Name: ________________________________________________________________

MIT Email: ____________________________________________________________

Instructions:

• This is an 80-minute exam, with a total of 80 points.

• The following is permitted:
  - two sheets of double-sided, letter-sized notes.

• Electronic devices are not allowed.

• You may use the scratch papers attached at the end of the exam booklet for calculations or drafts.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Points</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem 0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Problem 1</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Problem 2</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Problem 3</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td></td>
</tr>
</tbody>
</table>
Problem 0 (2 points) : Write your name on every page of this exam booklet.

Problem 1 (24 points) : For each of the statements below, circle whether they are true (T) or false (F). Please **read the statements carefully**, and in all cases **fully justify your answer**. Partial credit will **not** be given for answers lacking suitable justification.

(a) T F Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function. If a point $x^*$ is a local maximum, then it is also a global minimum.

**TRUE.** If $x^*$ is a local maximum, then the gradient vanishes. Since $f$ is convex, this implies that it is global minimum. In fact, in this case $f$ must be constant.

(b) T F When minimizing a *convex quadratic function* over a convex set $X$, the gradient projection method $x^{k+1} = [x^k - s\nabla f(x^k)]^+$ converges for any choice of constant stepsize $s$.

**FALSE.** We need the stepsize to be sufficiently small. Counterexamples are easy to construct.
(c) T   F   Let \( f(x, y) = \sin(x^2 + \gamma y^2) \), where \( \gamma > 0 \) is a given constant. Consider the steepest descent method (with the minimization rule), applied to the function \( f \). Assume that the (nonzero) initial condition \((x^0, y^0)\) is very small in magnitude, i.e., \( \|(x^0, y^0)\| \approx 0 \). Then, the method converges linearly, and the convergence rate is fastest when \( \gamma = 1 \).

**TRUE.** It follows from the properties discussed during the lecture. If \( \gamma = 1 \), then the condition number is 1 (smallest possible), and therefore the convergence is faster.

(d) T   F   Consider the convex function \( f(x) = x^4 \). To minimize \( f \), we apply Newton’s method with full stepsize, from the initial condition \( x^0 = 1 \). Then, the sequence \( \{ x^0, x^1, \ldots, x^k, \ldots \} \) generated by Newton’s method converges quadratically to the minimizer \( x = 0 \).

**FALSE.** The Newton iteration in this case is \( x_{k+1} = x_k - (4x_k^3)/(12x_k^2) = \frac{2}{3}x_k \), which converges only linearly. The reason why we don’t achieve quadratic convergence is because the hypothesis of the theorem are not satisfied – in particular, the minimum eigenvalue of the Hessian is not bounded away from zero.
Consider the convex set \( X = \{ x \in \mathbb{R}^n : |x_i| \leq 1 \} \), i.e., the unit cube in \( n \)-dimensional space. Then, given a point \( z = (z_1, \ldots, z_n) \in \mathbb{R}^n \) (where \( z \neq 0 \)), its Euclidean projection onto \( X \) is given by \( [z]^+ = z/\|z\|_\infty \), where \( \|z\|_\infty := \max_i |z_i| \).

**FALSE.** This is easy to see by considering, e.g., the point \( z = (2, 1) \). The given formula produces the point \( (1, 1/2) \), while the true projection is \( [z]^+ = (1, 1) \). This is easy to see geometrically.

Let \( X \) be a compact convex set. When minimizing a linear function \( f(x) = c^T x \) (where \( c \neq 0 \)) over the set \( X \), the minimum is always achieved on the boundary of \( X \). (Hint: use the optimality conditions).

**TRUE.** This follows directly from the optimality condition \( \nabla f(x^*)(x - x^*) \geq 0 \) for all \( x \in X \). If \( x^* \) is in the interior of \( X \), then the gradient should vanish, but on the other hand, \( \nabla f(x) = c \neq 0 \).
Problem 2 (20 points):
Consider the function \( f(x, y) = x^4 + 2y^2 - 4xy \), which is defined in all of \( \mathbb{R}^2 \).

We first consider \textit{unconstrained} minimization of \( f \):

(a) Find all the stationary points, local minima, and the global minima, if they exist.

(b) Is \( f \) a convex function?

(c) Write the optimality conditions. Are these necessary? Sufficient? Verify that the points you found satisfy these conditions.

Consider now the same objective function, with the additional constraint \( \{ y \geq 8 \} \).

(d) Consider the solutions you found in item (a). Do they satisfy the new constraint?

