10-725/36-725 Optimization Midterm Exam

November 6, 2012

NAME:

ANDREW ID:

Instructions:

This exam is 1hr 20mins long.

Except for a single two-sided sheet of notes, no other material or discussion is permitted.

There are 11 one-sided sheets. Please use the other side for rough work.

1 Warm start [Shiva, 25 points]

1.1 Multiple choice [15 points]

Circle the correct answer in each of the following questions.

1. Which of the following functions is convex in $x \in \mathbb{R}^n$?

   (a) $||x||_{1/2}$

   (b) $\sqrt{||x||_2}$

   (c) $\max_j \sqrt{x_j}$

   (d) $\min_i a_i^T x$

   [e] $\log \sum_j \exp(x_j)$

2. What is the normal cone of the nonnegative orthant $\{x : x_i \geq 0\}$ at the origin?

   (a) The nonnegative orthant $\{x : x_i \geq 0\}$

   [b] **The nonpositive orthant $\{x : x_i \leq 0\}$**

   (c) The line $y = x$

   (d) The line $y = -x$

   (e) The zero vector

3. In a nonsingular standard form linear program

$$\min c^T x \quad \text{s.t.} \quad Ax = b$$

   where $x, c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$, how many basic variables are there?

   (a) n

   [b] m

   (c) $n - m$

   (d) $n + m$

   (e) $\max\{n, m\}$

4. If the dual (maximization) LP is feasible and unbounded then the primal (minimization) LP is

    (a) nonsingular

    [b] infeasible

    (c) bounded

    (d) unbounded

    (e) zero

5. Given the general optimization program

$$\min f(x) \text{ s.t. } g(x) \leq 0, \ h(x) = 0$$

consider its Lagrangian $L(x, u, v)$ with u and v introduced for the inequality and equality constraints, respectively. Which of the following is true?

    (a) $L(x^*, u, v) = L(x, u^*, v^*)$ for primal-optimal x^*, dual-optimal (u^*, v^*), and all x, u, v

    (b) L is convex in x

    [c] L is concave in u and v

    (d) $f(x) \geq L(x, u, v)$ for all x, u, v

    (e) $u_i^* \cdot h_i(x) = 0$ at dual-optimal u^* for all i

1.2 Conjugate functions [5 points]

Derive the conjugate function $f^*(y)$ for each of the following functions:

1. $f(x) = 3x^2 + 4x$

    $f^*(y) = \max_x x \cdot y - (3x^2 + 4x)$. By stationarity $y - 6x - 4 = 0 \implies x = \frac{y-4}{6}$. Thus $f^*(y) = \frac{(y-4)^2}{12}$.

2. $f(x) = -\ln x + 2$

    $f^*(y) = \max_x x \cdot y + \ln(x) - 2$. If $y \geq 0$, then clearly $f^*(y) = \infty$. Otherwise, by stationarity $y + \frac{1}{x} = 0 \implies x = -\frac{1}{y}$. Thus $f^*(y) = -3 + \ln(-\frac{1}{y}) = -3 - \ln(y)$

1.3 Matrix differential [5 points]

Let $L = ||Ax - b||_2^2$ where A is a matrix and x and b are vectors. Derive dL in terms of dx.

$L = (Ax - b)^T(Ax - b) = x^T A^T Ax - x^T A^T b - b^T Ax + b^T b$ so $dL = x^T A^T dx - dx^T A^T b - b^T A dx$.

2   First and second order methods [Wooyoung, 20 points]

Consider the function $f(\mathbf{x}) = f(x_1, x_2) = (x_1 + x_2^2)^2$.

(a) [4 points] Derive the gradient of $f(x)$.
   ★SOLUTION:

$$\nabla f(x) = \begin{pmatrix} 2x_1 + 2x_2^2 \\ 4x_1 x_2 + 4x_2^3 \end{pmatrix}.$$

(b) [3 points] At the point $\mathbf{x}_0 = (0, 1)^T$, we consider the search direction $\mathbf{p} = (1, -1)^T$. Show that p is a descent direction.
   ★SOLUTION:
   $\nabla f(\mathbf{x}_0) = (2, 4)^T$.
   $\nabla f(\mathbf{x}_0)^T \mathbf{p} = 2 - 4 = -2 < 0$.
   Therefore, $\mathbf{p}$ is a descent direction at $\mathbf{x}_0$.

