

Exam of the Course: SOD314

2 May 2024

Notes: this exam lasts 2 hours, you can use the lecture notes that were distributed during the classes. There are 2 exercises. You can refer to Theorems that are present in the notes by indicating their number and page. In the interest of time, keep your answers concise and to the point.

Exercise 1. A faulty node

A number of nodes communicate as in Figure 1. As you can see, the node labeled with 8 is faulty and it cannot send messages to nodes 1 and 7, and it cannot receive anything from node 4. The nodes want to solve the problem,

$$\min_{x \in \mathbf{R}^n} \sum_{i=1}^N \|x - y_i\|_2^2,$$

with $N = 8$ and with $y_i \in \mathbf{R}^n$ being a datum belonging to node i .

1. Prove that the optimal solution is

$$x^* = \frac{1}{N} \sum_{i=1}^N y_i.$$

2. Use the fact that you know that the optimizer is the average of the data, to derive a first-order distributed algorithm for the nodes to reach x^* based on the communication topology in Figure 1. How much communication you need between iterations? What is the convergence like?
3. Consider now a communication topology where we *remove* the directed communication links going from node 1 and node 7 to node 8 and from node 8 to node 4. Derive a first-order distributed algorithm for the nodes to reach x^* based on this new reduced communication topology. How much communication you need between iterations? What is the convergence like?
4. What could be a good algorithmic strategy if the communication topology would oscillate (change at every time) between the topology of Figure 1 and the one of Point 3? Argue and derive the algorithm.

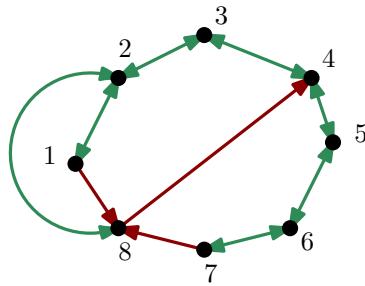


Figure 1: A communication topology with a faulty node.

5. Consider the topology of Point 3 and derive a peer-to-peer ADMM algorithm to solve the problem. What are the convergence guarantees that you can give? Can you find the best β parameter?

Exercise 2. Biased models through unfair selection

Consider a federated learning setting, where a number of users share their views on social matters on a social media platform. The platform collects their views and shares the averaged trend with them and with third-party organizations.

The problem that we are trying to solve is,

$$\min_{x \in \mathbf{R}^n} \mathbf{E}_{y \in \mathcal{Y}} \left[\frac{1}{2} \|x - y\|_2^2 \right] \approx \frac{1}{P} \sum_{p=1}^P \mathbf{E}_{y \in \mathcal{Y}_p} \left[\frac{1}{2} \|x - y\|_2^2 \right].$$

Each user $i \in \{1, \dots, P\}$ has a personal a-priori distribution \mathcal{Y}_p which shapes their own view x^p .

Consider the case in which the users are well separated on a particular matter. In particular, for the set of users \mathcal{C} , of cardinality $P/2$ (P is supposed even), their a-priori \mathcal{Y}_p 's have the same distribution, say a normal distribution with mean \bar{y}_C and variance σ^2 . The other users have all another distribution, say a normal distribution with mean \bar{w} and variance $100\sigma^2$.

1. Describe the FedAvg and Scaffold algorithms to solve the problem above in a federated learning setting. Which one is more adapted to solving this particular problem? Why?
2. Consider a setting where each user uses a constant learning rate $\gamma \in (0, \frac{1}{2}]$ and a local stochastic gradient descent, run for 1 epoch. Consider full participation. Prove that, with the FedAvg algorithm, each mixing leads to

$$x_{t+1} = \frac{1}{P} \sum_{i=1}^P x_{t+1}^p = (1 - \gamma)x_t + \gamma \frac{1}{P} \sum_{i=1}^P y_{p,t}, \quad y_{p,t} \sim \mathcal{Y}_p.$$

And that,

$$\lim_{t \rightarrow \infty} \lim_{P \rightarrow \infty} x_{t+1} = \sum_{\tau=0}^t (1 - \gamma)^{t-\tau} \gamma \lim_{P \rightarrow \infty} \frac{1}{P} \sum_{i=1}^P y_{p,\tau} = \frac{\bar{y}_C + \bar{w}}{2}.$$

How the local models x_∞^p look like at convergence?

3. Consider now the setting in which the platform chooses to average only the messages coming from the set \mathcal{C} , which are closer to its opinions. It can do it by looking at the sent messages and see which ones are closer to its opinions. Prove that the global model is then,

$$\lim_{t \rightarrow \infty} \lim_{P \rightarrow \infty} x_{t+1} = \bar{y}_C.$$

What about the local models?

4. For the latter setting, describe a differential privacy mechanism that could be used here to obfuscate the local models, so that a biased selection of the users cannot take place and bias is avoided.
5. What is peer-to-peer federated learning? Can we use that to avoid unfair selection?