ICML 2018 | Stockholm
35th International Conference on Machine Learning

Improving the Privacy and Accuracy of ADMM-Based Distributed Algorithms

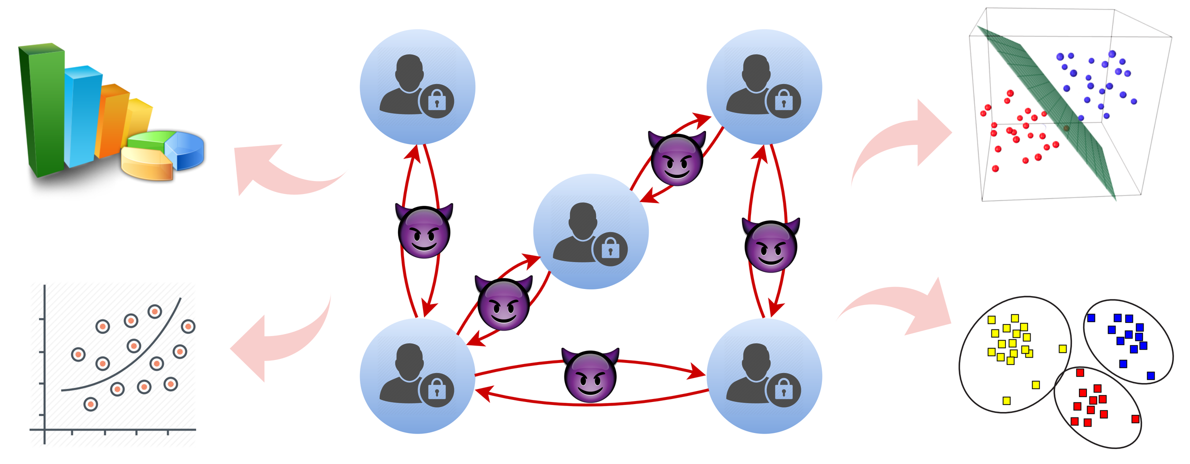
Xueru Zhang, Mohammad Mahdi Khalili and Mingyan Liu   University of Michigan

Objectives

Distributed learning tasks using distributed data:

- different data owners/locality
- common computational objective
- privacy concerns over sharing data

Goal: accomplish computational tasks distributedly while providing privacy guarantee.



Preliminaries

Regularized Empirical Risk Minimization:

$$\min_{f_c} O_{ERM}(f_c, \{D_i\}_{i=1}^N) = \sum_{i=1}^N O(f_c, D_i)$$

with $O(f_c, D_i) = \frac{C}{B_i}\sum_{n=1}^{B_i}\mathscr{L}(y_i^n f_c^T x_i^n) + \frac{\rho}{N}R(f_c)$

Decentralize ERM:

$$\min_{\{f_i\},\{w_{ij}\}} \tilde{O}_{ERM}(\{f_i\}_{i=1}^N, D_{all}) = \sum_{i=1}^N O(f_i, D_i)$$

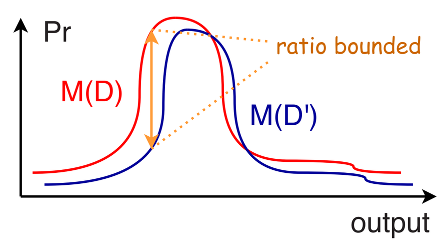
$$\text{s.t.} \quad f_i = w_{ij},\ w_{ij} = f_j,\quad i \in \mathscr{N}, j \in \mathscr{V}_i$$

Simplified Alternating Direction Method of Multiplier:

- Initialize dual variables: $\lambda_{ij}^a(0) = \lambda_{ij}^b(0) = 0$. Let $\lambda_i(t) = \sum_{j\in\mathscr{V}_i}\lambda_{ij}^a(t) = \sum_{j\in\mathscr{V}_i}\lambda_{ij}^b(t)$

Primal update: $$f_i(t+1) = \underset{f_i}{\text{argmin}}\{O(f_i, D_i) + 2\lambda_i(t)^T f_i + \eta\sum_{j\in\mathscr{V}_i}||f_i - \frac{1}{2}(f_i(t) + f_j(t))||_2^2\}\ ;$$

