DS801 Advanced Optimization for Data Science Assignment 2

Spring 2024

Out: 3rd April 2024

Due: 17th April 2024 at 11:59pm

Instructions

- Submit a PDF document with your solutions through the assignment portal on KLMS by the due date. Please ensure that your name and student ID are on the front page.
- **Late assignments will not be accepted** except in extenuating circumstances. Special consideration should be applied for in this case.
- It is **required** that you **typeset your solutions in LaTeX**. Handwritten solutions will not be accepted.
- Spend some time ensuring your arguments are **coherent** and your solutions **clearly** communicate your ideas.

Question:	1	2	3	4	5	6	Total
Points:	20	10	20	20	10	20	100

1. Recall that for a square matrix $A \in \mathbb{R}^d$, the Rayleigh quotinet is defined as

$$R(A, x) = \frac{x^\top Ax}{x^\top x} \quad \text{for any nonzero } x.$$

   (a) (10 points) Let $\lambda_{\max}$ denote the largest eigenvalue of A, and let $v_{\max}$ denote the associated eigenvector. Prove that $\max_{x \in \mathbb{R}^d : x \neq 0} R(A, x) = \lambda_{\max}$ and $\arg\max_{x \in \mathbb{R}^d : x \neq 0} R(A, x) = v_{\max}$.

   (b) (10 points) Let $\lambda_{\min}$ denote the largest eigenvalue of A, and let $v_{\min}$ denote the associated eigenvector. Prove that $\min_{x \in \mathbb{R}^d : x \neq 0} R(A, x) = \lambda_{\max}$ and $\arg\min_{x \in \mathbb{R}^d : x \neq 0} R(A, x) = v_{\max}$.

2. (10 points) Let $A \in \mathbb{R}^d$ be a symmetric matrix that is not necessarily positive semidefinite. Suppose that the eigenvalues $\lambda_1, \ldots, \lambda_d$ of A satisfy

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_d|.$$

   Let v_i denote the eigenvector of A associated with λ_i for $i \in [d]$. Moreover, let x_t denote the iterate generated by the power methods after t iterations where the initial point is x_0. Prove that the power method guarantees that for any $\epsilon > 0$, if

$$t \geq \frac{|\lambda_1|}{2(|\lambda_1| - |\lambda_2|)} \log\left(\frac{1}{\epsilon(v_1^\top x_0)^2}\right),$$

   then

$$(v_1^\top x_t)^2 \geq 1 - \epsilon.$$

3. (20 points) Prove that the unit nuclear norm ball is equivalent to the convex hull of rank 1 matrices, i.e.,

$$\left\{X \in \mathbb{R}^{n \times p} : \ \|X\|_* \leq 1\right\} = \operatorname{conv}\left\{uv^\top : \ \|u\|_2 = \|v\|_2 = 1, \ u \in \mathbb{R}^n, \ v \in \mathbb{R}^p\right\}.$$

4. (20 points) Let $A \in \mathbb{R}^{n \times p}$ be an $n \times p$ matrix. Let u and v be the top left and right singular vectors of $-A$, respectively. Prove that

$$k \cdot uv^\top \in \operatorname{argmin}\left\{A^\top X : \ \|X\|_* \leq k\right\}.$$

5. (10 points) We discussed an algorithm for computing an (ϵ, δ)-SOSP in Lecture 12. Recall that there is a step checking if $\nabla^2 f(x) \succeq -\delta I$ and that if it does not hold, we compute a unit vector v such that $v^\top \nabla^2 f(x) v < -\delta$. Provide an efficient algorithm computing such a vector **without singular value decomposition**.

6. (20 points) Suppose that a function $f : \mathbb{R}^d \to \mathbb{R}$ is β-smooth in the ℓ_2-norm. Then we apply adaptive gradient descent with an initial point x_1 with step size

$$\eta_t = C\left(\sum_{s=1}^{t} \|\nabla f(x_s)\|_2^2\right)^{-1/2}$$

   for $t \geq 1$ where $C \leq \|\nabla f(x_1)\|_2 / \beta$. Let $x_2, \ldots, x_T$ denote the iterates generated by the adaptive gradient descent after $T - 1$ iterations. Prove that

$$\min\left\{\|\nabla f(x_s)\|_2^2 : \ 1 \leq s \leq t\right\} \leq \frac{4(f(x_1) - f(x^*))^2}{C^2 T}$$

   where $x^* \in \arg\min_{x \in \mathbb{R}^d} f(x)$.