ECE 509 (Spring 2024) – Midterm Exam

March 29, 2024

Name: $\underbrace{\hspace{10cm}}$

By writing my name, I affirm *on my honor that I have neither received nor given any unauthorized assistance on this examination*.

Read (and comply with) all of the following information before starting:

- The exam is open book, open notes, and open to any other material, provided it is in non-electronic format. However, an exception is made for paper-like e-ink devices such as the reMarkable tablet and e-ink Kindle. The use of electronic devices, including cell phones, smart watches, tablets, laptops, etc., is strictly forbidden during the exam, with the exception of the specified e-ink devices. Please ensure that you only have the permitted items on your desk before starting the exam.
- Show all work, clearly and in order, if you want to get full credit. In addition, ***justify your answers*** to ensure full credit. I reserve the right to take off points if I cannot see how you arrived at your answer (even if your final answer is correct).
- Pages are provided at the end of the answer book for rough work and additional space for answers. If your answer spills over into these pages or other unused pages in the exam booklet, please clearly indicate the relevant page numbers to facilitate correct marking.
- This exam has 10 questions, for a total of 75 points and 0 bonus points. You have 80 minutes to complete it.
- Good luck!

Page:	1	2	3	4	5	6	7	8	Total
Points:	10	11	6	10	8	12	10	8	75
Bonus Points:	0	0	0	0	0	0	0	0	0

1. True or False? For each statement, mark it as true or false. **If the statement is false, provide the key fact that disproves the statement.** Note that there is a penalty of -0.25 points for each incorrect answer.

   (a) (1 point) Every subspace is a convex set.

   (b) (1 point) Every affine set is a convex set.

   (c) (1 point) Every linear function is a strictly convex function.

   (d) (1 point) Every linear function is a concave function.

   (e) (1 point) Every linear program is a convex program.

   (f) (1 point) It is possible for a concave function to have more than one maximizer $\mathbf{x} \in \mathbb{R}^n$.

   (g) (1 point) Any strictly convex function is guaranteed to have a unique $\mathbf{x} \in \mathbb{R}^n$ that serves as its minimizer.

   (h) (1 point) Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. It holds true that:

$$\arg\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) = \arg\min_{\mathbf{x}\in\mathbb{R}^n} -f(\mathbf{x})$$

   (i) (1 point) For a function $f : \mathbb{R}^n \to \mathbb{R}$, the following equality is valid:

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) = \min_{\mathbf{x}\in\mathbb{R}^n} -f(\mathbf{x})$$

   (j) (1 point) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. If $\nabla f(\mathbf{x}^*) = 0$ for some $\mathbf{x}^* \in \mathbb{R}^n$, then $\mathbf{x}^*$ is a minimizer of f.

(k) (1 point) Consider a convex function $f : \mathbb{R}^n \to \mathbb{R}$ with $\textbf{dom} f \subseteq \mathbb{R}^n$. If $\mathbf{x}^*$ is a minimizer of f, then $\nabla f(\mathbf{x}^*) = 0$.

(l) (1 point) Every convex function can be lower bounded by a quadratic function.

(m) (1 point) A convex function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying $m\mathbf{I} \preceq \nabla^2 f(\mathbf{x}) \preceq M\mathbf{I}$ for all $\mathbf{x} \in \textbf{dom} f$ can be bounded above and below by quadratic functions.

(n) (1 point) Every optimization method for convex optimization is guaranteed to be a descent method.

(o) (1 point) Gradient descent can be applied to solve any convex optimization problem.

(p) (1 point) The convergence rate of gradient descent can be significantly affected by the choice of coordinates for the function.

(q) (1 point) Steepest descent achieves linear convergence for any chosen norm when optimizing strongly convex functions.

(r) (1 point) When optimizing strongly convex functions, the convergence speed of steepest descent is universally faster than that of gradient descent for any choice of norm.

(s) (1 point) Newton's method is a descent method in the context of optimizing strongly convex functions.

(t) (1 point) Newton's method is classified as a second-order technique for solving optimization problems.

(u) (1 point) In the optimization of strongly convex functions, if $k = \ell$ is the first iteration where Newton's method selects $t = 1$ as the step-size using backtracking, then it will consistently use $t = 1$ as the step-size for all $k > \ell$.

(v) (1 point) During the quadratic convergence phase of Newton's method, the rate of convergence does not depend on the condition number of the function.

(w) (1 point) During the quadratic convergence phase of optimizing strongly convex functions, Newton's method's convergence speed remains largely unaffected by the choice of coordinates for the function.

2. You are given a budget of $1,000,000 ($B = \$1M$) to allocate across n different assets (stocks, bonds, etc.). The initial allocation in dollars to these assets is represented by the optimization variable $\mathbf{x} \in \mathbb{R}^n_+$, where $\mathbb{R}^n_+$ signifies that the allocations can only be non-negative, reflecting the amount of money invested in each asset.

   Objective: Maximize the objective function $R(\mathbf{x})$, representing the expected return on the initial allocation after 1 year.

   Constraints:

   1. The total budget constraint is given by $\mathbf{1}^\top \mathbf{x} = B$, ensuring that the sum of the allocations across all assets equals the total available budget of $1M.
   2. There is a risk constraint function $S(\mathbf{x}) \leq V$, where $S(\mathbf{x})$ measures the standard deviation of the return (volatility) for a given allocation $\mathbf{x}$ and is constrained to be less than or equal to $V = 100,000$, indicating a maximum allowable risk level in terms of volatility.

