Privacy-Preserving Distributed Optimization via Subspace Perturbation: A General Framework

Qiongxiu Li, Richard Heusdens, Mads Græsbøll Christensen
2020/08/03
Agenda

• Motivation
• Problem setup
• State of the art
• Proposed approach
• Numerical validation
• Conclusions & Future works
• Q&A
Motivation
Centralized system v.s. Distributed system

😊 Totally dependent on the authority
😊 Vulnerable to malicious attack

😊 No dependency on any single party
😊 More flexible system
😊 Robust to malicious attack
Privacy issue in smart meters  [Giaconi, 2018]

No real separation between your data and your identity.

When are you usually away from home?
Do you have an electronic alarm system? How often do you arm it?
How many people are at home? Are they on holiday?

Do you have disabilities/illnesses?
How often do you eat in? Do you eat hot/cold breakfasts?
Do you eat microwaved food?
Do you wake up at night? What are your sleep cycles?
Do you drive sleep deprived?
How many hours do you spend in front of TV?
How often do you entertain? Is any appliance inefficient?
Do you need to break speed limit to get to work on time?
Age, gender/sex, non-traditional family, race, ethnicity.

User Profiling
Discrimination
Targeted Marketing
Insurance Adjusting
Law Enforcement
Secure communication v.s. secure computation

**Secure communication**
- Only message passing
- No computation based on message
- Alice trusts Bob

**Secure computation**
- Computation based on message
- No trust
Intuitive examples (1)

- How to securely compute the average salary over a group of people while keeping each person’s own salary private from others?
Intuitive examples (2)

- Privacy-preserving machine learning over multiple parties
- Collaborative learning without revealing private data
Intuitive examples (3)

- Privacy-preserving distributed acoustic environment classification in WASNs
  - Privacy-preserving distributed clustering

Problem setup
Distributed convex optimization over a network

A graph $G = (\mathcal{N}, \mathcal{E})$ $\mathcal{N} = \{1, 2, ..., n\}$, $n = |\mathcal{N}|$, $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$, $m = |\mathcal{E}|$

Minimise $\min_{x_i} \sum_{i \in \mathcal{N}} f_i (x_i, s_i)$

Subject to $B_{i|j}x_i + B_{j|i}x_j = b_{ij}$ $\forall (i, j) \in \mathcal{E}$

Incidence matrix: $B \in \mathbb{R}^{m \times n}$

Main requirements for privacy-preserving distributed optimization

1. **Output correctness**: optimum result $x_i^*$ should be achieved
2. **Individual privacy**: the private data $s_i$ should not be revealed to others
Primal-Dual Methods of Multipliers (PDMM) [Sherson, 2018]

Extended augmented Lagrangian of PDMM

\[ L(x, \lambda) = f(x, s) + (P\lambda^{(k)})^T C x + \frac{c}{2} \| C x + PCx^{(k)} - 2d \|^2 \]

\( c > 0 \): constant for controlling convergence rate

\( \lambda \in \mathbb{R}^{2m} \): dual variables for constraints

Each edge \( e_l = (i, j) \in \mathcal{E} \) corresponds two dual variables: \( \lambda_{i|j} \) for node \( i \), \( \lambda_{j|i} \) for node \( j \)

\[
\lambda = \begin{pmatrix}
\lambda_{1|2} \\
\lambda_{1|3} \\
\lambda_{2|3} \\
\lambda_{2|4} \\
\lambda_{3|4} \\
\lambda_{4|5} \\
\lambda_{2|1} \\
\lambda_{3|1} \\
\lambda_{3|2} \\
\lambda_{4|2} \\
\lambda_{5|3} \\
\lambda_{5|4}
\end{pmatrix}
\]

Updating functions

\[
x^{(k+1)} = \arg \min_x L \left( x, x^{(k)}, \lambda^{(k)} \right)
\]

\[
\lambda^{(k+1)} = P\lambda^{(k)} + c(Cx^{(k+1)} + PCx^{(k)} - 2d)
\]

\[
C = \begin{pmatrix} B^+ \\ B^- \end{pmatrix}, \quad PC = \begin{pmatrix} B^- \\ B^+ \end{pmatrix} \in \mathbb{R}^{2m \times n}
\]
Why conventional approaches violate privacy?