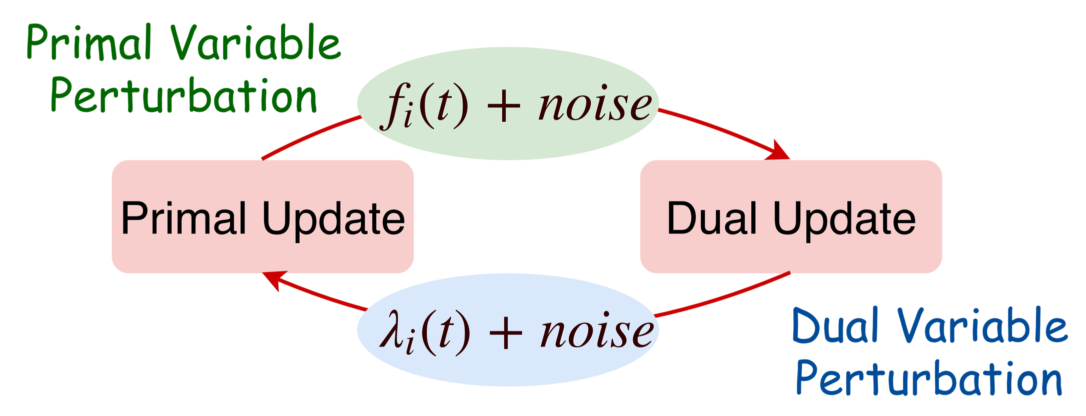
Dual update: $$\lambda_i(t+1) = \lambda_i(t) + \frac{\eta}{2}\sum_{j\in\mathscr{V}_i}(f_i(t+1) - f_j(t+1))\ .$$



Differential Privacy: obtain almost the same conclusion regardless of participation $\frac{\Pr(M(D)\in S)}{\Pr(M(D')\in S)} \leq \exp(\epsilon)$

Existing Work & Limitation

Two randomizations were proposed [1]:



Issues:

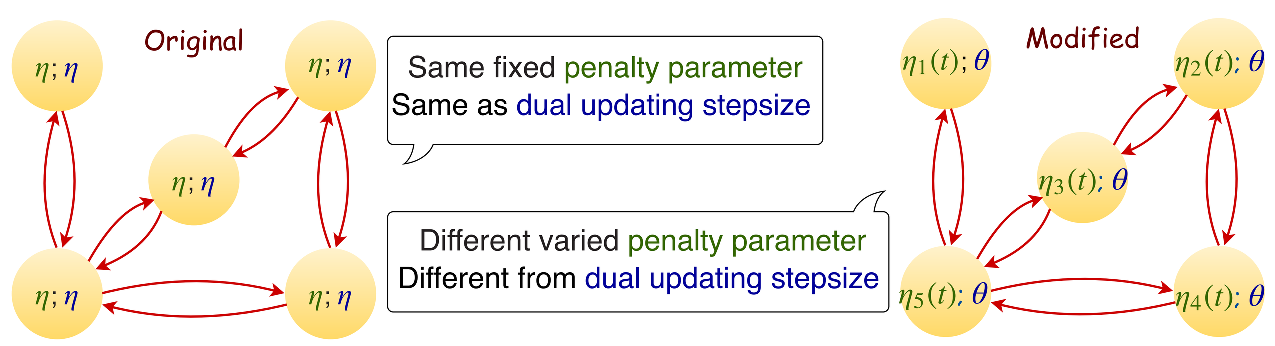
1. Privacy loss only evaluated for a single node for one iteration.
2. Privacy leakage accumulates over many iterations; hard to balance privacy and utility simply by summing up privacy losses.

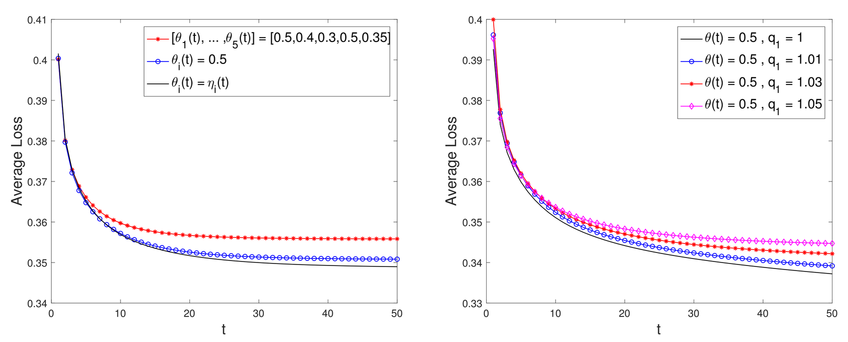
Modified ADMM & Penalty Perturbation

Modified ADMM: explore the use of penalty parameter η; allow this to be private information.

