Optimization and Computational Linear Algebra for Data Science
Final exam - December 17, 2019

- **Please justify your answers, proving the statements you make. You are allowed to refer to results shown in lectures/recitations/homeworks as long as you state them precisely, meaning that you should say exactly which hypothesis are needed in the result you use.**
- **This exam is open book/notes. You are allowed to consult notes and books you bring, but not allowed to use electronic devices.**
- **The exam has 2 pages. It has 6 question groups that together total 100 points plus extra credit. Extra credit points will be added to you grade (but your grade can not exceed 100).**

Problem 1 (16 points)**. True or false? [WITHOUT PROOF]** Let $A \in \mathbb{R}^{n\times m}$ and $y \in \mathbb{R}^n$. We assume that the equation $Ax = y$ admits one unique solution $x^* \in \mathbb{R}^m$.

Are the following statements true or false? You do not need to justify your answers in this exercise.

(a) $y \in \mathrm{Im}(A)$.

(b) $\mathrm{Ker}(A) = \{0\}$.

(c) $\mathrm{rank}(A) = m$.

(d) $n \geq m$.

Problem 2 (16 points)**. True or false? [WITH PROOF]** For each of the following, give a proof (if you think that the statement is true) or find a counterexample (if you think it is false). To give a counterexample, it suffices to find a value of n (the easiest way would probably be to take $n = 1$) and a function $f : \mathbb{R}^n \to \mathbb{R}$ for which the statement is not true.

(a) If $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function and if $\nabla f(x) = 0$ then x is a local maximum or a local minimum of f.

(b) If $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function and if $f(x) = f(x') = 0$ for some $x, x' \in \mathbb{R}^n$, then

$$f\left(\frac{x + x'}{2}\right) \leq 0.$$

Problem 3 (22 points)**.** Let $A \in \mathbb{R}^{n\times m}$, $x_0 \in \mathbb{R}^m$ and $w \in \mathbb{R}^n$. We define

$$y = Ax_0 + w,$$

i.e. y is a linear combination of the columns of A, plus the vector w that could for instance represent some noise. We assume that $n \geq m$ and $\mathrm{rank}(A) = m$, so that $A^\mathsf{T} A$ is invertible. We consider the least-squares problem

$$\text{minimize} \quad \|Ax - y\|^2 \quad \text{with respect to} \quad x \in \mathbb{R}^m. \tag{1}$$

(a) Show that the problem (1) has a unique solution $x^* = x_0 + (A^\mathsf{T} A)^{-1} A^\mathsf{T} w$.

(b) Let $A = U \Sigma V^\mathsf{T}$ be the singular value decomposition of A: $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{m\times m}$ are orthogonal matrices and

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_m \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{n\times m},$$

where $\sigma_1 \geq \cdots \geq \sigma_m$ are the singular values of A. Justify that $\sigma_m > 0$ and give the expression of the singular values of $(A^\mathsf{T} A)^{-1} A^\mathsf{T}$ in terms of the singular values of A.

(c) Show, using the previous questions, that $\|x^* - x_0\| \leq \dfrac{1}{\sigma_m}\|w\|$. We recall that the spectral norm of a matrix is equal to its largest singular value.

Problem 4 (16 points)**.** We admit that the constrained optimization problem

$$\text{minimize} \quad x^2+y^2+z^2 \quad \text{subject to} \quad x+y+z=1, \tag{2}$$

has a unique solution (x^*, y^*, z^*). Compute the values of x^*, y^* and z^*.

Problem 5 (16 points)**.** Assume that we have a dataset of n points $a_1, \dots, a_n$ in $\mathbb{R}^d$. We let A be the $n \times d$ matrix

$$A = \begin{pmatrix} - & a_1 & - \\ & \vdots & \\ - & a_n & - \end{pmatrix}.$$

We assume that the points have been centered ($\sum_{i=1}^n a_i = 0$) and we use Principal Component Analysis to obtain a dimensionally reduced dataset $b_1, \dots, b_n \in \mathbb{R}^k$ for some $k \le d$. Let $v_1, \dots, v_d \in \mathbb{R}^d$ be the right singular vectors of A.

Recall (without proof) the expression of b_i in terms of a_i and the k first right singular vectors $v_1, \dots, v_k$ of A. Then, prove that for all $i \in \{1, \dots, n\}$,

$$\|b_i\| \le \|a_i\|.$$

Problem 6 (8 points for **(a)** + 6 points for **(b)** + 5 extra-credit points for **(c)**.)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x) = x_1^2 + 4x_2^2 - 4x_1 - 8x_2 + 1, \qquad \text{for} \;\; x = (x_1, x_2) \in \mathbb{R}^2.$$

(**a**) Show that f is convex. Show that f admits a unique global minimizer x^* and give the coordinates of x^*.

(**b**) We would like to minimize f using gradient descent with constant step-size $\alpha > 0$:

$$\begin{aligned} &x(0) = (0,0) \\ &x(t+1) = x(t) - \alpha \nabla f(x(t)), \qquad \text{for all } t \ge 0. \end{aligned}$$

Let $w(t) = x(t) - x^*$ and let $w_1(t)$, $w_2(t)$ be the coordinates of $w(t)$. Show that for all $t \ge 0$:

$$\begin{cases} w_1(t+1) = (1-2\alpha) w_1(t) \\ w_2(t+1) = (1-8\alpha) w_2(t). \end{cases}$$

(**c**) **Extra-credit, 5 points**. Deduce that:

- if $0 < \alpha < 1/4$, gradient descent converge to x^*, that is, $w(t)$ goes to 0 as t goes to infinity.
- if $\alpha \ge 1/4$, gradient descent does not converge to x^*.

Problem 7 (**Extra-credit**, 10 points)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function, A be a $n \times m$ matrix and $\lambda > 0$. We assume that $x^* \in \mathbb{R}^m$ is a global minimizer of

$$f(Ax) + \lambda \|x\|^2.$$

Show that x^* belongs to the subspace S of $\mathbb{R}^m$ spanned by the rows of A, that is $S = \mathrm{Im}(A^\mathsf{T})$.
Hint: consider P_S, the orthogonal projection onto S.

Problem 8 (**Extra-credit**, 10 points)**.** We consider an unoriented graph of n nodes $1, 2, \dots, n$. Let $d \ge 0$ such that each node has a degree less or equal to d. Let $A \in \mathbb{R}^{n \times n}$ be the adjacency matrix of the graph. Show that if λ is an eigenvalue of A, then $|\lambda| \le d$.
Hint: consider an eigenvector $x \in \mathbb{R}^n$ of A associated to λ, and look at the coordinate i such that $|x_i|$ is maximal.