Talk 7: Gradient Tracking in Decentralized Optimization

JIN Zhengfen

2021-9-9

Main Resource: [1, 2, 3, 4, 5, 6]

Key points:

- decentralized, gradient tracking, compression
- (strongly) convex + smooth, non-convex + smooth
- neural networks, empirical risk minimization

Gradient Tracking

problem

$$\min_{\theta} f(\theta) = \frac{1}{n}\sum_{i=1}^{n} f_i(\theta)$$

DGD (decentralized gradient descent):

$$\theta_{k+1}^i = \sum_{r\in\mathcal{N}_i} w_{ir}\theta_k^r - \alpha_k \nabla f_i(\theta_k^i)$$

GT(gradient tracking)-DGD:

$$\theta_{k+1}^i = \sum_{r\in\mathcal{N}_i} w_{ir}\theta_k^r - \alpha_k d_k^i$$

$$\text{GT} \rightarrow d_{k+1}^i = \boxed{\sum_{r\in\mathcal{N}_i} w_{ir} d_k^r} + \nabla f_i(\theta_{k+1}^i) \boxed{-f_i(\theta_k^i)}$$

$\rightsquigarrow$ DSDG, GT-DSGD, with gradients $\nabla f_i(\theta_{k+1}^i)$, $f_i(\theta_k^i)$ replaced by stochastic gradients $\nabla f_{i,s_i^{k+1}}(\theta_{k+1}^i)$, $f_{i,s_i^k}(\theta_k^i)$ resp.

$\rightsquigarrow$ accelerated gradient tracking over **time-varying graphs**.

Question 0.1 (TODO) *Compare gradient tracking and variance reduction.*

Gradient compression and error feedback (ref. Talk 6),

- compressed gradient tracking [3]
- compressed push-pull [2]

References

[1] H. Li and Z. Lin, "Accelerated Gradient Tracking over Time-varying Graphs for Decentralized Optimization," *arXiv preprint arXiv:2104.02596*, 2021.

[2] Z. Song, L. Shi, S. Pu, and M. Yan, "Compressed Gradient Tracking for Decentralized Optimization Over General Directed Networks," *arXiv preprint arXiv:2106.07243*, 2021.

[3] Y. Liao, Z. Li, K. Huang, and S. Pu, "Compressed Gradient Tracking Methods for Decentralized Optimization with Linear Convergence," *2103.13748*, 2021.

[4] J. Zhang and K. You, "Decentralized Stochastic Gradient Tracking for Non-convex Empirical Risk Minimization," *arXiv preprint arXiv:1909.02712*, 2019.

[5] R. Xin, S. Kar, and U. A. Khan, "Gradient Tracking and Variance Reduction for Decentralized Optimization and Machine Learning," *arXiv preprint arXiv:2002.05373*, 2020.

[6] H. Sun, S. Lu, and M. Hong, "Improving the Sample and Communication Complexity for Decentralized Non-convex Optimization: Joint Gradient Estimation and Tracking," in *International Conference on Machine Learning*, pp. 9217–9228, PMLR, 2020.