This assignment is due at the **start** of your lecture on Friday, 20 March 2020.

1. [10 marks: 5 marks for the "if part" and 5 marks for the "only if part"]

   In Question 2 on last year's midterm test[1], the students were asked to prove a result that is equivalent to the following theorem.

   Theorem 1 *Assume that $f : \mathbb{D} \to \mathbb{R}$ is continuously differentiable for all $x \in \mathbb{D}$, where the domain $\mathbb{D}$ of f is an open, convex subset of $\mathbb{R}^n$.*

   Show that f is convex on $\mathbb{D}$ if and only if

   $$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

   for all x and $y \in \mathbb{D}$.

   You might have proven Theorem 1 as practice for this year's midterm test. If not, you can find it in many places on web.

   For this assignment, prove the following result.

   Theorem 2 *Assume that $f : \mathbb{D} \to \mathbb{R}$ is continuously differentiable for all $x \in \mathbb{D}$, where the domain $\mathbb{D}$ of f is an open, convex subset of $\mathbb{R}^n$.*

   Show that f is strictly convex on $\mathbb{D}$ if and only if

   $$f(y) > f(x) + \nabla f(x)^T (y - x)$$

   for all x and $y \in \mathbb{D}$ for which $x \neq y$.

   You can use Theorem 1 without proof in your proof of Theorem 2.

   Note that you were told in both Theorems 1 and 2 to assume that $\nabla f(x)$ exists and is continuous for all $x \in \mathbb{D}$. If you cannot prove either the "if part" or the "only if part" of Theorem 2 with this assumption alone, then you might choose to assume in addition that $\nabla^2 f(x)$ exists and is continuous for all $x \in \mathbb{D}$. However, if you make this additional assumption in either part, then there will be a two mark deduction in each part that you use this additional assumption. In particular, if you use this additional assumption in both parts, then there will be a total of a four mark deduction for this question.

   In proving Theorem 1, you might use the following result: if $a_k \geq a^*$ for all $k = 1, 2, 3, \ldots$ and $\lim_{k\to\infty} a_k$ exists, then $\lim_{k\to\infty} a_k \geq a^*$. In proving Theorem 2, it is tempting to assume that if $a_k > a^*$ for all $k = 1, 2, 3, \ldots$ and $\lim_{k\to\infty} a_k$ exists, then $\lim_{k\to\infty} a_k > a^*$. However, this is not always true. For example, if $a_k = 2^{-k}$ and $a^* = 0$, then $a_k > a^*$ for all $k = 1, 2, 3, \ldots$ and $\lim_{k\to\infty} a_k$ exists, but $\lim_{k\to\infty} a_k = a^* = 0$. So, it is <u>not</u> true that $\lim_{k\to\infty} a_k > a^*$ in this case.

[1] You can find last year's midterm test at `http://www.cs.toronto.edu/~krj/courses/466-2305/midterm.2018.pdf`.

2. [5 marks]

Question 3 on this year's midterm test[2] notes that when we were discussing in class the BFGS update

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$$

where

$$\begin{aligned} s_k &= x_{k+1} - x_k \\ y_k &= \nabla f(x_{k+1}) - \nabla f(x_k) \end{aligned}$$

I mentioned that, if B_0 is symmetric positive-definite and $y_k^T s_k > 0$ for all $k = 0, 1, 2, \ldots$, then B_k is symmetric positive-definite for all $k = 0, 1, 2, \ldots$

Prove that this claim is true. That is, prove that, if B_0 is symmetric positive-definite and $y_k^T s_k > 0$ for all $k = 0, 1, 2, \ldots$, then B_k is symmetric positive-definite for all $k = 0, 1, 2, \ldots$

3. [5 marks]

Question 3 on this year's midterm test goes on to give a condition that ensures that $y_k^T s_k > 0$.

Another set of conditions that ensures that $y_k^T s_k > 0$ is the following.

(a) $\nabla f(x_k) \neq 0$ (i.e., you are not already at a critical point), and

(b) $\nabla^2 f(x)$ exists, is continuous and is symmetric positive-definite for all $x \in \mathbb{D}$, where $\mathbb{D}$ is an open, convex set containing both x_k and x_{k+1}.

Show that, if

- (a) and (b) above are true,
- B_k is symmetric positive-definite,
- $x_{k+1} = x_k + p_k$, where $p_k = -B_k^{-1} \nabla f(x_k)$,

then $y_k^T s_k > 0$, where y_k and s_k are given in Question 2 above.

In both Questions 2 and 3 above, assume that $f : \mathbb{R}^n \to \mathbb{R}$. So, x_k, x_{k+1}, s_k, $\nabla f(x_k)$, $\nabla f(x_{k+1})$ and p_k are all vectors in $\mathbb{R}^n$ and B_0, B_k and B_{k+1} are real $n \times n$ matrices.

[2] You can find this year's midterm test at `http://www.cs.toronto.edu/~krj/courses/466-2305/midterm.2020.pdf`.