Local updating functions of PDMM

\[
x_i^{(k+1)} = \arg\min_{x_i} \left( f_i(x_i, s_i) + \sum_{j \in \mathcal{N}_i} \lambda_{j|i}^{(k)} B_{i|j} x_i + \frac{c}{2} \sum_{j \in \mathcal{N}_i} \| B_{i|j} x_i + B_{j|i} x_j^{(k)} - b_{i,j} \|_2^2 \right)
\]

\[
\forall j \in \mathcal{N}_i : \lambda_{i|j}^{(k+1)} = \lambda_{j|i}^{(k)} + c \left( B_{i|j} x_i^{(k+1)} + B_{j|i} x_j^{(k)} - b_{i,j} \right)
\]

\[
\partial f_i(x_i^{(k+1)}, s_i) \text{ is correlated with private data } s_i
\]

Information-theoretically: \( I(S_i; X_i^{(k+1)}) \neq 0 \)

Exchange of \( x_i^{(k+1)} \) violates individual privacy
State of the art
Existing approach (1) Homomorphic encryption [Freris, 2016]

Main idea: all computation are conducted on encrypted data

\[ Enc(x) \times Enc(y) = Enc(x + y) \]

Computational security model: based on computational hardness assumption

Pros and cons:
- ☺ No compromisation in algorithm accuracy
- ☹ Computationally complex
- ☹ The adversary is assumed computationally bounded
Existing approach (2) _Secret sharing_[Tjell, 2020]

Main idea: split each message into pieces and send to different parties

Information-theoretic security model: the adversary does not have enough information to infer the secret/private data

Pros and cons:
- 😊 The adversary is assumed computationally unbounded
- 😊 Computationally simple
- 😊 No compromisation in algorithm accuracy
- ☹️ Communication demanding
Main idea: obfuscate sensitive data before sharing to others

$$x' = x + r$$

Information-theoretic security model

Pros and cons:

😊 The adversary is assumed computationally unbounded
😊 Computationally simple
😊 Robust to n-1 number of corruptions
℅ Tradeoff between privacy and accuracy
Proposed approach
Information-theoretic security using noise insertion

Private data $x$, inserted noise $r$

The more noise inserted, the less privacy leakage

Normalized mutual information (i.e., information leakage) in terms of the amount of inserted noise for both additive and multiplicative cases.
Proposed approach

x-update of PDMM

\[ x_i^{(k+1)} = \arg \min_{x_i} \left( f_i(x_i, s_i) + \sum_{j \in N_i} \lambda_{j|i}^{(k)} \trans B_{i|j} x_i + \frac{c}{2} \sum_{j \in N_i} \| B_{i|j} x_i + B_{j|i} x_j^{(k)} - b_{i,j} \|^2_2 \right) \]

\[ 0 \in \partial f_i(x_i^{(k+1)}, s_i) + \sum_{j \in N_i} B_{i|j} \lambda_{j|i}^{(k)} + c \sum_{j \in N_i} (x_i^{(k+1)} - x_j^{(k)} - B_{i|j} b_{i,j}) \]

Motivation:

Instead of inserting additional noise, why not exploit the dual variable as noise?
Convergence behavior of dual variable

Consider two successive $\lambda$-updates

$$
\lambda^{(t+2)} = \lambda^{(t)} + c(Cx^{(t+2)} + 2PCx^{(t+1)} + Cx^{(t)})
$$

$$
H = \text{span}(C) + \text{span}(PC) \quad \lambda^{(t+2)} - \lambda^{(t)} \in H
$$

$$
\lambda^{(t)} = \Pi_H \lambda^{(t)} + (I - \Pi_H)\lambda^{(t)}
$$

$t \to \infty$ converge

Only be permuted at every iteration

$$
\lambda^* = P^{(t)}(I - \Pi_{\tilde{H}})\lambda^{(0)}
$$

Subspace noise

Non-convergence property will not affect the accuracy: \( x \to x^* \)

since \( (I - \Pi_{\tilde{H}})\lambda^{(0)} \)

