UNIVERSITY OF TEXAS AT AUSTIN

EE 381V - LARGE SCALE OPTIMIZATION

FALL 2012

EXAM 1

TUESDAY, OCTOBER 30, 2012

Name: ______________________________

Email: ______________________________

- You have 75 minutes for this exam.
- The exam is closed book and closed notes. You are allowed to have one standard letter-sized sheet, two 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 (20 pnts): ____________

Problem 2 (20 pnts): ____________

Problem 3 (20 pnts): ____________

Problem 4 (20 pnts): ____________

Problem 5 (20 pnts): ____________

Total (100 pnts) : ____________

Problem 1: (20 pnts) Consider gradient descent with fixed step size $\eta > 0$:

$$x_{k+1} = x_k - \eta \nabla f(x_k)$$

Given $a > 0$, make a convex function $f_a(x)$ so that, from *any* initial point, the above converges linearly for all step sizes $\eta < a$, and diverges for all $\eta > a$. Here linear convergence means that there exists some $c < 1$ such that $\|x_k - x_*\| \leq c^k$; of course this c will depend on η and a.

For full credit, you would need to prove both statements: that for $\eta > a$ it diverges from *any* initial point that is not already optimal, and for $\eta < a$ it converges linearly from any initial point.

Problem 2: (20 pnts) Recall we showed in class that steepest descent with the ℓ_1 norm $\|\cdot\|_1$ became the "greedy" coordinate descent algorithm. That is, for getting from x_k to x_{k+1}, the descent direction is along the coordinate vector e_{i^*}, where i^* is the "best" coordinate, given by $i^* := \arg\max_i |\nabla_i f(x_k)|$.

Consider now a square full-rank matrix A, and steepest descent with the norm $\|x\| := \|Ax\|_1$ (that is, the ℓ_1 norm of Ax).

Show that the steepest descent algorithm now chooses a column of A^{-1} for the descent direction. Also specify the rule it uses to choose the column.

Problem 3: (20 pnts) Consider the following problem, which is projecting a point $a \in \mathbb{R}^n$ onto the ℓ_1 ball:

$$\begin{aligned} \min_x \quad & \frac{1}{2}\|x - a\|_2^2 \\ s.t. \quad & \|x\|_1 \leq 1 \end{aligned}$$

(a) Write down the Lagrangian of this problem. Make sure to specify what the dual variable is, and whether it is positive, negative or unconstrained.

(b) Using the KKT conditions, find the optimal x for the primal problem. That is, specify what the optimum x^* is in terms of a.

(Hint: first solve for the best x_i's given a fixed dual variable λ, then find the best λ.)

Problem 4: (20 pnts) Suppose $A_1, \ldots, A_m$ and B are $n \times n$ symmetric matrices. Show that exactly one of the following can hold.

- There exists $x \geq 0$ such that

$$z^\top \left(B - \sum_i A_i x_i \right) z \geq 0, \quad \forall z \in \mathbb{R}^n.$$

- There exist non-positive scalars $\lambda_1, \ldots, \lambda_n \leq 0$ and vectors $y_1, \ldots, y_n \in \mathbb{R}^n$, such that

$$\sum_i \lambda_i y_i^\top B y_i > 0, \quad \text{and} \quad \sum_i \lambda_i y_i^\top A_j y_i \leq 0, \ \forall j.$$

(Hint: Recall that by the Spectral Theorem, any symmetric matrix X can be written as $X = \sum \mu_j v_j v_j^\top$.)

Problem 5: (20 pnts) Recall that in Newton's method, the update becomes:

$$x_{k+1} = x_k - \alpha_k [\nabla^2 f(x_k)]^{-1} \nabla f(x_k).$$

Recall also that, if f is quadratic, the Newton method makes the condition number 1 (i.e. best possible).

Because the inverse Hessian, $\nabla^2 f(x)^{-1}$, may be difficult to compute and invert, we considered various quasi-Newton methods of the form

$$x_{k+1} = x_k - \alpha_k D^{(k)} \nabla f(x_k).$$

One popular choice is to let $D^{(k)}$ be a *diagonal approximation* to the inverse Hessian

$$D^{(k)} = \begin{pmatrix} d_1^{(k)} & & & \\ & d_2^{(k)} & & \\ & & \ddots & \\ & & & d_n^{(k)} \end{pmatrix},$$

where

$$d_i^{(k)} = \left(\frac{\partial^2 f(x_k)}{(\partial x_i)^2} \right)^{-1}.$$

Find an example function f that has a bad condition number that cannot be improved by this diagonal approximation method. Prove this is the case.

(This shows that in the worst case, this method may not do any better than Gradient descent.)

(Hint: think of a quadratic function.)