Multiscale replay: A robust algorithm for stochastic variational inequalities with a Markovian buffer

Milind Nakul * Tianjiao Li * Ashwin Pananjady *, †

* Industrial & Systems Engineering, Georgia Institute of Technology
† Electrical & Computer Engineering, Georgia Institute of Technology

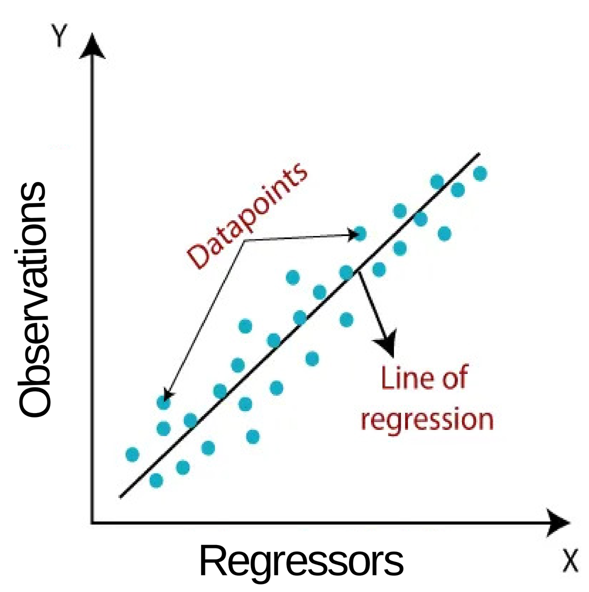
Problem of interest

Variational Inequality problem:

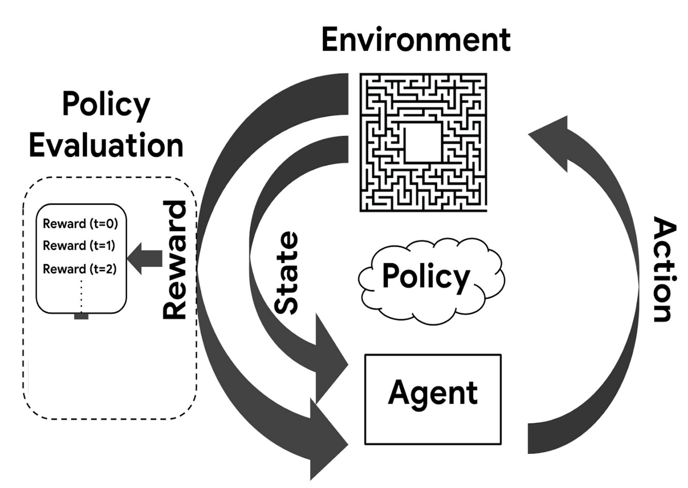
$$\text{Find } x^* \text{ such that } \langle F(x^*), x - x^* \rangle \geq 0. \quad \text{for all } x \in X.$$

- X: Closed convex feasible region. The set of optimal solutions X^* is nonempty.
- $F : X \to \mathbb{R}^n$ is a *Lipschitz continuous* operator, i.e., for some $L \geq 0$,
$$\|F(x_1) - F(x_2)\| \leq L\|x_1 - x_2\|, \quad \text{for all } x_1, x_2 \in X.$$
- F: satisfies a generalized strong monotonicity condition, i.e., for some $\mu > 0$,
$$\langle F(x), x - x^* \rangle \geq \mu \|x - x^*\|^2, \quad \text{for all } x \in X.$$

VIs with monotone operators have applications for policy evaluation in reinforcement learning (RL) and nonlinear signal estimation in generalized linear models (GLMs).



Signal Estimation in GLMs
$F(x) =_{a\sim\Pi} [af(a^T x)] -_{a\sim\Pi} [af(a^T x^\star)]$



Policy Evaluation in RL
$F(\theta) = \Psi\Pi(\Psi^T\theta - r - \gamma P \Psi^T \theta)$

