Solution Set #1

1. (Constrained optimization – 10 points) *Consider a constrained minimization problem given as*

$$\min_{(x,y):\, y+2x\geq 30} f(x,y)$$

where $f(x,y) = 5(x+y)^2 + (x-y)^2$.

(a) *Find a local minimum using the KKT condition.*

Solution

Let $h(x,y) = -2x - y + 30$. We can easily see all $(x,y) \in \mathbb{R}^2$ is regular. Therefore every candidate of local minimum satisfies the KKT condition. The generalized Lagrangian is given by

$$\begin{aligned} L(x,y,\mu) &= f(x,y) + \mu h(x,y) \\ &= 5(x+y)^2 + (x-y)^2 + \mu(-2x - y + 30) \end{aligned}$$

Then, the necessary condition yields

1. $\dfrac{\partial L(x,y,\mu)}{\partial x} = 10(x+y) + 2(x-y) - 2\mu = 12x + 8y - 2\mu = 0$
2. $\dfrac{\partial L(x,y,\mu)}{\partial y} = 10(x+y) - 2(x-y) - \mu = 8x + 12y - \mu = 0$
3. $\mu(-2x - y + 30) = 0$ or equivalently, $\mu = 0$ or $2x + y = 30$
4. $\mu \geq 0$
5. $y + 2x \geq 30$

Here the first two are from gradient of generalized Lagrangian and the third is given by complementary slackness. For $\mu = 0$, we have $x = y = 0$, but then $h(x,y) = 30 > 0$, which violates the constraint. For $2x + y = 30$, we have $\mu^* = \dfrac{600}{7}$ and $(x^*, y^*) = \left(\dfrac{120}{7}, -\dfrac{30}{7}\right)$ is the only candidate for a local minimum.

Finally, in order to verify that it is a local minimum of f, we use the second order sufficient KKT conditions. For homework, you do not need to provide this analysis. This is only for your reference. The second sufficiency condition is satisfied if $\mathbf{z}^T \nabla^2_{(xx)} L(x^*, y^*, \mu^*)\mathbf{z} > 0$ for all $\mathbf{z} = (z_1, z_2)^T \neq 0$ such that $\nabla h(x,y)^T \mathbf{z} = -2z_1 - z_2 = 0$. In this problem,

$$\mathbf{z}^T \nabla^2_{(xx)} L(x^*, y^*, \mu^*)\mathbf{z} = \mathbf{z}^T \begin{bmatrix} 12 & 8 \\ 8 & 12 \end{bmatrix} \mathbf{z} = 144z_1^2 - 192z_1 z_2 + 144z_2^2 = 1104z_1^2$$

since $z_2 = -2z_1$. Since $\mathbf{z} \neq 0$, it is greater than 0. Therefore, $(x^*, y^*) = \left(\dfrac{120}{7}, -\dfrac{30}{7}\right)$ is a strict local minimum of f. This is also the global minimum since there is only one local minimum.

(b) *Assume the constraint is changed to $y + 2x \leq 30$ and find a local minimum using the KKT condition.*

Solution
Let $h(x, y) = 2x + y - 30$. It is easy to see all $(x, y) \in \mathbb{R}^2$ is regular. Therefore every candidate of a local minimum satisfies the KKT condition. The generalized Lagrangian is given by

$$\begin{aligned} L(x, y, \mu) &= f(x, y) + \mu h(x, y) \\ &= 5(x + y)^2 + (x - y)^2 + \mu(2x + y - 30) \end{aligned}$$

Then, the necessary condition yields

1. $\dfrac{\partial L(x, y, \mu)}{\partial x} = 10(x + y) + 2(x - y) + 2\mu = 12x + 8y + 2\mu = 0$
2. $\dfrac{\partial L(x, y, \mu)}{\partial y} = 10(x + y) - 2(x - y) + \mu = 8x + 12y + \mu = 0$
3. $\mu(2x + y - 30) = 0$ or equivalently, $\mu = 0$ or $2x + y = 30$
4. $\mu \geq 0$
5. $y + 2x \leq 30$

For $\mu = 0$, we have $x = y = 0$. Different from (a), $(x^*, y^*) = (0, 0)$ does not violate the constraint. For $2x + y = 30$, we have $\mu^* = -\dfrac{600}{7}$ and $(x^*, y^*) = \left(\dfrac{120}{7}, -\dfrac{30}{7}\right)$. This violates the condition that $\mu^* \geq 0$, so it is not a candidate for local minimum. In the same way as (a), the sufficiency condition is satisfied for $(x^*, y^*) = (0, 0)$. This is also the global minimum since there is only one local minimum.

