Differentially Private Distributed Optimization

ZHENQI HUANG ∘ SAYAN MITRA ∘ NITIN VAIDYA

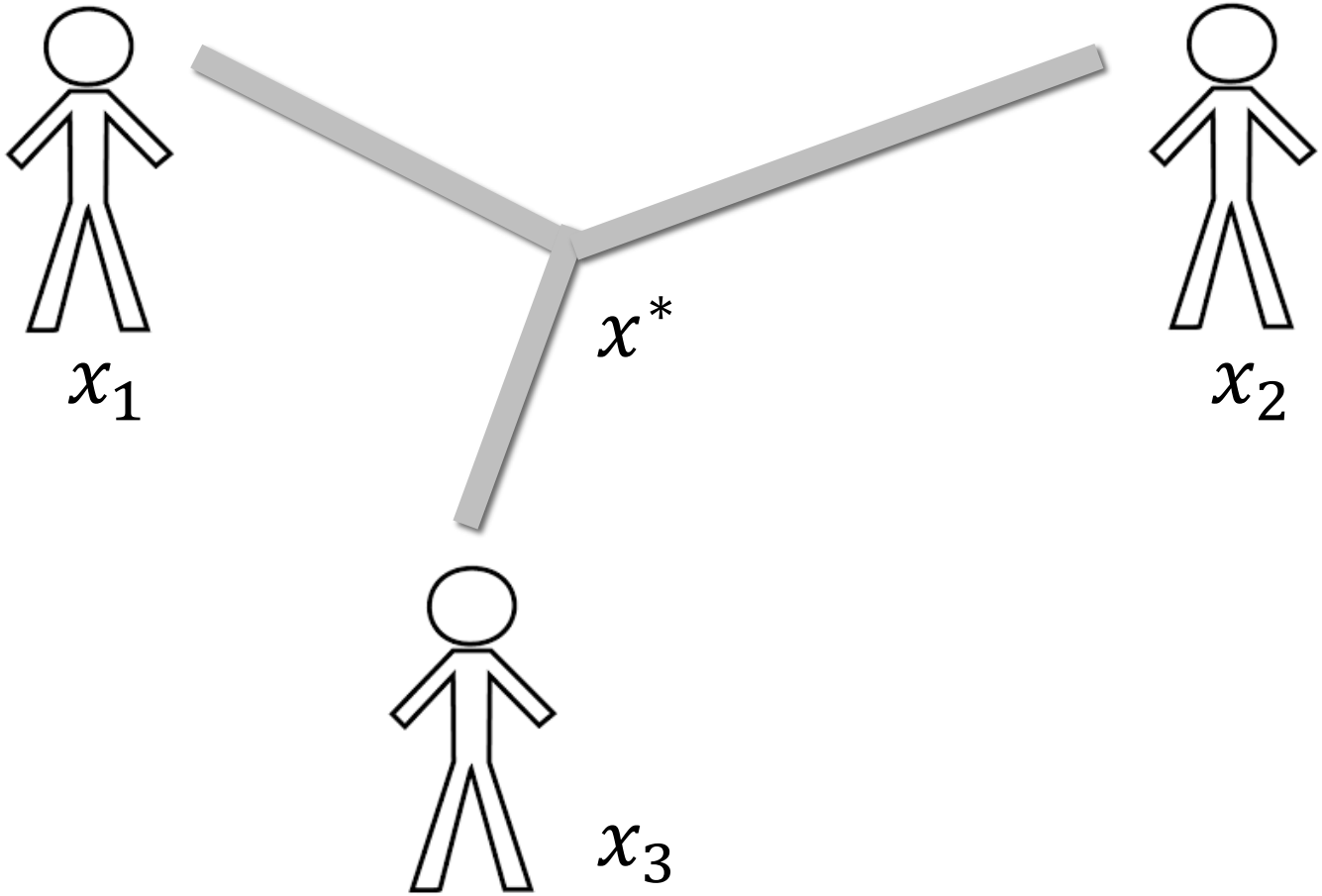
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

