MS&E 111X & 211X
Introduction to Optimization (Accelerated)
Homework 2
Course Instructor: Yinyu Ye
Due Date: 11:59 pm Oct 7, 2021

Please submit your homework through Gradescope. If you haven't already been added to Gradescope, you can use the entry code **2RJNKV** to join. Please note: late homework will not be accepted. Each problem will be graded out of 10 points.
Some problems allow group work. Groups should be no larger than 4. If you decide to work together, provide the names of those you worked with.

Problem 1 Convex sets and convex functions

Identify whether the following sets are convex.

(a) $F := \{x \in R^n : \; Ax = b, \; x \geq 0\}$, where data matrix $A \in R^{m \times n}$ and vector $b \in R^m$.

(b) Fix data matrix A and consider the b-data set for F defined above:

$$B := \{b \in R^m \;\; : \; F \text{ is not empty}\}$$

(c) $\{x : x^2 \geq 1\}$.

(d) $\{x : x^2 \leq 1\}$.

Identify whether the following functions are convex.

(e) *Negative entropy*

$$h(p) = \sum_{i=1}^{n} p_i log(p_i)$$

On $\{p \in R^n, p_i > 0 \; \forall \; i, e^T p = 1\}$, where $e \in R^n$ is the vector of ones.

(f) *Floor function*

$$f(x) = \lfloor x \rfloor = \max\{k \leq x, k \text{ is an integer}\}$$

(g) *Sum of largest components*

$$f(x) = \sum_{i=1}^{k} x_{(i)}$$

where $x \in R^n$, k is an integer between 1 and n and $x_{(i)}$ denote the i^{th} largest element of the vector x.

Solution

(a) Yes. Take any two points $x',\ x'' \in F$, that is, $Ax' = b$, $x' \geq 0$ and $Ax'' = b$, $x'' \geq 0$. Then, for any $\alpha \in [0, 1]$, we must have

$$ax' + (1 - \alpha)x'' \geq 0$$

Moreover,

$$A(\alpha x' + (1 - \alpha)x'') = \alpha Ax' + (1 - \alpha)Ax'' = \alpha b + (1 - \alpha)b = b$$

Thus, $\alpha x' + (1 - \alpha)x'' \in F$

(b) Yes. Take any two points $b, b'' \in B$. Then we must have $x' \geq 0$ and $x'' \geq 0$ such that $Ax' = b'$ and $Ax'' = b''$. Now we like to prove that the convex combination $\alpha b' + (1 - \alpha)b''$ is also in B. Consider the convex combination $x = \alpha x' + (1 - \alpha)x''$. Obviously, $x \geq 0$. Further more,

$$Ax = A(\alpha x' + (1 - \alpha)x'') = \alpha Ax' + (1 - \alpha)Ax'' = \alpha b' + (1 - \alpha)b''$$

which give the desired proof.

(c) No. Consider $x = 1$ and $x = -1$, the convex combination is not in the set.

(d) Yes. Level set of convex function is convex.

(e) Yes. Take the second order derivative, one have

$$f(p_i) = p_i log(p_i), \quad \nabla^2 f(p_i) = \frac{1}{p_i} > 0$$

And $h(p) = \sum_i f(p_i)$ is the sum of convex functions, which keeps convexity.

(f) No. Consider $x = 0.5$, $y = 1.5$ and $\alpha = 0.5$. The above instance does not satisfy definition on convexity.