\( L(x, \lambda) = f(x, s) + (P\lambda^{(k)})^\top Cx + \frac{c}{2} \Vert Cx + PCx^{(k)} - 2d \Vert^2_2 \)
Non-empty subspace always exists

\[ H = \text{span}(C) + \text{span}(PC) \]

Incidence matrix: \( B \in \mathbb{R}^{m \times n} \)

Incidence matrix:
\[
\begin{bmatrix}
C & PC
\end{bmatrix} = 
\begin{bmatrix}
B^+ & B^-
B^- & B^+
\end{bmatrix} \in \mathbb{R}^{2m \times 2n}
\]

As long as \( m \geq n \), non-zero subspace noise \((I - \Pi_H)\lambda^{(0)} \neq 0\) can be implemented without coordination between nodes:

- Each node randomly initializes its own dual variable with distributions having large variance.

Incidence matrix of a graph is always rank deficient \( \rightarrow \text{dim}(H) < 2n \)
Robustness against adversary models (1)

Eavesdropping adversary model
- It eavesdrops all communication channels between nodes

Channel encryption cost (only one iteration)
- Only the transmission of initialized dual variables needs channel encryption
Robustness against adversary models (2)

Passive (honest-but-curious) adversary model
- Corrupted nodes follow the protocol but share information together to infer the private data of honest nodes

Conditions for privacy guarantee
- One honest neighbor is required \( \mathcal{N}_{i,h} \neq \emptyset \)

\[
0 \in \partial f_i(x_i^{(k+1)}, s_i) + \sum_{j \in \mathcal{N}_i} B_{i|j} \lambda_{j|i}^{(k)} + c \sum_{j \in \mathcal{N}_i} (x_i^{(k+1)} - x_j^{(k)} - B_{i|j} b_{i,j})
\]

\[
\sum_{j \in \mathcal{N}_i} B_{i|j} \lambda_{j|i}^{(k)} = \sum_{j \in \mathcal{N}_{i,c}} B_{i|j} \lambda_{j|i}^{(k)} + \sum_{j \in \mathcal{N}_{i,h}} B_{i|j} \lambda_{j|i}^{(k)}
\]

Known to the corrupted nodes
Unknown to the corrupted nodes
What if the adversary knows the subspace?

The dual variables of honest nodes cannot be inferred even though the subspace (whole graph topology) is known to the adversary:

\[ \lambda^{(0)} \notin H \quad \leftrightarrow \quad \{ \lambda_{j|i} \}_{(i,j) \in \mathcal{N}_h \times \mathcal{N}_h, (i,j) \in \mathcal{E}} \] cannot be reconstructed

The proposed approach still preserves privacy even if the subspace is known to the adversary.
How about ADMM?

Augmented Lagrangian of ADMM

\[ L(x, \nu, z) = f(x) + \nu^T (Mx + Wz) + \frac{c}{2} \|Mx + Wz - 2d\|^2 \]

Updating functions

\[ x^{(k+1)} = \arg \min_x L(x, z^{(k)}, \nu^{(k)}) \]

\[ z^{(k+1)} = \arg \min_z L(x^{(k+1)}, z, \nu^{(k)}) \]

\[ \nu^{(k+1)} = \nu^{(k)} + c \left( Mx^{(k+1)} + Wz^{(k+1)} - 2d \right) \]

Incidence matrix

\[ [M \ W] = \begin{bmatrix} B^+ & -I \\ -B^- & -I \end{bmatrix} \in \mathbb{R}^{2m \times (m+n)} \]

Bipartite graph
The same applies to Dual ascent

Lagrangian of dual ascent

\[ L(x, u) = f(x) + u^\top (Bx - b) \]

Updating functions

\[ x^{(k+1)} = \arg \min_x L \left( x, u^{(k)} \right) \]

\[ u^{(k+1)} = u^{(k)} + t^{(k)} \left( Bx^{(k+1)} - b \right) \]
The proposed subspace perturbation also applies to other optimizers like ADMM and Dual Ascent
Applications:
applicable to all convex problems
Applications (1) _consensus

- How to securely compute the average salary over a group of people while keeping each person’s own salary private from others?

