Name: **Roll No:**

NATIONAL INSTITUTE OF TECHNOLOGY CALICUT

Department of Computer Science and Engineering

B.Tech Mid Semester Examination, Winter 2024-2025

CS4036E ALGORITHMS IN OPTIMIZATION (Elective)

Time: 120 Minutes **Maximum Marks: 36**

---

1. [CO1] Let $f : \mathbf{R}^n \mapsto \mathbf{R}$ satisfy $f(\alpha x + \beta y) > \alpha f(x) + \beta f(y)$ for all $x, y \in \mathbf{R}^n$ and $\alpha > 0, \beta > 0$ satisfying $\alpha + \beta = 1$. 9

   a. Suppose that $x_0 \in \mathbf{R}^n$ and $\delta > 0$ satisfy $f(x_0) > f(x_0 + d)$ whenever $||d|| < \delta$. Can we conclude either that $f(x_0) > f(y)$ or $f(x_0) < f(y)$ for all $y \in \mathbf{R}^n$? Either prove the appropriate statement or provide a counter example.
   Soln: $f(x_0) > f(y)$ must hold for all y. To prove this, assume for the purpose of contradiction that $f(y) < f(x_0)$. The given condition implies that:

$$\forall \epsilon > 0, f(x_0 + \epsilon(y - x_0)) > f(x_0) + \epsilon(f(y) - f(x_0)). \tag{1}$$

   If $f(y) > f(x_0)$, then $\epsilon(f(y) - f(x_0)) > 0$ for all $\epsilon > 0$ which would mean that $f(x_0) < f(x_0 + \epsilon(y - x_0)$. When ϵ is small enough to have $||\epsilon(y - x_0))|| < \delta$, this would contradict the given condition that $f(x_0) > f(x_0 + d)$ whenever $||d|| < \delta$.

   b. Suppose further that f is twice differentiable. Is it possible to conclude for all $x, y \in \mathbf{R}^n$ either that $f(y) - f(x) > \nabla f(x)(y - x)$ or $f(y) - f(x) < \nabla f(x)(y - x)$? Either prove the (appropriate) statement or provide a counterexample.
   Soln: Rearranging Equation 1 and taking limits, we see that the directional derviative of f along $y - x_0$ given by:

$$\nabla f(x_0)(y - x_0) = \lim_{\epsilon \to 0} \frac{f(x_0 + \epsilon(y - x_0)) - f(x_0)}{\epsilon} > (f(y) - f(x_0)). \tag{2}$$

   c. Suppose $\nabla f(x_0) = 0$. Is it possible to conclude that x_0 is a local/global maximum or local/global minimum of f? Either prove the (appropriate) statement or provide a counterexample.
   Soln: From Equation 2 we see that if $\nabla f(x_0) = 0$, then for all y, $f(x_0) > f(y)$. Hence x_0 must be a local as well as a global maximum.

2. [CO1] Suppose f, g be real valued functions on $\mathbf{R}$. Let $h(x) = f(g(x))$. Suppose that g is concave and f is convex, monotone decreasing. Is it possible to conclude that h is convex or concave? Either prove your claim or show a counterexample. 3

   Soln: Let $x, y \in \mathbf{R}$ and α, β be positive real number satisfying $\alpha + \beta = 1$. Let $z = \alpha x + \beta y$. By concavity of g, we have:

$$g(z) \geq \alpha g(x) + \beta g(y) \tag{3}$$

   Hence we have:

$$\begin{aligned} h(z) = f(g(z)) &\leq f(\alpha g(x) + \beta g(y)) \\ &\leq \alpha f(g(x)) + \beta f(g(y)) = \alpha h(x) + \beta h(x) \end{aligned}$$

   Here, the first equation was because f was monotone decreasing and the second due to its convexity.