Observation Models

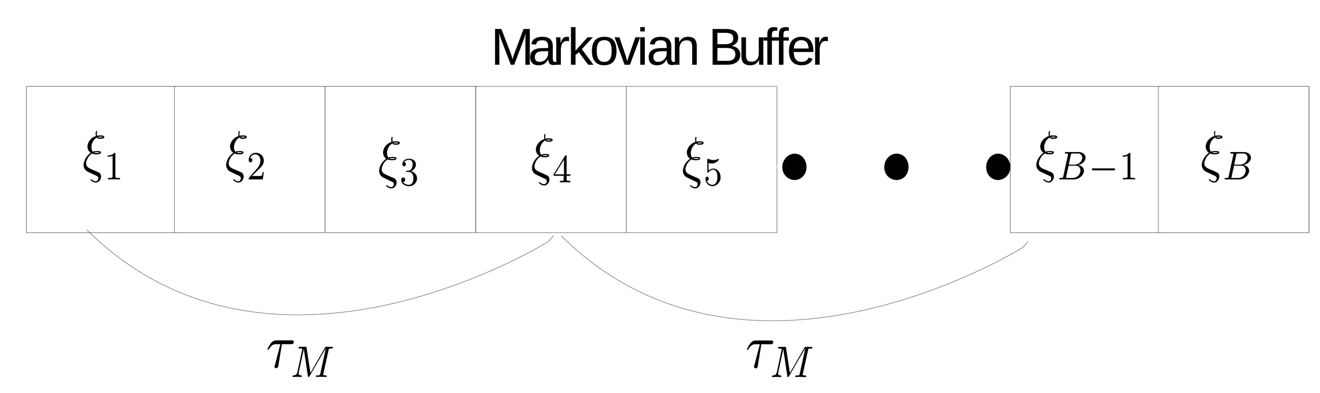
- We operate under a stochastic observation model in which we obtain inexact information about the operator F.
- The stochastic oracle generates an estimator $\widetilde{F}(x, \xi) \in \mathbb{R}^n$ for a query point $x \in X$, ξ being a random variable in Ξ with marginal distribution Π, and that
$$F(x) = \mathbb{E}[\widetilde{F}(x, \xi)] = \int_{\xi \in \Xi} \widetilde{F}(x, \xi) d\Pi(\xi). \tag{1}$$
- We have two observation models, the "i.i.d."" model and the Markovian noise model.
- **i.i.d.** model: The samples $\{\xi_0, \xi_1, ...\}$ are drawn i.i.d. from Π, so that we have access to the independent operators $\widetilde{F}(\cdot, \xi_1), \widetilde{F}(\cdot, \xi_2), \ldots$.
- **Markovian** model: We consider the more challenging *Markovian* setting, where $\{\xi_t : t = 1, 2, \ldots\}$ is a Markov process defined on Ξ, and Π denotes the unique stationary distribution.

Challenge with Markovian samples

- **Dependence:** Working with Markovian data iterative algorithms can get dependent on the Markov chain leading to sub-optimal convergence.
- In order to break this correlation a **memory buffer** was introduced.
- Random sampling from the buffer breaks the harmful correlation.

Markovian Buffer

- We assume that we are given a Markovian *buffer* of a fixed size from this Markov process which has the form:
$$\xi^B = \{\xi_k, 1 \leq k \leq B\}. \tag{2}$$
- Our goal is to design an algorithm that carefully utilizes the samples most effectively. In RL this buffer is called the "experience replay" [4].



Existing algorithm: Conditional Temporal Difference (CTD)

- The authors in [1] explored the following **stochastic approximation (SA) method with skipping** for solving stochastic VIs:
$$x_{t+1} = \arg\min_{x\in X} \eta \langle \widetilde{F}(x_t, \xi_{t\tau_M}), x\rangle + \frac{1}{2}\|x_t - x\|^2, \text{ for } t = 0, 1, ..., \tag{3}$$
where τ_M is the effective mixing time.
- The update achieves **i.i.d. rate of convergence for the stochastic error of** $\mathcal{O}(\frac{\sigma^2 \log k}{k})$, where k is the total iterations and σ^2 is the noise level.
- **Implementation requires mixing information**, which is computationally expensive and sample consuming to obtain.

Goal of the work

Design an algorithm that achieves the i.i.d. convergence rate for the stochastic error without estimating the mixing time.

