UNIVERSITY OF TEXAS AT AUSTIN

EE 381V - LARGE SCALE OPTIMIZATION

FALL 2012

FINAL EXAM

MONDAY, DEC 17, 2012

Name: ______________________________

Email: ______________________________

- You have 3 hours for this exam.
- The exam is closed book and closed notes. You are allowed to have three standard letter-sized sheet, 6 sides, of handwritten notes.
- Calculators, laptop computers, Palm Pilots, two-way e-mail pagers, etc. may not be used.
- Write your answers in the spaces provided. **SHOW YOUR REASONING FOR CREDIT.**
- **Please show all of your work. Answers without appropriate justification will receive very little credit.** If you need extra space, use the back of the previous page.

Problem 1 (10 pnts): ________	Problem 6 (15 pnts): ________
Problem 2 (15 pnts): ________	Problem 7 (10 pnts): ________
Problem 3 (10 pnts): ________	Problem 8 (10 pnts): ________
Problem 4 (10 pnts): ________	Problem 9 (10 pnts): ________
Problem 5 (10 pnts): ________	**Total (100 pnts)** : ________

Problem 1: (10 pnts) In this question we derive the projected gradient algorithm for the infinity norm constraint.

(a) Given a vector $a \in \mathbb{R}^n$ and a scalar $t > 0$, give a closed form equation for the optimum of the following optimization problem

$$\begin{aligned} \min_{u} \quad & \|u - a\|_2^2 \\ s.t. \quad & \|u\|_\infty \leq t \end{aligned}$$

(b) Give a closed form for the update rule for projected gradient descent, with step size η, for the optimization problem

$$\begin{aligned} \min_{x} \quad & f(x) \\ s.t. \quad & \|x\|_\infty \leq t \end{aligned}$$

(c) Give an closed form for the update rule for FISTA applied to this problem. Please show your reasoning.

Problem 2: (15 pnts) In this problem we explore an equivalence between constrained optimization and a particular penalty function. In the following, all functions f_i are convex.

Consider the convex optimization problem

$$\begin{aligned} \min_x \quad & f_0(x) \\ s.t. \quad & f_i(x) \leq 0 \text{ for all } 1 \leq i \leq m \end{aligned} \tag{1}$$

Suppose that its lagrange dual optimization problem is given by

$$\begin{aligned} \max_\lambda \quad & g(\lambda) \\ s.t. \quad & \lambda_i \geq 0 \text{ for all } 1 \leq i \leq m \end{aligned} \tag{2}$$

Assume also that we have strong duality, the value of the optimum of (both) problems is finite, and is attained by finite x and λ respectively.

We now consider a "penalized version" of the primal problem:

$$\min_x \quad f_0(x) + t \max_i f_i(x)^+ \tag{3}$$

where $f_i(x)^+ = \max\{f_i(x), 0\}$ and $t \geq 0$ is a penalty parameter. This can be rewritten as

$$\begin{aligned} \min_{x,y} \quad & f_0(x) + ty \\ s.t. \quad & f_i(x) \leq y \text{ for all } 1 \leq i \leq m \\ & y \geq 0 \end{aligned} \tag{4}$$

(a) Show that (3) is a convex optimization problem.

(b) Just as (2) is the dual problem of (1), find the dual problem of (4). Your answer should only involve $g(\cdot)$, λ and t.

(c) Let λ^* be the optimum of the original dual (2). Using part (b), show that for $t \geq \sum_i \lambda_i^*$, the optimal values of the original problem (1) and the penalized version (3) will be the same.

Problem 3: (10 pnts) Recall that, for a smooth function f, a random update rule $x^+ = x - \eta g(x)$ represents a unbiased stochastic gradient algorithm if $E[g(x)] = \alpha \nabla f(x)$ at every x, for some constant α.

We now develop a particular kind of stochastic gradient algorithm. Let $x \in \mathbb{R}^n$, and let $\mathcal{C}$ be a collection of subsets $C \subset [n]$. Assume that every coordinate i is in at least one set C in this collection. Consider the following update rule: sequentially, in every step, and for a fixed η,

1. Pick a C uniformly at random from the collection $\mathcal{C}$.
2. Update all its variables by (partial) gradient descent, i.e. $x_i^+ = x_i - \eta \nabla_i f(x)$ for all $i \in C$, and keep other variables $x_j, j \notin C$ fixed.

(a) Show, via a simple example, that the above may NOT be an unbiased stochastic gradient algorithm.

(b) Suppose now we could choose a different (but fixed) step size η_i for every coordinate i. Based on the set collection $\mathcal{C}$ as above, what is a choice of η_i's that will make the resulting algorithm an unbiased stochastic gradient algorithm ?

Problem 4: (10 pnts) We prove some properties of convex functions and their conjugates. Each of the parts of this question are independent of one-another.

(a) Show that the set of global minimizers of a convex function is a convex set.

(b) For any (not necessarily convex) function $f(x)$, let $f^*(\theta)$ be its Fenchel conjugate. Let $f^{**}(x)$ be the Fenchel conjugate of $f^*(\theta)$ (note: f^{**} is also called the bi-conjugate of f,) Show that $f^{**}(x) \leq f(x)$ for all x.

(c) Let $f(x)$, where $x \in \mathbb{R}^n$ be a convex function (not necessarily smooth). Define $g(x, y) = f(x + y)$ for all $x, y \in \mathbb{R}^n$. Is g a convex function on $\mathbb{R}^{2n}$? If yes, prove it. If no, give a simple counter-example.

