ISyE 6669 HW 4
Spring 2021

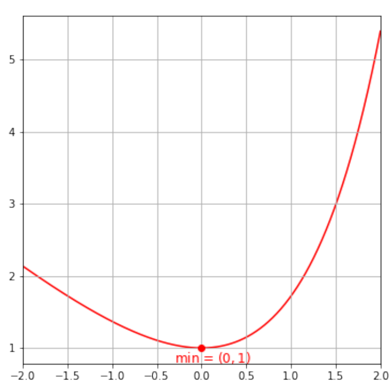
1 Convex optimization

Convex optimization is the backbone of modern optimization. We learned some simple algorithmic schemes such as gradient descent and the Newton method among others. These two algorithms are especially suited to minimize convex functions when they are continuously differentiable or have second-order derivatives.

1. Use Newton's method to Minimize the convex function $f(x) = e^x - x$ on the entire real line $x \in \mathbb{R}$. You need to first write down the first-order and second-order derivatives of $f(x)$. Then, write down a step in the Newton's method, e.g. $x^{k+1} = x^k - f'(x^k)/f''(x^k)$. You need to plug in the expressions of the derivatives. Carry out the Newton's method from the initial point $x^0 = -1$. Write down $x^k, f(x^k), f'(x^k)$ in each iteration, until you reach a solution with the first-order derivative $|f'(x^k)| < 10^{-5}$. You should only need less than 10 steps.

   Solution:

   We have convex function $f(x) = e^x - x$.



From plot above we see that this function attains minimum at point $x = 0$, and that its minimum value is $f(0) = 1$.

First and second order derivatives of the function are given by:

$f'(x) = e^x - 1$
$f''(x) = e^x$

Now let us use Newthon method to find minimum of the function:

Iteration 0:

$x^0 = -1$
$f(x^0) = 1.3678794411714423$
$f'(x^0) = -0.6321205588285577$
$f''(x^0) = 0.36787944117144233$
$|f'(x^0)| = 0.6321205588285577 \geq 10^{-5} \quad \Rightarrow \quad \text{continue}$

Iteration 1:

$x^1 = x^0 - \dfrac{f'(x^0)}{f''(x^0)} = 0.7182818284590451$
$f(x^1) = 1.332624544233456$
$f'(x^1) = 1.0509063726925012$
$f''(x^1) = 2.050906372692501$
$|f'(x^1)| = 1.0509063726925012 \geq 10^{-5} \quad \Rightarrow \quad \text{continue}$

Iteration 2:

$x^2 = x^1 - \dfrac{f'(x^1)}{f''(x^1)} = 0.20587112717830613$
$f(x^2) = 1.0227237341276951$
$f'(x^2) = 0.22859486130600137$
$f''(x^2) = 1.2285948613060014$
$|f'(x^2)| = 0.22859486130600137 \geq 10^{-5} \quad \Rightarrow \quad \text{continue}$

Iteration 3:

$$x^3 = x^2 - \frac{f'(x^2)}{f''(x^2)} = 0.019809091184598587$$
$f(x^3) = 1.0001975020028975$
$f'(x^3) = 0.02000659318749598$
$f''(x^3) = 1.020006593187496$
$|f'(x^3)| = 0.02000659318749598 \geq 10^{-5} \quad \Rightarrow \quad \text{continue}$

Iteration 4:

$$x^4 = x^3 - \frac{f'(x^3)}{f''(x^3)} = 0.00019491092231630272$$
$f(x^4) = 1.000000018996368$
$f'(x^4) = 0.00019492991868430565$
$f''(x^4) = 1.0001949299186843$
$|f'(x^4)| = 0.00019492991868430565 \geq 10^{-5} \quad \Rightarrow \quad \text{continue}$

Iteration 5:

$$x^5 = x^4 - \frac{f'(x^4)}{f''(x^4)} = 0.0000000189938997$$
$f(x^5) = 1.0000000000000002$
$f'(x^5) = 0.0000000189939000$
$f''(x^5) = 1.0000000189939$
$|f'(x^5)| = 0.0000000189939000 < 10^{-5} \quad \Rightarrow \quad \text{stop}$

Using Newton's method we get optimal solution $x = 0.0000000189938997$ and optimal value $f(x) = 1.0000000000000002$, which is quite close to the true values.

2. To minimize a convex function $f(x)$ without any constraint, it is equivalent to solving the first-order optimality condition $f'(x) = 0$, i.e. find the point x^* on the curve $y = f'(x)$ that crosses the horizon line $y = 0$. Such a x^* is also called a *zero* of the equation $f'(x) = 0$. The curve $y = f'(x)$ is usually a nonlinear curve. Think of Question 1. What Newton's method actually does is to solve this nonlinear equation by solving a sequence of linear equations that approximate this nonlinear equation. How do we approximate a nonlinear curve by a linear one? Right, use the tangent line at a point of the nonlinear curve, or equivalently, use the Taylor's expansion we all know from calculus.

