General Proximal Incremental Aggregated Gradient Algorithms: Better and Novel Results under General Scheme

Tao Sun, Yuejiao Sun, Dongsheng Li[a], Qing Liao

National University of Defense Technology & Harbin Institute of Technology (Shenzhen)

[a]Corresponding Author.

Overview

- We propose a general PIAG algorithm, which covers various classical algorithms.
- We apply the line search strategy to PIAG and prove its convergence.
- Compared with previous convergence analysis of PIAG, we use a new proof technique: Lyapunov function analysis.
- We proved better and novel results under general scheme.

Set up

Model: We consider the smooth nonconvex optimization problem

$$\min_{x\in\mathbb{R}^N} f(x)+g(x), \tag{1}$$

where $f(x)=\sum_{i=1}^{M} f_i(x)$

Algorithms: Consider a general proximal incremental aggregated gradient algorithm which performs as

$$\begin{cases}\mathbb{E}(v^k \mid \chi^k)=\sum_{i=1}^{m}\nabla f_i(x^{k-\tau_{i,k}})+e^k,\\ x^{k+1}=\mathbf{prox}_{\gamma_k g}[x^k-\gamma_k v^k],\end{cases} \tag{2}$$

where $\tau_{i,k}$ is the delay associated with f_i and e^k is the noise in the k-th iteration.

Difficulty

The difficulty of the study on the performance of general PIAG lies in the fact that the objective function values fail to be decreasing. It is thus quite difficult to establish the sublinear convergence rates.

Lyapunov function

The analysis in this section is heavily based on the following Lyapunov function and its variants:

$$\xi_k(\varepsilon,\delta)=\frac{L}{2\varepsilon}\sum_{d=k-\tau}^{k-1}(d-(k-\tau)+1)\|\Delta^d\|^2 + F(x^k)+\frac{1}{2\delta}\sum_{i=k}^{+\infty}\sigma_i^2-\min F,$$

where $\sigma_k := \left[\mathbb{E}(\|v^k-\sum_{i=1}^{m}\nabla f_i(x^{k-\tau_{i,k}})\|^2 \mid \chi^k)\right]^{\frac{1}{2}}$.

Problem

Assumption: The objective function f_i is differentiable and gradient is L Lipschitz, the function g is proximable. The convergence analysis in the paper depends on the square summable assumption on $(\sigma_k)_{k\geq 1}$, i.e., $\sum_i \sigma_i^2 < +\infty$.

Selected Results

Theorem 1: Assume the gradient of f is Lipschitz continuous with L and g is convex, and $\mathbf{prox}_g(\cdot)$ is bounded. Choose the step size $\gamma_k \equiv \gamma = \frac{2c}{(2\tau+1)L}$ for arbitrary fixed $0<c<1$. Let $(x^k)_{k\geq 0}$ be generated by the general proximal incremental aggregated gradient algorithm. And the $\sigma_k \sim O(\zeta^k)$, where $0<\zeta<1$. Then, we have $\mathbb{E}F(x^k)-\min F \sim O(\frac{1}{k})$. Further more if function F satisfies PŁ condition and $\sigma_k \sim O(\zeta^k)$, where $0<\zeta<1$. Then, we have $\mathbb{E}F(x^k)-\min F\sim O(\omega^k)$, for some $0<\omega<1$.

Theorem 2: Let f be a convex function with L-Lipschitz gradient and g is convex, and finite $\min F$. Let $(x^k)_{k\geq 0}$ be generated by the proximal incremental aggregated gradient algorithm with line search, and $\max_{i,k}\{\tau_{i,k}\}\leq\tau$. Choose the parameter $c_2 \geq \frac{(2\tau+1)L}{2c}$ and $0<c<1$. If F is coercive, $F(x^k)-\min F \sim O(\frac{1}{k})$. If F satisfies PŁ condition, $F(x^k)-\min F\sim O(\tilde{\omega}^k)$ for some $0<\tilde{\omega}<1$.