CS157 Homework 8

BY SILAO_XU AND TC28

Problem 4

Convexity:

In this problem we will be exploring various aspects of convexity. Recall the definition of a convex function (http://en.wikipedia.org/wiki/Convex_function): a function f from n-dimensional space to the real numbers is convex if for any two inputs x and y, the line segment in the graph of f from $(x, f(x))$ to $(y, f(y))$ lies on or above the graph of f. (The picture is easiest to draw for dimension $n=1$.) To express this slightly more formally, for any interpolation parameter between 0 and 1, we have the condition $\lambda f(x)+(1-\lambda)f(y)\geqslant f(x+(1-\lambda)y)$. Make sure you understand what this means in one dimension before moving on.

1. (5 points) Recall the golden section search algorithm from the optimization lab, for minimizing convex functions f in one dimension. At each moment in the algorithm, you will be considering the function value at three unequally spaced points, $x_1 < x_2 < x_3$ where x_2 is ϕ times closer to one endpoint than than the other, where ϕ, the golden ratio, is $\frac{\sqrt{5}+1}{2}$. Consider the case where x_2 is closer to x_1 than x_3. Also, by the induction hypothesis of the algorithm, $f(x_2)$ is less than or equal to $f(x_1)$ and $f(x_3)$.

   Clearly the minimum value of f on this interval is at most $f(x_2)$. Find the best *lower bound* for f, in terms of $f(x_1)$, $f(x_2)$, and $f(x_3)$, given that f is convex. Explain.

   - *Case 1*: x locates in $[x_2, x_3]$

     According to the definition of convexity, we know that $\frac{x_2-x_1}{x_3-x_2}=\frac{\lambda}{1-\lambda}$. Also, according to the golden ratio, the $\frac{|x_1-x_2|}{|x_2-x_3|}=\frac{1}{\phi}$ and thus we find the relation between ϕ and λ, that is

$$\begin{aligned} \frac{\lambda}{1-\lambda} &= \frac{1}{\phi} \\ \lambda &= \frac{1}{\phi^2} \end{aligned} \tag{1}$$

     Similarly, we could also find the relation between $1-\lambda$ and ϕ, which is

$$\begin{aligned} 1-\lambda &= 1-\frac{1}{\phi^2} \\ &= \frac{\phi^2-1}{\phi^2} \\ &= \frac{\phi}{\phi^2} \\ &= \frac{1}{\phi} \end{aligned} \tag{2}$$

     The condition of convexity constraints the point $(x,\, y)$ for the best *lower bound*, lie on the same line as $(x_2, f(x_2))$ and $(x_1, f(x_1))$, and that is

$$\begin{aligned} (1-\lambda)f(x_1)+\lambda y &= f(x_2) \\ y &= \frac{f(x_2)-(1-\lambda)f(x_1)}{\lambda} \end{aligned} \tag{3}$$

     Plugging (1) and (2) into (3), we get the possible best *lower bound* for case 1

$$y = \phi^2 f(x_2)-\phi f(x_1)$$

- *Case 2*: x locates in $[x_1, x_2]$

  Similarly, the condition of convexity constraints the point (x', y') for the best *lower bound*, locating in $[x_1, x_2]$, lie on the same line as $(x_2, f(x_2))$ and $(x_3, f(x_3))$, that is

$$\begin{aligned}(1-\lambda)y' + f(x_3)\lambda &= f(x_2)\\ y' &= \frac{f(x_2) - f(x_3)\lambda}{1-\lambda}\end{aligned} \tag{4}$$

  Plugging (1) and (2) into (4), we get the possible best *lower bound* for case 2

$$y' = \phi^2 f(x_2) - \phi f(x_3)$$

- *Summary*

  The best *lower bound* taking *Case 1* and *2* into account both would be $\min(y, y')$.

2. (3 points) Suppose we want our golden section search algorithm to return a value x such that $f(x)$ is within of the global minimum. What *stopping condition* for the algorithm do your results from the previous part suggest? Explain.

   Using the *Golden Ratio Search* method, each step we would derive a new searching point x_4 where $x_4 = x_1 + x_3 - x_2$. Also, we could predict the best *lowest point* within $[x_1, x_3]$ of the convex function would be either y or y'.

   Now we could define the *stopping condition* as $|y - f(x_4)| \leqslant \epsilon$ (for *Case 1*) or $|y' - f(x_4)| \leqslant \epsilon$ (for *Case 2*).

3. Given a convex function in two dimensions, $f(x, y)$, if we define a new function $g(x) = \min_y f(x, y)$ to be the minimum of f along its second dimension, for each value of x, then g is a convex function. Prove this.

