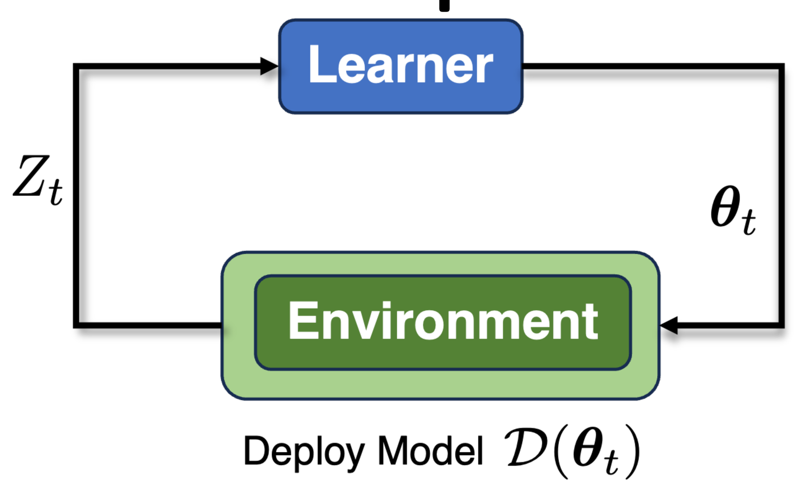
# Privacy-Efficacy Tradeoff of Clipped SGD with Decision-dependent Data

Qiang Li* Michal Yemini† Hoi-To Wai*
*Dept. of SEEM, CUHK; †Fac. of Engineering, Bar-Ilan University.

## Privacy Concerns in Model Training

- The training of prediction models hinges on the use of **private and sensitive** user data such as credit history.
- **Risk**: model inversion attack [Ghosh et al., 2009] exposes sensitive user data using just the training history of SGD.
- **Distribution Shift**: user reacts to the changing models, also known as **performative prediction problem**.




- **Performative Prediction** [Perdomo et al., 2020]

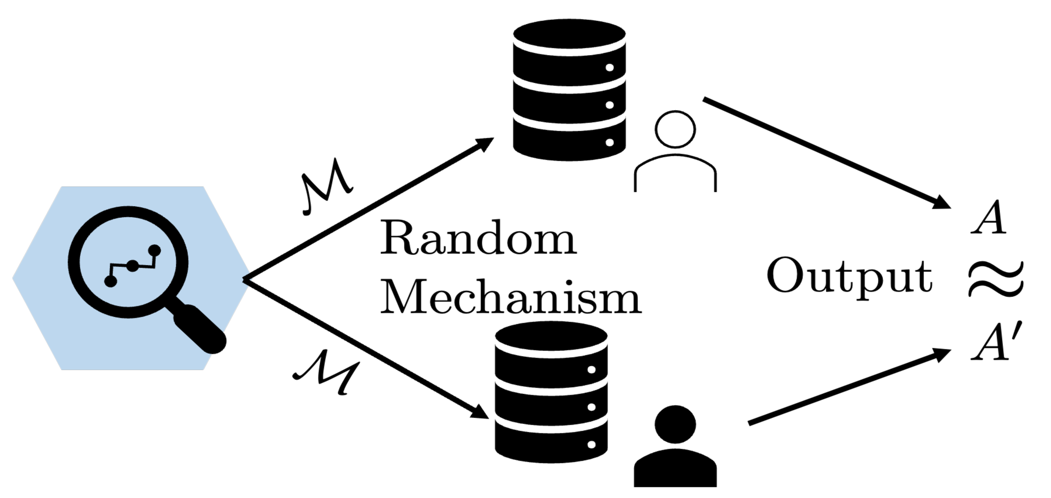
$$\min_{\theta \in \mathcal{X}} \mathbb{E}_{Z \sim \mathcal{D}(\theta)}[\ell(\theta; Z)],$$

- Dist. shifts also affects the convergence of SGD and their efficacy since the distribution of gradient estimates vary.