Modified primal update: $$f_i(t+1) = \underset{f_i}{\text{argmin}}\{O(f_i, D_i) + 2\lambda_i(t)^T f_i + \eta_i(t+1)\sum_{j\in\mathscr{V}_i}||f_i - \frac{1}{2}(f_i(t) + f_j(t))||_2^2\}\ ;$$

Modified dual update: $$\lambda_i(t+1) = \lambda_i(t) + \frac{\theta}{2}\sum_{j\in\mathscr{V}_i}(f_i(t+1) - f_j(t+1))\ .$$





Penalty Perturbation (PP): generalizes Dual Variable Perturbation (DVP)

$$f_i(t+1) = \underset{f_i}{\text{argmin}}\{O(f_i, D_i) + 2\lambda_i(t)^T f_i + \eta_i(t+1)\sum_{j\in\mathscr{V}_i}||\text{noise} + f_i - \frac{1}{2}(f_i(t) + f_j(t))||_2^2\}$$

Main Results

Convergence Analysis (non-private)

- **Condition for convergence of modified ADMM:** $0 < \theta \leq \eta_i(t) \leq \eta_i(t+1) < +\infty, \forall i, t$
- **Quantify lower bound on the convergence rate:** increasing $\eta_i(t)$ slows down the rate

Privacy Analysis (private): the total privacy loss during T iterations

$$\ln\left(\frac{\Pr(\{\{f_i(t)\}_{i=1}^N\}_{t=0}^T \in S|D)}{\Pr(\{\{f_i(t)\}_{i=1}^N\}_{t=0}^T \in S|D')}\right) \leq \max_{i\in\mathscr{N}}\left\{\sum_{t=1}^T \frac{C(1.4c_1 + \alpha_i(t))}{\eta_i(t)V_i B_i}\right\}$$

Numerical Results

Real world dataset: *Adult* dataset

Accuracy: $L(t) := \frac{1}{N}\sum_{i=1}^N \frac{1}{B_i}\sum_{n=1}^{B_i}\mathscr{L}(y_i^n f_i(t)^T x_i^n)$

Privacy: $P(t) := \max_{i\in\mathscr{N}}\{\sum_{r=1}^t \frac{C(1.4c_1+\alpha_i(r))}{\eta_i(r)V_i B_i}\}$

Parameter setting:

$$\eta_i(t+1) = \theta q_1^t = \begin{cases} q_1 = 1, \text{ DVP} \\ q_1 \geq 1, \text{ PP} \end{cases}\quad ;\ \alpha_i(t+1) = \alpha q_2^t$$

Network: small (five-node) and large (hundreds of nodes); with even and uneven distributed sample.