   From the convext function $f(x, y)$, we know that

$$\lambda f(x_1, y_1) + (1-\lambda) f(x_2, y_2) \geqslant f(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda) y_2) \tag{5}$$

   where $\lambda \in [0, 1]$.

   We let $f(x_1, y_{\min})$ be the minimum of $f(x_1, y)$ along f's second dimension, and $f(x_2, y'_{\min})$ be the minimum of $f(x_2, y)$ along its second dimension. We have

$$\begin{aligned}g(x_1) &= \min{}_y f(x_1, y)\\ &= f(x_1, y_{\min})\end{aligned}$$

   and

$$\begin{aligned}g(x_2) &= \min{}_y f(x_2, y)\\ &= f(x_2, y'_{\min})\end{aligned}$$

   also according to the convexity condition and substitute $f(x_2, y'_{\min})$ and $f(x_1, y_{\min})$ with $g(x_2)$ and $g(x_1)$ respectively, we have

$$\begin{aligned}\lambda f(x_1, y_{\min}) + (1-\lambda) f(x_2, y'_{\min}) &\geqslant f(\lambda x_1 + (1-\lambda)x_2, \lambda y_{\min} + (1-\lambda) y'_{\min})\\ \lambda g(x_1) + (1-\lambda) g(x_2) &\geqslant f(\lambda x_1 + (1-\lambda)x_2, \lambda y_{\min} + (1-\lambda) y'_{\min})\end{aligned} \tag{6}$$

Fixing the first dimension of f and let it be $\lambda x_1 + (1-\lambda)x_2$, we could define a new function $h(y) = f(\lambda x_1 + (1-\lambda)x_2, y)$ where y is variable. Based on the fact that $f(x, y)$ is a convext function in two dimensions, for any two inputs y_1 and y_2, the line segment in the graph of f from $h(y_1)$ to $h(y_2)$ lies on or above the graph of f, and thus lies on or above the graph of h. So $h(y)$ is also a convex function, and from the defintion of convexity we could derive

$$h(\lambda y_{\min} + (1-\lambda)y'_{\min}) \leqslant \lambda h(y_{\min}) + (1-\lambda)h(y'_{\min})$$

where $\lambda y_{\min} + (1-\lambda)y'_{\min}$ is a minimum value of f over its second dimension for value $x = \lambda x_1 + (1-\lambda)x_2$. So $f(\lambda x_1 + (1-\lambda)x_2, \lambda y_{\min} + (1-\lambda)y'_{\min}) = g(\lambda x_1 + (1-\lambda)x_2)$ and plug it into the right-hand side of (6), we get

$$\lambda g(x_1) + (1-\lambda)g(x_2) \geqslant g(\lambda x_1 + (1-\lambda)x_2)$$

According to the formal convexity condition, we conclude that g is also a convex function.

4. (6 points) A further complication with this kind of recursive optimization procedure has to do with numerical precision. The golden section search algorithm will not find the exact minimum value of its input function, but will, rather, return a value which may be larger than the minimum. This error may make subsequent layers of the recursion have even larger errors.

   Show that the first time the algorithm makes a bad decision, recursing to an interval that does not contain the global minimumis happens, the true global minimum of g is at least $\tilde{g}(x_2) - \phi^2\epsilon$.

   In this case, we could apply our extimated best *lower bound* from part 1, such that

$$g_{\min} \geqslant \phi^2 g(x_4) - \phi g(x_2) \tag{7}$$

   Also $\tilde{g}(x_2)$ must be within the interval $[g(x), \epsilon + g(x)]$, that is $\tilde{g}(x_2) > g(x_2)$, plugging it into (7) we get

$$g_{\min} \geqslant \phi^2 g(x_4) - \phi \tilde{g}(x_2) \tag{8}$$

   We know that $\tilde{g}(x_4) - g(x_4) \leqslant \epsilon$, that is $\tilde{g}(x_4) - \epsilon \leqslant g(x_4)$ and given $\tilde{g}(x_2) < \tilde{g}(x_4)$ we have

$$g(x_4) \geqslant \tilde{g}(x_2) - \epsilon \tag{9}$$

   Plugging (9) into (8), we could draw the conclusion that

$$\begin{aligned} g_{\min} &\geqslant \phi^2(\tilde{g}(x_2) - \epsilon) - \phi \tilde{g}(x_2) \\ &= (\phi^2 - \phi)\tilde{g}(x_2) - \phi^2\epsilon \\ &= \tilde{g}(x_2) - \phi^2\epsilon && (1 + \phi = \phi^2) \end{aligned}$$