Home Exercise 9: Gradient method with line search

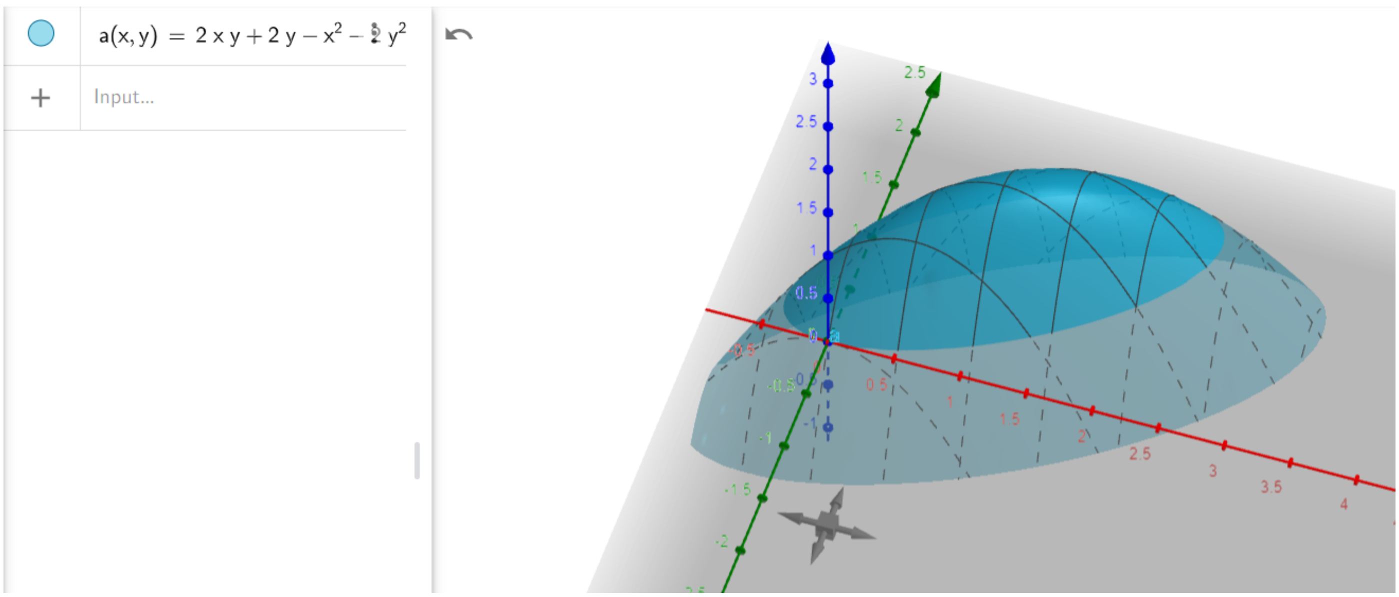
Perform one iteration of the gradient method to solve

$$\max . f(x_1, x_2) = 2x_1x_2 + 2x_2 - x_1^2 - 2x_2^2$$

from the initial point $x_0 = (0.5, 0.5)$. Use the bisection method to find the optimal step size with interval [0,2] and tolerance $\varepsilon = 0.01$. Is the new point obtained optimal (considering the tolerance of $\varepsilon = 0.01$)?

Hint. Do it by hand and notice it is a maximisation.

3D graph by GeoGebra



Graphically, we can observe that max .f(x1, x2) = 1, at [1, 1]

Gradient (descent) method

Algorithm Gradient descent method

1: **initialise.** tolerance $\epsilon > 0$, initial point x_0, iteration count $k = 0$.
2: **while** $||\nabla f(x_k)|| > \epsilon$ **do**
3: $\quad d_k = -\nabla f(x_k)$.
4: $\quad \overline{\lambda} = \text{argmin}_{\lambda \in \mathbb{R}} \{f(x_k + \lambda d_k)\}$.
5: $\quad x_{k+1} = x_k + \overline{\lambda} d_k$.
6: $\quad k \leftarrow k + 1$.
7: **end while**
8: **return** x_k.

