Reverse Experience Replay: A Way to Learn From Dependent Data

Praneeth Netrapalli

Google Research India

Joint work with

Naman Agarwal

Prateek Jain

Google Research

Suhas Kowshik, Dheeraj Nagaraj

Guy Bresler

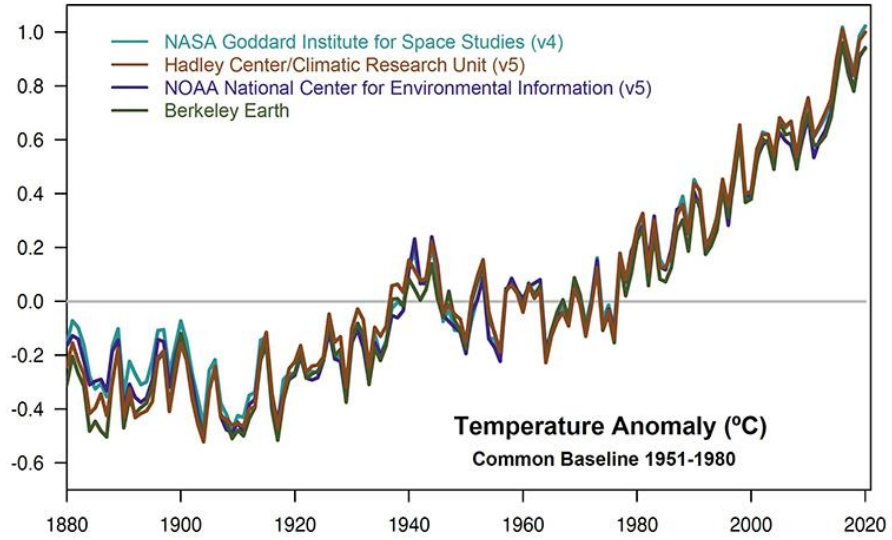
MIT

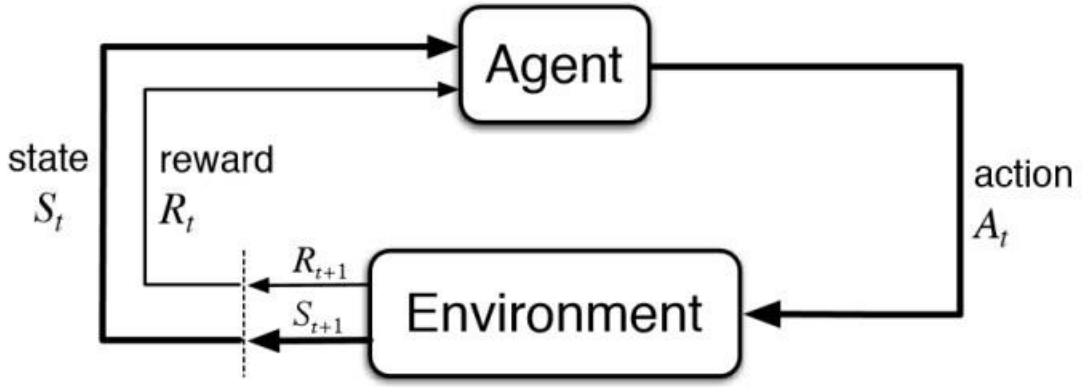
Syomantak Chaudhuri

Carrie Wu

UC Berkeley

Learning with dependent data





- Examples: time series forecasting, reinforcement learning, prediction using time series data.
- Usually modeled as a Markov chain/decision process.
- We wish to design streaming, SGD style algorithms.

