Final Exam

Instructions:

• This is an *open book* exam. You can use the course textbook, other optimization books, the homework and recitation solutions, or your own class notes. Please decide what you will use prior to starting the exam. *Write this down* on the first page of your exam.

• You *cannot* discuss or talk about the content of the exam (or any other aspect of it) with other people.

• You *cannot* use the Web during the exam (for purposes related to the exam, of course).

• Violating these rules goes against MIT’s academic honesty policies. We take these quite seriously, and will enforce them.

• There are five problems in the exam. There is one computational problem (P4). Please plan accordingly, to make sure you will have access to a computer (e.g., Athena cluster). If for any reason you won’t be able to do this, please let us know immediately.

• Please hand in the completed exam before Saturday 19, at 1PM, in 32D-630 (Stata building, Prof. Ozdaglar’s office).

Problem 1 (22 points):

(1) For each of the following functions \( f : \mathbb{R}^n \to \mathbb{R} \), show that they are convex and compute a subgradient at an arbitrary \( x \in \mathbb{R}^n \):

(a) \( f(x) = \sup_{0 \leq t \leq 1} p(t) \), where \( p(t) = x_1 + x_2 t + \cdots + x_n t^{n-1} \).

(b) \( f(x) = \inf_{A y \leq b} \| x - y \|_2 \), where \( A \) is an \( m \times n \) matrix and \( b \in \mathbb{R}^m \).

(c) \( f(x) = \| x_1 A_1 + \cdots + x_n A_n \| \), where \( A_i \in \mathbb{R}^{m \times p} \) are real matrices, and \( \| \cdot \| \) is the \( \ell_2 \) induced norm (i.e., the maximum singular value).
Consider the following notion of approximate subgradient: Given a convex function \( f : \mathbb{R}^n \to \mathbb{R} \) and a positive scalar \( \epsilon \), we say that a vector \( d \in \mathbb{R}^n \) is an \( \epsilon \)-subgradient of \( f \) at a point \( x \in \mathbb{R}^n \) if

\[
f(z) \geq f(x) + (z - x)'d - \epsilon, \quad \forall z \in \mathbb{R}^n.
\]

The set of all \( \epsilon \)-subgradients of a convex function \( f \) at some \( x \in \mathbb{R}^n \) is called the \( \epsilon \)-subdifferential of \( f \) at \( x \) and is denoted by \( \partial_\epsilon f(x) \).

(a) Consider the function \( f(x) = \max_{i=1,\ldots,m} \{a_i'x + b_i\} \), where \( a_i \in \mathbb{R}^n \) and \( b_i \in \mathbb{R} \) for all \( i = 1, \ldots, m \). Find an \( \epsilon \)-subgradient at an arbitrary \( x \in \mathbb{R}^n \).

(b) Interpret an \( \epsilon \)-subgradient geometrically in terms of the epigraph of \( f \) for a general function \( f \).

(c) Show that \( 0 \in \partial_\epsilon f(x) \) if and only if

\[
f(x) \leq \inf_{z \in \mathbb{R}^n} f(z) + \epsilon.
\]

(d) Consider the following iterative algorithm:

- At the current iterate, check whether \( 0 \in \partial_\epsilon f(x) \); if so, stop.
- Otherwise, move along the direction opposite to the vector of minimum norm in \( \partial_\epsilon f(x) \) using some stepsize.

Assume that \( \inf_{z \in \mathbb{R}^n} f(z) > -\infty \). Show that with some appropriate stepsize rule, the algorithm terminates with an \( \epsilon \)-optimal solution.

**Hint:** Show that at each iteration, by moving along the direction opposite to the vector of minimum norm in \( \partial_\epsilon f(x) \), we can guarantee a cost improvement of at least \( \epsilon \).

**Problem 2 (20 points):** We consider the problem of finding a sequence of points \( \{x^k\} \) that converges to the intersection \( C = \cap_{i=1}^m C_i \) of some convex sets \( C_i \). We assume that the set \( C \) is nonempty.

(a) Consider a method where, at every step, the current point is projected onto any set \( C_i \) not containing the point. Give an example showing that such an algorithm can fail, i.e., we may have

\[
\lim_{k \to \infty} \inf \text{dist}(x^k, C) > 0.
\]

(b) Assume next that there exists some \( B \) such that for all \( i = 1, \ldots, m \) and all \( k \geq 1 \), we have \( x^j \in C_i \) for some \( j \in \{k + 1, \ldots, k + B\} \). In other words, in each block of \( B \) successive iterates of the algorithm, the iterates visit each of the sets. Informally
speaking, this means that the projections can be chosen in any order as long as each set is taken into account every so often. Show that:

$$\lim_{k \to \infty} \text{dist}(x^k, C) = 0.$$ 

*Hint:* What can you say about the squared-distances of the iterates to a point in the intersection $C$? Use the bound $B$ to show that the $\text{dist}(x^k, C_i)$ goes to 0 as $k \to \infty$ for each $i$.