So suppose we want to solve $f'(x) = 0$ and suppose the second-order derivative $f''(x)$ always exists. First, write down the tangent line of the curve $y = f'(x)$ at a point x^k. Hint: the line passing through the point $(x = a, y = b)$ with slope k has the equation $y = k(x - a) + b$.

Suppose your equation for the tangent line is $y = k(x - a) + b$. Of course, you need to fill in k, a, b and express them in $x^k, f'(x^k), f''(x^k)$. Then, solve the equation $k(x - a) + b = 0$ and write down the expression of the solution in k, a, b. The solution is the next iteration x^{k+1}. If you have done everything correctly, you should recover the Newton's iterate. The next step of Newton's method starts from x^{k+1}, forms the tangent line of the curve $y = f'(x)$ at x^{k+1}, and finds the zero of this linear equation, and continues.

Plot $y = f'(x)$ for $f(x) = e^x - x$. Start from $x^0 = -1$. Draw the tangent line at x^0, find the zero, and continue, until you reach x^2. You should see the same sequence as you find in the first question, and this should give you a geometric understanding of Newton's method.

Solution: In this part we will search for minimum by trying to solve $f'(x) = 0$. In order to do that we will plot tangent to $f'(x) = e^x - 1$ at x^i. Using formula $y = k(x - a) + b$ and substituting $k = f''(x^i), a = x^i, b = f'(x^i)$ we get that equation of the tangent line at point x^i is given by $y = f''(x^i)x + f'(x^i) - f''(x^i)x^i$. In order to find next point, we find intercept of the tangent line and $x-$ axis.

$$f''(x^i)x + f'(x^i) - f''(x^i)x^i = 0 \quad \Rightarrow \quad x = x^i - \frac{f'(x^i)}{f''(x^i)}$$

This is the same formula we used in previous part.

As before we start at the point $x^0 = -1$. Equation of a tangent to a $f'(x) = e^x - 1$ at point $x^0 = -1$ is given by

$$y = 0.367879441171442\,33 \times x - 0.264241117657115\,33$$

We find point x^1 by finding intercept of the tangent line and $x-$ axis:

$$0.36787944117144233 \times x - 0.26424111765711533 = 0 \quad \Rightarrow \quad x^1 = 0.7182818284590451.$$

At point x^1 we get tangent line:

$$y = 2.050906372692501 * x - 0.4222224066833764.$$

We get next point by solving:

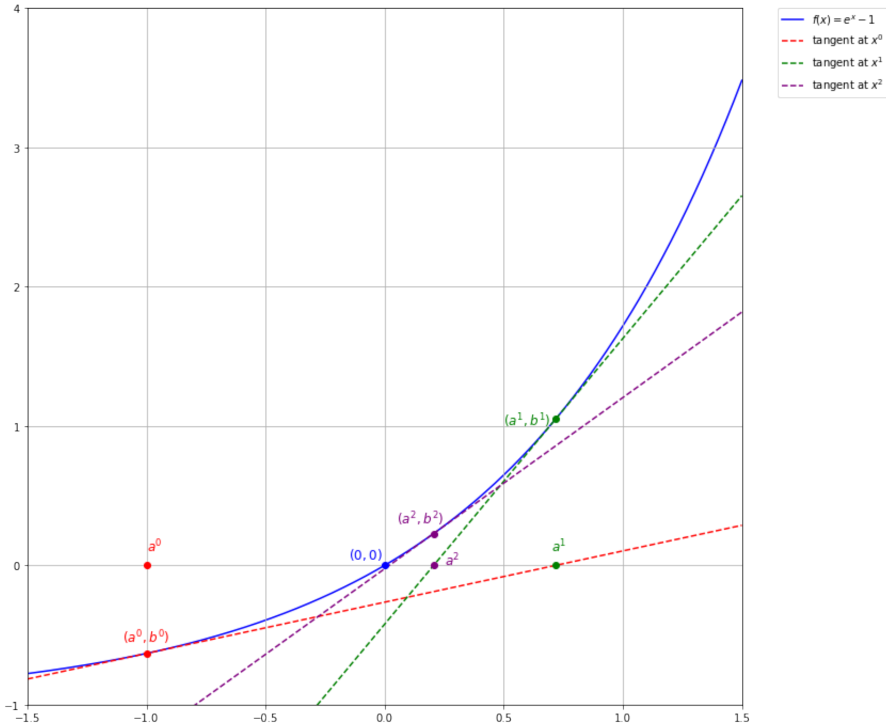
$$2.050906372692501 * x - 0.4222224066833764 = 0 \quad \Rightarrow \quad x^2 = 0.20587112717830613.$$

At point x^2, equation of a tangent line is given by:

$$y = 1.2285948613060014 * x - 0.02433734763653983.$$

We could continue this process, until we reach stopping criteria as before, but we will stop here. As expected , points x^0, x^1, x^2 we obtained this way, are same points we obtained in Q1.1.

