MoTEF

June 21, 2024

1 Algorithm in terms of clients updates

Algorithm 1 MoTEF

Input: $\mathbf{x}_i^0 = \mathbf{x}^0, \mathbf{g}_i^0, \mathbf{h}_i^0, \mathbf{v}_i^0, \gamma, \eta, \lambda$, and $\mathcal{C}_\alpha$
for $t = 0, 1, 2, \dots$ **do** in parallel for each client $i \in [n]$
  for neighbors $j : (i, j) \in E$ **do**
    send $\mathbf{q}_{h,j}^t, \mathbf{q}_{g,j}^t$
    receive $\mathbf{q}_{h,j}^t, \mathbf{q}_{g,i}$
    $\mathbf{h}_j^t = \mathbf{h}_j^{t-1} + \mathbf{q}_{h,j}^t$
    $\mathbf{g}_j^t = \mathbf{g}_j^{t-1} + \mathbf{q}_{g,j}^t$
  $\mathbf{x}_i^{t+1} = \mathbf{x}^t + \gamma \sum_{j:(i,j)\in E} w_{ij}(\mathbf{h}_j^t - \mathbf{h}_i^t) - \eta \mathbf{v}_i^t$
  $\mathbf{q}_{h,i}^{t+1} = \mathcal{C}_\alpha(\mathbf{x}_i^{t+1} - \mathbf{h}_i^t)$
  $\mathbf{h}_i^{t+1} = \mathbf{h}_i^t + \mathbf{q}_{h,i}^{t+1}$
  $\mathbf{m}_i^{t+1} = (1 - \lambda)\mathbf{m}_i^t + \lambda \nabla f_i(\mathbf{x}_i^{t+1}, \xi_i^{t+1})$
  $\mathbf{v}_i^{t+1} = \mathbf{v}_i^t + \gamma \sum_{j:(i,j)\in[n]} w_{ij}(\mathbf{g}_j^t - \mathbf{g}_i^t) + \mathbf{m}_i^{t+1} - \mathbf{m}_i^t$
  $\mathbf{q}_{g,i}^{t+1} = \mathcal{C}_\alpha(\mathbf{v}_i^{t+1} - \mathbf{g}_i^t)$
  $\mathbf{g}_i^{t+1} = \mathbf{g}_i^t + \mathbf{q}_{g,i}^{t+1}$

Client i maintains:

- local state $\mathbf{x}_i^t$
- local approximation of the global gradient (momentum) $\mathbf{v}_i^t$
- local approximation of $\mathbf{x}_i^t$ as $\mathbf{h}_i^t$
- local approximation of $\mathbf{v}_i^t$ as $\mathbf{g}_i^t$
- all of its neighbors $j : (i, j) \in E$: $\mathbf{h}_j^t$ and $\mathbf{g}_j^t$