SGD for Markovian data

$$\min_{\theta} \ell(\theta) \coloneqq \mathbb{E}_{x\sim\pi}[\ell_x(\theta)]$$

- Observe $x_1, x_2, \cdots$ from a Markov chain; π is the stationary distribution.
- SGD: $\theta_{t+1} = \theta_t - \eta\nabla\ell_{x_t}(\theta_t)$ is biased i.e., $\mathbb{E}\left[\nabla\ell_{x_t}(\theta_t)\right] \neq \nabla\ell(\theta_t)$ since x_t and θ_t are dependent on each other.
- Significant work on analyzing SGD/designing new algorithms but
  - Most works give asymptotic results only.
  - Best known non-asymptotic results match a simple baseline.

SGD with data drop – a simple baseline

- Let τ_{mix} be the mixing time of the Markov chain.
- $x_0, x_{\tau_{\text{mix}}}, x_{2\tau_{\text{mix}}}, \cdots$ are all approximately independent.
- Do SGD using only such a subset of samples.
- Number of effective samples $= \frac{n}{\tau_{\text{mix}}}$, for a total of n samples.
- Convergence rate is $\sqrt{\frac{\tau_{\text{mix}}}{n}}$ or $\frac{\tau_{\text{mix}}}{n}$ depending on the setting.
- Best known results [1] do no better than this baseline.

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

Can we do better?

No! (in a minimax sense)

Theorem [BJNNW, NeurIPS 2020]

There exists a linear regression problem, where the best convergence rate is $\widetilde{\Omega}\left(\frac{\tau_{\text{mix}}}{n}\right)$.

- The bound is information theoretic.
- In contrast, for i.i.d. data, the rate is $O\left(\frac{1}{n}\right)$.

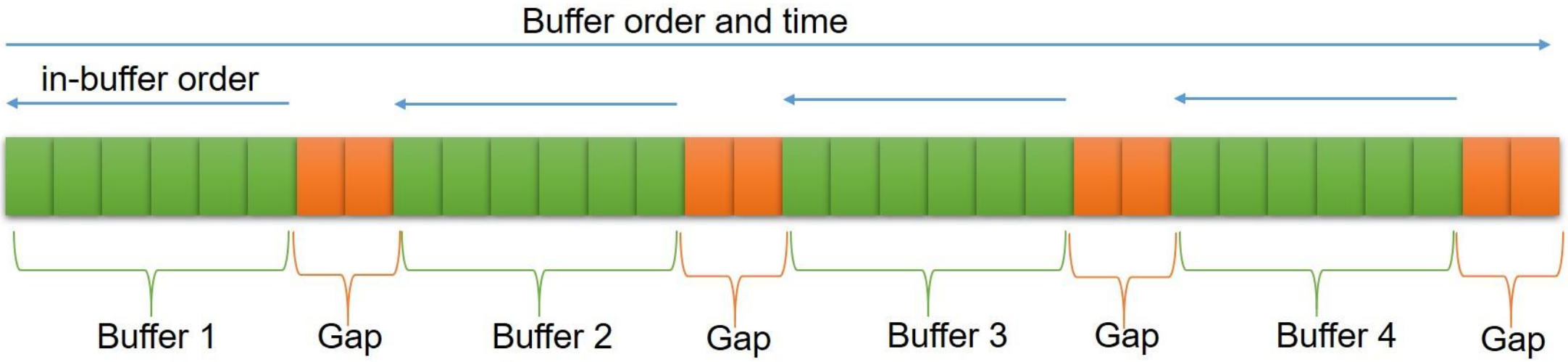
Can we do better under more assumptions?

- We consider linear least squares regression.

$$\min_{\theta} \ell(\theta) \coloneqq \mathbb{E}_{(x,y)\sim\pi}\left[(y - \langle\theta, x\rangle)^2\right]$$

- Information theoretically: Yes, $O\left(\frac{1}{n}\right)$ rate under realizability assumption!
- Realizability assumption: $\exists\theta^*$ such that $y = \langle\theta^*, x\rangle + \epsilon$, where ϵ is an independent zero mean random variable.
- Is there a streaming SGD style algorithm that achieves $O\left(\frac{1}{n}\right)$ rate?