## Privacy Preserving Algorithm

- $(\varepsilon, \delta)$**-DP** (privacy budget, leakage probability)

$$\Pr[\mathcal{M}(\mathrm{D}) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(\mathrm{D}') \in S] + \delta \text{ [Dwork et al., 2014]}$$




- **Projected clipped SGD algorithm** [Abadi et al., 2016]:

$$\theta_{t+1} = \mathcal{P}_{\mathcal{X}}\left(\theta_t - \gamma_{t+1}\mathsf{clip}_c\,(\text{stoc. grad}) + \zeta_{t+1}\right)$$

where $\mathcal{P}(\cdot)$ is projection operator, $\zeta_{t+1}$ is Gaussian noise,

$$\mathsf{clip}_c(\boldsymbol{g}) : \boldsymbol{g} \in \mathbb{R}^d \mapsto \min\left\{1, \frac{c}{\|\boldsymbol{g}\|_2}\right\} \boldsymbol{g},$$

is designed to reduce gradient exposure.

- **Research Question**: What effect does performativity have on bias and convergence of clipped SGD algorithms?

**Our Answer**: PCSGD converges to a biased solution in expectation, bias $\propto \mathcal{O}(1/\text{dist. shift sensitivity})$.

## PCSGD Algorithm

- **Update Rule**: PCSGD scheme:

$$\theta_{t+1} = \mathcal{P}_{\mathcal{X}}\big(\theta_t - \gamma_{t+1}(\mathsf{clip}_c(\nabla \ell(\theta_t; Z_{t+1})) + \zeta_{t+1})\big),$$

- **Greedy deployment** sampling scheme: $Z_{t+1} \sim \mathcal{D}(\theta_t)$.
- ✿ **Difficulty**: clipping operator is non-smooth and leads to

$$\mathbb{E}_{Z \sim \mathcal{D}(\theta)} \mathsf{clip}_c(\nabla \ell(\theta; Z)) \neq \mathbb{E}_{Z \sim \mathcal{D}(\theta)}(\nabla \ell(\theta; Z))$$

## Main Results

$f(\theta_1, \theta_2) := \mathbb{E}_{Z \sim \mathcal{D}(\theta_2)}[\ell(\theta_1; Z)], \nabla f(\theta_1, \theta_2) := \mathbb{E}_{Z \sim \mathcal{D}(\theta_2)}[\nabla \ell(\theta_1; Z)]$.

- **A1**: $\mu$-strongly convex of $f(\theta_1; \theta_2)$ w.r.t. $\theta_1$.
- **A2**: Maps $\nabla f(\cdot; \bar{\theta})$ and $\nabla \ell(\bar{\theta}; \cdot)$ are $L$-Lipschitz.
- **A3**: Wasserstein-1 Dist.: $\mathcal{W}_1(\mathcal{D}(\theta), \mathcal{D}(\theta')) \leq \beta \|\theta - \theta'\|$.
- **A4**: Uniform bound: $\sup_{\theta \in \mathcal{X}, z \in \mathsf{Z}} \|\nabla \ell(\theta; z)\| \leq G$
  $\rightarrow$ reasonable, since $\mathcal{X}$ is a compact set

**Theorem 1**: (Upper bound) Under **A1-4**. Suppose that $\beta < \mu/L$, the step sizes $\{\gamma_t\}_{t \geq 1}$ are non-increasing and sufficient small. Then, for any $t \geq 1$,

$$\mathbb{E}\left\|\tilde{\theta}_{t+1}\right\|^2 \lesssim \prod_{i=1}^{t+1}(1 - \tilde{\mu}\gamma_i)\left\|\tilde{\theta}_0\right\|^2 + \frac{c_1}{\tilde{\mu}}\gamma_{t+1} + \frac{\max\{G - c, 0\}^2}{(\mu - L\beta)^2},$$

where $\tilde{\theta}_t := \theta_t - \theta_{PS}$, $\tilde{\mu} = \mu - L\beta$. Note $(c, \beta \rightarrow$ **Bias**$)$

- When $c \geq G$, then bias vanishes. Our convergence rate $\mathcal{O}(\gamma_t)$ coincides with prior works.
- When $c < G$, to achieve *minimum bias*, the opt. constant stepsize is $\gamma^\star = \mathcal{O}(1/(\tilde{\mu}T))$.

