MS&E 111X & 211X
Introduction to Optimization (Accelerated)
Homework 4
Course Instructor: Yinyu Ye
Due Date: 11:59 pm Nov 30, 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

For parts a)-c) below, label them as True or False. If true, provide a short reason; if false, provide reasoning or a counter example.

a) *True or False:* The simplex method (with cycle breaking rules) for a general linear program with n variables always converges to the optimal solution after a finite amount of steps and it takes at most a polynomial number of steps in n to converge.

b) Consider the following optimization problem:

$$\min f(x_1, x_2) = -\log(x_1 x_2) + x_1^2 + x_2 + 3(x_1 - x_2)^4.$$

True or False: The gradient descent procedure with line search on this problem is guaranteed to converge to the global minimum.

c) Consider the following optimization problem:

$$\min f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 2)^2.$$

True or False: Newton's method on this problem will converge to the global minimum faster than the gradient descent method provided that we use line search.

Solution:

a) False. In the worst case, the simplex method can take an exponential number of steps to reach the optimal solution.

b) True. The gradient descent procedure converges in this case since the objective function is convex. It will converge to the global minimum.

c) False. Since the given function is quadratic, we know Newton's method will converge in exactly 1 step. The given function also represents a circle with center at (1, 2). The gradient at any point on the level set of this objective function, will pass through the center of the circle. The center of the circle is the global minimum. Thus, by choosing an appropriate step size in line search, the gradient method will also converge in exactly 1 step. Thus, the rate of convergence of Newton's method will be equal to that of the gradient method.

Problem 2

Solve the transportation problem described in pages 12-14, Lecture 10 using the simplex method. Specifically, write down the shadow prices, reduced costs, the termination test and the new BFS for each iteration. Clearly state your optimal solution at the end. [Hint: You can begin with the BFS provided in page 21, Lecture 10.]

Solution:
The transportation problem is

$$\begin{aligned}
\min \quad & \sum_i \sum_j c_{ij} x_{ij} \\
\text{subject to} \quad & \sum_j x_{ij} = s_i, \quad \forall i = 1, \ldots, m \\
& \sum_i x_{ij} = d_j, \quad \forall j = 1, \ldots, n \\
& x_{ij} \geq 0, \quad \forall i, j
\end{aligned}$$

where $m = 3$, $n = 4$, and

$$c = \begin{bmatrix} 12 & 13 & 4 & 6 \\ 6 & 4 & 10 & 11 \\ 10 & 9 & 12 & 4 \end{bmatrix}, \quad s = \begin{bmatrix} 500 \\ 700 \\ 800 \end{bmatrix}, \quad d = \begin{bmatrix} 400 \\ 900 \\ 200 \\ 500 \end{bmatrix}.$$

We begin the initial BFS found in the lecture notes:

$$\begin{aligned}
x_{11} &= 400 \\
x_{12} &= 100 \\
x_{22} &= 700 \\
x_{32} &= 100 \\
x_{33} &= 200 \\
x_{34} &= 500
\end{aligned}$$

and $x_{ij} = 0$ elsewhere. After running the simplex algorithm for two iterations, you should find the optimal solution

$$\begin{aligned} x_{11} &= 300 \\ x_{13} &= 200 \\ x_{22} &= 700 \\ x_{31} &= 100 \\ x_{32} &= 200 \\ x_{34} &= 500 \end{aligned}$$

and $x_{ij} = 0$ elsewhere.

