Locally Differentially Private Decentralized Stochastic Bilevel Optimization with Guaranteed Convergence Accuracy

Ziqin Chen[1] and Yongqiang Wang[1]

1. Department of Electrical and Computer Engineering, Clemson University, Clemson, SC, 29634, United States.

Correspondence to: Yongqiang Wang <yongqiw@clemson.edu>.

Abstract

Decentralized stochastic bilevel optimization (DSBO) based machine learning techniques are achieving remarkable success. However, the intensive exchange of information (involving nested-loops of consensus or communication iterations) in existing DSBO algorithms leads to a great challenge to ensure rigorous differential privacy. By proposing a new decentralized stochastic bilevel-optimization algorithm which avoids nested-loops of information-exchange iterations, we achieve, for the first time, both differential privacy and accurate convergence in decentralized bilevel optimization. This is significant since even for single-level decentralized optimization and learning, existing differential-privacy solutions have to sacrifice convergence accuracy for privacy. Besides characterizing the convergence speed under nonconvex/convex/strongly convex conditions, we also rigorously quantify the price of differential privacy in the convergence rate.

Problem Statement

The bilevel optimization problem:

$$\min_{x\in\mathbb{R}^p} F(x), \quad F(x)=\frac{1}{m}\sum_{i=1}^{m} f_i(x,y^*(x)), \qquad\qquad (1)$$
$$\text{s.t. } y^*(x)=\text{argmin}_{y\in\mathbb{R}^q} g(x,y):=\frac{1}{m}\sum_{i=1}^{m} g_i(x,y),$$

with $f_i(x,y)=\mathbb{E}_{\varphi_i}[h(x,y;\varphi_i)]$ and $g_i(x,y)=\mathbb{E}_{\xi_i}[l(x,y;\xi_i)]$.

Existing DSBO algorithms cannot protect the privacy of participating agents. As differential privacy (DP) is evolving as the de facto standard for privacy, it is of great interest to achieve differential privacy in DSBO.

Challenges

1) Existing DSBO algorithms involve nested-loops of consensus iterations, which will result in an exploding cumulative privacy budget as iteration proceeds, leading to diminishing privacy protection in the long run.
2) Maintaining the accuracy of DSBO algorithms under the constraint of DP is challenging. In fact, even for the simpler single-level decentralized optimization problem, existing DP solutions have to trade optimization accuracy for privacy.
3) The major challenge in solving DSBO lies in the lack of explicit knowledge of $y^*(x)$, which makes it impossible for individual agents to evaluate the hypergradient $\nabla F(x,y^*(x))$:

$$\nabla F(x)=\frac{1}{m}\sum_{i=1}^{m}\nabla_x f_i(x,y^*(x))-\frac{1}{m}\sum_{i=1}^{m}\nabla^2_{xy} g_i(x,y^*(x))$$
$$\times\boxed{\begin{array}{c}\left[\frac{1}{m}\sum_{i=1}^{m}\nabla^2_{yy} g_i(x,y^*(x))\right]^{-1}\frac{1}{m}\sum_{i=1}^{m}\nabla_y f_i(x,y^*(x)).\\ \textbf{Hessian-inverse-vector product}\end{array}}$$

Methods and Results

Methods

We first introduce an approach for individual agents to locally estimate Hessian-inverse-vector product under the constraint of LDP. Using it as a subroutine, we then propose our differentially private DSBO algorithm.