Distributed average consensus

\[
\min_{\mathbf{x}_i} \sum_{i \in \mathcal{N}} \frac{1}{2} \| \mathbf{x}_i - s_i \|_2^2
\]

s.t. \( \mathbf{x}_i = \mathbf{x}_j, \forall (i, j) \in \mathcal{E}. \)
Applications(2) _machine learning

- Privacy-preserving machine learning over multiple parties
- Collaborative learning without revealing private data

\[
\min_{\mathbf{x}_i} \sum_{i \in N} \frac{1}{2} \left\| y_i - Q_i \mathbf{x}_i \right\|_2^2 \\
\text{s.t. } \mathbf{x}_i = \mathbf{x}_j, \forall (i, j) \in \mathcal{E},
\]
Applications(3) _sparsity related

- Privacy-preserving distributed compressed sensing

\[
\min_{x_i} \sum_{i \in N} \frac{1}{2} \| y_i - Q_i x_i \|_2^2 \\
\text{s.t. } x_i = x_j, \forall (i, j) \in \mathcal{E},
\]

*Distributed least squaures and Lasso have similar problem setup, the former assumes an overdetermined system and the latter assumes an underdetermined one.
Numerical results
Network setup

The connectivity of nodes is enabled if their distance is within a radius $2\sqrt{\frac{\log n}{n}}$ to have a connected graph with high probability.

A random connected geometric graph with 50 nodes
Experimental results (Average consensus)

Proposed approaches

**PDMM**

**ADMM**

![Graphs showing convergence](image)

Fig.2: Convergence of the primal variable, the converging component and non-converging component of the dual variable in PDMM and ADMM with two different initializations.
Experimental results (Least square & Lasso)

Fig. 3: Distributed least squares with two different initializations of the dual variable with a variance of $10^6$: convergence of the optimization variable, the convergent and non-convergent component of the dual variable of (a) dual ascent, (b) ADMM and (c) PDMM.

Fig. 4: Distributed LASSO with two different initializations of the dual variable with a variance of $10^6$: convergence of the optimization variable, the convergent and non-convergent component of the dual variable of (a) dual ascent, (b) ADMM and (c) PDMM.
Comparison with existing methods

Convergence of the proposed PDMM and state-of-the-art algorithms under three different noise levels for distributed average consensus.

DP: differential privacy
[Nozari, 2017]
CNI: correlated noise insertion
[He, 2019]
Lower bound of information leakage

Sometimes it is impossible to have zero information leakage

- The optimum solution itself may reveal some private information (unavoidable if perfect accuracy is preserved)

Distributed average consensus

Zero privacy leakage and perfect accuracy are sometimes impossible to achieve

Fig. 7: Normalized mutual information of an arbitrary node \( i \) (i.e., \( \frac{I(S_i;X_i^{(k+1)})}{I(S_i;S_i)} \)) using the proposed p-PDMM and non-private PDMM (n-PDMM) for each iteration.
Conclusions and future works
Conclusions & future works

Conclusions

• A new subspace perturbation approach based on distributed convex optimization
• Generally applicable to all convex problems
• Both computationally and communication efficient (compared to SMPC)
• No tradeoff between privacy and accuracy (compared to differential privacy)
• Convergence rate is not affected
• Require one honest neighbor

Future works:

• Optimization in terms of practical constraints for example quantization
• Apply to distributed federated learning
Q&A

thank you!