Algorithm: Multiscale Experience Replay (MER)

We propose the following algorithm:

Algorithm 1 Multiscale Experience Replay (MER)

Input: Memory Buffer $\{\xi_1, \xi_2, \ldots, \xi_B\}$, Total number of epochs K.
for $k = 1, \ldots, K$ **do**
  Re-initialization of the variable $x_1^{(k)}$.
  Define the replay gap, $\tau_k = \frac{B}{2^k}$.
  Number of samples used in kth epoch $T_k = 2^k$.
  for $t = 1, \ldots, T_k$ **do**
    Using the sample $\xi_{t\tau_k}$, perform one step of SA:
$$x_{t+1}^{(k)} = \arg\min_{x\in X} \eta_k \langle \widetilde{F}(, \xi_{t\tau_k}), x\rangle + \frac{1}{2}\|x - x_t^{(k)}\|^2 \tag{4}$$
  end for
end for

- MER algorithm optimizes sample usage by keeping **the "active" samples used in the SA updates as far apart as possible**.
- Implementation of **MER requires no mixing information**.

Theorem (Convergence guarantees for MER):

Let $\{x_{t+1}^{(k)}\}$ be generated by the MER method and $\bar{\sigma}^2$ be the effective noise level. With proper parameters $\{\eta_k\}$ and a **condition on the buffer size** $B \geq \widetilde{\mathcal{O}}\left(\frac{\tau_M \bar{L}^2}{\mu^2}\right)$, **the deterministic error converges at a linear rate** and we have the following two regimes for the convergence of stochastic error:

1. $\tau_k \geq \tau_M$: the stochastic error converges at the rate of $\mathcal{O}\left(\frac{\bar{\sigma}^2 \log T_k}{\mu^2 T_k}\right)$.
2. $\tau_k \leq \tau_M$: the stochastic error converges at the rate of $\mathcal{O}\left(\frac{\bar{\sigma}^2 \log T_k}{\mu^2 T_k}\left(\frac{\tau_M}{\tau_k} + 1\right)\right)$.

- Epochs where **the replay gap is greater than the effective mixing time the MER algorithm recovers the i.i.d. rate of convergence for the stochastic error.**
- Epochs when replay gap is smaller than the effective mixing time **MER outperforms the standard Markovian SA in [1] by a multiplicative factor of the replay gap** τ_k.

Applications

- **GLMs:** For finding an ϵ-optimal solution for signal estimation in GLMs, where ϵ is the error, the required iteration complexity is:
$$\mathcal{O}\left(\max\left((\tau_M/\tau_k + 1)\frac{\bar{L}^2}{\mu_f^2}\log\frac{(\|x_1^{(k)} - x^*\|^2 + 1}{\epsilon}, \frac{\bar{\sigma}^2(\tau_M/\tau_k + 1)}{\epsilon\mu_f^2}\log\frac{1}{\epsilon}\right)\right).$$
- **Policy Evaluation in RL:** We use MER algorithm for policy evaluation in RL with function approximation. When the buffer size is large enough, MER outperforms the well known TD-algorithm by a multiplicative factor of τ_k.
- In both settings we are able to recover the **i.i.d. rate of convergence** for the stochastic error from Markovian samples when $\tau_k \geq \tau_M$.
- **These complexity bounds are nearly optimal**, up to a logarithmic factor, in terms of the dependence on ϵ.

Discussions and Future direction

Conclusion:

- We proposed a novel MER algorithm to solve stochastic VIs in settings where samples are drawn from a Markovian buffer.
- MER demonstrates how "experience replay" heuristic from RL can be applied in a provable and statistically efficient manner with dependent data.

Future Directions:

- Provide optimal dependence on μ and $\bar{L}$ by combining MER with operator extrapolation [1].

References

[1] Georgios Kotsalis, Guanghui Lan, and Tianjiao Li.
Simple and optimal methods for stochastic variational inequalities, ii: Markovian noise and policy evaluation in reinforcement learning.
SIAM Journal on Optimization, 32(2):1120–1155, 2022.

[2] Tianjiao Li, Guanghui Lan, and Ashwin Pananjady.
Accelerated and instance-optimal policy evaluation with linear function approximation.
SIAM Journal on Mathematics of Data Science, 5(1):174–200, 2023.

[3] Wenlong Mou, Ashwin Pananjady, Martin J. Wainwright, and Peter L. Bartlett.
Optimal and instance-dependent guarantees for markovian linear stochastic approximation, 2024.

[4] R.S. Sutton and A.G. Barto.
Reinforcement Learning, second edition: An Introduction.
Adaptive Computation and Machine Learning series. MIT Press, 2018.