Problem

$$\min_{x\in\mathbb{R}^d} f(x)$$

$$x^{k+1} = x^k - \lambda_k \nabla f(x^k)$$

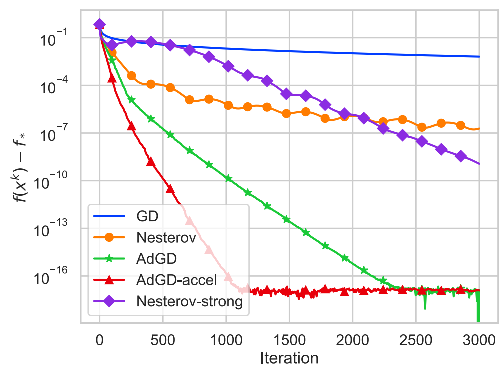
Our method

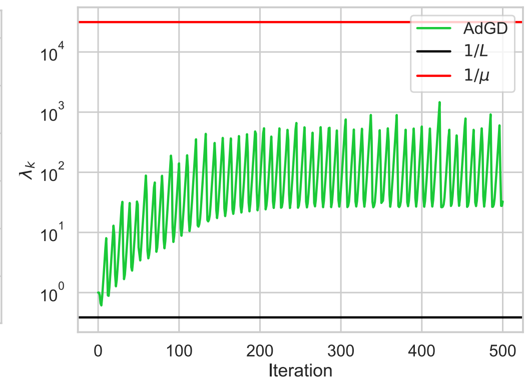
$$L_k = \frac{\|\nabla f(x^k) - \nabla f(x^{k-1})\|}{\|x^k - x^{k-1}\|}$$

$$\lambda_k = \min\left\{\sqrt{1 + \frac{\lambda_{k-1}}{\lambda_{k-2}}}\lambda_{k-1}, \frac{1}{2L_k}\right\}$$

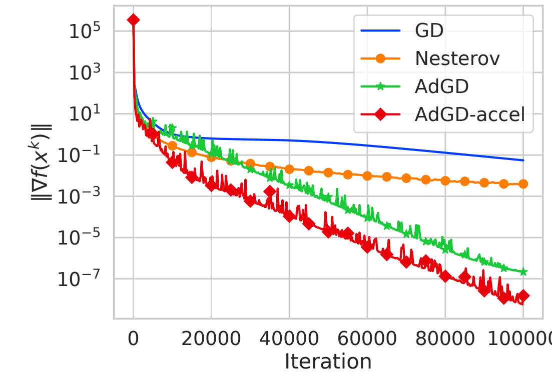
$$\frac{1}{2L} \le \lambda_k \le \frac{1}{2\mu}$$

Logistic regression





Matrix factorization



Convergence

$$f(\hat{x}^k) - f(x^*) = \mathcal{O}\left(\frac{1}{\sum_{t=1}^k \lambda_t}\right) = \mathcal{O}\left(\frac{1}{k}\right)$$

SGD updates on ResNet-18

$$L_k = \frac{\|\nabla f_i(x^k) - \nabla f_i(x^{k-1})\|}{\|x^k - x^{k-1}\|} \qquad \lambda_k = \min\left\{\sqrt{1 + 0.02\theta_{k-1}}\lambda_{k-1}, \frac{1}{L_k}\right\}$$

$$x^{k+1} = x^k - \lambda_k \nabla f_i(x^k)$$