3. [CO2] Consider the set of equations: $3x + 2y = 7$, $5x + 9y = 8$, $8x + 3y = 21$ and $4x + 7y = 10$. Clearly, with two variables, we cannot satisfy for equations. Hence, we try to find values of x, y to minimize the absolute error given by: $|3x + 2y - 7| + |5x + 9y - 8| + |8x + 3y - 21| + |4x + 7y - 10|$. 9

   a Is this minimization problem a convex optimization problem? Justify your answer.
   Soln: $f(x, y) = |ax + by + c|$ for any real values of a, b, c is convex because norm composition of a linear function with the convex function $||$ is convex. Finally, if f, g are convex, then $af + bg$ is convex if $a, b > 0$. Hence, the sum of convex functions is convex.

   b Suppose instead of minimizing the absolute error, we decide to minimize the maximum deviation: $\max\{|3x + 2y - 7|, |5x + 9y - 8|, |8x + 3y - 21|, |4x + 7y - 10|\}$, will the optimization problem be convex? Justify your answer?
   Soln: The function is convex. To prove this, by the previous sub question, $f(x, y) = |ax + by + c|$ is convex for any a, b, c. It sufficies to prove that if f, g are convex, then $h(x, y) = \max\{f(x, y), g(x, y)\}$ is convex (why?). It suffices to prove that the graph of h, G_h is convex.
   Let $(x, y, s), (x', y', t) \in G_h$. Then $f(x, y) \leq s \geq g(x, y)$ and $f(x', y') \leq t \geq g(x'y')$. Let $\alpha, \beta \geq 0$ be any real values such that $\alpha + \beta = 1$. By convexity, $f(\alpha x + \beta x', \alpha y + \beta y') \leq \alpha s + \beta t \geq g(\alpha x + \beta x', \alpha y + \beta y')$. Hence, $\max\{f(\alpha x + \beta x', \alpha y + \beta y'), g(\alpha x + \beta x', \alpha y + \beta y')\} \leq \alpha s + \beta t$. That is, $(\alpha x + \beta x', \alpha y + \beta y', \alpha s + \beta t) \in G_h$.

c Does the Gradient descent algorithm, as discussed in the class, be useful to solve the first optimization problem? Can it be applied to the second? In each case justify as to why (or why not) it could be applied.

Soln: No. The version of gradient descent done in the class requires differentiablity of the function, which does not exist in the above cases.

4. [CO2] Suppose $f : \mathbf{R}^n \mapsto \mathbf{R}$ is **concave**, differentiable. Assume that $||\nabla f|| \leq B$ for some positive real B. Further assume that there exists a maximizing $x_0 \in \mathbb{R}^n$ for f such that $||x_0|| \leq R$ for some $R > 0$. Is it possible to use gradient descent to find a point $x \in \mathbf{R}^n$ such that $|f(x_0) - f(x)| \leq \epsilon$? Answer YES or NO first. If YES, explain how. If NO explain why not. $\boxed{3}$

   Soln: YES. $-f$ is convex and satisfies $||\nabla(-f)|| = || - \nabla f|| \leq B$. x_0 is the minimum of $-f$ and $||x_0|| < R$. This satisfies all the conditions necessary for gradient descent.

5. [CO3] Suppose that the gradient descent algorithm is used to optimize a differentiable convex objective function $f : \mathbf{R^n} \mapsto \mathbf{R}$. Assume that the optimal solution $||x_0||$ satisfies $||x_0|| \leq 1$. Let $||\nabla f|| \leq 2$. Assume that the step parameter is chosen to be $\gamma = \frac{1}{\sqrt{T}}$, where T is the number of iterations. With the above fixed, find the number of iterations T in order to achieve error bound of $\frac{1}{2^6}$. Show your calculations precisely. Equations derived in the class can be stated and used without proof. $\boxed{3}$

   Soln: When $||\nabla f|| \leq B$ and $||x_0|| \leq R$, it sufficies to start iterations from the origin, and it was seen in the class that after t iterations with step parameter γ, (that is, with $x_{t+1} = x_t - \gamma\nabla(f(x_t))$ as the update rule), $||f(x_T) - f(x_0)|| \leq \frac{\gamma B^2}{2} + \frac{R^2}{2\gamma T}$. Setting $\gamma = \frac{1}{\sqrt{T}}$, this value is $\frac{B^2}{2\sqrt{T}} + \frac{R^2}{\sqrt{T}}$. For this value to be less than $\epsilon = \frac{1}{2^6}$ with $B = 2$ and $R = 1$, we have $\frac{3}{\sqrt{T}} \leq \frac{1}{2^6}$, $T \geq 9.2^{12}$