(g) Yes. The problem could be formulate as

$$f(x) = \max_{p} \ x^T p$$

$$s.t. \quad e^T p = k, \quad 0 \le p \le e$$

Hence the $f(x)$ is convex in x. To see this, note

$$f(\alpha x + (1-\alpha) y) = \max_{p,\ e^T p = k, 0 \le p \le e} (\alpha x + (1-\alpha) y) p$$

$$\le \alpha \max_{p,\ e^T p = k, 0 \le p \le e} x^T p \quad + \quad (1-\alpha) \max_{p,\ e^T p = k, 0 \le p \le e} y^T p = \alpha f(x) + (1-\alpha) f(y)$$

Problem 2

Consider the two-variable linear program with 6 inequaltiy constraints:

$$\max \ 3x_1 + 5x_2$$

$$s.t. \quad x_1 \ge 0$$

$$x_2 \ge 0$$

$$-x_1 + x_2 \le 2.5$$

$$x_1 + 2x_2 \le 9$$

$$x_1 \le 4$$

$$x_2 \le 3$$

(a) Plot the lines $x_j = 0,\ j = 1, \dots 6$ (all on the same two-dimensional graph of x_1 and x_2) where for $j = 3, 4, 5, 6,\ x_j$ denotes the slack variable in j^{th} constraint in the LP standard form .

(b) Identify the extreme points of the feasible region as intersections of suitable lines $x_j = 0$.

(c) For each pair of adjacent extreme points of the feasible region, describe how each direction of the edge between them can be generated by increasing the value of a single variable chosen from $\{x_1, x_2, x_3, x_4, x_5, x_6\}$.

Solution (a) Omitted

(b)

Extreme points	Defining Equations
$O:(0,0)$	$x_1 = x_2 = 0$
$A:(0,2.5)$	$x_1 = x_3 = 0$
$B:(0.5,3)$	$x_3 = x_6 = 0$
$C:(3,3)$	$x_4 = x_6 = 0$
$D:(4,2.5)$	$x_4 = x_5 = 0$
$E:(4,0)$	$x_2 = x_5 = 0$

(c)

From	To	Increasing
O	A	x_2
A	B	x_1
B	C	x_3
C	D	x_6
D	E	x_4
E	O	x_5
O	E	x_1
E	D	x_2
D	C	x_5
C	B	x_4
B	A	x_6
A	O	x_3

Problem 3 Phase One Problem

Consider a system of m linear equations in n non-negative variables, say

$$Ax = b, \quad x \geq 0$$

Assume the right-hand side vector b is non-negative. Now consider the related linear program

$$\min \quad e^T y$$

$$s.t. \quad Ax + Iy = b$$

$$x \geq 0, y \geq 0$$

where e is the vector of all ones, and I is the $m \times m$ identity matrix. This linear program is called a Phase One problem.

(a) Write the Lagrange function of the Phase One problem and the dual of Phase One Problem.

(b) Write the complementary slackness conditions for the Phase One problem.

(c) True of False Questions

1. The Phase One problem always has a basic feasible solution.

2. The Phase One Problem always has an optimal solution.

3. $\{x : \ Ax = b, \quad x \geq 0\} \neq \emptyset$ if and only if the optimal value of the objective function in the corresponding Phase One problem is zero.

Solution

(a)

$$L(x, y, u)_{x\geq 0, y\geq 0} = e^T y - \mu^T (Ax + Iy - b)$$

$$\max \ b^T \mu$$

$$s.t. \ A^T \mu \leq 0$$

$$\mu \leq e$$

(b) Primal complementary Slackness:

$$\mu_i (Ax + Iy - b)_i = 0 \quad \forall \ i$$

Dual complementary slackness:

$$x_j (-A^T \mu)_j = 0 \quad \forall \ j = 1, \ldots, n$$

$$y_i (1 - \mu_i) = 0 \quad \forall i = 1, \ldots, m$$

(c) (1) True. $[x; y] = [0; b]$ is a basic solution to the Phase One Problem; since b is non-negative by assumption, it is also a feasible solution.

(2) True. Since the Phase One problem is feasible, and its objective value is bounded from below by 0, the problem always has an optimal solution

(3) True. If the optimal value of the Phase One problem is zero, then we must have also the optimal solution $(x \geq 0, y = 0)$ and that $Ax = b$,that is, $\{x \ : \ Ax = b, \ x \geq 0\} \neq \emptyset$. Conversely, if $\{x \ : \ Ax = b, \ x \geq 0\} \neq \emptyset$, then for any x satisfies $Ax = b, x \geq 0$ (we know such x exists as the set is non-empty), $[x; y] = [x; 0]$ is an optimal solution to the Phase One Problem with optimal value 0 (it is feasible as $Ax + Iy = b$, and $x \geq 0, y \geq 0$, with objective value 0 and no other solution can achieve a lower value. Note $e^T y \geq 0$ for all $y \geq 0$, so no other solution can achieve a lower value.)