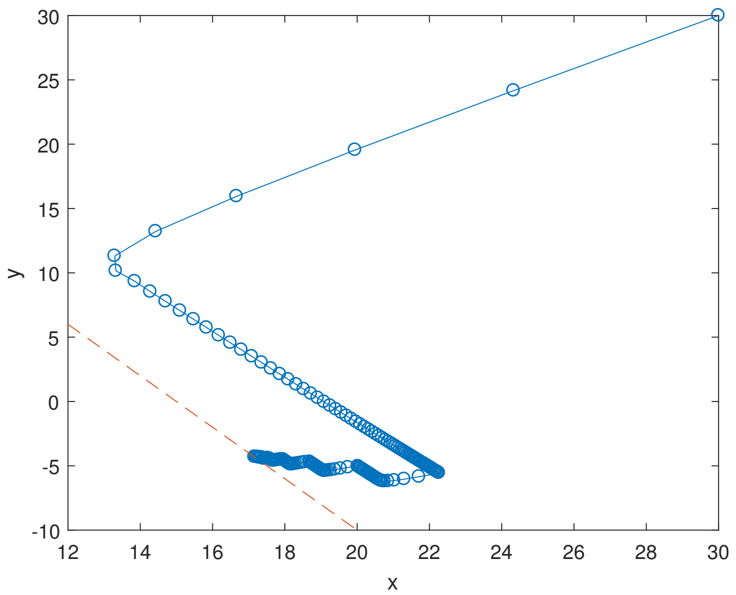
2. (Barrier and penalty methods – 10 points) *Consider a constrained minimization problem given as*

$$\min_{(x,y):\, y+2x \geq 30} f(x, y)$$

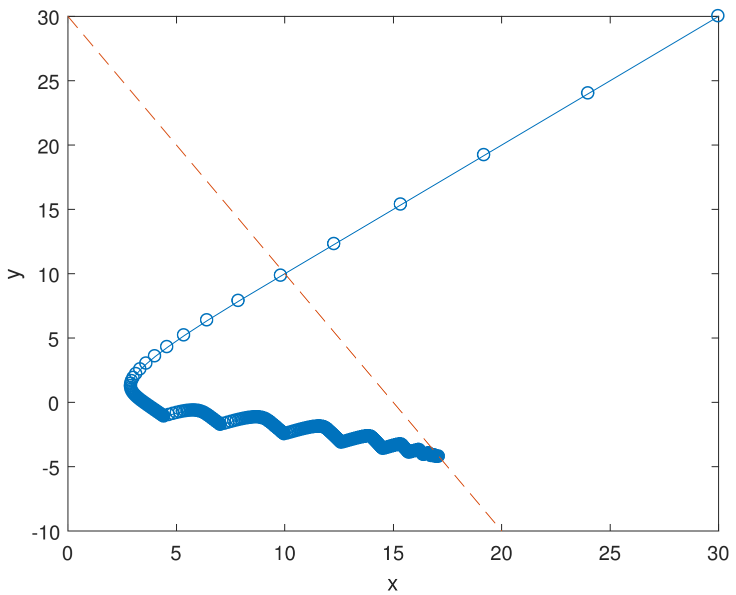
where $f(x, y) = 5(x + y)^2 + (x - y)^2$. Implement gradient descent with barrier and penalty methods and compare the results. You may use any programming language. Explain how your codes work. Submit your codes and plots.

Solution

The following figure shows the trajectory of the barrier method. You can download the sample code from KLMS. The final output of this code is $(17.1540, -4.3079)$, which is close to the true answer $(120/7, -30/7) \sim (17.1429, -4.2857)$.



The following figure shows the trajectory of the penalty method. You can download the sample code from KLMS. The final output of this code is $(17.1142, -4.2286)$, which is close to the true answer $(120/7, -30/7) \sim (17.1429, -4.2857)$.



3. (Worst-case error – 10 points) *Consider a problem of estimating θ from a set of samples $\{x^{(1)}, \ldots, x^{(m)}\}$ generated* i.i.d. *according to Bernoulli(θ), where $0 \le \theta \le 1$. Assume $m = 1$. Can you think of an estimator that minimizes the worst-case absolute error over all θ? Namely, find an estimator $\bar{\theta}_1$ based on $x^{(1)}$ that minimizes the following.*

$$\max_{0 \le \theta \le 1} \mathbb{E}[|\bar{\theta}_1 - \theta|]$$

What is the minimum worse-case absolute error $\mathbb{E}[|\bar{\theta}_1 - \theta|]$ that the estimator achieves?

Solution

The estimator $\bar{\theta}_1$ is a function of $x^{(1)}$. Thus we can write $\bar{\theta}_1$ as $\bar{\theta}_1(x^{(1)})$. Because $x^{(1)} \in \{0, 1\}$, there are two values for the estimator, $\bar{\theta}_1(0)$ and $\bar{\theta}_1(1)$. Let θ_0 and θ_1 denote $\bar{\theta}_1(0)$ and $\bar{\theta}_1(1)$, respectively. Then the mean absolute error of the estimator is

$$\begin{aligned}\mathbb{E}[|\bar{\theta}_1 - \theta|] &= \sum_{x\in\mathcal{X}} |\bar{\theta}_1(x) - \theta| p(x) \\ &= |\theta_0 - \theta|(1-\theta) + |\theta_1 - \theta|\theta \\ &=: g(\theta, \theta_0, \theta_1)\end{aligned}$$