Algorithm 1 Subroutine for Estimating Hessian-Inverse-Vector Product for Agent i, $i\in[m]$

```
1: Input: Parameters $x_{i,t}$, $y_{i,t}$, and $z_{i,t}$; Data samples $\{\varphi_{i,k}\}_{k\in[0,t]}$ and $\{\xi_{i,t}\}_{k\in[0,t]}$; Stepsize $\lambda_{z,t}=\frac{\lambda_{z,0}}{(t+1)^{v_z}}$ with $\lambda_{z,0}>0$ and $v_z\in(0,1)$; DP-noise $\vartheta_{i,t}$.
2: $H_{i,t}z_{i,t}=\nabla^2_{yy}g_{i,t}(x_{i,t},y_{i,t})z_{i,t}$.
3: $b_{i,t}=\nabla_y f_{i,t}(x_{i,t},y_{i,t})$.
4: $\nabla_z\phi_{i,t}(z_{i,t})=H_{i,t}z_{i,t}-b_{i,t}$.
5: $z_{i,t+1}=z_{i,t}+\sum_{j\in\mathcal{N}_i}w_{ij}(z_{j,t}+\vartheta_{j,t}-z_{i,t})-\lambda_{z,t}\nabla_z\phi_{i,t}(z_{i,t})$.
6: Output: $z_{i,t+1}$ on agent $i$.
```

Estimate Hessian-inverse-vector product corresponding to solve the following optimization problem:

$$\min_{z\in\mathbb{R}^q}\frac{1}{m}\sum_{i=1}^{m}\phi_i(z), \quad \phi_i(z)=\frac{1}{2}z^T H_i z-b_i^T z. \qquad (2)$$

Algorithm 2 LDP Design for DSBO Algorithm for Agent i, $i\in[m]$

```
1: Input: Random initialization $x_{i,0}\in\mathbb{R}^p$, $y_{i,0}\in\mathbb{R}^q$, and $z_{i,0}\in\mathbb{R}^q$ for each agent $i\in[m]$. Stepsizes $\lambda_{x,t}=\frac{\lambda_{x,0}}{(t+1)^{v_x}}$ and $\lambda_{y,t}=\frac{\lambda_{y,0}}{(t+1)^{v_y}}$ with $\lambda_{x,0}>0$, $\lambda_{y,0}>0$, and $v_x,v_y\in(0,1)$; DP-noises $\chi_{i,t}$ and $\zeta_{i,t}$.
2: for $t=0,1,\cdots,T-1$ do
3:     Acquire current data $\varphi_{i,t}$ and $\xi_{i,t}$.
4:     $y_{i,t+1}=y_{i,t}+\sum_{j\in\mathcal{N}_i}w_{ij}(y_{j,t}+\zeta_{j,t}-y_{i,t})-\lambda_{y,t}\nabla_y g_{i,t}(x_{i,t},y_{i,t})$.
5:     Run Algorithm 1 and obtain the output $z_{i,t+1}$.
6:     $u_{i,t}=\nabla_x f_{i,t}(x_{i,t},y_{i,t})-\nabla^2_{xy}g_{i,t}(x_{i,t},y_{i,t})z_{i,t}$.
7:     $x_{i,t+1}=x_{i,t}+\sum_{j\in\mathcal{N}_i}w_{ij}(x_{j,t}+\chi_{j,t}-x_{i,t})-\lambda_{x,t}u_{i,t}$.
8: end for
9: Output: $x_{i,T}$ on agent $i$.
```

Algorithm 1 enables all agents to collaboratively find the optimal solution z^* to problem (2).

Results

Algorithm	Decentralized?	Computational Complexity	Jacobian	DP	Privacy Budget
BSA (Ghadimi & Wang, 2018)	No	$\mathcal{O}(\delta^{-3}+(q^2\log(\delta^{-1})+pq)\delta^{-2})$	YES	No	$\mathcal{O}(\delta^{-3})$
SPDB (Lu et al., 2022)	YES	$\mathcal{O}(\max\{p,q\}\log(\delta^{-1})\delta^{-2})$	No	No	$\mathcal{O}(\delta^{-2})$
VRDSBO (Gao et al., 2023)	YES	$\mathcal{O}((pq+q^2)\delta^{-\frac{3}{2}})$	YES	No	$\mathcal{O}(\delta^{-\frac{3}{2}})$
DSBO-JHIP (Chen et al., 2022)	YES	$\mathcal{O}(pq\log(\delta^{-1})\delta^{-3})$	YES	No	$\mathcal{O}(\delta^{-3})$
GBDSBO (Yang et al., 2022)	YES	$\mathcal{O}((q^2\log(\delta^{-1})+pq)\delta^{-2})$	YES	No	$\mathcal{O}(\delta^{-2})$
MA-DSBO (Chen et al., 2023)	YES	$\mathcal{O}(\max\{p,q\}\log(\delta^{-1})\delta^{-2})$	No	No	$\mathcal{O}(\delta^{-2})$
Our Algorithm	YES	$\mathcal{O}(\max\{p,q\}\delta^{-2.6})$	No	YES	$\mathcal{O}(1)$

Experimental results:

➢ Decentralized meta learning under LDP constraints.

