $F(x, y) = \nabla_x f(x, y) - \nabla^2_{xy} g(x, y)[\nabla^2_{yy} g(x, y)]^{-1} \nabla_y f(x, y)$

Finding Fixed Points with Stochastic Samples

- We only have **stochastic samples** and the problems are **coupled**.
- Let ξ_{k+1} denotes the random 'seed' at iteration k, and $F(\cdot;\xi_{k+1})$, $G(\cdot;\xi_{k+1})$ denote the stochastic samples of F, G, respectively.
- If x **is fixed** and under suitable conditions, the recursion

$$y_{k+1} = y_k + \beta_k G(x, y_k; \xi_{k+1}) \stackrel{k\to\infty}{\longrightarrow} y^\star(x) \text{ s.t. } G(x, y^\star(x)) = 0.$$

- Furthermore, the recursion

$$x_{k+1} = x_k + \alpha_k F(x_k, y^\star(x_k); \xi_{k+1}) \stackrel{k\to\infty}{\longrightarrow} x^\star \text{ s.t. } F(x^\star, y^\star(x^\star)) = 0.$$

- If one could run the two recursions $\Rightarrow$ fixed point, but the y_k recursion requires x to be fixed; and x_k recursion requires $y^\star(x_k)$.

Suggesting a double-loop algorithm? e.g., [Ghadimi and Wang, 2018].

Two Timescale Stochastic Approximation (TTSA)

- Consider a **single-loop, two timescale** algorithm [Borkar, 1997]:

$$x_{k+1} = x_k + \alpha_k F(x_k, y_k; \xi_{k+1})$$
$$y_{k+1} = y_k + \beta_k G(x_k, y_k; \xi_{k+1})$$

- We require that

$$\lim_{k \to \infty} \frac{\alpha_k}{\beta_k} = 0$$

  x-update is at slow timescale; while y-update is at fast timescale.

- **Intuition**: when updating y_k, as $\alpha_k \ll \beta_k$, then x_k is *almost static*; when updating x_k, the used y_k have *almost converged* to $y^\star(x_k)$.

TTSA for Tackling (Bi): The Algorithm

TTSA Algorithm for (Bi)

Follow the recursion:

$$\begin{aligned} x_{k+1} &= x_k - \alpha_k h_f^k \qquad [h_f^k \approx F(x^k, y^k)] \\ y_{k+1} &= y_k - \beta_k \nabla_y g(x_k, y_k; \zeta_{k+1}) \end{aligned} \qquad \text{(TTSA-Bi)}$$

- x_k update uses **decision-dependent data** via y_k driven by x_{k-1}.
- **Two timescale** step sizes to *balance* upper and lower level updates.

- **Challenge**: *easy* to estimate $G(\cdot) = \nabla_y g(\cdot)$, but $F(\cdot)$ is non-trivial since

$$F(x, y) = \nabla_x f(x, y) - \nabla^2_{xy} g(x, y) \quad \underbrace{[\nabla^2_{yy} g(x, y)]^{-1}}_{\text{can't replace by } \nabla^2_{yy} g(x, y; \zeta)} \quad \nabla_y f(x, y)$$

Biased estimate is possible.

This Work

Characterize the **rate of convergence** for TTSA when:

- the inner objective $g(x, y)$ is strongly convex in y, and
- the outer objective $\ell(x)$ is *smooth, convex, strongly convex*.

$\ell(x)$	Constraint	Step Size (α_k, β_k)	Rate (outer)	Rate (Inner)
SC	$X \subseteq \mathbb{R}^{d_1}$	$\mathcal{O}(k^{-1})$, $\mathcal{O}(k^{-2/3})$	$\mathcal{O}(K^{-2/3})$	$\mathcal{O}(K^{-2/3})$
WC	$X \subseteq \mathbb{R}^{d_1}$	$\mathcal{O}(K^{-3/5})$, $\mathcal{O}(K^{-2/5})$	$\mathcal{O}(K^{-2/5})$	$\mathcal{O}(K^{-2/5})$

Prior Works — many and

- *Linear TTSA* $\approx$ *solving* **quadratic** *upper/lower level*
  - Dalal et al. [2018, 2019] obtained h.p. bounds with a **projection** step.
- *Finite-time Analysis of Bilevel Stochastic Optimization*
  - [Couellan and Wang, 2016, Ghadimi and Wang, 2018] – double loop SA & recently [Yang et al., 2021, Chen et al., 2022a, Guo and Yang, 2021].
  - constrained bilevel problem [Xiao et al., 2022], relax strong convexity [Chen et al., 2023].

Why two-timescale?