Problem 3

Recall the logistic regression problem in Homework 2 Problem 6 whose objective is to minimize the function

$$f(\boldsymbol{x}, x_0) = \sum_i \log(1 + \exp(-\boldsymbol{a}_i^T \boldsymbol{x} - x_0)) + \sum_i \log(1 + \exp(\boldsymbol{b}_i^T \boldsymbol{x} + x_0))$$

with training data $\boldsymbol{a}_1 = (0; 0)$, $\boldsymbol{a}_2 = (1; 0)$, $\boldsymbol{a}_3 = (0; 1)$, $\boldsymbol{b}_1 = (0; 0)$, $\boldsymbol{b}_2 = (-1; 0)$, $\boldsymbol{b}_3 = (0; -1)$, where $\boldsymbol{a}_i$ are points we negatively label and $\boldsymbol{b}_i$ are points we positively label. We consider numerically solving this problem using gradient descent. You are free to use whatever programming language you prefer, but you cannot use any off-the-shelf implementation such as `scipy.optimize` in python. You may work in teams for this problem.

a) We fix step size as 0.1 and we stop iterating at the earliest round when the Euclidean norm of gradient drops below 10^{-3} or we reach 1000th round. Write code for fixed step size gradient descent for the problem in Homework 2 Problem 6. You are free to choose your starting point of $(\boldsymbol{x}^{(0)}, x_0^{(0)})$. After running the algorithm, please produce and report the following 2 plots:

   1) horizontal axis: the iterations k, vertical axis: Euclidean norm of $[\boldsymbol{x}; x_0]$
   2) horizontal axis: the iterations k, vertical axis: the value of negative log-likelihood function $f(\boldsymbol{x}^{(k)}, x_0^{(k)})$;

b) Modify your code to add the regularization proposed by Homework 2 Problem 6 Part (c), which is

$$\min \quad f(\boldsymbol{x}, x_0) + \mu ||\boldsymbol{x}||^2.$$

   Produce the same plots as Part (a) with $\mu_1 = 0.01$ and $\mu_2 = 1$ (i.e. please report 4 plots for this subsection, 2 plots for each regularization scale).

c) Compare the results in Parts (a) and (b) and summarize what effects of regularization you notice.

d) Attach/print your code for implementations of this problem.

Solution:

a) See Figures 1 and 2. We chose initial point $(\boldsymbol{x}^{(0)}, x_0^{(0)}) = (0, 0, 0)$

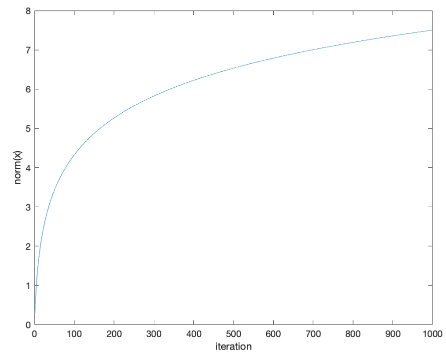

Figure 1: Norm of gradient descent iterates on unregularized problem

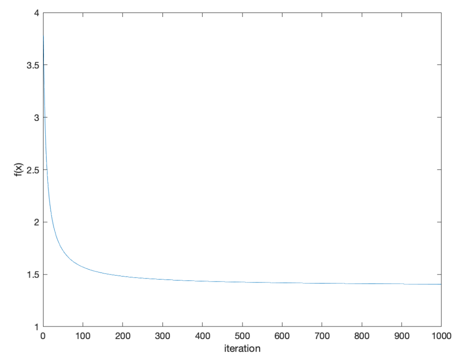

Figure 2: Unregularized objective value

b) See Figures 3, 4, 5, 6. We chose initial point $(\boldsymbol{x}^{(0)}, x_0^{(0)}) = (0, 0, 0)$

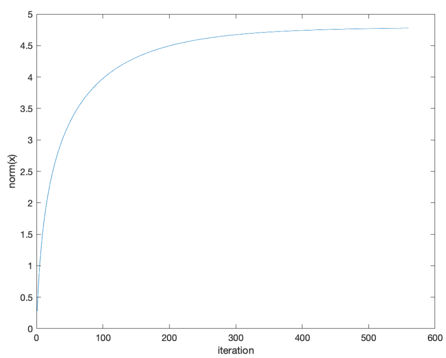

Figure 3: Norm of gradient descent iterates on regularized problem, $\mu_1 = 0.01$

