IEEE MAG 4: Gradient Tracking Algorithms

Sunday, April 7, 2024 8:00 PM

Assume that the agents are aware that they comprise a system but are not aware of the structure of the graph nor the global objective function.
In addition to mixing their variables, they also use and mix directions that serve as estimates of the global objective gradient.

$$v_i^{(k)} = \sum_{j=1}^{m} a_{ij} x_j^{(k)}$$

$$x_i^{(k+1)} = v_i^{(k)} - \frac{\alpha}{m} \sum_{i=1}^{m} \nabla f_i(v_i^{(k)})$$

α: fixed step size

Without knowing all of the functions fi, each agent can track the aggregate function gradients by using a direction di^k and tracking the average of these directions.
At time k+1, the agents have xi^k and di^k and exchange these with their neighbors. They then compute the following:

(16) Works on undirected graphs

$$x_i^{(k+1)} = \sum_{j=1}^{m} a_{ij} x_j^{(k)} - \alpha d_i^{(k)}$$

$$d_i^{(k+1)} = \underbrace{\sum_{j=1}^{m} a_{ij} d_j^{(k)}}_{\text{tracks the sum of the gradients of the overall system function objective.}} + \nabla f_i(x_i^{(k+1)}) \underbrace{- \nabla f_i(x_i^{(k)})}_{\text{Subtracted to ensure that only new gradient information is captured}}$$

- Requirements:

→ $d_i^{(0)} = \nabla f_i(x_i^{(0)}) \quad \forall i \in [m]$

→ A → column stochastic

→ $d_i^{(k+1)}$: $\sum_{i=1}^{m} d_i^{(k+1)} = \sum_{i=1}^{m} \nabla f_i(x_i^{(k+1)})$

- Average Iterates:

$x^{(-k)} = \frac{1}{m} \sum_{i=1}^{m} x_i^{(k)}$ satisfies

$$x^{(-k+1)} = x^{(-k)} - \frac{\alpha}{m} \sum_{i=1}^{m} \nabla f_i(x_i^{(k)})$$

↓ erroneous centralized GD:

$$x^{(-k+1)} = x^{(-k)} - \frac{\alpha}{m} \sum_{i=1}^{m} \nabla f_i(x_i^{(k)}) + e_k$$

$$e_k = \frac{\alpha}{m} \sum^{m} \left(\nabla f_i(x^{(-k)}) - \nabla f_i(x_i^{(k)})\right)$$

$$x^{(-k+1)} = x^{(-k)} - \frac{\alpha}{m} \sum_{i=1}^{m} \nabla f_i(x_i^{(k)}) + e_k$$
$$e_k = \frac{\alpha}{m} \sum_{i=1}^{m} \left(\nabla f_i(x^{(-k)}) - \nabla f_i(x_i^{(k)}) \right)$$

For **smooth convex functions** with **Lipschitz continuous gradients**, the convergence of the method is in the order of **O(1/k)**

Lipschitz: $\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\| \quad \forall x, y \in \mathbb{R}^n$

If fi is **strongly convex**, and with a carefully chosen alpha, the method produces iterates:

$x_i^{(k)}$ that converges to x^* minimizing global objective function:

$\frac{1}{M} \sum_{i=1}^{m} f_i(x)$ with geometric rate:

$\|x_i^{(k)} - x^*\| \leq q^{(k)} C \quad \forall$ agents i and all $k \geq 0$

$q \in (0,1)$, C constant depends on G, A, α, other parameters

Thus, the **distributed** method with the **gradient tracking mechanism matches** the fastest convergence rate of **centralized** gradient method **(geometric)**

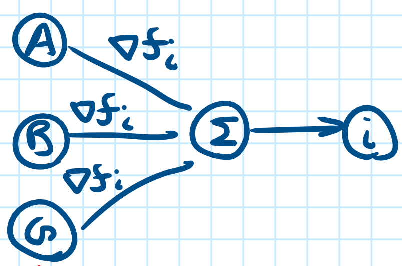
- **Augmented Gradient Push Method**

If G is a directed graph, the gradient-push method can also be augmented to include a gradient tracking mechanism.

$$v_i^{(k+1)} = \sum_{j \in N_i^{in} \cup \{i\}} \left(a_{ij} v_j^{(k)} - \alpha d_j^{(k)} \right)$$

$$y_i^{(k+1)} = \sum_{j \in N_i^{in} \cup \{i\}} a_{ij} y_j^{(k)}$$

$$x_i^{(k+1)} = \frac{v_i^{(k+1)}}{y_i^{(k+1)}}$$



(17)
Works on directed graphs

$$d_i^{(k+1)} = \underbrace{\sum_{j \in N_i^{in} \cup \{i\}} a_{ij} d_j^{(k)}}_{\text{sum of gradients of neighbors of agent } i} + \underbrace{\nabla f_i(x_i^{(k+1)}) - \nabla f(x_i^{(k)})}_{\text{error : difference of current gradient to previous gradient}}$$

Also has a geometric convergence rate.
(16) works on undirected graphs, (17) works on directed graphs.
Undirected graph -> Doubly stochastic A.
Directed graph -> Column stochastic A.