[ML20] Assignment 3

Max Machalek

Due: Feb 10 (before class)

[1] Let $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$. If $f(x) = x^T M$, calculate $\frac{\partial}{\partial x} f(x)$. Elaborate your arguments.

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = \begin{bmatrix} x_1 M_{11} + x_2 M_{21} & x_1 M_{12} + x_2 M_{22} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \end{bmatrix} \begin{bmatrix} x_1 M_{11} + x_2 M_{21} & x_1 M_{12} + x_2 M_{22} \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

[2] Let $f(x) = x^2 - 4x + 1$. Implement gradient descent (GD) to identify its minimum.[1] Requirements are

(i) Write your update formula. Let λ denote the step-size.

$$x = x - \lambda * \delta \tag{1}$$

Where δ is $f'(x) = 2x - 4$ with current x plugged in (gradient).

(ii) Explain your stopping criterion. For example, "I stop the update when ..." I specified a set number of iterations for the algorithm to run. Smaller λ values required higher arbitrary iteration values.

(iii) Initially set $x = 50$. (In practice you should initialize randomly. Here we fix it for easier grading.)

(iv) Show two figures. Each figure shows a *convergent* [2] curve of $f(x)$, with x-axis being the number of GD updates and y-axis being $f(x)$. Choose two λ's for the two figures, respectively, such that one choice allows f to converge faster but not very close to its true minimum, while the other choice allows f to get closer to its minimum but has a more slowly convergence rate.

[1] You need to implement from scratch. You cannot use any optimization library in Python.
[2] This means you need to run sufficient number of updates until convergence of the curve is observed.

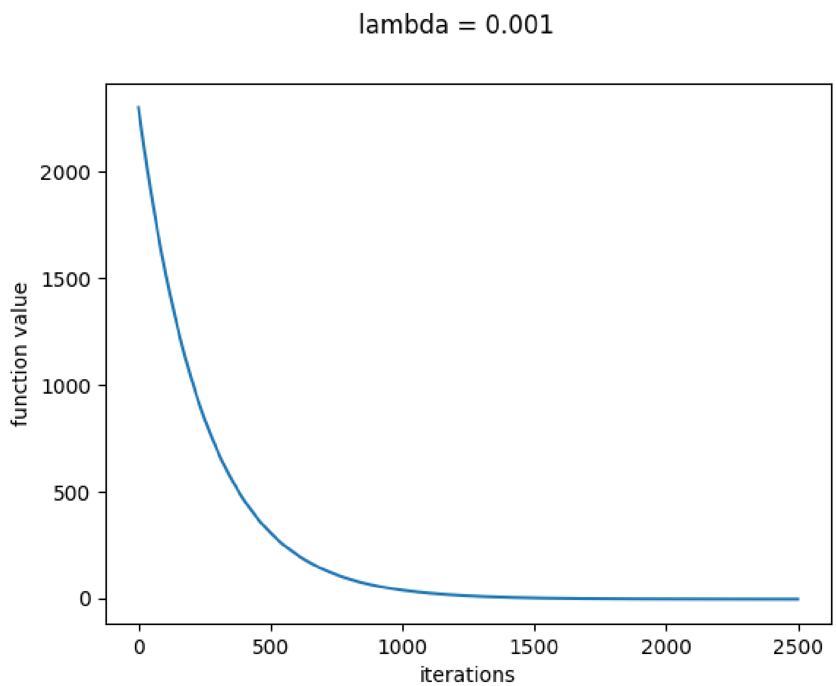

Fig. 1. Convergence Curve with $\lambda = 0.001$. After 2500 iterations, the x value was 2.32 and appeared to be converging toward 2

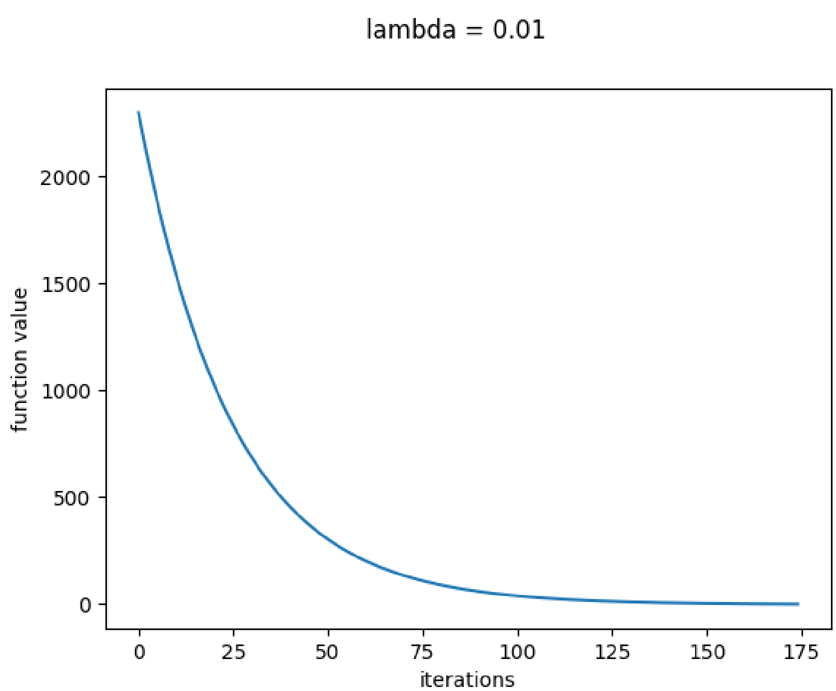

Fig. 2. Convergence Curve with $\lambda = 0.01$. After 175 iterations, the x value was 3.4 and (in a naive setting) may have been converging to 3 as far as any user might tell, not knowing the true value is 2.

[3] Let $f(x) = 3x^2 - 6x - 7$. Apply the *Lagrange multiplier* technique to solve $\min_x f(x)$, s.t. $2x^2 \leq 5$. Elaborate your arguments based on the following five steps.

Step 1. Construct Lagrange Function F.

$F = 3x^2 - 6x - 7 + \alpha(2x^2 - 5)$

Step 2. Minimize F over x (using the critical point method if you can)

$\frac{\partial}{\partial x}F(x) = 6x - 6 + 4\alpha x$

$6x - 6 + 4\alpha x = 0$

$x = \frac{3}{3+2\alpha}$

Step 3. Plug the optimal x back to F to get a dual function G.

$G = 3(\frac{3}{3+2\alpha}) - 6(\frac{3}{3+2\alpha}) - 7 + \alpha(2(\frac{3}{3+2\alpha})^2 - 5) = \frac{-9}{3+2\alpha} - (5\alpha + 7)$

Step 4. Maximize G over any newly introduced parameter(s) (using the critical point method if you can).

$G'(x) = \frac{18}{(3+2\alpha)^2} - 5$

Now set $G' = 0$ and see if any solutions are above 0 . . .

$\frac{18}{(3+2\alpha)^2} - 5 = 0$

$\alpha = \frac{-3\pm\sqrt{\frac{18}{5}}}{2}$

Both zeros are negative values so the constraint is redundant and we may calculate $F(x)$ with α as 0, which is $f(x)$.

Step 5. Plug that optimal new parameter back to x and finally to $f(x)$.

Plug $\alpha = 0$ in to $x = \frac{3}{3+2\alpha}$:

$x = \frac{3}{3+2(0)}$
$x = 1$
Now plug x = 1 in to original $f(x)$:

$f(x) = 3(1)^2 - 6(1) - 7 = -10$

So the min point is at $x = 1, f(x) = -10$