(c) [3 points] Find the stepsize α that minimizes $f(\mathbf{x}_0 + \alpha\mathbf{p})$; that is, what is the result of this exact line search? Report $f(\mathbf{x}_0 + \alpha\mathbf{p})$.
   ★SOLUTION:
   We need to find α that minimizes

$$\begin{aligned} f(\mathbf{x}_0 + \alpha\mathbf{p}) &= f((0, 1)^T + (\alpha, -\alpha)^T) = f((\alpha, 1 - \alpha)^T) = (\alpha + (1 - \alpha)^2)^2 \\ &= (\alpha^2 - \alpha + 1)^2 = ((\alpha - \frac{1}{2})^2 + \frac{3}{4})^2 \end{aligned}$$

Minimizing $f(\mathbf{x}_0 + \alpha\mathbf{p})$ with respect to α is equivalent to minimizing $g(\alpha) = (\alpha - \frac{1}{2})^2 + \frac{3}{4}$ since $g(\alpha) > 0$ and $\alpha = \frac{1}{2}$ is the minimizer.

Therefore, $f(\mathbf{x}_0 + \alpha\mathbf{p}) = (\frac{3}{4})^2 = \frac{9}{16}$.

(d) [5 points] Derive the Hessian of $f(x)$.
★SOLUTION:
$$\nabla^2 f(x) = \begin{pmatrix} 2 & 4x_2 \\ 4x_2 & 4x_1 + 12x_2^2 \end{pmatrix}.$$

(e) [5 points] Run one Newton's step with fixed stepsize $t = 1$ starting from $\mathbf{x}_0 = (0, 1)^T$
to compute $\mathbf{x}_1$. Show $\mathbf{x}_1$ and $f(\mathbf{x}_1)$.
Hint: The inverse of a 2×2 matrix is as follows
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

★SOLUTION:

$$\nabla^2 f(\mathbf{x}_0) = \begin{pmatrix} 2 & 4 \\ 4 & 12 \end{pmatrix}.$$

$$\Delta x_{nt} = -(\nabla^2 f(\mathbf{x}_0))^{-1} \nabla f(\mathbf{x}_0) = -\frac{1}{8} \begin{pmatrix} 12 & -4 \\ -4 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

Therefore, $\mathbf{x}_1 = \mathbf{x}_0 + t\Delta x_{nt} = (0, 1)^T + (-1, 0)^T = (-1, 1)^T$ and $f(\mathbf{x}_1) = 0$.

3 Projection and convexity [Aadi, 25 points]

3.1 Projecting onto nullspaces

The nullspace of a matrix A ($\in \mathbb{R}^{m \times n}, m < n$) is defined to be the set of all $x \in \mathbb{R}^n$ such that $Ax = 0$. The projection of a point z onto a convex set S is defined as the point $x^* \in S$ which is closest in Euclidean distance to z. Find a closed form solution for the projection of z onto the convex set $\{x \mid Ax = 0\}$. (You can assume A is full rank, i.e., rank m.) To solve the problem, set it up as a constrained optimization, write out the Lagrangian, and derive the KKT conditions. [13 points]

Solution $\min_x \|x - z\|_2^2$ such that $Ax = 0$. Lagrangian $L(x, \lambda) = \|x - z\|_2^2 + \lambda^\top Ax$. KKT conditions give $Ax^* = 0$ and $x^* - z + A^\top \lambda^* = 0$. The second condition (on multiplying by A) yields $Az = AA^\top \lambda^*$, implying $\lambda^* = (AA^\top)^{-1}Az$, yielding $x^* = z - A^\top(AA^\top)^{-1}Az$.

3.2 Understanding convexity

Consider the five functions x, x^2, x^3, x^4, x^5. Which of these functions are convex on $\mathbb{R}$? Which are strictly convex on $\mathbb{R}$? Which are strongly convex on $\mathbb{R}$? Which are strongly convex on $[0.5, 4.5]$? NO EXPLANATIONS REQUIRED! [12 points]

Convex: [3 pts] x, x^2, x^4

Strictly convex: [3 pts] x^2, x^4

Strongly convex: [3 pts] x^2

Strongly convex on $[0.5, 4.5]$: [3 pts] x^2, x^3, x^4, x^5

4 The Dantzig Selector [Kevin, 25 Points]

The Dantzig selector is another regression technique for when the problem dimension is larger than the number of data points, like LASSO. That is, we wish to learn a sparse $w \in \mathbb{R}^p$ such that $y \approx f(x) = w^T x$.

