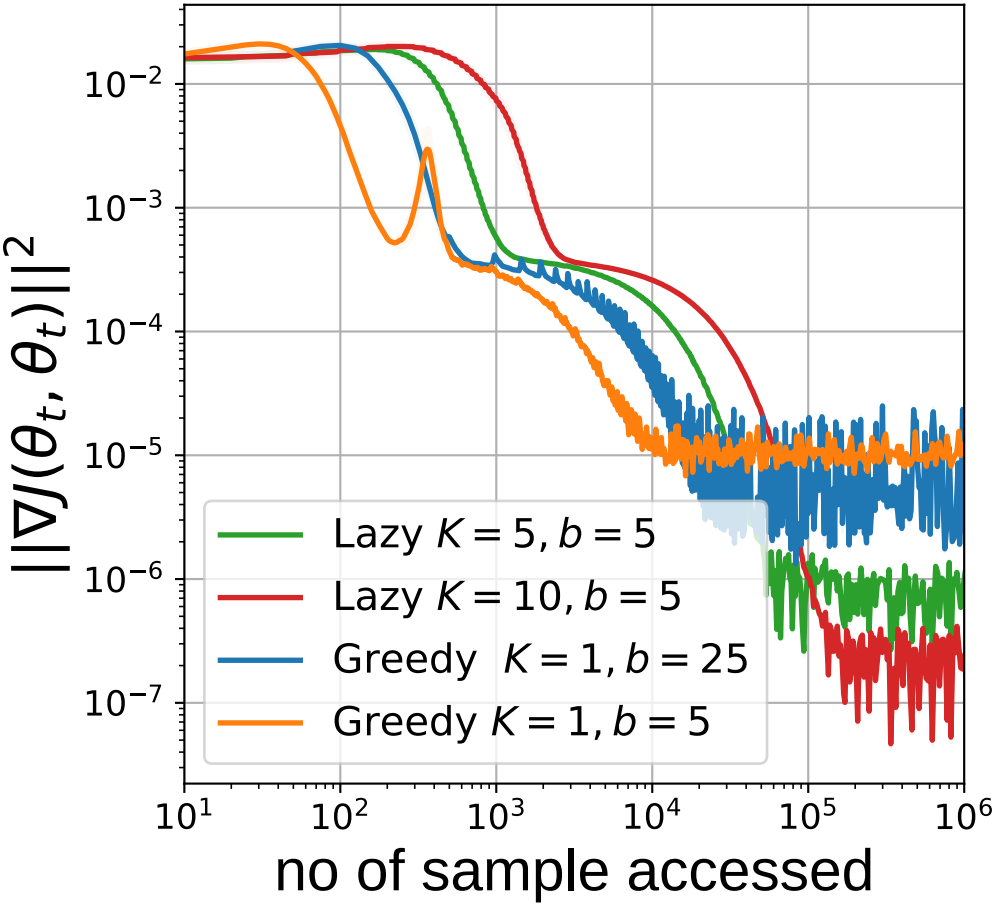
$||\nabla J(\theta_t, \theta_t)||^2$
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^1
10^2
10^3
10^4
10^5
10^6
no of sample accessed
Lazy $K = 5, b = 5$
Lazy $K = 10, b = 5$
Greedy $K = 1, b = 25$
Greedy $K = 1, b = 5$