UNIVERSITY OF ILLINOIS AT URBANA CHAMPAIGN

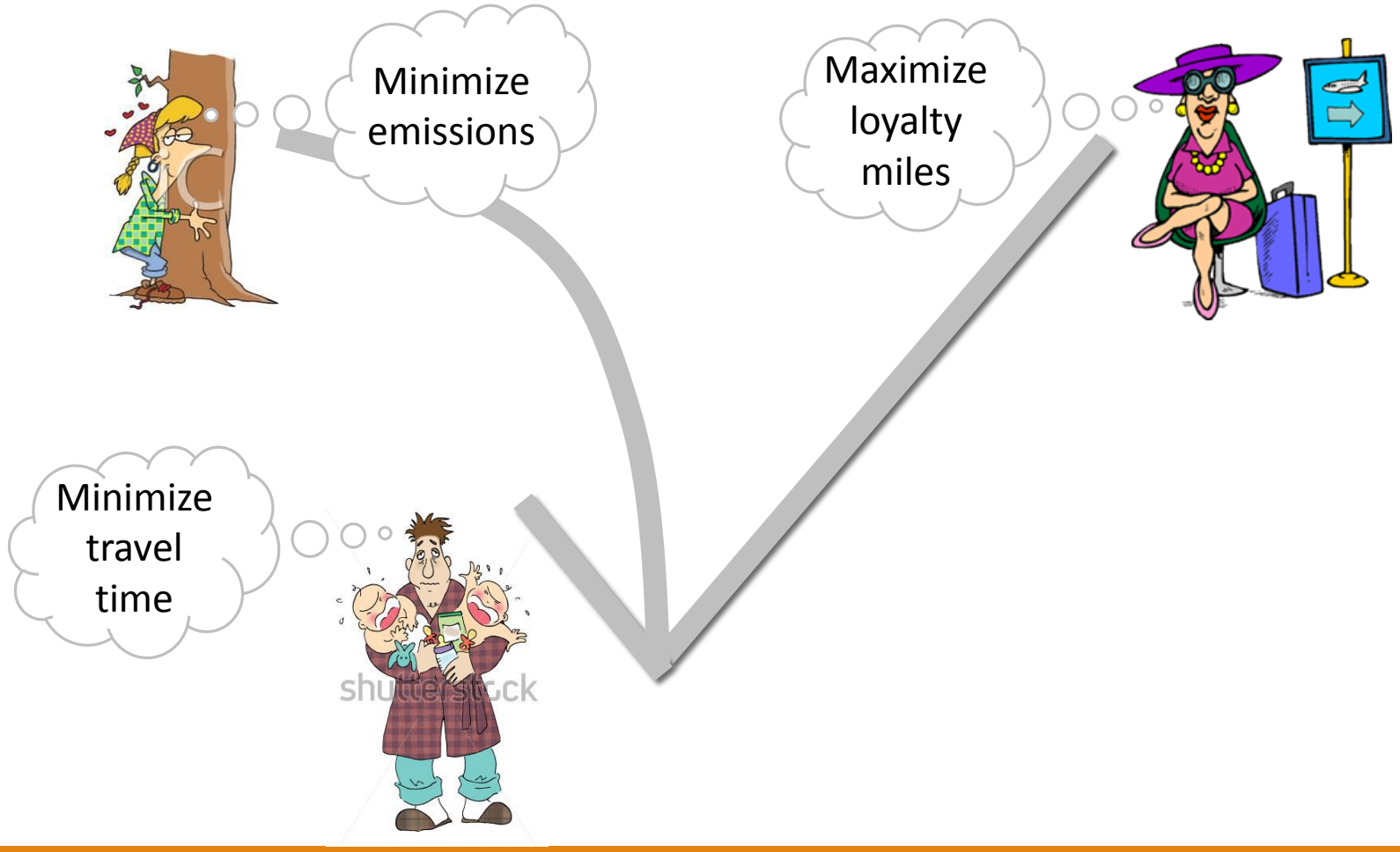
INTERNATIONAL CONFERENCE ON DISTRIBUTED COMPUTING AND NETWORKING 2015

Search for an address or place

Distributed optimization: Rendezvous



Rendezvous with private preferences



Outline

1. Introduction
2. The Private Distributed Optimization Problem
3. Privacy
4. Convergence
5. Accuracy
6. Conclusions

Background : iterative consensus & optimization

Distributed consensus:

- Initially $x_1(0), \dots, x_N(0) \in \mathbb{R}$, desire that $\lim_{t\to\infty} x_i(t) = \overline{x}(0)$
- Simple algorithms: $x_{i,j}(t+1) = \frac{x_i(t)+x_j(t)}{2}$ **[DeGroot 74]**
- $x(t+1) = A(t)x(t)$ stochastic matrices