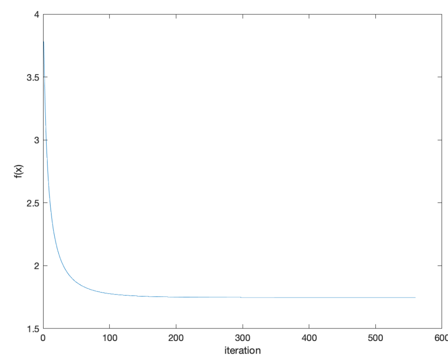

Figure 4: Regularized objective value, $\mu_1 = 0.01$

c) Gradient descent on the unregularized problem only terminates because it meets the maximum iterate, not because the norm of its gradient reaches the threshold. Even though the objective plateaus near the minimum, the norm of the solution keeps increasing without bound. The regularized problem with $\mu = 0.01$ reaches the termination criteria before the maximum iterate with a minimum objective value comparable to the unregularized optimal objective

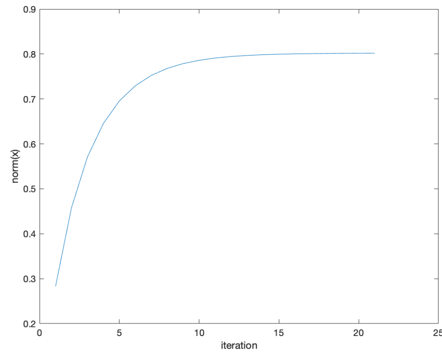

Figure 5: Norm of gradient descent iterates on regularized problem, $\mu_1 = 1$

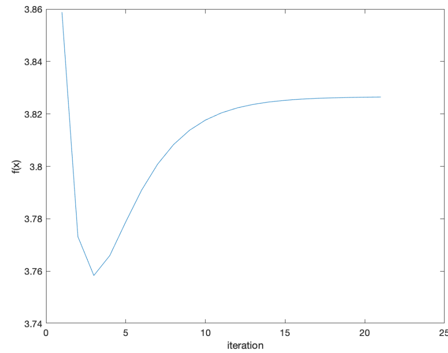

Figure 6: Regularized objective value, $\mu_1 = 1$

and also has a plateauing solution norm. Likewise, the regularized problem with $\mu = 1$ terminates early, however its objective is far from the optimum and in fact it increases after a few iterations. This shows that moderate regularization helps in controlling the size of the solution and in achieving faster convergence, but high regularization can be more detrimental than no regularization.

d) Omitted.

Problem 4

Consider the following variant of the portfolio management quadratic program in Lecture 12 Slide 16:

$$\begin{aligned} \min \quad & x_1^2 + 2x_2^2 + 3x_3^2 - x_1x_2 - x_2x_3 - x_1x_3 - x_1 - 2x_2 - 3x_3 \\ \text{subject to} \quad & x_1 + x_2 + x_3 = 1 \end{aligned}$$

We consider numerically solving this problem using Newton's method. You are free to use whatever programming language you prefer, but you cannot use any off-the-shelf implementation such as `scipy.optimize` in python. Write code to perform Newton's method and report your final (x_1, x_2, x_3).

Solution:
In standard quadratic form, this problem is

$$\begin{aligned} \min \quad & \frac{1}{2}x^TQx + c^Tx \\ \text{subject to} \quad & Ax = b \end{aligned}$$

where

$$Q = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 4 & -1 \\ -1 & -1 & 6 \end{pmatrix}, \quad c = \begin{pmatrix} -1 \\ -2 \\ -3 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \end{pmatrix}.$$

Therefore, the Newton update is

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \\ y' \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ y \end{pmatrix} - \nabla g(x_1, x_2, x_3, y)^{-1} g(x_1, x_2, x_3, y)$$

where

$$\begin{aligned} g(x_1, x_2, x_3, y) &= \begin{pmatrix} Q & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ y \end{pmatrix} + \begin{pmatrix} c \\ -b \end{pmatrix} \\ \nabla g(x_1, x_2, x_3, y) &= \begin{pmatrix} Q & -A^T \\ A & 0 \end{pmatrix}. \end{aligned}$$

