## Graded Homework 1

Rasmus Kyng Graded Homework 1 - Wednesday, March 18

- The solution is due on Sunday, April 5, 2020 by 12:00 noon. Please email your solution, which should be written in $\mathrm{EA}_{\mathrm{E}} \mathrm{X}$, to abdolahad.noori@inf.ethz.ch. We will send out a confirmation that we have received your file. Make sure you receive this confirmation within the day of the due date, otherwise complain timely.
- We would like to stress that the ETH Disciplinary Code applies to this special assignment as it constitutes part of your final grade. The only exception we make to the Code is that we encourage you to verbally discuss the tasks with your colleagues. It is strictly prohibited to share any (hand)written or electronic (partial) solutions with any of your classmates. We are obligated to inform the Rector of any violations of the Code.


## Problem 1: Strongly Convex Functions

Let $f: \mathbb{R}^{n} \rightarrow \mathbb{R}$ be a convex function. Assume $f$ is twice continuously (Frechét) differentiable and that its first and second (Frechét) derivatives are integrable (basically, don't worry that weird stuff is happening with the derivatives).
Assume that for all $\boldsymbol{x}$, we have for some constant $\mu>0$, that $\lambda_{\min }\left(H_{f}(\boldsymbol{x})\right) \geq \mu$. When this holds, we say that $f$ is $\mu$-strongly convex.

Part A. (5 points) Prove that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{n}$

$$
f(\boldsymbol{y}) \geq f(\boldsymbol{x})+\boldsymbol{\nabla} f(\boldsymbol{x})^{\top}(\boldsymbol{y}-\boldsymbol{x})+\frac{\mu}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_{2}^{2}
$$

Part B. (5 points) Prove that there is value $L \in \mathbb{R}$ such that for all $\boldsymbol{x} \in \mathbb{R}^{n}$, we have $f(\boldsymbol{x}) \geq L$. In other words, the function is not unbounded below.

Part C. (5 points) Prove that $f$ is strictly convex as per Definition 2.8 in Lecture 3. Prove also that the minimizer $\boldsymbol{x}^{*} \in \arg \min _{\boldsymbol{x} \in \mathbb{R}^{n}} f(\boldsymbol{x})$ of $f$ is unique.

Part D. ( 20 points) Let $x_{0}$ be a given starting point and $\boldsymbol{x}^{*}$ be the minimizer of $f$. Suppose we have an algorithm FunnyDescent which takes a starting point $\boldsymbol{x}_{0}$, and a step count $t \in \mathbb{N}$. FunnyDescent $\left(\boldsymbol{x}_{0}, t\right)$ runs for $t$ steps and returns $\tilde{\boldsymbol{x}} \in \mathbb{R}^{n}$ such that

$$
f(\tilde{\boldsymbol{x}})-f\left(\boldsymbol{x}^{*}\right) \leq \frac{\gamma\left\|\boldsymbol{x}_{0}-\boldsymbol{x}^{*}\right\|_{2}^{2}}{t+1}
$$

where $\gamma>0$ is a positive number.

Assume that the cost of running FunnyDescent for $t$ steps is $t$. Explain how, with a total cost of at most $\frac{8 \gamma}{\mu} \log \left(\left\|\boldsymbol{x}_{0}-\boldsymbol{x}^{*}\right\|_{2} / \delta\right)$, we can produce a point $\widehat{\boldsymbol{x}} \in \mathbb{R}^{n}$ such that $\left\|\widehat{\boldsymbol{x}}-\boldsymbol{x}^{*}\right\|_{2} \leq \delta$ for $\delta>0$.

Part E. (10 points) Consider a function $h: \mathbb{R}^{n} \rightarrow \mathbb{R}$ which is both $\mu$-strongly convex and $\beta$-gradient Lipschitz. Give an algorithm that returns $\boldsymbol{x}^{\prime}$ with

$$
h\left(\boldsymbol{x}^{\prime}\right)-h\left(\boldsymbol{x}^{*}\right) \leq \epsilon
$$

by computing the gradient of $h$ at at most $\frac{32 \beta}{\mu} \log \left(\beta\left\|\boldsymbol{x}_{0}-\boldsymbol{x}^{*}\right\|_{2} / \epsilon\right)$ points.