**Theorem 2**: (Lower bound) For any $c \in (0, G)$, $\exists \ell(\theta; Z)$ and $\mathcal{D}(\theta)$ satisfying A1-4, s.t. for fixed-points of PCSGD $\theta_\infty$ satisfying $\mathbb{E}_{Z \sim \mathcal{D}(\theta_\infty)}[\mathsf{clip}_c(\nabla \ell(\theta_\infty; Z))] = \mathbf{0}$, it holds

$$\|\theta_\infty - \theta_{PS}\|^2 = \Omega\big(1/(\mu - L\beta)^2\big).$$

- Provided that $\beta < \frac{\mu}{L}$, Theorems 1 and 2 show that PCSGD admits an unavoidable bias of $\Theta(1/(\mu - L\beta)^2)$.

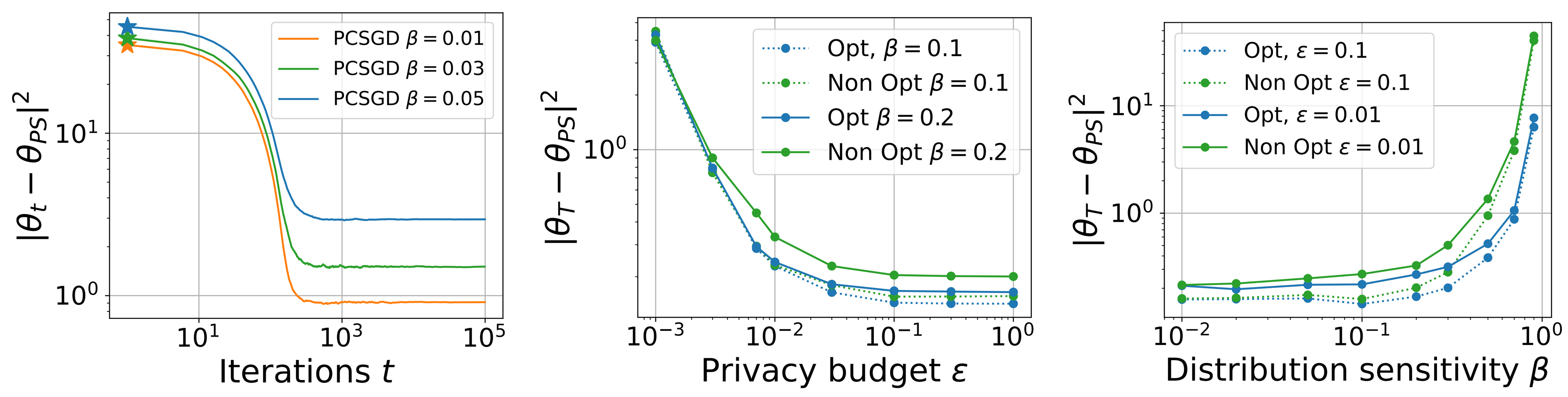
**Corollary 1**: (Differential Privacy Guarantee) For any $\varepsilon \leq T/m^2$, $\delta \in (0, 1)$, and $c > 0$, PCSGD with greedy deployment is $(\varepsilon, \delta)$-DP after $T$ iterations if we let

$$\sigma_{\mathsf{DP}} \geq c\sqrt{T \log(1/\delta)}/(m\varepsilon).$$

## Numerical Simulation: Quadratic Minimization

$\min_{\theta \in \mathcal{X}} \mathbb{E}_{z \sim \mathcal{D}(\theta)}[(\theta + az)^2/2], \mathcal{D}(\theta) = \{b\tilde{Z}_i - \beta\theta\}_{i=1}^m$ where $\tilde{Z}_i \sim \mathcal{B}(p)$ is Bernoulli. Note $\theta_{PS} = \frac{-\bar{p}a}{1 - a\beta}$.

- **Observations**: PCSGD cannot converge to $\theta_{PS}$ due to bias which increases as $\beta \uparrow$.
- **Effect of stepsize on bias**: Optimal stepsize, $\gamma^\star$, minimizes bias. (non-opt stepsize $\gamma = \frac{\log(1/\Delta(\mu))}{\mu T}$).




- As the privacy budget decreases $\varepsilon \downarrow 0$ or $\beta \uparrow \frac{\mu}{L}$, the bias $\uparrow$.

Contact E-mail: {liqiang, htwai}@se.cuhk.edu.hk, michal.yemini@biu.ac.il