**Problem 3 (20 points):**

(1) Consider the following optimization problem:

$$\begin{align*}
\text{minimize} & \quad e^{-x_2} \\
\text{subject to} & \quad \|x\| - x_1 \leq u, \quad X = \{(x_1, x_2) \mid x_2 \geq 0\}.
\end{align*}$$

(a) Sketch the feasible region for $u = 0$ and $u > 0$.

(b) Assume now that $u = 0$. What is the optimal value of problem (1)? What is the optimal value of the corresponding dual problem? Explain why sufficient conditions for no duality gap do not apply here.

(2) Consider the problem

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in X, \quad g_j(x) \leq 0, \quad j = 1, \ldots, r,
\end{align*}$$

where $X$ is a nonempty subset of $\mathbb{R}^n$, and $f : \mathbb{R}^n \mapsto \mathbb{R}$, $g_j : \mathbb{R}^n \mapsto \mathbb{R}$ are given functions. We assume that the problem is feasible and the optimal value $f^*$ is bounded from below, i.e., $f^* > -\infty$.

Let $\bar{r}$ be an integer with $1 \leq \bar{r} < r$, and consider the set

$$\bar{X} = \{x \in X \mid g_{\bar{r}+1}(x) \leq 0, \ldots, g_r(x) \leq 0\},$$

and the problem

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in \bar{X}, \quad g_j(x) \leq 0, \quad j = 1, \ldots, \bar{r}.
\end{align*}$$

Let $q^*$ and $\bar{q}^*$ be the optimal dual values of problems (2) and (3), respectively.

(a) Show that $q^* \leq \bar{q}^* \leq f^*$. 

3
(b) Show that if problem (2) has no duality gap, the same is true for problem (3). Furthermore, if \((\mu_1, \ldots, \mu_r)\) is a Lagrange multiplier for problem (2), then \((\mu_1, \ldots, \mu_r)\) is a Lagrange multiplier for problem (3).

(c) Construct an example to show that we may have \(q^* < \bar{q}^* = f^*\).

Problem 4 (23 points):
For this problem, you will need the dataset available in Stellar, at the URL

http://stellar.mit.edu/S/course/6/sp07/6.252/

Please download it, and make sure you can use it, as soon as you read this.

The MATLAB file (or the plain text files) contains matrices \(A, b\), where \(A \in \mathbb{R}^{m \times n}\) and \(b \in \mathbb{R}^m\) (in the problem, \(n = 50, m = 100\)). Your task is to develop a numerical algorithm to solve the following optimization problem:

\[
\min_x \quad \|Ax - b\|_3 + \sum_{i=1}^{n} x_i \log x_i \\
\text{subject to} \quad 1 \leq x_i \leq 5.
\]

Here \(\| \cdot \|_3\) denotes the \(\ell_3\) norm [sic], i.e., \(\|x\|_3 = (\sum_{i=1}^{n} |x_i|^3)^{\frac{1}{3}}\).

(a) Is the objective function differentiable? Convex? Is the feasible set convex?

(b) What are the KKT optimality conditions? Are they necessary? Sufficient?

(c) Do Lagrange multipliers exist? Why?

(d) Design, thoroughly describe, and implement a projected gradient method to solve this problem.

(e) Describe, and clearly justify, how to do the projection onto the feasible set.

(f) Discuss your choice of stepsize rules, and the associated convergence properties that you should expect from the theory.

(g) Implement your method (e.g., in MATLAB). Plot the current objective value \(f - f^*\) as a function of the iteration number, for 10 different initial conditions. Use a log-linear scaling in the axes, so you can make conclusions on the observed convergence rate.

(h) What is the best objective value you have found? Plot the best candidate solution \(x^*\) found, making sure that it is feasible. Does this solution satisfy (approximately) the KKT conditions?
Problem 5 (15 points):
Consider a noncooperative game with $n$ players, each with strategy set $C_i \subset \mathbb{R}^{n_i}$. Let $C = C_1 \times \cdots \times C_n$ and let $u_i : C \mapsto \mathbb{R}$ denote the utility function of player $i$, i.e., for a given strategy profile $x = (x_i, x_{-i}) \in C$ (where $x_{-i} = [x_j]_{j \neq i}$), the utility of player $i$ is $u_i(x_i, x_{-i})$. A strategy profile $x^*$ is a (pure strategy) Nash equilibrium if
\[
u_i(x^*_i, x^*_{-i}) \geq u_i(x_i, x^*_{-i}), \quad \text{for all } i \text{ and all } x_i \in C_i,
\]
i.e., no player can gain by unilaterally deviating from his/her strategy.
Assume that each $C_i$ is a closed convex subset of $\mathbb{R}^{n_i}$. Assume also that for each $i$, the utility function $u_i(x_i, x_{-i})$ is concave and continuously differentiable in $x_i$ for all $x_{-i} \in C_{-i}$ (where $C_{-i} = \prod_{j \neq i} C_j$). Show that $x^*$ is a Nash equilibrium if and only if it is the solution of a variational inequality problem.