This is a **closed-book test**: no books, no notes, no computers (of any kind), no tablets, no calculators, no phones, etc. allowed. The only aid permitted is an 8.5 by 11 inch **aid sheet**. You may write (as small as you like) on both sides of the aid sheet, but you cannot have any "attachments" to the aid sheet.

Do **NOT** turn this page over until you are **TOLD** to start.

Duration of the test: 50 minutes (1:10 to 2:00 PM).

Write your answers to **ALL** questions in the test booklets provided.

Please fill-in **ALL** the information requested on the front cover of **EACH** test booklet that you use.

The test consists of 3 pages, including this one. Make sure you have all 3 pages.

The test consists of 3 questions. **Answer ALL questions**.

The test was written with the intention that you would have ample time to complete it. You will be rewarded for concise well-thought-out answers, rather than long rambling ones. **We seek quality rather than quantity**.

Moreover, an answer that contains relevant and correct information as well as irrelevant or incorrect information will be awarded fewer marks than one that contains the same relevant and correct information only.

Write legibly. Unreadable answers are worthless.

1. [10 marks: 5 marks for each part]

Consider the set

$$\mathbb{D} = \{x \in \mathbb{R}^n : 0 < x_i < 1 \text{ for } i = 1, 2, \ldots, n\} \tag{1}$$

and the function $f : \mathbb{D} \to \mathbb{R}$ defined by

$$f(x) = \sum_{i=1}^{n} x_i \log(x_i) \tag{2}$$

where log is the natural logarithm (i.e., the logarithm to the base e).

(a) Show that $\mathbb{D}$ given in (1) is an open, convex subset of $\mathbb{R}^n$.

(b) Show that the function $f(x)$ given in (2) is strictly convex.

In answering part (b), you can assume that $\mathbb{D}$ given in (1) is an open, convex subset of $\mathbb{R}^n$ even if you didn't prove it in part (a).

Hint: in answering this question, you can use without proof any of the results that we proved in class or you were asked to prove on the course assignments.

2. [5 marks part]

Consider the quadratic function

$$f(x) = \frac{1}{2} x^T A x - x^T b + c \tag{3}$$

where $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$ is a symmetric positive-definite matrix, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Suppose that we apply Newton's method to $f(x)$ starting from any initial guess $x_0 \in \mathbb{R}^n$. That is, we compute the search direction

$$p_0^N = -\left(\nabla^2 f(x_0)\right)^{-1} \nabla f(x_0)$$

and set

$$x_1 = x_0 + p_0^N$$

Show that $x_1 = x^*$, where x^* is the unique minimizer of $f(x)$.

That is, Newton's method converges to the unique minimizer of $f(x)$ in one step.

3. [5 marks]

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 $k = 0, 1, \ldots,$ then B_k is symmetric positive-definite for all $k = 0, 1, \ldots.$

There are several sets of conditions that ensure that $y_k^T s_k > 0$ for all $k = 0, 1, \ldots.$ One such set of condition is the following.

Assume that $\nabla f(x)$ exists at x_k and x_{k+1}. Moreover, assume p_k is a *descent direction* (i.e., $p_k^T \nabla f(x_k) < 0$), $x_{k+1} = x_k + \alpha_k p_k$ for $\alpha_k > 0$, and the step from x_k to x_{k+1} satisfies the *strong Wolfe conditions*

$$\begin{aligned} f(x_k + \alpha_k p_k) &\leq f(x_k) + c_1 \alpha_k p_k^T \nabla f(x_k) \\ |p_k^T \nabla f(x_k + \alpha_k p_k)| &\leq c_2 \, |p_k^T \nabla f(x_k)| \end{aligned}$$

for real constants c_1 and c_2 satisfying $0 < c_1 < c_2 < 1$.

Show that, under the assumptions in the paragraph above, $y_k^T s_k > 0$.