SI251 - Convex Optimization homework 3

Deadline: 2024-4-17 23:59:59

1. You can use Word, Latex or handwriting to complete this assignment. If you want to submit a handwritten version, scan it clearly.
2. The **report** has to be submitted as a PDF file to Gradescope, other formats are not accepted.
3. The submitted file name is **student_id+your_student_name.pdf**.
4. Late policy: You have 4 free late days for the quarter and may use up to 2 late days per assignment with no penalty. Once you have exhausted your free late days, we will deduct a late penalty of 25% per additional late day. Note: The timeout period is recorded in days, even if you delay for 1 minute, it will still be counted as a 1 late day.
5. You are required to follow ShanghaiTech's academic honesty policies. You are not allowed to copy materials from other students or from online or published resources. Violating academic honesty can result in serious sanctions.

Any plagiarism will get Zero point.

1. (50 pts) **L-smooth functions**. Suppose the function $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable. Please prove that the following relations holds for all $x, y \in \mathbb{R}$ if f with an L-Lipschitz continuous conditions,

$$[1] \Rightarrow [2] \Rightarrow [3]$$

$$\begin{aligned}
&[1] \quad \langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L\|x - y\|^2, \\
&[2] \quad f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|y - x\|^2, \\
&[3] \quad f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{1}{2L}\|\nabla f(y) - \nabla(x)\|^2, \forall x, y,
\end{aligned}$$

2. (50 pts) **Backtracking line search.** Please show the convergence of backtracking line search on a m-strongly convex and M-smooth objective function f as

$$f\left(x^{(k)}\right) - p^\star \leq c^k \left(f\left(x^{(0)}\right) - p^\star\right)$$

where $c = 1 - \min\{2m\alpha, 2\beta\alpha m/M\} < 1$.

Algorithm 9.2 *Backtracking line search.*

given a descent direction Δx for f at $x \in$ **dom** f, $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$.
$t := 1$.
while $f(x + t\Delta x) > f(x) + \alpha t \nabla f(x)^T \Delta x, \quad t := \beta t$.
