Advanced Deep Learning for Physics (IN2298)
Exercise 2
Convergence rate and Momentum

In this exercise we are going to take a look at the optimization with the Gradient Descent algorithm to get a general feel for it. There will be no programming tasks in this assignment.

(1) Convergence Rate

In this problem, we investigate the Gradient Descent algorithm. The optimization variables x are iteratively updated by using:

$$x[i+1] = x[i] - \eta \cdot \frac{\partial L}{\partial x} \tag{1}$$

Here, $x[i]$ denotes the value of x at iteration i, η stands for the learning rate, and we require a loss function L and a starting point $x[0]$ to apply the algorithm. One way to measure how fast an optimization algorithm approaches a minimum x_* is by using the convergence rate. If we can find a number c fulfilling the following inequality for any chosen $x[i]$, we call this c convergence rate.

$$||x[i+1] - x_*|| \leq c \cdot ||x[i] - x_*|| \tag{2}$$

Here, $||\cdot||$ denotes a norm. For $0 \leq c < 1$ the algorithm is called convergent. A better convergence rate means c has a smaller value, which guarantees that for a generic starting point $x[0]$, the minimum is approached faster.

(a) Compute the convergence rate of Gradient Descent for the loss function $L(x) = ax^2$ with $a > 0$. For which choices of the learning rate is the algorithm convergent? What choice of the learning rate gives the best convergence rate? What is the value of the best convergence rate?

(b) We now consider a two-dimensional loss function $L(x_1, x_2) = ax_1^2 + bx_2^2$ with $0 < a < b$. Compute the convergence rate of Gradient Descent for the loss function L. What choice of the learning rate gives the best convergence rate? What is the value of the best convergence rate?

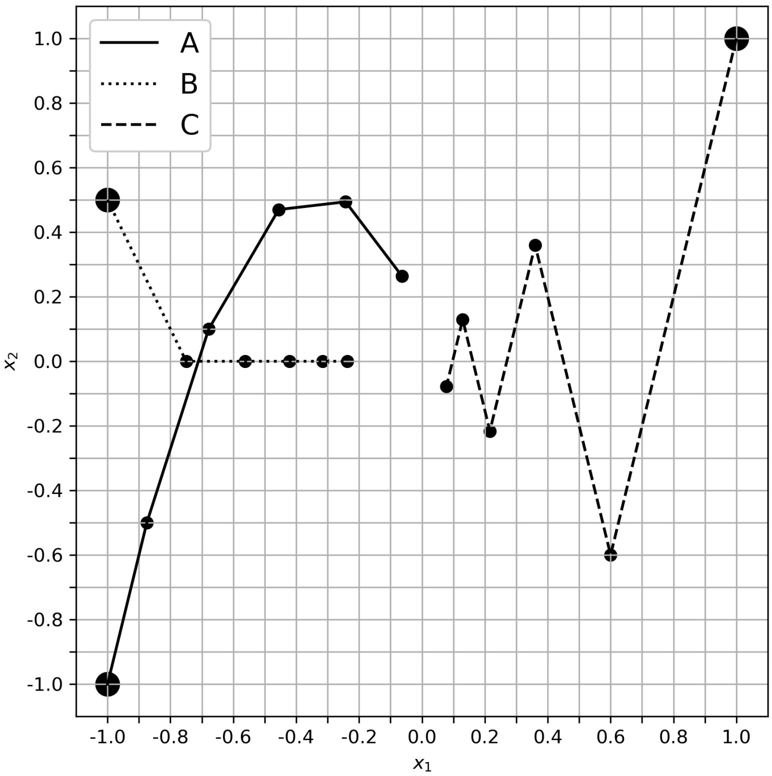
(2) Gradient Descent and its Acceleration with Momentum

Momentum is an efficient and effective way to improve the convergence rate of Gradient Descent. One variant of such an optimization method is given by these equations:

$$\begin{aligned} v[0] &= 0 \\ v[i+1] &= m \cdot v[i] - \eta \cdot \frac{\partial L}{\partial x} \\ x[i+1] &= x[i] + v[i+1] \end{aligned} \tag{3}$$

We used the notation $x[i]$ and $v[i]$ to denote the value of x and v at the i-th iteration step. m is called momentum parameter and can take values $0 < m < 1$. η denotes the learning rate, and a loss function L and a starting point $x[0]$ are required run the algorithm.

The next figure shows three optimization trajectories (named A, B and C) for three different methods with three different starting positions and the same loss $L(x_1, x_2) = x_1^2 + 4 \cdot x_2^2$. Two of the three trajectories were obtained by using Gradient Descent and the third one by using the momentum algorithm. The three big black dots mark the starting position for each method. For A, we have $x[0] = (-1, -1)$; for B, we have $x[0] = (-1, 0.5)$; for C, we have $x[0] = (1, 1)$. Subsequent $x[i]$ are given by the smaller black dots and connected by the respective linestyle of each method. For example, for C, $x[1] = (0.6, -0.6)$.



(a) Which of the three trajectories is from the momentum algorithm? Explain your decision briefly.

(b) Determine the used learning rate for each of the three trajectories.

(c) Determine the momentum parameter for the momentum trajectory.

Submission instruction

Please upload a single PDF file containing your results along with your code for implementation tasks or your derivation for non-implementation tasks (LaTex typesetting). The uploaded PDF should only include the final code, so please trim empty spaces and your intermediate work before submitting.

The easiest way to generate such a PDF is by using Jupyter notebooks and LaTex (we recommend MiKTeX for Windows users). With Jupyter and LaTeX installed, you can create a PDF from your notebook by running *jupyter nbconvert –to pdf your-notebook.ipynb*

Additional information

This is an individual assignment. Plagiarism will result in the loss of eligibility for the bonus this semester.

If you have any questions about the exercises, please contact us via the forum on Moodle. If you need further face-to-face discussion, please join our weekly online Q&A session (every Monday at 15:00 and 16:00 via BBB).