CSCI 1520: Algorithmic Aspects of Machine Learning (Spring 2025)

Written Assignment 3

Due at 11:59pm ET, Thursday, April 17

1. (4 points) Compute the derivative $\frac{\partial f}{\partial x}$ of the following vector-to-scalar functions.
   All other variables are constants (i.e., not a function of x).

   (1) $f(x) = \langle c, x \rangle = c^\top x$. $\quad$ ($c \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$)

   (2) $f(x) = x^\top M x$. $\quad$ ($x \in \mathbb{R}^n$ and $M \in \mathbb{R}^{n \times n}$)

   (3) $f(x) = \|Ax - b\|_2^2$. $\quad$ ($A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$)

   (4) $f(x) = \left((a^\top x)^2 - z\right)^2$. $\quad$ ($a \in \mathbb{R}^n$, $x \in \mathbb{R}^n$, and $z \in \mathbb{R}$)

2. (4 points) Compute the derivative $\frac{\partial f}{\partial X}$ of the following matrix-to-scalar functions.
   All other variables are constants (i.e., not a function of X).

   (1) $f(X) = \langle C, X \rangle = \mathrm{tr}(C^\top X)$. $\quad$ ($C \in \mathbb{R}^{m \times n}$ and $X \in \mathbb{R}^{m \times n}$)

   (2) $f(X) = \mathrm{tr}(AXB)$. $\quad$ ($A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{n \times p}$, and $B \in \mathbb{R}^{p \times m}$)

   (3) $f(X) = \|Y - DX\|_F^2$. $\quad$ ($Y \in \mathbb{R}^{m \times p}$, $D \in \mathbb{R}^{m \times n}$, and $X \in \mathbb{R}^{n \times p}$)

   (4) $f(X) = \left\|XX^\top - M\right\|_F^2$. $\quad$ ($X \in \mathbb{R}^{n \times r}$ and $M \in \mathbb{R}^{n \times n}$)

3. (6 points) Consider the problem of finding the best rank-one approximation of a matrix M. We focus on the case where $M \in \mathbb{R}^{n \times n}$ is positive semi-definite (PSD). That is, $M = M^\top$ and $x^\top M x \geq 0$ for all $x \in \mathbb{R}^n$.

   Consider the (non-convex) loss function $f(x) = \left\|M - xx^\top\right\|_F^2$ where $x \in \mathbb{R}^n$.

   (1) Let $\nabla f(x) \in \mathbb{R}^n$ and $\nabla^2 f(x) \in \mathbb{R}^{n \times n}$ be the gradient and Hessian of f at x, respectively. Show that

   $$\nabla f(x) = 4(xx^\top - M)x \quad \text{and} \quad \nabla^2 f(x) = 4((x^\top x)I + 2xx^\top - M)\ .$$

   (2) Suppose M has a unique largest eigenvalue λ with corresponding (unit) eigenvector z. Prove that the only second-order stationary points of f are $\pm\sqrt{\lambda}z$.
   (Hint: If x is a second-order stationary point of f, then $\nabla f(x) = 0$ and $z^\top \nabla^2 f(x) z \geq 0$. The following fact may be helpful: For any symmetric matrix A, if $Av_1 = \lambda_1 v_1$ and $Av_2 = \lambda_2 v_2$ with $\lambda_1 \neq \lambda_2$, then $v_1^\top v_2 = 0$.)

4. (1 bonus point) Consider the problem of matrix sensing, where the goal is to recover a hidden rank-r matrix M from linear measurements. We focus on the case where the hidden matrix $M \in \mathbb{R}^{n\times n}$ is symmetric. The input includes the rank $r > 0$, a list of sensing matrices $A_1, \dots, A_m \in \mathbb{R}^{n\times n}$, and the corresponding linear measurements $b_i = \langle A_i, M\rangle$.

   Consider the loss function $f(X) = \sum_{i=1}^m \left(\langle A_i, XX^\top\rangle - b_i\right)^2$ where $X \in \mathbb{R}^{n\times r}$. We will prove all local optima of f are globally optimal.

   Fix any $U \in \mathbb{R}^{n\times r}$ such that $UU^\top = M$. Suppose X is a second-order stationary point of f. Let $R \in \mathbb{R}^{r\times r}$ be an orthogonal matrix that minimizes $\|X - UR\|_F$. Let $\Delta = X - UR$.

   You can use the following facts without proving them:

   - The first and second-order optimality condition of X implies that $\langle \nabla f(X), \Delta\rangle = 0$ and $\Delta : \nabla^2 f(X) : \Delta \geq 0$.
   - $\left\|\Delta\Delta^\top\right\|_F^2 \leq 2\left\|XX^\top - M\right\|_F^2$.

   (1) Show that $\langle \nabla f(X), \Delta\rangle = 0$ is equivalent to

   $$\sum_{i=1}^m \left[\left(\langle A_i, XX^\top\rangle - b_i\right)\langle A_i, X\Delta^\top + \Delta X^\top\rangle\right] = 0.$$

   (2) Show that $\Delta : \nabla^2 f(X) : \Delta \geq 0$ is equivalent to

   $$\sum_{i=1}^m \left[2\left(\langle A_i, XX^\top\rangle - b_i\right)\langle A_i, \Delta\Delta^\top\rangle + \langle A_i, X\Delta^\top + \Delta X^\top\rangle^2\right] \geq 0.$$

   (Hint: You can use Taylor expansion to derive the first and second-order optimality conditions in the direction Δ.)

   (3) Suppose the sensing matrices satisfy the $(\frac{1}{10}, 2r)$-restricted isometry property (RIP) as defined below. Prove that any second-order stationary point of f recovers M exactly.

   (Hint: You can follow a similar approach to the matrix completion proof we discussed in class.)

   Definition (Matrix RIP)**.** We say a list of matrices $A_1, \dots A_m$ satisfies (δ, r)-RIP if the following condition holds for all matrices M with $\text{rank}(M) \leq r$:

   $$(1-\delta)\|M\|_F^2 \leq \frac{1}{m}\sum_{i=1}^m \langle A_i, M\rangle^2 \leq (1+\delta)\|M\|_F^2 .$$

5. (1 bonus point) Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Let $U \in \mathbb{R}^{n\times r}$ be any matrix such that $UU^\top = M$. Let $X \in \mathbb{R}^{n\times r}$ be any matrix. Let $R \in \mathbb{R}^{r\times r}$ be an orthogonal matrix that minimizes $\|X - UR\|_F$. Let $Z = UR$ and $\Delta = X - Z$.

   We will prove that $\left\|\Delta\Delta^\top\right\|_F^2 \leq 2\left\|XX^\top - M\right\|_F^2$.

   (1) Assume $X^\top Z$ is PSD. Prove that

   $$\left\|(X-Z)(X-Z)^\top\right\|_F^2 \leq 2\left\|XX^\top - ZZ^\top\right\|_F^2 .$$

   (2) Prove that $X^\top Z$ is indeed PSD.