Gradient type methods

- $\min \sum f_i(x); x \leftarrow x - \gamma \nabla f(x)$

Differential Privacy for Databases

- **[Dinur & Nissim 03] [Dwork, Nissim, McSherry & Smith 06]**
- Privacy for query on databases $Q(D) \to y$

Distributed Optimization Problem (DOP)

N participating agents

$f_i: \mathbb{R}^n \rightarrow \mathbb{R}$ agent $i's$ cost function

- Convex, smooth

$f(x) = \sum_{i=1}^{N} f_i(x)$ global cost function

- $x^* = argmin_x f(x)$

$\boldsymbol{A} = A(1), A(2), \ldots$ sequence of **robustly connected doubly stochastic** matrices define the allowed communication among agents in each round

- $A(t) = \{a_{ij}(t)\}$

P

Iterative algorithms for DOP

At agent i:

$x_i(t) \in \mathbb{R}^n$: agent $i's$ estimate of x^* at the beginning of round t

$y_i(t) \leftarrow x_i(t) + \omega(t)$

$Broadcast(y_i(t)); Receive(\{y_j(t)\})$

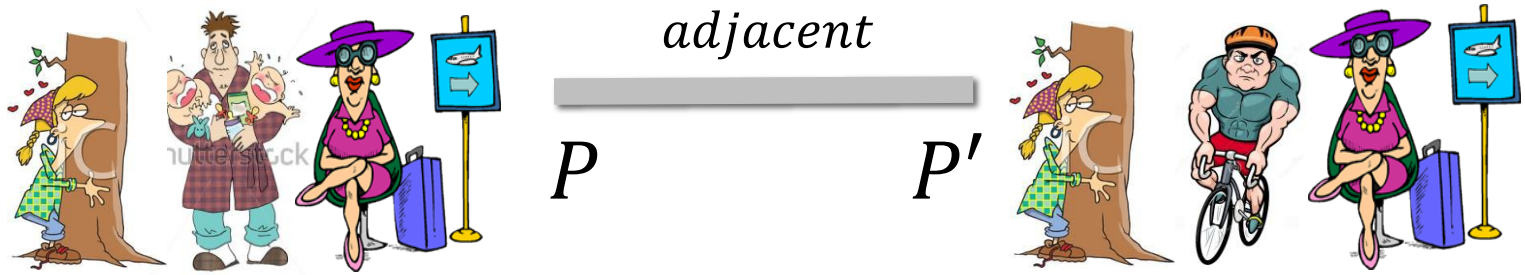
$z_i(t) \leftarrow \sum_j a_{ij}(t)\, y_j(t)$

$x_i(t+1) \leftarrow z_i(t) - \gamma_t \nabla f_i(z_i(t))$

γ_t: step size geometrically decaying with t

Adjacent PDOPs

Two DOPs P and P' are ***adjacent*** if all their components are identical except for one particular agent's cost function. That is, there exits i such that $f_i \neq f_i'$ and for all other j, $f_j = f_j'$ and for all t $A(t) = A'(t)$.



An iterative algorithm is ***ϵ*-differentially private** ($\epsilon > 0$) if for any adjacent pair of DOPs P and P', any set of observation sequences Y:

$$\frac{Prob[Obs_P^{-1}(Y)]}{Prob[Obs_{P'}^{-1}(Y)]} \leq e^{\epsilon}$$

Iterative algorithms for DOP

At agent i:

$x_i(t) \in \mathbb{R}^n$: agent $i's$ estimate of x^* at the beginning of round t

$y_i(t) \leftarrow x_i(t) + Lap(M_t)$

$Broadcast(y_i(t)); Receive(\{y_j(t)\})$

$z_i(t) \leftarrow \sum_j a_{ij}(t)\, y_j(t)$

$x_i(t+1) \leftarrow Proj(z_i(t) - \gamma_t \nabla f_i(z_i(t)))$

M_t and γ_t are sequences that are geometrically decaying with t

$x(t) = [x_1(t), \ldots, x_N(t)]$

$y(t), z(t)$ similarly defined

Execution

$\langle x(1),$ ...