Plugging in, we see that the initial point does not matter and we get the optimal solution with a single Newton update

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \\ y' \end{pmatrix} = -\begin{pmatrix} Q & -A^T \\ A & 0 \end{pmatrix}^{-1} \begin{pmatrix} c \\ -b \end{pmatrix} = \begin{pmatrix} 0.2535 \\ 0.3521 \\ 0.3944 \\ -1.2394 \end{pmatrix}.$$

Problem 5

Consider the LP problem

$$\begin{array}{lll} \text{minimize} & x_1 + x_2 & \\ \text{subject to} & x_1 + x_2 + x_3 & = 1, \\ & (x_1, x_2, x_3) & \geq 0. \end{array}$$

a) Formulate the barrierred problem using a logarithmic barrier and derive the KKT optimality conditions for the barrierred problem. Resolve the optimal solution in terms of the barrier parameter $\mu : x_1^*(\mu)$, $x_2^*(\mu)$ and $x_3^*(\mu)$.

b) The trajectory defined by $(x_1^*, x_2^*, x_3^*)(\mu)$ is known as the central path to the optimum of the problem. Find the location of 3 points on this path by evaluating the solution to the barrierred problem for $\mu \to \infty$, $\mu = 1$, and $\mu \to 0$. Show that the solution obtained with $\mu \to 0$ is the optimal solution of the original problem using LP duality.

Solution:

a) The barrierred problem takes the following form:

$$\begin{aligned} \min \quad & x_1 + x_2 - \mu(\log x_1 + \log x_2 + \log x_3) \\ \text{subject to} \quad & x_1 + x_2 + x_3 = 1 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

where $\mu \geq 0$. Deriving the KKT conditions, we get that a point on the central path must satisfy:

$$\begin{aligned} 1 - \frac{\mu}{x_1} + \lambda = 0 \\ 1 - \frac{\mu}{x_2} + \lambda = 0 \\ -\frac{\mu}{x_3} + \lambda = 0 \\ x_1 + x_2 + x_3 = 1 \end{aligned}$$

This implies the following forms for x:

$$\begin{aligned} x_1^* &= \frac{\mu}{1+\lambda} \\ x_2^* &= \frac{\mu}{1+\lambda} \\ x_3 &= \frac{\mu}{\lambda} \end{aligned}$$

and, plugging these into the constraint $x_1^* + x_2^* + x_3^* = 1$, we get

$$\lambda = \frac{(3\mu - 1) + \sqrt{9\mu^2 - 2\mu + 1}}{2}$$

because $x_3 \geq 0$ enforces that $\lambda > 0$. Plugging back to (x_1^*, x_2^*, x_3^*) we will get $(x_1^*(\mu), x_2^*(\mu), x_3^*(\mu))$.

b) As $\mu \to \infty$, $(x_1^*, x_2^*, x_3^*) = (1/3, 1/3, 1/3)$.
As $\mu = 1$, $(x_1^*, x_2^*, x_3^*) = (\frac{2-\sqrt{2}}{2}, \frac{2-\sqrt{2}}{2}, \sqrt{2} - 1)$.
As $\mu \to 0$, $(x_1^*, x_2^*, x_3^*) = (0, 0, 1)$.
Note that the primal objective using $(x_1, x_2, x_3) = (0, 0, 1)$ is 0. Since the dual problem is

$$\begin{aligned} \max \quad & z \\ \text{subject to} \quad & z \leq 1 \\ & z \leq 1 \\ & z \leq 0 \end{aligned}$$

whose optimal solution is obviously $z = 0$ with objective value 0. Therefore, we have zero duality gap with $(x_1, x_2, x_3, z) = (0, 0, 1, 0)$. Hence the central path leads to the optimal solution.