Let X is an n by p matrix of n examples, y is an n-dimensional vector of targets, and $\epsilon > 0$ is a supplied constant. The Dantzig selector is a solution to the following convex optimization problem:

$$\min_{w} \ \|w\|_1 \quad \text{subject to:} \tag{1}$$
$$\|X^T(y - Xw)\|_\infty \le \epsilon \tag{2}$$

The $L1$-norm objective attempts to keep the weight vector w sparse, like in the LASSO optimization. The constraints impose that no feature has high dot product with the residual. That is, either the prediction error is low, or no single feature can explain the remaining error.

Let's reformulate the optimization problem as a linear program by expanding the norms and introducing an auxilary vector $t \in \mathbb{R}^p$. Here, 1 is the vector of all ones.

$$\min_{w,t} \ 1^T t, \qquad \text{subject to:} \tag{3}$$
$$X^T(y - Xw) \le \epsilon 1, \tag{4}$$
$$-X^T(y - Xw) \le \epsilon 1, \tag{5}$$
$$w \le t, \tag{6}$$
$$-w \le t. \tag{7}$$

(a) [6 points] Write the dual linear program of the reformulated program. Use $\alpha, \beta, \gamma, \delta \ge 0$ as your dual variables for the four sets of constraints (in that order).

Solution:

$$\max_{\alpha,\beta,\gamma,\delta \ge 0} \ y^T X(\alpha - \beta) - \epsilon 1^T(\alpha + \beta) \qquad \text{subject to:} \tag{8}$$
$$X^T X(\alpha - \beta) = \gamma - \delta, \tag{9}$$
$$\delta + \gamma = 1. \tag{10}$$

(b) [6 points] Rewrite this dual program by replacing $\alpha - \beta = q$ and $\gamma - \delta = z$. Your program will no longer be a linear program. You do not need to justify your answer.
Hint 1: Given the constraint $z_i = \gamma_i - \delta_i$ and the constraints you derived on γ_i and δ_i, what is the largest that z_i could be? What is the smallest that z_i could be?
Hint 2: Think about the norms and their duals in the original program.

Solution:

$$\max_{q,z} \ y^T X q - \epsilon \|q\|_1 \qquad \text{subject to:} \tag{11}$$

$$X^T X q = z, \tag{12}$$

$$\|z\|_\infty \le 1. \tag{13}$$

(c) [1 points] You should have one remaining linear constraint involving $X^T X$. Use it to eliminate z.

Solution:

$$\max_{q} \ y^T X q - \epsilon \|q\|_1 \qquad \text{subject to:} \tag{14}$$

$$\|X^T X q\|_\infty \le 1. \tag{15}$$

Let $\tilde{w}(\epsilon)$ be a LASSO solution with regularization parameter ϵ,

$$\tilde{w}(\epsilon) = \arg\min_{w} \|y - Xw\|^2/2 + \epsilon \|w\|_1 \tag{16}$$

and write $w(\epsilon)$ for the solution to the Dantzig selector program with the same value of ϵ. We will show how to relate $\tilde{w}(\epsilon)$ to $w(\epsilon)$.

(d) [6 points] Write the KKT (i.e., first-order optimality) conditions for the LASSO optimization at $\tilde{w}(\epsilon)$.

Solution:

$$\frac{X^T(y - X\tilde{w}(\epsilon))}{\epsilon} \in \partial \|\tilde{w}(\epsilon)\|_1 \tag{17}$$

(e) [6 points] Using the KKT conditions, show that $\tilde{w}(\epsilon)$ is a feasible point for the Dantzig selector optimization with parameter ϵ.

Solution:

For all $j \in \{1, 2, \ldots, p\}$,

$$\frac{|X_j^T(y - X\tilde{w}(\epsilon))|}{\epsilon} \le 1 \tag{18}$$

$$|X_j^T(y - X\tilde{w}(\epsilon))| \le \epsilon \tag{19}$$

$$\max_{j} |X_j^T(y - X\tilde{w}(\epsilon))| \le \epsilon \tag{20}$$

$$\|X^T(y - X\tilde{w}(\epsilon))\|_\infty \le \epsilon \tag{21}$$

$$\tag{22}$$

We've now shown that the Dantzig selector finds solutions whose $L1$-norm is at least as small as the LASSO solution.

GRADING

Shiva:

Wooyoung:

Aadi:

Kevin:

TOTAL: / 95