(d) Consider the problem of *maximizing* a strictly convex function over a closed, compact set. Show that the optimum is always attained at an extreme point. (Recall: an extreme point of a set is a point which cannot be expressed as a convex combination of two points in the set)

Problem 5: (10 pnts) Consider a discrete-time Markov chain $X(t)$, with probability transition matrix P, whose entries are defined to be

$$P_{ij} \;:=\; \mathbf{prob}[X(t+1) = j | X(t) = i]$$

i.e., the probability of seeing state j after state i. Consider the problem of estimating P from data, i.e. from an observed sequence $x(1), \ldots, x(m)$.

(a) Show that this can be formulated as a convex optimization problem. Specify both the objective function and the constraints. Let n_{ij} be the number of times we see a transition from i to j. Your answer should be in terms of these n's.

(b) Show that the optimum is given by $\widehat{P}_{ij} = \frac{n_{ij}}{\sum_k n_{ik}}$. *(Hint: decouple)*

Problem 6: (15 pnts) INDEPENDENTSET. Consider a graph $G = (V, E)$. A subset $S \subset V$ of the nodes is called an independent set, if it contains no two vertices joined by an edge. In class we considered two different binary linear programming formulations for the problem of finding the maximum independent set, and also its relaxation. In this problem we show that this problem can be written exactly *as a convex optimization problem over two scalar variables*.

(a) Recall the set of totally positive matrices:

$$\mathcal{P} = \{\sum x_i x_i^\top \,|\, x_i \in \mathbf{R}^n_{\geq 0}\}.$$

Show that the dual cone is equal to the set of copositive matrices:

$$\begin{aligned} \mathcal{P}^* &= \{M \,|\, \langle M, P\rangle \geq 0,\ \forall P \in \mathcal{P}\} \\ &= \mathcal{C}_+ \\ &= \{M \,|\, x^\top M x \geq 0,\ \forall x \geq 0\}. \end{aligned}$$

(b) Let J_n be the all 1's matrix. Show that the following problem has value at least equal to $\alpha(G)$, the size of the maximum independent set.

$$\begin{aligned} \max: \quad & \langle J_n, X\rangle \\ \text{s.t.}: \quad & \mathrm{Tr}(X) = 1 \\ & x_{ij} = 0,\ \forall (i,j) \in E \\ & X \in \mathcal{P}. \end{aligned}$$

(c) Write the dual to this problem, and show that it has the same value as the following two-variable problem:

$$\begin{array}{ll} \min: & t \\ \text{s.t.}: & tI_n + zA_G - J_n \in \mathcal{C}_+ \\ & t, z \in \mathbb{R}, \end{array}$$

where I_n denotes the identity matrix, and A_G is the adjacency matrix of the graph:

$$(A_G)_{ij} = \begin{cases} 1, & (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

(d) Now use the Motzkin-Strauss Theorem and weak duality to conclude that the primal above must have a value equal to the size of the maximum independent set. The Motzkin-Strauss Theorem says (use this, do not prove it):

$$\frac{1}{\alpha(G)} = \min\{x^\top (A_G + I_n)x \,:\, x \geq 0, \sum x = 1\}.$$

Problem 7: (10 pnts) Suppose f is an unknown function on $\mathbb{R}^n$. We are given m points and their function evaluations $\{x_i, f(x_i)\}$, and we want to determine if f is convex or not.

Write down a set of linear constraints which are such that *(a)* they will have a feasible point if f is convex, and *(b)* if they do have a feasible point then there exists a convex f that is consistent with the n points and values given.

After writing down the constraints, for proving *(a)*, construct a feasible point using properties of convex functions. For proving *(b)*, make a convex function from the feasible point.

Problem 8: (10 pnts) Consider the problem:

$$\min_{x \in \mathcal{X}} : \; f(x),$$

where $\mathcal{X}$ is a convex set, and f is a convex function. Suppose that we have a black box that gives *noisy* sub gradients; in particular, at the point x, it returns a g that satisfies

$$f(y) \geq f(x) + \langle g, y - x \rangle - \delta, \quad \forall y \in \mathbb{R}^n.$$

Here $\delta > 0$ is a universal constant . Suppose, moreover, that we have: $\|g\|_2 \leq G$, uniformly.

Show that with fixed stepsize, projected noisy-sub-gradient: $x_{k+1} = x_k - h_k \langle g_k, x_k \rangle$, converges to a δ-suboptimal solution, with convergence rate similar to that of ordinary sub gradient descent: $O(1/\varepsilon^2)$.

Hint: start by showing the a version of the basic Lyapunov inequality:

$$\|x_{k+1} - x^*\|_2^2 - \|x_k - x^*\|_2^2 \leq 2h_k(f(x_k) - f^*).$$

that now includes a δ term.

Problem 9: (10 pnts)

(a) Consider the optimization problem:

$$\min: \; f_1(x) + f_2(x).$$

Assuming f_1 and f_2 are both closed convex functions, show that this is equivalent to the problem:

$$\min: \; f_1^*(\lambda) + f_2^*(-\lambda).$$

(Hint: dual decomposition.)

(b) Consider the proximal point problem update:

$$x_{k+1} \in \arg\min \left\{ f(x) + \frac{1}{2h_k}\|x - x_k\|_2^2 \right\}.$$

Assuming f is a closed convex function, apply the result of part **(a)** to show that the optimization is equivalent to

$$\min: \; f^*(\lambda) - \langle x_k, \lambda \rangle + \frac{h_k}{2}\|\lambda\|_2^2.$$