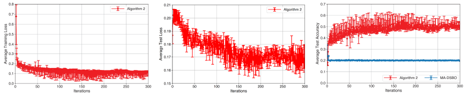

Fig. 1. Comparison by using the “CIFAR-10” dataset.

➢ Decentralized hypergradient optimization under LDP constraints.

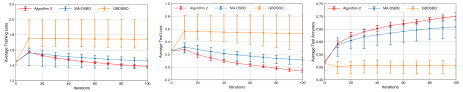

Fig. 2. Comparison by using the synthetic dataset.

Theoretical results:

- Strongly convex case: $\mathbb{E}\left[\|x_{i,T}-x^*\|^2\right]\le\mathcal{O}\left(T^{-\beta_1}\right)$
- Convex case: $\frac{1}{T+1}\sum_{t=0}^{T}\mathbb{E}\left[F(x_{i,t})-F(x^*)\right]\le\mathcal{O}\left(T^{-(1-v_x)}\right)$.
- Nonconvex case: $\frac{1}{T+1}\sum_{t=0}^{T}\mathbb{E}\left[\|\nabla F(x_{i,t})\|^2\right]\le\mathcal{O}\left(T^{-(1-v_x)}\right)$.
- The cumulative privacy budget is finite even when the number of iteration tends to infinity.

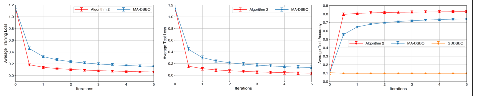

Fig. 3. Comparison by using the “MNIST” dataset.

Contributions

- We propose a differentially private DSBO algorithm that can ensure both accurate convergence and rigorous LDP, with the cumulative privacy budget bounded even when the number of iterations tends to infinity. To the best of our knowledge, no such results have been reported before. Moreover, by employing the LDP framework, our results can be applied to the fully decentralized setting where no data aggregator or mediator exists to gather data or assist privacy design.
- Our new algorithm successfully circumvents nested-loops of consensus, which makes it possible to alleviate the growth of the cumulative privacy budget as the number of iterations increases. In fact, given that using intensive (nested-loops of) consensus or communication rounds is the only approach to ensuring accurate convergence when the objective functions are heterogeneous across the agents, our algorithm is of independent interest even if privacy is not of concern.
- We establish the convergence rate of our algorithm for nonconvex/convex/strongly convex objective functions.
- Despite retaining accurate convergence, our algorithm does pay a price for obtained DP in convergence rate.
- We conduct experiment evaluation using several machine learning problems.

Conclusion

In this paper, we proposed a decentralized stochastic bilevel algorithm that can simultaneously ensure both accurate convergence and rigorous differential privacy. This is significant because even for the simpler problem of single-level decentralized optimization/learning, existing differential-privacy solutions have to sacrifice convergence accuracy for privacy. Lying at the core of our approach is a new algorithm for decentralized stochastic bilevel optimization that avoids any nested-loops of consensus (communication) iterations. This is important since all existing decentralized algorithms for bilevel optimization rely on nested-loops of consensus iterations, which, unfortunately, constitutes an obstacle for achieving differential privacy because the intensive consensus operations lead to an exploding cumulative privacy budget. We systematically characterized the convergence performance of our algorithm under both nonconvex and convex objective functions, and quantified the price and tradeoff in the convergence rate.

Acknowledgement

We thank the National Science Foundation for supporting this work under Grants ECCS-1912702, CCF-2106293, CCF-2215088, CNS-2219487, and CCF 2334449.