University of Waterloo

CO 367 – Fall 2018 – Practice final

Question 1 Let $A, B \in \mathbb{R}^{n\times n}$ be symmetric, positive semidefinite matrices, and let $C \in \mathbb{R}^{n\times n}$ be any matrix. Prove that $C^T(A+B)C$ is symmetric and positive semidefinite.

Question 2 Consider $f : \mathbb{R}^n \to \mathbb{R}$ where $f \in C^1(D)$ and $D \subseteq \mathbb{R}^n$ is a convex set. Note that f is not necessarily C^2-smooth. Prove that if

$$(\nabla f(x) - \nabla f(y))^T(x-y) \geq 0, \quad \forall x, y \in D,$$

then f is convex over D.

Question 3 Give a counter-example to the following **wrong** claim:
If $f : \mathbb{R}^n \to \mathbb{R}$ is a C^1-smooth function over $\mathbb{R}^n$ and

$$||\nabla f(y) - \nabla f(x)||_2^2 \geq \ell |f(y) - f(x)|, \quad \text{for all } x, y \in \mathbb{R}^n,$$

for some constant $\ell > 0$, then f is strongly convex.

Question 4 Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = x^TQx$ where $Q \in \mathbb{R}^{n\times n}$ and $\frac{1}{2}(Q+Q^T)$ is a symmetric, positive definite matrix. Prove that f is coercive.

Question 5 Let $q : \mathbb{R}^n \to \mathbb{R}$ be defined by $q(x) = x^TAx + b^Tx + c$ where the eigenvalues of A are $\lambda_1 \leq \cdots \leq \lambda_n$. Define two quadratic functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ such that

$$f(||x||_2) \leq q(x) \leq g(||x||_2)$$

for all $x \in \mathbb{R}^n$.

Question 6 Indicate whether each of the following statements is true or false. Write "True" or "False" next to each item. No explanation is necessary.

- If a function $f \in C^2(\mathbb{R}^n)$ is such that $\nabla^2 f(x)$ is symmetric for all $x \in \mathbb{R}^n$, then it is convex.
- Let $f : \mathbb{R}^n \to \mathbb{R}$. If $f(x+p) < f(x)$, then p is a descent direction at x.
- If $f \in C^1(\mathbb{R}^n)$ has a gradient ∇f that is Lipschitz continuous over $\mathbb{R}^n$ and f is unbounded below, then the steepest descent method with an Armijo line search starting from any x^0 is guaranteed to produce a sequence $x^0, x^1, \ldots$ such that $\lim_{k\to\infty} f(x^k) \to -\infty$.
- Let f be a strongly convex function with a unique global minimizer $x^* \in \mathbb{R}^n$ that satisfies the sufficient conditions for local optimality. If the trust region method is initialized with a trust region radius $\delta^0 > 0$ at a point $x^0 \in B_{\delta^0}(x^*)$, then $x^1 = x^*$.
- We consider a neural network whose purpose is to classify input data points in $\mathbb{R}^n$, with $n = 1000$, into $k = 8$ categories. Let $F : \mathbb{R}^n \to \mathbb{R}^k$ be such that $F(x)$ represents the output vector of our neural network for an input data point $x \in \mathbb{R}^n$. Let $\bar{x}$ be an input data point that we (manually) assign to category 5. If $\bar{x}$ is part of the training set and training reaches a global minimizer, then $F(\bar{x}) = e_5$.
- In the stochastic gradient descent method, the step is always a descent direction.

Question 7 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1-smooth function. Consider the sequence of points x^k generated by a steepest descent method, i.e.,

$$x^{k+1} = x^k - \alpha^k \nabla f(x^k).$$

The step length α^k is chosen to satisfy sufficient decrease condition. Assume that, for our particular function f, we have $\alpha^k \geq C$ for some constant $C > 0$ independent of k. Prove that if $\lim_{k\to\infty} ||\nabla f(x^k)||_2 = D > 0$, then $f(x^k)$ goes to minus infinity as k grows.

Question 8 Consider a function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \frac{1}{3}x^3 - 2x$.
(i) Determine all values of x^k such that x^{k+1} is well defined with Newton's method. Among those, find one that is a local minimizer.
(ii) Determine a value $r > 0$ such that if $x^0 \in B_r(x^*)$, then Newton's method converges quadratically to x^*.
(iii) Give a counter-example to the following (wrong) claim: With Newton's method, for all $x^k \in \mathbb{R}$, $||x^{k+1} - x^*||_2 < ||x^k - x^*||_2$.

Question 9 Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix, let $b \in \mathbb{R}^n$, let $D \subseteq \mathbb{R}$, and let $f : D \to \mathbb{R}$ be defined by

$$f(x) = \left|\left| -\frac{1}{2}(A + xI)^{-1}b \right|\right|_2 .$$

(a) Determine the domain $D \subseteq \mathbb{R}$ over which f is defined.
(b) Prove that f is a decreasing function over its domain D.

Question 10 Consider the problem (NLP) defined by

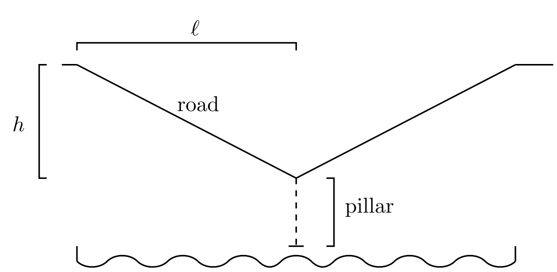
$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h(x) = 0 \end{aligned} \tag{NLP}$$

where $f, g \in C^1(\mathbb{R}^n)$. Let us fix $x^0 \in \mathbb{R}^n$ and let

$$x^1 = \operatorname*{argmin}_{x\in\mathbb{R}^n} \left\{ f(x) - 3h(x) + 5\left(h(x)\right)^2 \right\}.$$

Prove that if $h(x^1) = 0$, then x^1 is a KKT point of (NLP).

Question 11 We want to build a bridge over a river of width 2ℓ, with a pillar in the middle of the river. The bridge is symmetric and drops (linearly) to a minimum height of h meters below the initial level of the road, as depicted here:



which lets us save on the costs of building the pillar. The resulting bridge costs A dollars per meter of paved road, plus the central pillar costing B dollars per meter of height. Given the constants ℓ, A and B, we want to determine the ideal height drop h for the bridge.
(a) Formulate the problem as a conic optimization problem.
(b) Write the dual of the formulation in (a).
(c) Use the dual from (b) to determine the optimal cost savings of the project, compared to a totally flat bridge ($h = 0$).