Bellow you can see plot of the function $f'(x)$ and first three tangents.



2 Nonconvex optimization

The following problems involve optimization of nonlinear nonconvex functions with some simple constraints, such as an interval or a box in higher dimension. To minimize each of the following functions, you can use the command `minimize` from `scipy.optimize`. Due to the nonconvex and sometimes nonsmooth (i.e. not differentiable at some points) nature of the objective function, you need to be careful about the starting point and the constraints you set. For example, you may need to set the box small enough to help the solver find a good local optimum. You need to provide a printout of your code, along with the solution.

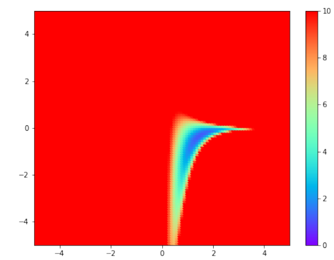
1. $f(x_1, x_2) = (1 - x_1 + x_1 x_2)^2 + (2 - x_1 + x_1^2 x_2)^2 + (3 - x_1 + x_1^3 x_2)^2$ over the box $-5 \leq x_1, x_2 \leq 5$. Start from $(0, 0)$. Plot the function $f(x_1, x_2)$ over the box $[-5, 5]^2$ using both a 2D contour plot and a 3D plot.

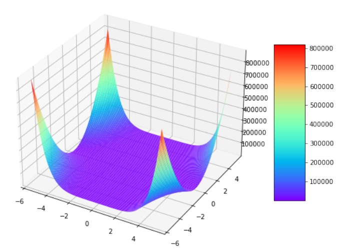
   Solution:

   Optimal value: 1.1684123071910781

   Optimal solution: $x_1 = 1.51907761, x_2 = -0.27672452$

   Bellow are 2D and 3D plots of the function. From 2D plot bellow we see that solution is where we would expect it to be.





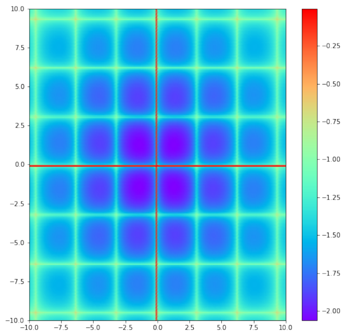
2. Consider the following function

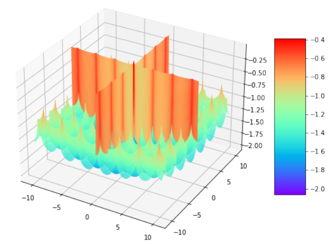
$$f(x_1, x_2) = -0.0001\left(\left|\sin(x_1)\sin(x_2)\exp\left(\left|100 - \frac{\sqrt{x_1^2 + x_2^2}}{\pi}\right|\right)\right| + 1\right)^{0.1}.$$

   This function has a lot of local minima and global minima. Plot $f(x_1, x_2)$ over the box $[-10, 10]^2$ using both a 2D contour plot and a 3D plot. Try to find at least two different global minima and one local minimum that is not a global minimum. Hint: You can start the algorithm that you choose from different starting points.

   Solution:

   Bellow you can see 2D and 3D plots of the function. We see that function has multiple local and global minima.





Global minima:

Global minimum obtained using starting point $(-1,-1)$: -2.0626118707233263
Optimal solution: $x_1 = -1.3493852, x_2 = -1.3493852$

Global minimum obtained using starting point $(1,1)$: -2.062611870736794
Optimal solution: $x_1 = 1.3493867, x_2 = 1.3493867$

Global minimum obtained using starting point $(-1,1)$: -2.0626118707360184
Optimal solution: $x_1 = -1.3493867, x_2 = 1.34938652$

Global minimum obtained using starting point $(1,-1)$: -2.0626118707360184
Optimal solution: $x_1 = 1.34938652, x_2 = -1.3493867$

Local minima different than global minima:

Global minimum obtained using starting point $(-5,-1)$: -1.8899380457820998
Optimal solution: $x_1 = -4.41908156, x_2 = -1.47062333$

Global minimum obtained using starting point $(-1,-1)$: -1.554463337514644
Optimal solution: $x_1 = 7.63258721, x_2 = -7.63253296$