Problem 4

Consider the following problem for a parameter $\kappa > 0$:

$$\min\ (x_1 - 1)^2 + x_2^2$$

$$s.t.\ -x_1 + \frac{x_2^2}{\kappa} \geq 0$$

(a) Is $x = 0$ a first order KKT solution?

(b) Is $x = 0$ a second order KKT solution for some value of κ?

Solution

Define $f(x) = (x_1 - 1)^2 + x_2^2$, $c(x) = -x_1 + \frac{x_2^2}{\kappa}$. Then the Lagrangian function for this problem is

$$L(x, y) = f(x) - yc(x) = (x_1 - 1)^2 + x_2^2 - y(-x_1 + \frac{x_2^2}{\kappa}),\ y \geq 0$$

(a) Firstly, $x = 0$ is feasible with $c(x) = 0$, moreover,

$$\nabla f(0) = (-2; 0), \quad \nabla c(0) = (-1; 0)$$

Thus $y = 2$ makes $\nabla f(0) = 2\nabla c(0)$ so that $x = 0$ is a first order KKT solution.

(b) The hessian is given by

$$H = \begin{pmatrix} \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} \\ \frac{\partial^2 L}{\partial x_2 \partial x_1} & \frac{\partial^2 L}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 - \frac{2\lambda}{\kappa} \end{pmatrix}$$

We then need to have $2 - \frac{2\lambda}{\kappa} \geq 0 \Rightarrow k \geq 2$

Problem 5

Consider the Support Vector Machine problem

$$\min\ \beta + \mu\|x\|^2$$

$$s.t.\ a_i^T x + x_0 + \beta \geq 1,\ \forall i$$

$$b_j^T x + x_0 - \beta \leq -1,\ \forall j$$

$$\beta \geq 0$$

with $\beta, x_0 \in$ and $x = [x_1; \ldots; x_n], a_i, b_j \in R^n$ for all i, j, and $\|x\|^2 = \sum_{i=1}^n x_i^2$

(a) Write out the Lagrangian function of SVM problem.

(b) Suppose that we have 6 training data in R^2: $a_1 = (0; 0)$, $a_2 = (1; 0)$, $a_3 = (0; 1)$, $b_1 = (0; 0)$, $b_2 = (-1; 0)$, $b_3 = (0; -1)$. Use the optimality conditions to verify whether the following solutions is optimal.

(1) for $\mu = 0$, $x = [1; 1]$, $x_0 = 0$, $\beta = 1$

(2) For $\mu = 10$, $x = [1; 1]$, $x_0 = 0$, $\beta = 1$

(c) Use solver to find the optimal solution for $\mu = 10$.

Bonus question: When $\mu = 0$, is the optimal solution unique? When $\mu = 10$, is the optimal solution unique? Prove your claim.

Solution

(a) Let the multipliers for a_i constraints be $y_i^a \geq 0$ and those for b_j constraints be $y_j^b \leq 0$ and $\beta \geq 0$ be $y^\beta \geq 0$. Then the Lagrangian function is

$$L(x, x_0, \beta, y^a, y^b, y^\beta)_{y_i^a, y_j^b, y^\beta, \beta \geq 0} = \beta + \mu\|x\|^2 - \sum_i y_i^a(a_i^T x + x_0 + \beta - 1) - \sum_j y_j^b(-b_j^T x - x_0 + \beta - 1)$$

1) Yes. For $\mu = 0$, any point of the form $\beta = 1$, $x = (t; t)$,$x_0 = 0$ with $t \geq 0$ is optimal, as the objective value is 1 and the constraints are satisfied. So the optimal solution is not unique.