Reverse experience replay (RER)



- Store a buffer of examples of size $\gtrsim \tau_{\text{mix}}$.
- Process the examples in the **reverse** order.
- Discard $\approx \tau_{\text{mix}}$ samples between buffers $\Rightarrow$ buffers are independent.
- Originally proposed in [2] showing good empirical performance.

[2] Reverse Experience Replay by E. Rotinov arXiv:1910.08780

Key intuition

$$\theta_{t+1} - \theta^* = (\theta_t - \theta^*) - \eta\left(\langle \theta_t - \theta^*, x_{i_t}\rangle - \epsilon_{i_t}\right)x_{i_t}$$

$$\theta_{t+1} - \theta^* = \left(I - \eta x_{i_t} x_{i_t}^\top\right)(\theta_t - \theta^*) + \eta\epsilon_{i_t} x_{i_t}$$

$$\theta_{t+1} - \theta^* = \left(I - \eta x_{i_t} x_{i_t}^\top\right)\left(I - \eta x_{i_{t-1}} x_{i_{t-1}}^\top\right)(\theta_{t-1} - \theta^*)$$

Geometrically decaying term

$$+\left(I - \eta x_{i_t} x_{i_t}^\top\right)\eta\epsilon_{i_{t-1}} x_{i_{t-1}} + \eta\epsilon_{i_t} x_{i_t}$$

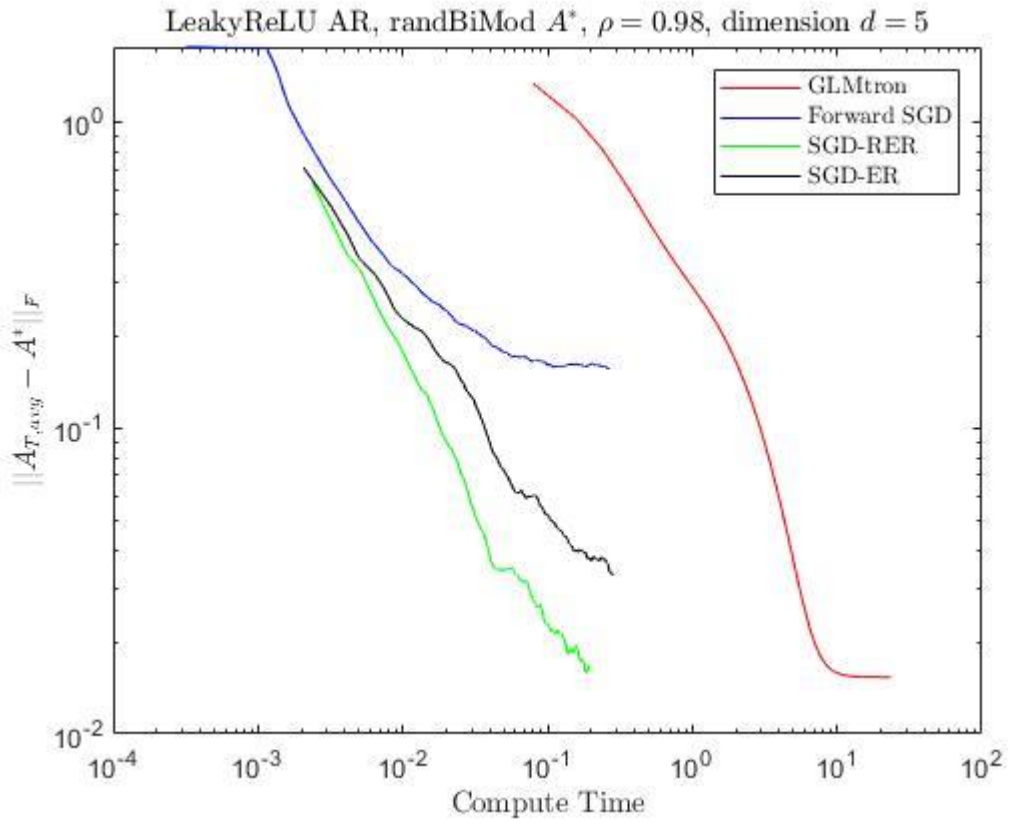
Zero mean terms due to realizability assumption

