MTH702 Optimization

Assignment 2

MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

1. You can discuss the ideas with classmates, however, any write up you made during discussions should be discarded after the meeting. You are not allowed to solve the problems itself, but only discuss the high-level problems and discuss possible directions to solve the problem.
2. You have to acknowledge all resources you have used and all classmates you discussed the problems with.
3. This applies both for theoretical and programming part. Under no circumstances you can share codes! If violation of these rules are discovered, both parties will receive 0 points.

1. Theoretical Assignment [100 pts]

Task 1 (50 pts)**.** *Consider the problem of*

$$\min_{x\in\mathbb{R}^d} \frac{1}{2}x^T Ax + q^T x + r,$$

where A is a symmetric matrix. Show that

- *5pts* *if $A \not\succeq 0$ then the problem is unbounded and the objective function is non-convex;*
- *5pts* *Assume that $A \succeq 0$. Is the objective function convex?*
- *10pts* *If $A \succeq 0$ and following problem $Ax = -q$ has no solution then the problem is unbounded below.*
- *30pts* *Assume that $A \succeq 0$, $\operatorname{rank}(A) < d$ and problem $Ax = -q$ has a solution. Show that the objective function is not strongly convex but gradient descent would still convergence linearly. Derive the convergence rate.*

Task 2 (20pts)**.** *Assume a feasible region $\Delta = \{x \in \mathbb{R}^d : x \geq 0, x^T\mathbf{1} = 1\}$. Show that the function*

$$f(x) = \sum_{i=1}^{d} x_i \log x_i$$

is strongly convex with respect to the norm $\|x\|_1$ on Δ. What is the strong convexity parameter?

Task 3 (30pts)**.** *Recall the analysis of SGD for strongly convex function. We have chosen step-size $\gamma_t = \frac{2}{\mu(t+1)}$ and also we proved the convergence for some specially weighted iterate.*

Assume that we want to derive some convergence guarantee for $\mathbb{E}[f(\frac{1}{T}\sum_{t=???}^{???} x_t) - f^*]$ (fill the ???). Could you define a new step-size that would allow you to derive such a convergence? Hint: think how to get the telescoping sequence and also finish the proof.