Exam of the Course: SOD314

28 March 2023

Notes: this exam lasts 2 hours, you can use the material that was distributed during the classes. There are 4 exercises. You can refer to Theorems that are present in the slides by indicating their number and slide page. In the interest of time, keep your answers concise and to the point.

Exercise 1. Consider the convex problem

$$(P) \qquad \min_{x\in\mathbf{R}} \sum_{i=1}^{N} f_i(x),$$

for $f_i : \mathbf{R} \to \mathbf{R}$, which is μ-strongly convex and L-smooth.

1. Describe the idea of DGD with constant stepsize α. In particular, show that DGD with constant stepsize α is equivalent to solving the following penalized problem,

$$(P) \qquad \min_{x_1,\dots,x_N\in\mathbf{R}} \sum_{i=1}^{N} f_i(x_i) + \frac{1}{2\alpha} \sum_{i=1}^{N} \left(\|x_i\|^2 - \sum_{j=1}^{N} w_{ij} x_i x_j \right)$$

   where w_{ij} are the coefficients of a doubly-stochastic matrix $[W]_{ij} = w_{ij}$.

2. Consider now $N = 2$ for simplicity, and a different penalization. In particular, consider the problem,

$$(P') \qquad \min_{x_1,x_2\in\mathbf{R}} \sum_{i=1}^{2} f_i(x_i) + \epsilon|x_1 - x_2|.$$

   (a) Apply a (sub)-gradient descent to it using a constant stepsize α, and remember that $\partial_{x_1}|x_1 - x_2| = \text{sign}(x_1 - x_2)$. Show that the resulting algorithm is distributed (in the sense that f_i is not shared).

   (b) What is the convergence and convergence rate you can expect for the algorithm developed at the previous point?

   (c) Show that, if you assume that $\|\nabla f_i(x)\| \leq B$ for all $x \in \mathbf{R}$, and you choose $\epsilon > B$, then at convergence, $x_1^* = x_2^*$, and therefore the penalization is exact. This means that you have no error.

   (Hint) Use the optimality conditions.

   (d) **[Bonus question]** A sub-gradient algorithm is very slow. Write a proximal gradient algorithm and show that is distributed and converges linearly.

Exercise 2. Resource allocation.

Consider the convex problem,

$$(P) \qquad \min_{x_1,x_2,\dots,x_N\in\mathbf{R}} \sum_{i=1}^{N} f_i(x_i), \quad \text{subject to } \sum_{i=1}^{N} x_i = 1,$$

for $f_i : \mathbf{R} \to \mathbf{R}$, which is μ-strongly convex and L-smooth.

Consider a dual decomposition approach to solve (P) in a distributed way. Introduce the Lagrangian function,

$$\mathcal{L}(x, \lambda) = \sum_{i=1}^{N} f_i(x_i) + \lambda(\sum_{i=1}^{N} x_i - 1).$$

1. Write the resulting dual decomposition algorithm in a cloud-based setting (i.e., having only one $\lambda \in \mathbf{R}$ that is shared with the cloud).

2. Let $q(\lambda)$ be the dual function; prove that the dual problem can be written as,

$$\min_{\lambda} \sum_{i=1}^{N} -q_i(\lambda), \quad q_i(\lambda) := \min_{x_i}\{f_i(x_i) + \lambda(x_i - 1/N)\}.$$

3. Write a distributed (peer-to-peer) ADMM algorithm to solve the above dual problem (giving a copy λ_i of the dual variables to each node..), considering an arbitrary communication graph between the nodes $1, \dots, N$. Which convergence, and convergence rate should I expect?

4. **[Bonus question]** Write the ADMM as completely as possible when $f_i(x_i) = \frac{p_i}{2} x_i^2$ for $p_i > 0$.

Exercise 3. Consider the standard FedAvg algorithm to solve the problem $\min_x \frac{1}{P} \sum_{i=1}^{P} f_i(x)$.

1. Write the algorithm, describe it briefly, and highlight that the client update iterations are in fact an application of SGD.

2. Describe the idea behind SAGA. Why SCAFFOLD looks like FedAvg when instead of SGD we employ SAGA?

3. Consider now yet another variant of FedAvg, when instead of SGD, we employ the following single-epoch client update iterations:

$$x_{t+1}^i = \arg\min_x \{ f_i(x) + \langle \nabla f(x_t) - \nabla f_i(x_t), x - x_t \rangle + \frac{\mu}{2} \|x - x_t\|^2 \}, \qquad \forall i \in \{1, \dots, P\},$$

   where f_i is the local functions (harbouring the local data), $f(x) = \frac{1}{P} \sum_{i=1}^{P} f_i(x)$, and $\mu > 0$.
   This update is not local, since there is a coupling term, which one?

4. Devise a method that requires an extra mixing step among randomly selected clients, to make the above iterations local.

Exercise 4. In this exercise, we consider Class 5 and the issues related to privacy. In particular, we look at the optimization problem,

$$(P) \qquad \min_{x \in \mathbf{R}} \sum_{i=1}^{N} f_i(x),$$

for $f_i : \mathbf{R} \to \mathbf{R}$, which is μ-strongly convex and L-smooth.

1. Describe the concept of ϵ-Differential Privacy. What is a Laplacian mechanism?

2. Consider the following variant of DGD-DP (see slide 118 for notation):

$$x_{k+1}^i = x_k^i - \sum_{j \in N_i} w_{ij} (x_k^j + \zeta_k^j - x_k^i) - \alpha_k \nabla f_i(x_k^i)$$

   Would this algorithm verify the assumptions of Theorem 13 and 14?

3. How does the DGD-DP that we have seen in class enforce optimality and privacy at the same time?