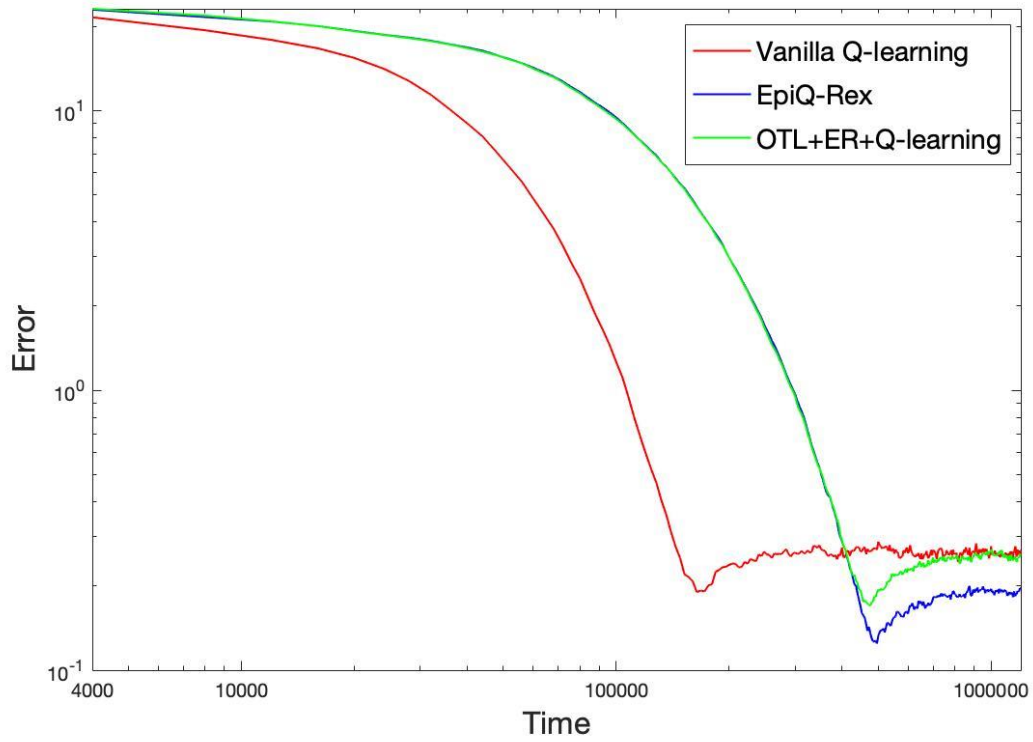
Results

- Nonlinear dynamical systems: $x_{t+1} = \sigma(A^* x_t) + \epsilon_t$
  - Task: learn A^* from a single trajectory $x_1, x_2, \cdots x_n$.
  - σ: $\mathbb{R} \to \mathbb{R}$ is an elementwise monotonic function with $0 < c \leq \sigma'(\cdot) \leq C$.
  - $\sigma = \text{Identity} \Leftrightarrow$ Linear dynamical system.
  - **Result**: SGD-RER achieves optimal statistical rates of estimation.

- Q-learning for linear MDPs
  - Task: learn the optimal Q-value function using a given exploration policy.
  - **Result**: Strong non-asymptotic convergence rate for Q-learning with online-target learning and RER.

Empirical results





Nonlinear dynamical system

Q-learning

Conclusions and future directions

- Learning with dependent data an important problem.
- Standard SGD is biased.
- Simple baseline of using only 1 out of every τ_{mix} samples is minimax optimal.
- Under **realizability**, **reverse experience replay** helps!

Open questions

- Extending beyond linear least squares regression.
- Combining exploration + RER in reinforcement learning.