Part F. (5 points) Suppose we have an algorithm FunnierDescent $\left(\boldsymbol{x}_{0}, t\right)$ which takes a starting point $\boldsymbol{x}_{0}$, and a step count $t$. Then, it runs for $t$ steps and returns $\tilde{\boldsymbol{x}} \in \mathbb{R}^{n}$ such that

$$
f(\tilde{\boldsymbol{x}})-f\left(\boldsymbol{x}^{*}\right) \leq \frac{\gamma\left\|\boldsymbol{x}_{0}-\boldsymbol{x}^{*}\right\|_{2}^{2}}{(t+1)^{2}}
$$

Explain how, with a total cost of at most $8 \sqrt{\frac{\gamma}{\mu}} \log \left(\gamma\left\|\boldsymbol{x}_{0}-\boldsymbol{x}^{*}\right\|_{2} / \epsilon\right)$, we can produce a point $\boldsymbol{x}^{\prime}$ such that

$$
f\left(\boldsymbol{x}^{\prime}\right)-f\left(\boldsymbol{x}^{*}\right) \leq \epsilon .
$$

## Problem 2: Linear Equation Solvers using Accelerated Gradient Descent

Part A. (15 points) Suppose $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is positive definite with minimum eigenvalue $\lambda_{1}(A)>0$ and maximum eigenvalue $\lambda_{n}(A)$. Consider a function $f: \mathbb{R}^{n} \rightarrow \mathbb{R}$ given by $f(\boldsymbol{x})=\frac{1}{2} \boldsymbol{x}^{\top} \boldsymbol{A} \boldsymbol{x}+\boldsymbol{b}^{\top} \boldsymbol{x}+c$ for some $\boldsymbol{b} \in \mathbb{R}^{n}$ and $c \in \mathbb{R}$. Answer the following questions:

1. Is $f$ convex?
2. Is $f \mu$-strongly convex for some $\mu$ ? If yes, what is the value of $\mu$ ?
3. Is $f \beta$-gradient Lipschitz for some $\beta$ ? If yes, what is the value of $\beta$ ?
4. Is there a unique minimizer of $f$ ? Can you give an explicit equation for it? If yes, state the equation.

Part B. (5 points) Consider $\boldsymbol{A}$ as defined above. Suppose $\boldsymbol{A} \boldsymbol{x}^{*}=\boldsymbol{d}$. Consider the function

$$
h(\boldsymbol{x})=\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{x}^{*}\right)^{\top} \boldsymbol{A}\left(\boldsymbol{x}-\boldsymbol{x}^{*}\right) .
$$

Answer the following questions (as above):

1. Is $h$ convex?
2. Is $h \mu$-strongly convex for some $\mu$ ? If yes, what is the value of $\mu$ ?
3. Is $h \beta$-gradient Lipschitz for some $\beta$ ? If yes, what is the value of $\beta$ ?
4. Is there a unique minimizer of $h$ ? Can you give an explicit equation for it? If yes, state the equation.

Part C. ( 30 points) Consider $h$ as defined above. We want to compute an $\tilde{\boldsymbol{x}}$ s.t.

$$
\begin{equation*}
h(\tilde{\boldsymbol{x}}) \leq \epsilon \cdot \frac{1}{2}\left(\boldsymbol{x}^{*}\right)^{\top} \boldsymbol{A} \boldsymbol{x}^{*} \tag{1}
\end{equation*}
$$

Assume we know $\lambda_{1}$ and $\lambda_{n}$, and assume we know the vector $\boldsymbol{d} \in \mathbb{R}^{n}$ such that $\boldsymbol{A} \boldsymbol{x}^{*}=\boldsymbol{d}$. Suppose we have an algorithm MatApply that given a vector $\boldsymbol{y} \in \mathbb{R}^{n}$ returns $\boldsymbol{z}=\boldsymbol{A} \boldsymbol{y}$. Assume we have no other way to access $\boldsymbol{A}$. Let $\kappa=\frac{\lambda_{n}(\boldsymbol{A})}{\lambda_{1}(\boldsymbol{A})}$. Explain how to obtain an algorithm that produces an $\tilde{\boldsymbol{x}}$ satisfying Equation (11) using at most $O(\sqrt{\kappa} \log (\kappa / \epsilon))$ calls to MatApply.

Hint: Problem 1 may be relevant. There's also a hint in the title of this Problem.

## Problem 3: Laplacian Eigenvalue of a Tree

Part A. (15 points) Let $P_{n}$ be the path from vertex 1 to $n$ and $G_{1, n}$ be the graph with only the edge between vertex 1 and $n$. Furthermore, assume that the edge between vertex $i$ and $i+1$ has positive weight $w_{i}$ for $1 \leq i \leq n-1$. Prove that

$$
G_{1, n} \preceq\left(\sum_{i=1}^{n-1} \frac{1}{w_{i}}\right) \sum_{i=1}^{n-1} w_{i} G_{i, i+1} .
$$

Part B. (35 points) In Lecture 5, we proved that

$$
\lambda_{2}\left(T_{d}\right) \geq \frac{1}{(n-1) \log _{2} n} .
$$

Improve this bound to $\lambda_{2}\left(T_{d}\right) \geq 1 / c n$ for some constant $c>0$.