- In the upper-level update,

$$x_{k+1} = x_k - \alpha_k h_f^k \leftarrow \text{note } h_f^k \approx F(x,y) \neq \nabla \ell(x)$$

- We recall that $\|F(x,y) - \nabla \ell(x)\| = \mathcal{O}(\|y - y^\star(x)\|)$.
- Need $\alpha_k \leq c_0 \beta_k^{3/2}$ to balance the errors, leading to the step sizes

$$\text{strongly convex } \ell(x): \quad \alpha_k \asymp k^{-1}, \beta_k \asymp k^{-2/3}$$
$$\text{weakly convex } \ell(x): \quad \alpha_k \asymp K^{-3/5}, \beta_k \asymp K^{-2/5}$$

- The convergence rate is **limited** by the 'faster' timescale[11].
- For 1-level problem, even naive SGD achieves $\mathbb{E}[\tilde{\Delta}_x^K] = \mathcal{O}(1/K^{1/2})$.

[11] [Kaledin et al., 2020] M. Kaledin et al. Finite time analysis of linear two-timescale stochastic approximation with markovian noise. COLT 2020.

Main Results (Weakly Convex ℓ)

Theorem

Under TT1, TT2, suppose that $\mu_\ell \in \mathbb{R}$. Set $\mathsf{K} \sim \mathcal{U}\{0, ..., K-1\}$ and $\alpha_k \asymp K^{-3/5}, \beta_k \asymp K^{-2/5}$. For sufficiently large $K \geq 1$, it holds

$$\mathbb{E}[\|\nabla \ell(x^{\mathsf{K}})\|^2] \lesssim \left[L^2\left(\Delta^0 + \frac{\sigma^2}{\mu_g^2}\right) + \mu_g \sigma^2\right] \frac{K^{-\frac{2}{5}}}{|\mu_\ell|^2},$$

$$\mathbb{E}[\|y^{\mathsf{K}} - y^\star(x^{\mathsf{K}-1})\|^2] \lesssim \left[\frac{\Delta^0}{\mu_g} + \frac{\sigma^2}{\mu_g^2} + \frac{\mu_g \sigma^2}{L^2}\right] K^{-\frac{2}{5}},$$

where Δ^0 depends on the initialization, the inequality is up to constants not depending on k (exact expressions can be found in the paper)

- **Consequence**: we get $\mathbb{E}[\tilde{\Delta}_x^{\mathsf{K}}] = \mathcal{O}(1/K^{2/5})$, $\mathbb{E}[\Delta_y^{\mathsf{K}}] = \mathcal{O}(1/K^{2/5})$.
- **Note**: $\tilde{\Delta}_x^{\mathsf{K}}$ is a stationarity measure for x^{K} related to Moreau envelope.
- Actor-critic requires slightly different algorithm than (TTSA-Bi) thru exploiting structure; but similar analysis applies.
- Single-timescale algo.? [Khanduri et al., 2021, Chen et al., 2022a].

Agenda

Summary

We have studied variants of SA with decision dependent data:

$$\underline{\text{SA}}: \quad \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_{t+1} H(\boldsymbol{\theta}_t; X_{t+1}),$$

where X_{t+1} is not i.i.d., and depends on $\boldsymbol{\theta}_t$ (via a controlled MC).

- *General SA* with possibly non-gradient $H(\boldsymbol{\theta}; X)$:
  $\Rightarrow$ convergence to stationary point $\mathbb{E}[\|h(\boldsymbol{\theta}_T)\|^2] = \mathcal{O}(\log T/\sqrt{T})$.
  $\Rightarrow$ application to online policy gradient.
- *Performative Prediction* through SA:
  $\Rightarrow$ modelling stateful population through controlled MC.
  $\Rightarrow$ convergence to PS solution $\mathbb{E}[\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{PS}\|^2] = \mathcal{O}(1/t)$.
- *Bilevel optimization* via TTSA:
  $\Rightarrow$ utilizes two timescales for coupled SAs & application to actor-critic.
  $\Rightarrow$ convergence rates to stationary solution.

Take-away Point: SA with data-dependent distribution is everywhere. Don't panic when your application has it.

Perspectives

SA with decision dependent data:

$$\underline{\text{SA}}: \quad \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_{t+1} H(\boldsymbol{\theta}_t; X_{t+1}),$$

where X_{t+1} is not i.i.d., and depends on $\boldsymbol{\theta}_t$ via a controlled MC.