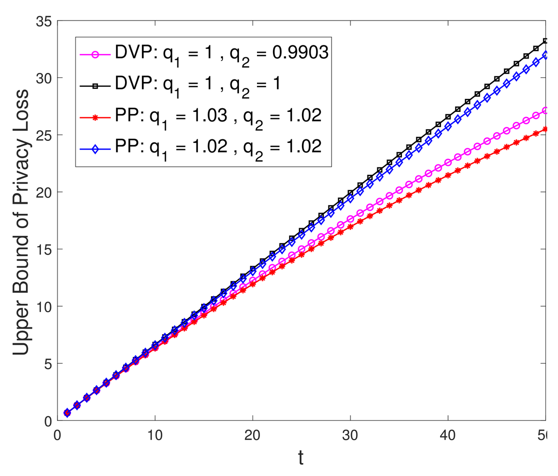


Figure 1: Privacy: $\alpha = 3$

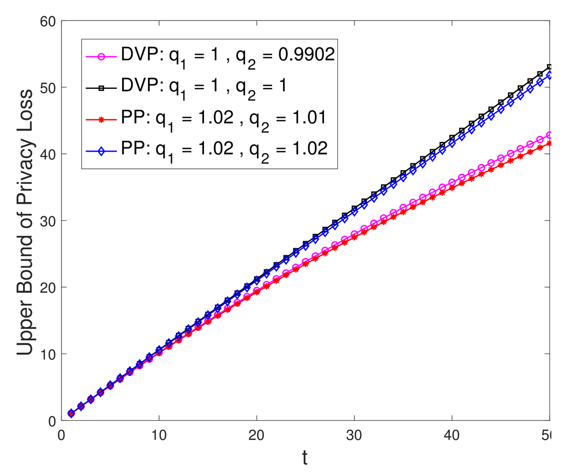


Figure 2: Privacy: $\alpha = 5$

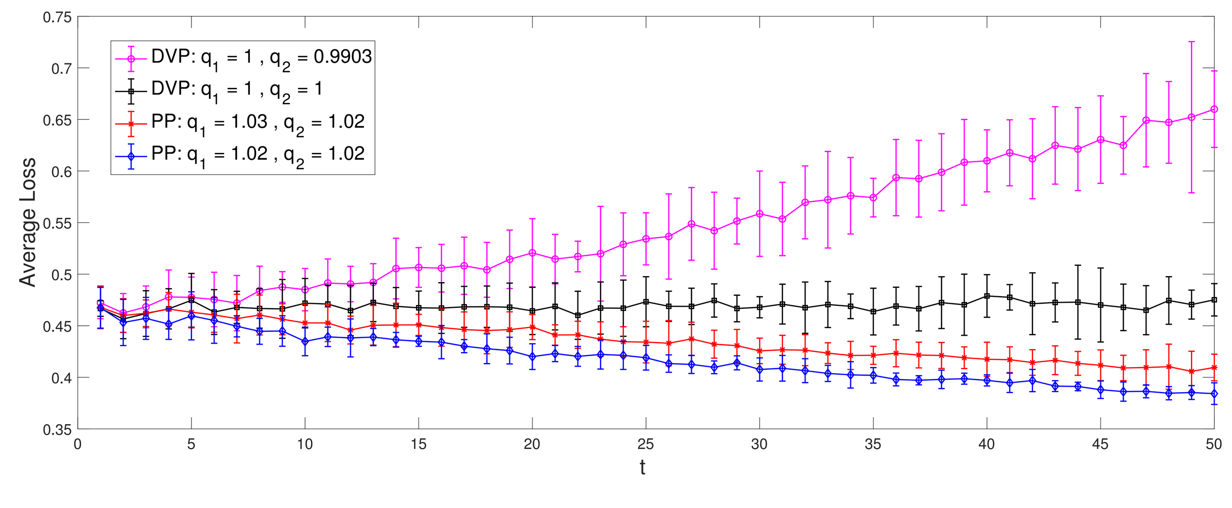


Figure 3: Accuracy: $\alpha = 3$

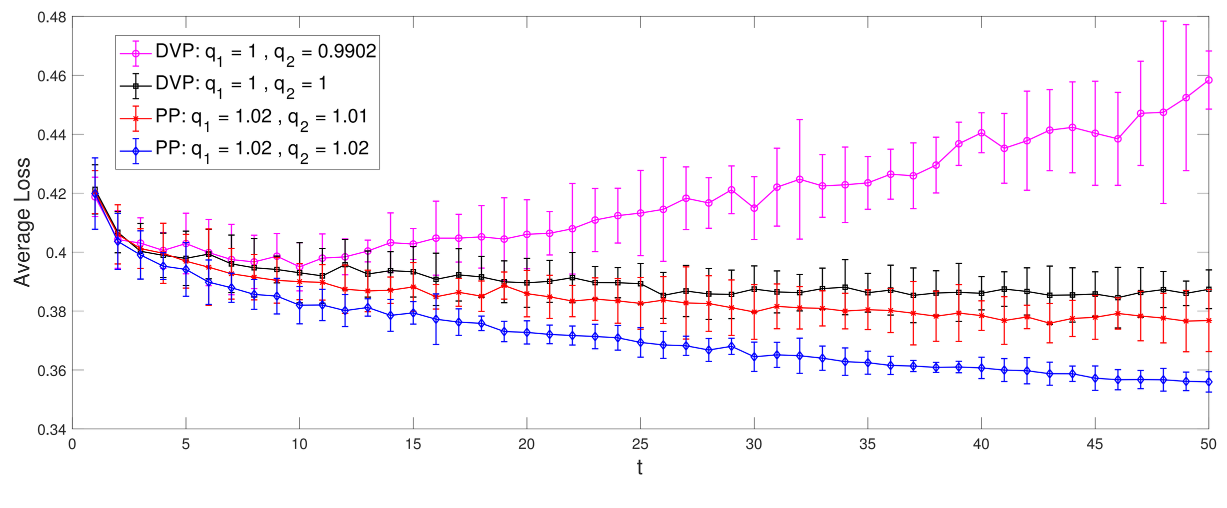


Figure 4: Accuracy: $\alpha = 5$

Conclusions

Better performance and stronger privacy can be obtained simultaneously by increasing $\eta_i(t)$.

The improvement is more significant with higher privacy requirement.

References

[1] Tao Zhang and Quanyan Zhu. Dynamic differential privacy for admm-based distributed classification learning. *IEEE Transactions on Information Forensics and Security*, 12(1):172–187, 2017.