Multi-agent Performative Prediction with Greedy deployment and Consensus Seeking Agents

Qiang Li, Chung-Yiu Yau, Hoi-To Wai Dept. of SEEM, The Chinese University of Hong Kong

Decentralized Optimization

- Decentralized learning uses networked **agents** to solve an optimization problem **cooperatively**, e.g., consensus-seeking.
- **Motivating Example**: learning from clinical data.

Local Performativity

- Supervised learning: static + i.i.d. data.
- Decision can cause distribution shift.
- **Performative Prediction**: data distribution depends on decision variables.

Goal of Agent i: min performative risk

$$\min_{\theta} \mathbb{E}_{Z\sim\mathcal{D}_i(\theta)}\left[\ell(\theta; Z)\right]$$

- Greedy Deploy [Mendler-Dünner, 2020]:
  - *Sampling*: $Z_{k+1} \sim \mathcal{D}_i(\theta_k)$
  - *SGD*: $\theta_{k+1} = \theta_k - \gamma_{k+1}\nabla\ell(\theta_k; Z_{k+1})$,
- If $\epsilon_i < \mu/L$, θ_k converges to performative stable point:

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

- **What if** $\epsilon_i > \mu/L$? Cooperation !

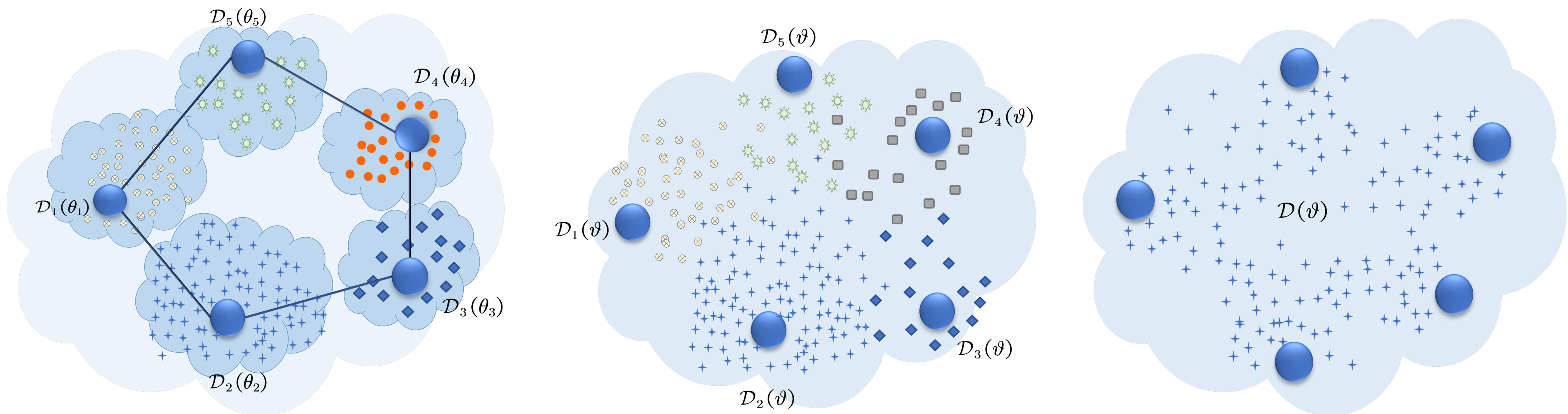
Multi-agent Performative Prediction

Goal of Multi-PfD: find a common decision vector for avg. loss.

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

- $\mathcal{M}(\boldsymbol{\theta}) := \arg\min_{\boldsymbol{\theta}'\in\mathbb{R}^d} \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{Z_i\sim\mathcal{D}_i(\boldsymbol{\theta})}[\ell(\boldsymbol{\theta}'; Z_i)]$.
- Multi-PS sol. $\boldsymbol{\theta}^{PS}$: fix point of $\mathcal{M}(\boldsymbol{\theta})$.