Now, we want to find θ_0^* and θ_1^* that minimize $\max_{0\le\theta\le 1} g(\theta,\theta_0,\theta_1)$, i.e.

$$(\theta_0^*, \theta_1^*) = \underset{(\theta_0,\theta_1)}{\text{argmin}} \max_{0\le\theta\le 1} g(\theta,\theta_0,\theta_1)$$

First, we can see that

$$\begin{aligned}\min_{0\le\theta_0,\theta_1\le 1}\max_{0\le\theta\le 1} g(\theta,\theta_0,\theta_1) &\overset{(a)}{\ge} \min_{0\le\theta_0,\theta_1\le 1}\max_{\theta\in\{0,\frac{1}{2},1\}} g(\theta,\theta_0,\theta_1) \\ &= \min_{0\le\theta_0,\theta_1\le 1} \max\left\{\theta_0, \frac{|\theta_0-\frac{1}{2}|+|\theta_1-\frac{1}{2}|}{2}, 1-\theta_1\right\} \\ &\overset{(b)}{=} \min_{0\le u,v\le 1} \max\left\{u, \frac{|u-\frac{1}{2}|+|v-\frac{1}{2}|}{2}, v\right\} \\ &\overset{(c)}{=} \min_{0\le u\le v\le 1} \max\left\{\frac{|u-\frac{1}{2}|+|v-\frac{1}{2}|}{2}, v\right\} \\ &\overset{(d)}{\ge} \frac{1}{4},\end{aligned}$$

where (a) follows since a function maximized over a larger set must be bigger than or equal to that maximized over a smaller set, and in this case, $\{0,\frac{1}{2},1\} \subseteq [0,1]$, (b) follows by defining $u = \theta_0$ and $v = 1-\theta_1$, (c) follows since we only need to consider $v \ge u$ due to symmetry (i.e., the case $u \ge v$ can be analyzed similarly and will give the same answer), and finally (d) follows since we can split the minimization in the right hand side of (c) into two domains for v, namely $v \ge \frac{1}{4}$ and $v < \frac{1}{4}$. For the former, $\min_{0\le u\le v\le 1,\, v\ge\frac{1}{4}} \max\left\{\frac{|u-\frac{1}{2}|+|v-\frac{1}{2}|}{2}, v\right\} \ge \min_{0\le u\le v\le 1,\, v\ge\frac{1}{4}} v = \frac{1}{4}$. For the latter, it can be easily seen that $\min_{0\le u\le v<\frac{1}{4}} \max\left\{\frac{|u-\frac{1}{2}|+|v-\frac{1}{2}|}{2}, v\right\} > \frac{1}{4}$. Therefore, (d) holds.

On the other hand, we get

$$\begin{aligned}\min_{0\le\theta_0,\theta_1\le 1}\max_{0\le\theta\le 1} g(\theta,\theta_0,\theta_1) &\overset{(e)}{\le} \max_{0\le\theta\le 1} g(\theta,\frac{1}{4},\frac{3}{4}) \\ &\overset{(f)}{\le} \frac{1}{4},\end{aligned}$$

where (e) follows since fixing $\theta_0 = \frac{1}{4}$ and $\theta_1 = \frac{3}{4}$ can only increase the value of the left hand side of (e) and (f) follows since $g(\theta,\frac{1}{4},\frac{3}{4}) \le \frac{1}{4}$ for all θ (you need to check this also).

Therefore, by combining steps (a) $\sim$ (f), we conclude that

$$\min_{0\le\theta_0,\theta_1\le 1}\max_{0\le\theta\le 1} g(\theta,\theta_0,\theta_1) = \frac{1}{4}$$

is the minimum possible worst-case mean absolute error and $\theta_0^* = \frac{1}{4}, \theta_1^* = \frac{3}{4}$ achieve it.

4. (Underfitting and overfitting – 10 points) Write a program that can produce figures similar to those in pages 12 and 14 of lecture notes #5. Your figures do not need to look exactly like those in the lecture notes, but try to make them look similar. Training and test data should be generated as follows.

- Generate 10 *i.i.d.* training examples (x_1, y_1), (x_2, y_2), …, (x_{10}, y_{10}), where x_i's are uniform between -1 and 3 and $y_i = ax_i^2 + bx_i + c + z_i$, $i = 1, 2, \ldots, 10$, where z_i's are *i.i.d.* uniform noise between -0.2 and 0.2 and are independent of x_i's. Pick a, b and c appropriately so that your figure looks similar to that in page 12 of lecture notes #5.

- Generate 100 test examples such that x values are uniformly spaced between -1 and 3 and $y = ax^2 + bx + c$, where a, b and c are the same as the ones you chose for generating the train data. Note that we are not including noise in the test examples, i.e., no z_i's.

You may use any programming language. Explain how your codes work. Submit your codes and plots.

Solution

Here are some plots from an example code. The example code can be downloaded from KLMS.