Convex Optimization Homework 12

Yongpeng Jiang
ID 2019010465

December 6, 2022

1 Proximal Gradient Descent

(a) The derivative of f can be derived as

$$\nabla_x f = \nabla_x \left(\frac{1}{2} (Ax - b)^T (Ax - b) \right) \tag{1}$$
$$= A^T (Ax - b) \,. \tag{2}$$

Assume $A \in \mathbb{R}^{m \times n}, x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, and the singular values for $A^T A$ are $\sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_n$. Then for $\forall x \neq y$, we have

$$\|\nabla f(x) - \nabla f(y)\|_2 = \|A^T A (x - y)\|_2 \tag{3}$$
$$\leq \sigma_k \|x - y\|_2 \,. \tag{4}$$

The equality is achieved when $x - y$ is aligned with the right singular vector corresponding to σ_n. Thus ∇f is proved to be Lipschitz and the smallest positive constant scalar M such that $\|\nabla f(x) - \nabla f(y)\|_2 \leq M \|x - y\|_2, \forall x, y$ equals to the largest singular value of $A^T A$, which is

$$\underline{M} = \sigma_n = \max_v \frac{\|Av\|}{\|v\|} \tag{5}$$

(b) The proximal operator

$$\mathrm{Prox}_{\alpha g} = (I + \alpha \partial g)^{-1} = \begin{cases} x - \alpha, & x > \alpha \\ 0, & \text{otherwise} \\ x + \alpha, & x < -\alpha \end{cases} , \tag{6}$$

then the iteration could be derived as

$$x_{k+1} = \mathrm{Prox}_{\alpha g} \circ (I - \alpha \nabla f) x_k \tag{7}$$
$$= \mathrm{sgn}\left(\tilde{x}_k\right) \cdot \max\left(|\tilde{x}_k| - \alpha, 0\right) \,. \tag{8}$$

where $\tilde{x}_k = x_k - \alpha A^T (A x_k - b)$. We implement the iteration with $\alpha = \frac{1}{M}$.

(c) Solve the problem with A, b in A1.csv and b1.csv, the results are shown as (1).
Solve the problem with A, b in A2.csv and b2.csv, the results are shown as (2).

(d) For the same problem with Proximal Gradient Method, the case with A1.csv converged much faster than the case with A2.csv. Since A1.csv is full column rank, the Lasso converges with

$O(\rho^{-k})$, which resembles $O(-e^{\log k} \cdot \log \rho)$ in logarithm plot. By contrast, the Lasso with non full-rank matrix A2.csv converges with $O(\frac{1}{k})$ and $O(\frac{1}{k^2})$ when accelerated. The convergence speed resembles $O(-\log k)$ in logarithm plot. Exponential speed is obviously faster than linear speed.

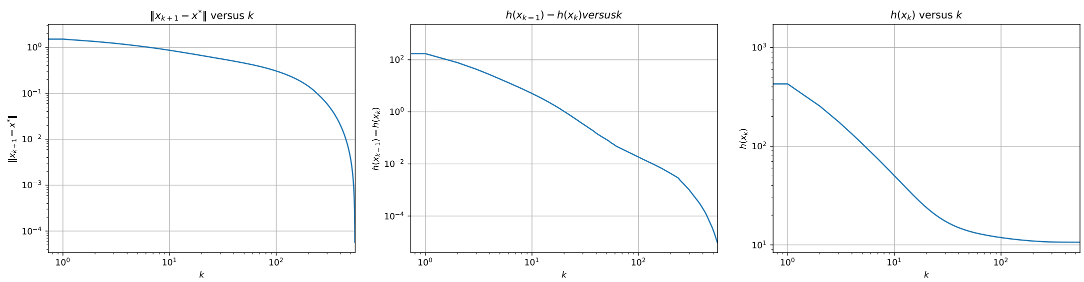

Figure 1: result with A1.csv and b1.csv

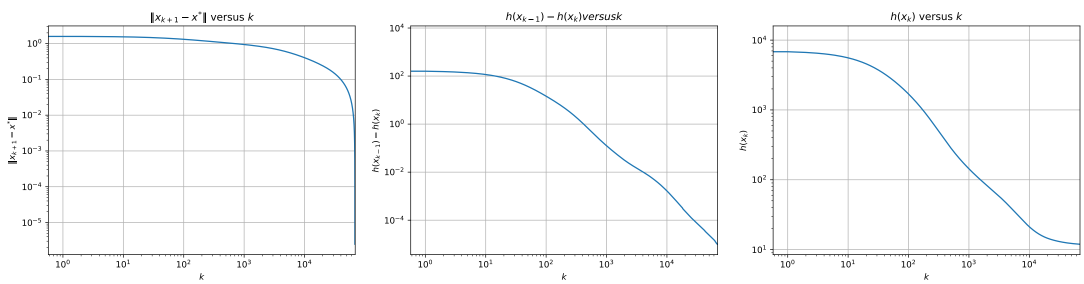

Figure 2: result with A2.csv and b2.csv