   Given are four different allocations $\mathbf{x}$ returned by an optimization algorithm:

   - $\mathbf{x}_1 \;\Rightarrow\; R(\mathbf{x}_1) = \$100,000$ and $S(\mathbf{x}_1) = \$20,000$
   - $\mathbf{x}_2 \;\Rightarrow\; R(\mathbf{x}_2) = \$100,000$ and $S(\mathbf{x}_2) = \$30,000$
   - $\mathbf{x}_3 \;\Rightarrow\; R(\mathbf{x}_3) = \$500,000$ and $S(\mathbf{x}_3) = \$100,000$
   - $\mathbf{x}_4 \;\Rightarrow\; R(\mathbf{x}_4) = \$10,000$ and $S(\mathbf{x}_4) = \$1,000$

   (a) (2 points) Between the allocations $\mathbf{x}_1$ and $\mathbf{x}_2$, which one is better in terms of the optimization problem as it is posed?

   (b) (2 points) Which allocation is the best one in terms of the optimization problem, as it is posed, and why?

3. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) := (x+1)^2 - 4$. Evaluate:
   (a) (2 points) $\min_{x \in \mathbb{R}} f(x)$
   (b) (2 points) $\arg\min_{x \in \mathbb{R}} f(x)$

4. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) := -\exp(-x)$. For each of the following constrained optimization problems, determine the respective infimum or supremum, and indicate whether there exists an x within the constraint set that achieves this extremum.
   (a) (2 points) $\inf_{x \geq 0} f(x)$
   (b) (2 points) $\inf_{x > 0} f(x)$
   (c) (2 points) $\sup_{x > 0} f(x)$

5. Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(\mathbf{x}) := \exp\left(-\frac{1}{2}\mathbf{x}^\top \mathbf{P}\mathbf{x}\right)$, with $\mathbf{P}$ belonging to the set of symmetric matrices $\mathbb{S}^n$. Compute:
   (a) (3 points) The gradient $\nabla f(\mathbf{x})$
   (b) (5 points) The Hessian $\nabla^2 f(\mathbf{x})$

6. Given the vectors $\mathbf{x}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$, and $\mathbf{x}_3 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ in $\mathbb{R}^2$:
   (a) (2 points) Sketch the convex hull of these three vectors.
   (b) (2 points) Identify the affine hull of these three vectors.
   (c) (2 points) What is the dimension of the affine hull? Provide justification for your answer.

7. Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ with domain f being $\mathbb{R}^n$, defined as:

$$f(\mathbf{x}) := \frac{1}{2}(\mathbf{x}^\top \mathbf{a})^2.$$

   (a) (3 points) Discuss whether $f(\mathbf{x})$ is a convex function, providing reasoning for your conclusion.
   (b) (3 points) Define $g(\mathbf{x}) := f(\mathbf{x}) + \|\mathbf{x}\|_2^2$. Determine if $g(\mathbf{x})$ exhibits strong convexity. If it does, identify the strong convexity parameter for $g(\mathbf{x})$. If it does not, provide a rationale for your conclusion.

8. Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ that is m-strongly convex with strong convexity parameter $m = 0.5$. Let $p^* := \min_{\mathbf{x}} f(\mathbf{x})$ and $\mathbf{x}^* := \arg\min_{\mathbf{x}} f(\mathbf{x})$.

   (a) (3 points) Determine the upper bound for the gradient norm $\|\nabla f(\mathbf{x}^{(k)})\|_2$ that ensures the inequality $f(\mathbf{x}^{(k)}) - p^* \leq 10^{-12}$ holds for an iterative algorithm generating the sequence $\{\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots\}$.

   (b) (3 points) If the objective is to achieve convergence in terms of the solution's precision, to the effect that $\|\mathbf{x}^{(k)} - \mathbf{x}^*\|_2 \leq 10^{-12}$, how would this requirement affect the gradient norm bound established in part (a)?

9. (4 points) While both steepest descent and Newton's method are designed to achieve faster convergence compared to gradient descent, gradient descent is often the preferred method in many practical applications. Discuss the reasons behind this preference.

10. Let $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$ be two functions characterized by the expressions $f_1(\mathbf{x}) = \mathbf{x}^\top \mathbf{P}_1 \mathbf{x}$ and $f_2(\mathbf{x}) = \mathbf{x}^\top \mathbf{P}_2 \mathbf{x}$, respectively. Here, the matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ are elements of the set of positive definite matrices $\mathbf{S}^n_{++}$. The eigenvalues of $\mathbf{P}_1$ and $\mathbf{P}_2$ are ordered such that:

$$1.1 = \lambda_1(\mathbf{P}_1) \geq \ldots \geq \lambda_n(\mathbf{P}_1) = 0.1$$
$$1000 = \lambda_1(\mathbf{P}_2) \geq \ldots \geq \lambda_n(\mathbf{P}_2) = 500$$

Suppose we employ gradient descent with backtracking to optimize both functions. For notation simplicity, we denote the initial points for both functions as $\mathbf{x}^{(0)}$, with the understanding that although they are different points in $\mathbb{R}^n$, they yield the same initial objective function values, i.e., $f_1(\mathbf{x}^{(0)}) = f_2(\mathbf{x}^{(0)})$. This is a notational abuse for the sake of comparison.

(a) (3 points) Considering the eigenvalue distributions of $\mathbf{P}_1$ and $\mathbf{P}_2$, which function is expected to require more iterations to converge, and for what reason?

(b) (5 points) Assuming the convergence criterion $f_i(\mathbf{x}^{(k)}) \leq 10^{-12}$ for both functions $i = 1, 2$, what are the upper bounds on the number of iterations k for $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$, given that $f_1(\mathbf{x}^{(0)}) = f_2(\mathbf{x}^{(0)}) = 1000$?

—Scratch Pages—

Question:	1	2	3	4	5	6	7	8	9	10	Total
Points:	23	4	4	6	8	6	6	6	4	8	75
Score:											