Theory: also see our recent overview [Dieuleveut et al., 2023],

- Current results require 'strong' assumptions on MC which usually makes sense for finite-state space only, see [Durmus et al., 2021b].
- Strong convergence, e.g., with high probability [Durmus et al., 2021a].
- Avoid saddle point in non-convex problems? [Lee et al., 2019]

Applications/Algorithmic:

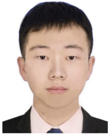
- Decentralized & federated learning; see [Wai, 2020].
- Beyond reinforcement learning & performative prediction — Langevin Monte-carlo [De Bortoli et al., 2021], inducing equilibrium thru TTSA and pricing [Liu et al., 2022], etc.

Most importantly, thanks to ...

Belhal Karimi
(Baidu Research)

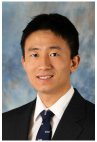
Blazej Miasojedow
(U of Warsaw)

Eric Moulines
(Ecole Polytechnique)



Qiang Li
(CUHK)

Chung-Yiu Yau
(CUHK)



Mingyi Hong
(UMN)

Zhuoran Yang
(Yale)

Zhaoran Wang
(Northwesthern)

Thank you! Questions?

For more info: http://www1.se.cuhk.edu.hk/~htwai/

References I

Yves F Atchadé, Gersende Fort, and Eric Moulines. On perturbed proximal gradient algorithms. *The Journal of Machine Learning Research*, 18(1):310–342, 2017.

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

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.

Vivek S Borkar. Stochastic approximation with two time scales. *Systems & Control Letters*, 29(5): 291–294, 1997.

Gavin Brown, Shlomi Hod, and Iden Kalemaj. Performative prediction in a stateful world. In *AISTATS*, 2022.

Lesi Chen, Jing Xu, and Jingzhao Zhang. On bilevel optimization without lower-level strong convexity. *arXiv preprint arXiv:2301.00712*, 2023.

Shuhang Chen, Adithya Devraj, Ana Busic, and Sean Meyn. Explicit mean-square error bounds for monte-carlo and linear stochastic approximation. In *International Conference on Artificial Intelligence and Statistics*, pages 4173–4183. PMLR, 2020.

Tianyi Chen, Yuejiao Sun, Quan Xiao, and Wotao Yin. A single-timescale method for stochastic bilevel optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 2466–2488. PMLR, 2022a.

Yudong Chen, Qiaomin Xie, et al. Bias and extrapolation in markovian linear stochastic approximation with constant stepsizes. *arXiv preprint arXiv:2210.00953*, 2022b.

Nicolas Couellan and Wenjuan Wang. On the convergence of stochastic bi-level gradient methods. *Optimization*, 2016.

References II

Gal Dalal, Gugan Thoppe, Balázs Szörényi, and Shie Mannor. Finite sample analysis of two-timescale stochastic approximation with applications to reinforcement learning. In *Conference On Learning Theory*, pages 1199–1233, 2018.

Gal Dalal, Balazs Szorenyi, and Gugan Thoppe. A tale of two-timescale reinforcement learning with the tightest finite-time bound. *arXiv preprint arXiv:1911.09157*, 2019.

Valentin De Bortoli, Alain Durmus, Marcelo Pereyra, and Ana F Vidal. Efficient stochastic optimisation by unadjusted langevin monte carlo. *Statistics and Computing*, 31(3):1–18, 2021.

Aymeric Dieuleveut, Gersende Fort, Eric Moulines, and Hoi-To Wai. Stochastic approximation beyond gradient for signal processing and machine learning. *arXiv preprint arXiv:2302.11147*, 2023.

Thinh T Doan. Finite-time analysis of markov gradient descent. *IEEE Transactions on Automatic Control*, 2022.

Randal Douc, Eric Moulines, Pierre Priouret, and Philippe Soulier. *Markov chains*. Springer, 2018.

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

Alain Durmus, Eric Moulines, Alexey Naumov, Sergey Samsonov, Kevin Scaman, and Hoi-To Wai. Tight high probability bounds for linear stochastic approximation with fixed stepsize. *Advances in Neural Information Processing Systems*, 34, 2021a.

Alain Durmus, Eric Moulines, Alexey Naumov, Sergey Samsonov, and Hoi-To Wai. On the stability of random matrix product with markovian noise: Application to linear stochastic approximation and td learning. In *Conference on Learning Theory*, pages 1711–1752. PMLR, 2021b.

Alain Durmus, Eric Moulines, Alexey Naumov, and Sergey Samsonov. Finite-time high-probability bounds for polyak-ruppert averaged iterates of linear stochastic approximation. *arXiv preprint arXiv:2207.04475*, 2022.

References III

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

Saeed Ghadimi and Mengdi Wang. Approximation methods for bilevel programming. *arXiv preprint arXiv:1802.02246*, 2018.