Obse

$y($

On s

Lap

seq

initia

Sensitivity to private information

Sensitivity: Maximum change in the system state (at time t) per unit change in private information, for the same observations. In this case, the change in agent cost functions is the change in the input.

$$\Delta(t) = \sup_{\rho} \sup_{\substack{x(t) \in Obs_P^{-1}(\rho) \\ x'(t) \in Obs_{P'}^{-1}(\rho)}} \left|\left|x(t) - x'(t)\right|\right|_1$$

For an observation ρ let $\alpha = Obs_P^{-1}(\rho)$ and $\alpha' = Obs_{P'}^{-1}(\rho)$

Along these executions $y(t) = y'(t)$, it follows that $z(t) = z'(t)$

$$\left|\left|x(t) - x'(t)\right|\right|_1 = ||z_i(t) - \gamma_t \nabla f_i\big(z_i(t)\big) - z_i(t) - \gamma_t \nabla f'_i\big(z_i(t)\big)||_1 =$$

Theorem. If $\sum \frac{\Delta(t)}{M_t} \leq \epsilon$ then the algorithm is ϵ-Differentially Private.

Consider two adjacent DOPs and a set of observations Y.

$$\frac{Prob[Obs_P^{-1}(Y)]}{Prob[Obs_{P'}^{-1}(Y)]} = \frac{\int_{\alpha \in Obs_P^{-1}(Y)} P[\alpha] d\mu}{\int_{\alpha' \in Obs_{P'}^{-1}(Y)} P[\alpha'] d\mu} = \frac{\int_{\alpha \in Obs_P^{-1}(Y)} P[\alpha] d\mu}{\int_{\alpha \in Obs_P^{-1}(Y)} P[B(\alpha)] d\mu}$$

$$= \prod_{i \in [N]} \frac{Lap(M_t)(y_i(t) - x_i(t))}{Lap(M_t)(y_i'(t) - x'_i(t))}$$

$$= \prod_{i \in [N]} \exp\left(\frac{|y_i(t) - x_i(t) - y_i'(t) + x_i'(t)|}{M_t}\right)$$

$$= \prod_{i \in [N]} \exp\left(\frac{|x_i'(t) - x_i(t)|}{M_t}\right)$$

$$= \exp\left(\sum_{i \in [N]} \frac{|x_i'(t) - x_i(t)|}{M_t}\right) \leq \exp\left(\sum_{i \in [N]} \frac{\Delta(t)}{M_t}\right) \leq \epsilon$$

Convergence in Expectation

We show convergence of the algorithm in expectation

Theorem. The iterative algorithm converges in expectation. That is,

$$\lim_{t\to\infty} E\left|\left|x_i(t) - x_j(t)\right|\right| = 0.$$

Key property of robustly connected doubly stochastic matrices: for any i, j and any natural numbers $t > s$ each term in the product $A(s +$

Accuracy

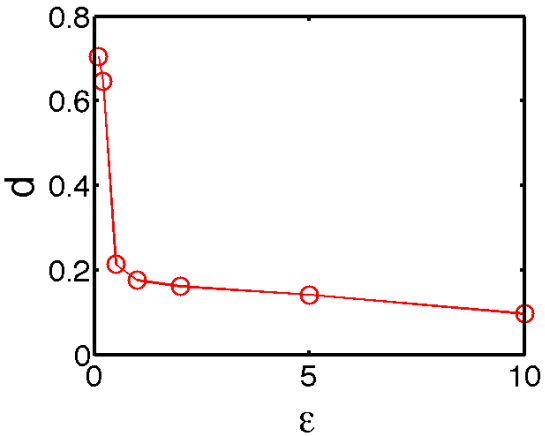
For a given execution, we define the average at time t as

$$\overline{x}(t) = \frac{1}{N}\sum_{i} x_i(t)$$

For $d \geq 0$ an iterative distributed optimization algorithm is ***d*-accurate** if $\lim_{t\to\infty} E\left|\left|\overline{x}(t) - x^*\right|\right|^2 \leq d.$

Higher level of privacy ⇒ More noise ⇒ Quadratic decay in accuracy

Fixing all other parameters $d \sim O(\frac{1}{\epsilon^2})$



Summary & Future directions

Participants might be willing to sacrifice optimality provided this loss is commensurate with the gain privacy

We characterize this trade-off for an iterative optimization algorithm

Generalization to nonconvex functions and sub-gradient methods

Open problem: Establishing a lower-bound. What is the best accuracy we can get for a given level of privacy **[Wang et al. CDC 2014]**