**CS726, Fall 2016**
**Midterm Examination**
Thursday, October 27, 2016, 7:15pm–9:15pm.

Answer ALL questions below. One handwritten sheet of notes (written front and back) is allowed. EXPLAIN ALL YOUR ANSWERS.

1. Consider the unconstrained optimization problem $\min_{x \in \mathbb{R}^n} f(x)$, for smooth convex $f$. Suppose that the problem has a solution $x^*$, with optimal value $f^* = f(x^*)$. Suppose that we have an algorithm that yields the following guaranteed improvement at each iteration:
$$f(x^{k+1}) - f^* \le \left(1 - \frac{\hat{c}}{k+1}\right)(f(x^k) - f^*), \quad k = 0, 1, 2, \dots,$$
for some constant $\hat{c} \in (0, 1)$. Find, approximately, the number of iterations $K$ required to satisfy $f(x^K) - f^* \le \epsilon$, for a given $\epsilon > 0$. (Hint: You can use the fact that $\sum_{i=1}^{K}(1/i) \approx \log K$, for large integers $K$.)
2. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth convex function, and consider the optimization problem $\min_{x \in \mathbb{R}^n} f(x)$. Suppose that an algorithm generates a sequence of iterates $\{x^k\}$ such that $f(x^{k+1}) < f(x^k)$ for all $k$, and $\nabla f(x^k) \to 0$.
   (a) Can we say that "the sequence $\{x^k\}$ converges to a solution of the optimization problem"? Prove this claim if it is true, and give a counterexample if it is false.
   (b) If $f$ is *strongly convex*, we know that the optimization problem has a unique solution $x^*$, and that the level set $\mathcal{L} := \{x \mid f(x) \le f(x^0)\}$ is compact. With the additional assumption that $f$ is strongly convex, can we say that $\lim_{k \to \infty} x^k = x^*$? Explain.
3. Consider the unconstrained optimization problem $\min_{x \in \mathbb{R}^n} f(x)$, for smooth $f$. Suppose we implement a line search method in which at iterate $x^k$, the search direction $p^k$ is defined to have a single nonzero element, that is the *negative of the element of largest absolute value in* $\nabla f(x^k)$. That is, we choose $i_k \in \{1, 2, \dots, n\}$ such that
$$|[\nabla f(x^k)]_{i_k}| \ge |[\nabla f(x^k)]_i|, \quad \text{for all } i = 1, 2, \dots, n,$$
and set
$$p^k = -[\nabla f(x^k)]_{i_k} e_{i_k},$$
where $e_{i_k}$ is the vector which has all zero components except for a 1 in position $i_k$. Find a constant $\bar{\epsilon} > 0$ such that
$$\nabla f(x^k)^T p^k \le -\bar{\epsilon} \|p^k\| \|\nabla f(x^k)\|.$$
4. Suppose that for some algorithm for minimizing a smooth convex function $f$, whose minimizer is $x^*$, there is a constant $\gamma > 0$ such that
$$f(x^{k+1}) \le f(x^k) - \gamma \|\nabla f(x^k)\|^2.$$
Suppose too that there is a constant $c > 0$ such that
$$\|\nabla f(x)\| \ge c[f(x) - f(x^*)]^{1/2},$$

where $x^*$ is the solution of the problem. Find a value of the constant $r \in (0, 1)$ such that

$$f(x^{k+1}) - f(x^*) \leq r \left( f(x^k) - f(x^*) \right).$$

5. Suppose that $f$ is a quadratic function of the form

$$f(x) = \frac{1}{2} x^T A x - b^T x,$$

where $A$ is symmetric positive definite. Suppose that for some starting point $x^0$, we initialize the conjugate gradient method by setting

$$r^0 = Ax^0 - b, \quad p^0 = -r^0.$$

Suppose that the conjugate gradient method subsequently generates search directions $p^1, p^2, \ldots$.

(a) The search directions $p^0, p^1, p^2$ are known to be *conjugate*. Write down the conjugacy condition. (You just need to state the definition; you don't need to prove that this property holds.)

(b) Consider the problem of minimizing $f$ over the subspace spanned by the first $k$ search directions, where $k < n$. To do this, we define a function $h(\sigma)$ for $\sigma = (\sigma_0, \sigma_1, \ldots, \sigma_{k-1})^T \in \mathbb{R}^k$ by

$$h(\sigma) = f(x^0 + \sigma_0 p^0 + \sigma_1 p^2 + \ldots + \sigma_{k-1} p^{k-1}).$$

Show that $h : \mathbb{R}^k \to \mathbb{R}$ is a quadratic function, of the form

$$h(\sigma) = \frac{1}{2} \sigma^T B \sigma - c^T \sigma + \beta,$$

and write down the matrix $B$, the vector $c$, and the scalar $\beta$.

(c) Assuming that $p^i \neq 0$ for all $i = 0, 1, \ldots, k-1$, write down the optimal values of $\sigma_i$, $i = 0, 1, \ldots, k-1$, that is the values that minimize $h$.