Homework 2: Programming Gradient Methods

Due Date: 22 January, 2021

Problem 1

Consider the function $f(x_1, x_2, \ldots, x_m) = \sum_{i=1}^{m} a_i \cdot (x_i - b_i)^2 + 3$

1. Implement gradient descent for this function.
2. Implement gradient descent with backtracking (backtracking parameters $\alpha = 0.5, \beta = 0.5$)
3. Assume that $m = 500$, $a_i = 1, \forall i$, and b_i is chosen randomly (and uniformly) in $[0, 100]$. (a) Run both gradient descent and gradient descent with backtracking with initial point $x_i = 0, \forall i$. (b) Explain your choise of stepsize and stopping condition; (c) compare the convergence speed of the two algorithms.
4. Assume again that $m = 500$, but pick both a_i and b_i uniformly in $[1, 100]$. (a) Run both gradient descent and gradient descent with backtracking with initial point $x_i = 0, \forall i$. (b) Explain your choise of stepsize and stopping condition; (c) compare the convergence speed of the two algorithms; (d) explain the differences in running speed with the previous question (if any).

Problem 2

Consider the function $f(x_1, x_2, \ldots, x_m) = \sum_{i=1}^{m} a_i \cdot (x_i - b_i)^2 + 3$, with constraints $x_i \geq 0, \forall i$, $\sum_i x_i \leq 100$.

1. Write down the KKT conditions for this problem, and derive analytical expressions for the optimal primal and dual variables (or give insights as to how these could be calculated, like the examples we did in class).
2. Solve the problem using dual ascent. Assume again that $m = 500$, and pick both a_i and b_i uniformly in $[1, 100]$. Explain again your step size choice, stopping condition, and convergence rate observed.
3. Could the solution be parallelized? If so, estimate how many iterations a fully parallel algorithm would take (based on the number of iterations of the previous (non-parallelized) question.