Instead of argmin, it is argmax since we need to maximize the function
Firstly, the gradient is $\nabla f(x_1, x_2) = [2x_2 - 2x_1, 2x_1 - 4x_2 + 2]^T$
Initial point: $x_0 = [0.5, 0.5]$
=> $\nabla f(0.5, 0.5) = [0, 1]^T$

Since norm($\nabla f(0.5, 0.5)$) > ε => $d_0 = \nabla f(0.5, 0.5) = [0, -1]^T$
$f(\lambda) = \text{argmax}_\lambda$ ($f([0.5, 0.5]^T + \lambda[0, 1]^T)$) = $f(0.5, 0.5 + \lambda)$
$= -2\lambda^2 + \lambda + 0.75$
=> $f'(\lambda) = -4\lambda + 1$
Now we will use bisection method to find roots of $f'(\lambda)$

Algorithm Bisection method (minimisation)

1: **initialise.** tolerance $l > 0$, $[a_0, b_0] = [a, b]$, $k = 0$.
2: **while** $b_k - a_k > l$ **do**
3: $\lambda_k = \frac{(b_k + a_k)}{2}$ and evaluate $f'(\lambda_k)$.
4: **if** $f'(\lambda_k) = 0$ **then** return λ_k.
5: **else if** $f'(\lambda_k) > 0$ **then**
6: $a_{k+1} = a_k$, $b_{k+1} = \lambda_k$.
7: **else**
8: $a_{k+1} = \lambda_k$, $b_{k+1} = b_k$.
9: **end if**
10: $k = k + 1$.
11: **end while**
12: **return** $\overline{\lambda} = \frac{a_k + b_k}{2}$.

Remark: if maximising, the condition in Line 5 must be replaced with $f'(x) < 0$ and concavity is presumed.

Initialize: tolerance ε = 0.01, [a0, b0] = [0,2], k = 0

- First iteration: $[a_0, b_0] = [0, 2]$, $k = 0$

$b_0 - a_0 = 2 > \varepsilon$ => $\lambda_0 = (2+0)/2 = 1$
$f'(\lambda) = -4\lambda + 1$ => $f'(\lambda_0) = -3 < 0$
=> $a_1 = 0$, $b_1 = 1$

- Second iteration: $[a_1, b_1] = [0, 1]$, $k = 1$

$b_1 - a_1 = 1 > \varepsilon$ => $\lambda_1 = (1+0)/2 = 0.5$
$f'(\lambda) = -4\lambda + 1$ => $f'(\lambda_1) = -1 < 0$
=> $a_2 = 0$, $b_2 = 0.5$

- Third iteration: $[a_2, b_2] = [0, 0.5]$, $k = 2$

$b_2 - a_2 = 0.5 > \varepsilon$ => $\lambda_2 = (0+0.5)/2 = 0.25$
$f'(\lambda) = -4\lambda + 1$ => $f'(\lambda_2) = 0$ (stop)
=> $\lambda_2 = 0.25$ is the root of $f'(\lambda) = -4\lambda + 1$ and is the optimal step in the range [0, 2] (answer)

Now we return to the gradient method:
$x_1 = x_0 + \lambda d_0 = [0.5, 0.5]^T + 0.25 * [0, -1]^T = [0.5, 0.75]^T$
Second iteration of gradient method:
$\nabla f(x_1) = [0.5, 0]^T$, norm($[0.5, 0]^T$) = 0.5 > ε = 0.01
=> The algorithm of gradient descent has not stopped and we haven't arrived at the desired optimal solution. In fact, graphically the optimal solution is [1, 1], f(x) = 1
=> The new obtained point $x_1 = [0.5, 0.75]^T$ is not optimal (tolerance of ε = 0.01) (Answer)