2) No. For $\mu > 0$, a point is optimal if and only if $\beta = 1$ and $x = 0$. Hence we obtain a unique optimal solution $\beta = 1$,$x = 0$ and $x_0 = 0$.

(c) *Omitted*

Bonus Question: Firstly, we show that for the set of a_i,b_j given in this problem, any feasible β satisfies $\beta \geq 1$. To see this, suppose on the contrary that $\beta < 1$. Then for $a_1 = b_1$, we have

$$a_1^T x + x_0 \geq 1 - \beta > 0 > -1 + \beta \geq b_1^T x + x_0$$

which is a contradiction. Hence the optimal value $\beta + \mu\|x\|^2$ of the primal objective function is at least 1. Moreover, it can always be achieved by simply setting $\beta = 1$, $x = 0$ and $x_0 = 0$. Hence we know that the optimal value is always 1 no matter whether $\mu = 0$ or not. Then, we know for $\mu > 0$ the optimal solution is always unique, as for any $x \neq 0$, we could deviate to $x = 0$ and strictly decrease the objective value. For $\mu = 0$, the optimal solution is not unique. both $x = [0;0], x_0 = 0, \beta = 1$ and $x = [1;1], x_0 = 0, \beta = 1$ are optimal solutions.

Problem 6

In logistic regression, we determine x_0 and x by maximizing

$$\left(\Pi_{i,c_i=1} \frac{1}{1+\exp(-a_i^T x - x_0)}\right)\left(\Pi_{i,c_i=-1} \frac{1}{1+\exp(b_i^T x + x_0)}\right)$$

which is equivalent to minimize the log-likelihood loss

$$\sum_{i,c_i=1} log(1+\exp(-a_i^T x_i - x_0)) + \sum_{i,c_i=-1} log(1+\exp(b_i^T x + x_0))$$

Suppose that we have 6 training data, with $a_1 = (0;0)$, $a_2 = (1;0)$, $a_3 = (0;1)$, $b_1 = (0;0)$, $b_2 = (-1;0)$, $b_3 = (0;-1)$, where a_i are points we label as spam and b_i are points we label as non-spam. (You may view a_i has label $c_i = 1$ and b_i has label $c_i = -1$).

(a) What is the KKT Condition for logistic regression?

(b) Use the KKT condition to verify the whether the following solution is optimal: $x = [0;0]$ and $x_0 = 0$.

(c) Typically, we add a regularization to the objective

$$\min \sum_{i,c_i=1} log(1+\exp(-a_i^T x_i - x_0)) + \sum_{i,c_i=-1} log(1+\exp(b_i^T x + x_0)) + \mu \sum_{i=0}^{n} x_i^2$$

what is the KKT Condition for the regularized logistic regression?

Bonus Teamwork : solve the problem using any non-linear programming solver for the non-regularized logistic regression.

Solution.
(a)

$$\sum_{i,c_i=1} \frac{exp(-a_i^T x_i - x_0)}{1+\exp(-a_i^T x - x_0)}[-a_i;-1] + \sum_{i,c_i=-1} \frac{exp(b_i^T x + x_0)}{1+exp(b_i^T x + x_0)}[b_i;1] = [0;\ldots;0]$$

(b) plug in $x = [0; 0]$ and $x_0 = 0$ one have

$$\frac{1}{2}([0;0;-1] + [-1;0;-1] + [0;-1;-1]) + \frac{1}{2}([0;0;-1] + [-1;0;-1] + [0;-1;-1]) \neq [0;0;0]$$

so $x = [0; 0]$ and $x_0 = 0$ is not an optimal solution.

(c)

$$\sum_{i,c_i=1} \frac{1}{1+\exp(a_i^T x + x_0)}[-a_i;-1] + \sum_{i,c_i=-1} \frac{1}{1+exp(-b_i^T x - x_0)}[b_i;1] + 2[x_1;\ldots;x_n;x_0] = 0$$

(d) *Omitted*