6. [CO3] Let $p_1, p_2 \ldots p_n$ and $q_1, q_2 \ldots q_n$ be positive real numbers satisfying $\Sigma_i p_i = \Sigma_i q_i = 1$. Show that $\Sigma_i p_i \log \frac{q_i}{p_i} \leq 0$. $\boxed{3}$

   Soln: $f(x) = \log x$ is concave. Hence by Jensen's inequality:

$$\sum_i p_i \log \frac{q_i}{p_i} \leq \log(\sum_i p_i \frac{q_i}{p_i}) = \log \sum q_i = \log 1 = 0. \tag{4}$$

7. [CO3] Given a "training set" $x_1, x_2, \ldots x_n$" of vectors in $\mathbf{R}^n$ with their "labels" $y_1, y_2 \ldots y_n$ in $\{\pm 1\}$, the (offline) Perceptron algorithm finds a vector $w \in \mathbf{R}^n$ such that $y_i(w, x_i) > 0$ for all i, provided at least one such vector exists. Recall that the analysis of the algorithm defined $B = min\{||w|| : y_i(w, x_i) \geq 1 \text{ for all } i\}$ and $R = \max_i ||x_i||$, and showed that the number of iterations required for convergence is bounded by $(RB)^2$. $\boxed{6}$

   a) Can it happen that there exists a vector w such that $y_i(w, x_i) > 0$ for all i, but no vector w exists such that $y_i(w, x_i) \geq 1$ for all i? That is, can it happen that B is not well defined? Justify your answer?

   Soln: B will be always well defined. It is assumed that $y_i(w, x_i) > 0$ for each i. Now w can always be chosen large enough to ensure that the above product at least 1. Formally, let $t = \min_i y_i(w, x_i)$. Then, for all i, $y_i(w, x_i) \geq t$. Note that t could be less than 1. However, we conclude that $y_i(\frac{w}{t}, x_i) \geq 1$. Thus $\frac{1}{t}w$ works.

   b) Suppose the update rule of the Perceptron in during the t^{th} iteration is changed to:

   - WHILE there is some i such that $y_i(w_t, x_i) < 0$ DO
   - $$w_{t+1} = w_t + \gamma y_i x_i$$
   - ENDWHILE

   for some positive real number γ, how will the analysis of the number of iterations required for convergence change? Clearly show the steps of your analysis up to the final conclusion.

   Soln: Let w^* be a solution of minimum weight. Assume that $||w^*|| = B$ and $||x_i|| \leq R$. With the above update rule, we have:

$$(w^*, w_{t+1}) - (w^*, w_t) = (w^*, w_{t+1} - w_t) = (w^*, \gamma y_i x_i) = \gamma y_i (w^*, \gamma y_i x_i) \geq \gamma \tag{5}$$

   After iterations 1 to T, starting with initial $w_0 = 0$,

$$(w^*, w_{T}) = \sum_{t=1}^{T} (w^*, w_t) - (w^*, w_{t-1}) \geq \gamma T \tag{6}$$

   On the other hand,using the fact that $y_i(w_t, x_i) < 0$, we get:

$$||w_{t+1}||^2 = ||w_t + \gamma y_i x_i||^2 = ||w_t||^2 + 2\gamma y_i(w_t, x_i) + y_i^2\gamma^2||x_i||^2 \leq ||w_t||^2 + \gamma^2|x_i||^2. \tag{7}$$

Now, using the fact that $||x_i|| \leq R$, and iterating over all T, we get

$$||w_T||^2 \leq \gamma^2 T||R||^2 \text{ or } ||w_T|| \leq \gamma R\sqrt{T} \tag{8}$$

Now, since $||w^*|| = B$, the Cauchy Schrawrtz inequality yields,

$$\gamma T \leq |(w^*, w^T)| \leq ||w^*||||w_T|| \leq \gamma R\sqrt{T}B \tag{9}$$

This essentially yields the same bound $T \geq (BR)^2$. Thus, the step size doesn't affect convergence.