Alternative Multi-PfD Formulations



- This work (left): Heter. + locally influenced distributions **with consensus** among agents.
- [Narang et al., 2022] (middle): Heter. + globally influenced distributions $\mathcal{D}_i(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_n)$.
- [Piliouras and Yu, 2022] (right): Homo. + globally influenced distribution $\mathcal{D}(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_n)$.

Main Results

- A1. $\ell(\boldsymbol{\theta}; z)$ is μ-strongly convex. A2. $\ell(\boldsymbol{\theta}; z)$ has L-Lipschitz gradient.
- A3. (ϵ−sensitivity) $W_1(\mathcal{D}_i(\theta), \mathcal{D}_i(\theta')) \le \epsilon_i\|\theta - \theta'\|$, $\forall\, \theta, \theta' \in \mathbb{R}^d$,
- A4. (Mixing matrix) $\exists\, \rho \in (0, 1]$ s.t. $\left\|W - (1/n)\mathbf{1}\mathbf{1}^\top\right\|_2 \le 1 - \rho$.
- A5. σ-perturbation. A6. Heterogeneity ς

Proposition 1: **Existence and Uniqueness of** $\boldsymbol{\theta}^{PS}$
Multi-PfD admits a unique fixed point $\boldsymbol{\theta}^{PS} = \mathcal{M}(\boldsymbol{\theta}^{PS})$ if and only if $\epsilon_{avg} := 1/n\sum_{i=1}^n \epsilon_i < \mu/L$.

- Consensus improved robustness to sensitive local distribution shifts.

Decentralized SGD-Greedy Deployment (DSGD-GD)

$$Z_i^{t+1} \sim \mathcal{D}_i(\boldsymbol{\theta}_i^t) \quad \Big| \quad \boldsymbol{\theta}_i^{t+1} = \sum_{j=1}^n W_{ij}\boldsymbol{\theta}_j^t - \gamma_{t+1}\nabla\ell(\boldsymbol{\theta}_i^t; Z_i^{t+1}),$$

Theorem 1: Under some mild assumptions. Let $\epsilon_{\text{avg}} < \frac{\mu}{(1+\delta)L}$, $\exists\, \mathbb{C}$ s.t.

$$\mathbb{E}[\|\overline{\boldsymbol{\theta}}^t - \boldsymbol{\theta}^{PS}\|^2] \lesssim \underbrace{\prod_{i=1}^t\left(1 - \frac{\tilde{\mu}\gamma_i}{2}\right) + \frac{L(\sigma^2+\varsigma^2)}{n\delta\tilde{\mu}\rho^2\epsilon_{\text{avg}}}\gamma_t^2}_{\text{Transient}} + \underbrace{\frac{\sigma^2}{n\tilde{\mu}}\gamma_t}_{\text{Fluctuation}},$$

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\|\boldsymbol{\theta}_i^t - \overline{\boldsymbol{\theta}}^t\|^2\right] \lesssim \left(1 - \frac{\rho}{2}\right)^t + \frac{(\sigma^2+\varsigma^2)}{\rho^2}\gamma_t^2,$$

where δ is a parameter to be determined, $\tilde{\mu} := \mu - (1+\delta)\epsilon_{\text{avg}}L$.

- Squared distance $\sim \mathcal{O}(\gamma_t)$, consensus error$\sim \mathcal{O}(\gamma_t^2)$.