Jesús Gómez-Gardenes, Irene Reinares, Alex Arenas, and Luis Mario Floría. Evolution of cooperation in multiplex networks. *Scientific Reports*, 2(1):1–6, 2012.

Zhishuai Guo and Tianbao Yang. Randomized stochastic variance-reduced methods for stochastic bilevel optimization. *arXiv preprint arXiv:2105.02266*, 2021.

Mingyi Hong, Hoi-To Wai, Zhaoran Wang, and Zhuoran Yang. A two-timescale framework for bilevel optimization: Complexity analysis and application to actor-critic. *SIOPT*, 2023.

Zachary Izzo, Lexing Ying, and James Zou. How to learn when data reacts to your model: Performative gradient descent. In *ICML*, 2021.

Maxim Kaledin, Eric Moulines, Alexey Naumov, Vladislav Tadic, and Hoi-To Wai. Finite time analysis of linear two-timescale stochastic approximation with markovian noise. In *Conference on Learning Theory*, pages 2144–2203. PMLR, 2020.

Belhal Karimi, Blazej Miasojedow, Eric Moulines, and Hoi-To Wai. Non-asymptotic Analysis of Biased Stochastic Approximation Scheme. In *Conference on Learning Theory*, 2019.

Prashant Khanduri, Siliang Zeng, Mingyi Hong, Hoi-To Wai, Zhaoran Wang, and Zhuoran Yang. A near-optimal algorithm for stochastic bilevel optimization via double-momentum. *arXiv preprint arXiv:2102.07367*, 2021.

Vijay Konda and John Tsitsiklis. Actor-critic algorithms. *Advances in neural information processing systems*, 12, 1999.

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

References IV

Jason D Lee, Ioannis Panageas, Georgios Piliouras, Max Simchowitz, Michael I Jordan, and Benjamin Recht. First-order methods almost always avoid strict saddle points. *Mathematical programming*, 176(1):311–337, 2019.

Qiang Li and Hoi-To Wai. State dependent performative prediction with stochastic approximation. In *International Conference on Artificial Intelligence and Statistics*, pages 3164–3186. PMLR, 2022.

Qiang Li, Chung-Yiu Yau, and Hoi-To Wai. Multi-agent performative prediction with greedy deployment and consensus seeking agents. In *NeurIPS*, 2022.

Boyi Liu, Jiayang Li, Zhuoran Yang, Hoi-To Wai, Mingyi Hong, Yu Marco Nie, and Zhaoran Wang. Inducing equilibria via incentives: Simultaneous design-and-play finds global optima. In *NeuIPS*, 2022.

Zhi-Quan Luo, Jong-Shi Pang, and Daniel Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996. doi: 10.1017/CBO9780511983658.

Debmalya Mandal, Stelios Triantafyllou, and Goran Radanovic. Performative reinforcement learning. *arXiv preprint arXiv:2207.00046*, 2022.

Gauthier Gidel Mehrnaz Mofakhami, Ioannis Mitliagkas. Performative prediction with neural networks. In *AISTATS*, 2023.

Celestine Mendler-Dünner, Juan Perdomo, Tijana Zrnic, and Moritz Hardt. Stochastic optimization for performative prediction. *Advances in Neural Information Processing Systems*, 33:4929–4939, 2020.

John P Miller, Juan C Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. In *International Conference on Machine Learning*, pages 7710–7720. PMLR, 2021.

References V

Wenlong Mou, Chris Junchi Li, Martin J Wainwright, Peter L Bartlett, and Michael I Jordan. On linear stochastic approximation: Fine-grained polyak-ruppert and non-asymptotic concentration. *arXiv preprint arXiv:2004.04719*, 2020.

Juan C. Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *ICML*, 2020.

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

H. Van Stackelberg. *The theory of market economy*. Oxford University Press, 1952.

Tao Sun, Yuejiao Sun, and Wotao Yin. On Markov chain gradient descent. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 9918–9927. Curran Associates, Inc., 2018. URL http://papers.nips.cc/paper/8195-on-markov-chain-gradient-descent.pdf.

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

Hoi-To Wai. On the convergence of consensus algorithms with markovian noise and gradient bias. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 4897–4902. IEEE, 2020.

Quan Xiao, Han Shen, Wotao Yin, and Tianyi Chen. Alternating implicit projected sgd and its efficient variants for equality-constrained bilevel optimization. *arXiv preprint arXiv:2211.07096*, 2022.

Junjie Yang, Kaiyi Ji, and Yingbin Liang. Provably faster algorithms for bilevel optimization. *arXiv preprint arXiv:2106.04692*, 2021.

Yulai Zhao. Optimizing the performative risk under weak convexity assumptions. *arXiv preprint arXiv:2209.00771*, 2022.