(e) Find all the points that satisfy the Karush-Kuhn-Tucker first-order necessary conditions.

(f) What’s the value of the Lagrange multipliers? Are the signs correct?

Solution:

(a) The stationary points are \((0, 0), (1, 1)\) and \((-1, -1)\). The first is a saddle point with \( f(0, 0) = 0 \), and the other two are global minima with \( f(1, 1) = f(-1, -1) = -1 \).

(b) \( f \) is not convex. This can be easily seen, for instance, since the Hessian at \((0, 0)\) is not positive semidefinite.

(c) The optimality condition is \( \nabla f(x^*) = 0 \). Checking the eigenvalues of the Hessian...

(d) The solutions computed earlier violate the constraint \( g(x, y) = 8 - y \leq 0 \).

(e) Solving the KKT equations

\[
\nabla f(x) + \lambda \nabla g(x) = \begin{bmatrix} 4x^3 - 4y \\ -4x + 4y - \lambda \end{bmatrix} = 0
\]

and the complementarity equation \( \lambda (8 - y) = 0 \), we have that \( x = 2, y = 8 \), and \( \lambda = 24 \). We also have \( f(2, 8) = 80 \).
Page left blank for calculations.
Page left blank for calculations.
Problem 3 (20 points):

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. For simplicity, we assume that all the eigenvalues of $A$ are positive and distinct, i.e., $\lambda_1 > \cdots > \lambda_n > 0$ (this implies that $A$ is positive definite). Consider the optimization problem

$$\text{maximize } x^T A x \text{ subject to } x \in X,$$

where $X$ is the unit ball of the Euclidean norm in $\mathbb{R}^n$, i.e., $X = \{ x \in \mathbb{R}^n : \|x\|^2 \leq 1 \}$. Notice that the feasible set is convex, and that this is a maximization problem.

(a) Write the optimality conditions for this problem.

(b) What are the stationary (KKT) points? What is the global maximum?

(c) Consider the conditional gradient (or Frank-Wolfe) method to solve this problem. Write an explicit iteration that implements this algorithm. What happens if you use the “minimization rule” to choose the stepsize?

(d) Consider now the gradient projection method. Write an explicit iteration that implements this algorithm.

(e) Compare the form of two methods. Describe choices of parameters (e.g., stepsize) to ensure convergence properties.

Remark: If possible, make the simplest possible choices (e.g., when selecting stepsizes), in order to guarantee that the resulting methods are “nice.” When writing your algorithm, keep in mind that as given, this is a maximization problem.

Solution:

(a) The KKT optimality conditions (removing the factor 2) are:

$$Ax - \lambda x = 0, \quad \lambda (1 - \|x\|^2) = 0.$$

(b) The possible KKT points are either $(\lambda, x)$, where the components are an eigenvalue / normalized eigenvector pair, or $(0, 0)$. The global maximum is $\lambda_1$.

(c) To apply the conditional gradient method, we need to be able to maximize a linear function over $X$. But this is easy, since the maximizer of $c^T x$ over $X$ is $c/\|c\|$. Thus, since $\nabla f(x) = 2Ax$, the iterations take the form

$$x_{k+1} = x_k + \alpha_k (Ax_k/\|Ax_k\| - x_k)$$

$$= (1 - \alpha_k)x_k + \alpha_k (Ax_k/\|Ax_k\|),$$

Since the objective function is quadratic and positive definite, the best choice of stepsize will be $\alpha_k = 1$, in which case the method reduces to the iteration

$$x_{k+1} = Ax_k/\|Ax_k\|,$$

which is the standard power iteration algorithm.
(d) For the gradient projection method, we need to be able to compute the projection \([z]^+\) onto \(X\). It is easy to see that \([z]^+ = z/\|z\|\) if \(\|z\| \geq 1\), or \([z]^+ = z\) otherwise.

Then, we have:

\[
x_{k+1} = [x_k + s_k \nabla f(x_k)]^+
= [x_k + s_k 2Ax_k]^+
= [(I + 2s_k A)x_k]^+
= (I + 2s_k A)x_k / \|(I + 2s_k A)x_k\|,
\]

where the last equation follows because the iterates will (eventually) always satisfy \(\|(I + 2s_k A)x_k\| \geq 1\) if \(A\) is positive semidefinite.

It can be see that this is just a (scaled and shifted) version of the standard power iteration algorithm.
Page left blank for calculations.
Page left blank for calculations.
(SCRATCH PAPER. Do not detach. Writings on this paper will not be graded.)