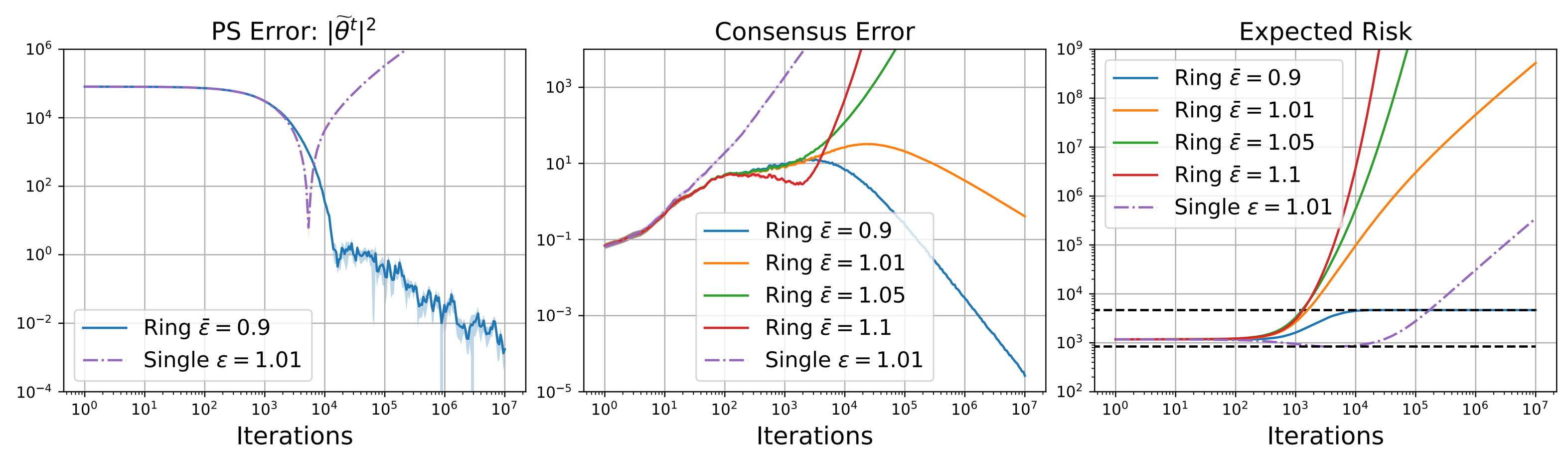
Additional Contributions

- B-connected graph: Analysis of DSGD-GD also works on time-varying graph.

Numerical Experiments

Multi-agent Gaussian Mean Estimation:

- Setup: $n = 25$-agent right graph.
- Quadratic Loss: $\ell(\boldsymbol{\theta}_i; Z_i) = (\boldsymbol{\theta}_i - Z_i)^2/2$.
- Local distribution: $\mathcal{D}_i(\boldsymbol{\theta}_i) \equiv \mathcal{N}(\bar{z}_i + \epsilon_i\boldsymbol{\theta}_i, \sigma^2)$, where $\bar{z}_i$ is the mean value to be estimated.
- Multi-PS sol. $\boldsymbol{\theta}^{PS} = \sum_{i=1}^n \bar{z}_i/[n(1-\epsilon_{\text{avg}})]$, if $0 < \bar{\epsilon} = \epsilon_{\text{avg}} < 1$.



- Proposition 1 ✓ (left) when $\epsilon_{avg} < 1$ converge,
  - (right) when $\epsilon_{avg} > 1$, diverge.
- Theorem 1 ✓ (left) $|\tilde{\boldsymbol{\theta}}^t|^2$ decays at $\mathcal{O}(1/t)$
  - (middle) $\|\Theta_o^t\|^2$ decays at $\mathcal{O}(1/t^2)$.
- (dash-dotted) agent i ($\epsilon_i = 1.01$) disconnects and performs greedy deployment *individually*, its performative risk diverges.

Reference

- Perdomo, Juan, et al. *Performative prediction*, ICML 2020.
- Mendler-Dünner, et al. *Stochastic optimization for performative prediction* NeurIPS 2020.
- Narang, A., Faulkner, E., Drusvyatskiy, D., Fazel, M., and Ratliff, L. J. (2022). Multiplayer performative prediction: Learning in decision-dependent games. In AISTATS.
- Piliouras, G. and Yu, F.-Y. (2022). Multi-agent performative prediction: From global stability and optimality to chaos.

Contact E-mail: {liqiang,cyyau,htwai}@se.cuhk.edu.hk

Acknowledgement: CUHK Direct Grant #4055113.