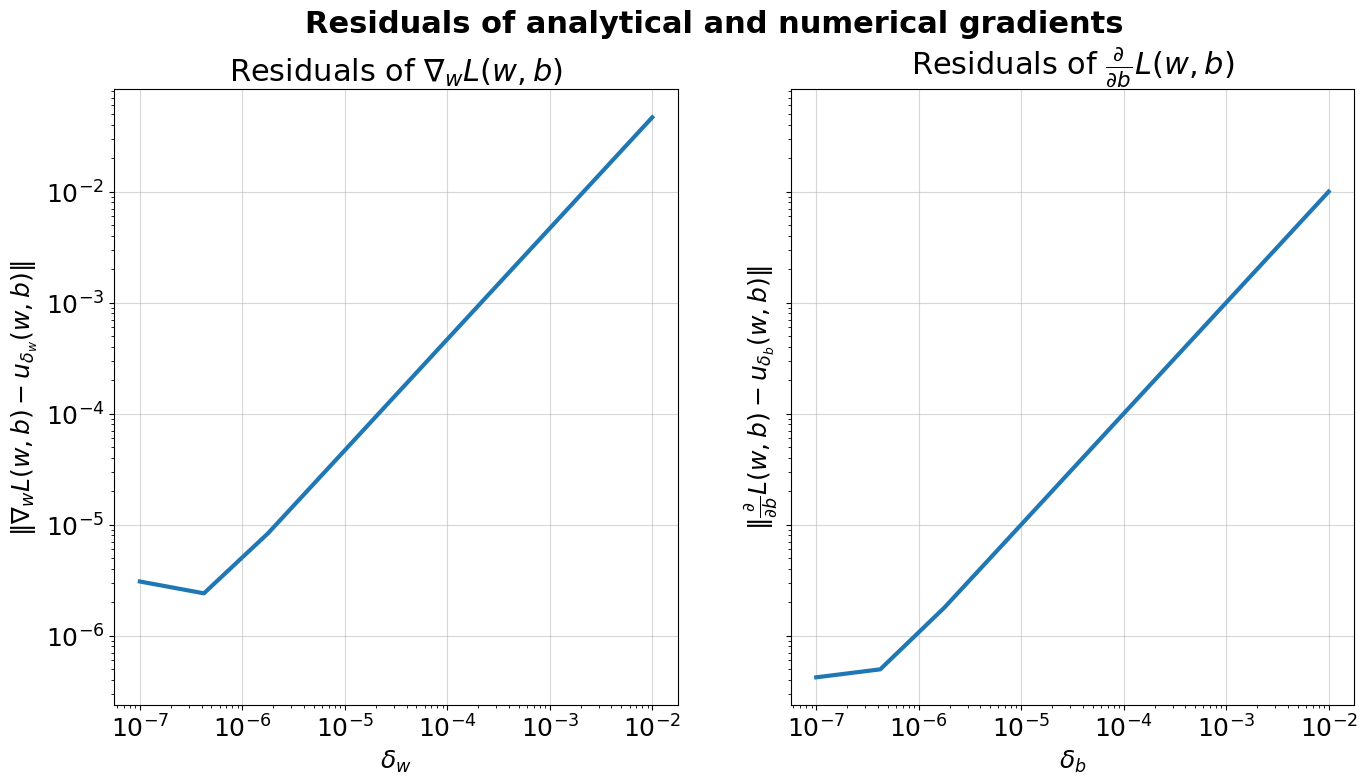
Major HW 3 – Regression

Orad Barel, 311288203, [oradbarel@campus.technion.ac.il](mailto:oradbarel@campus.technion.ac.il)

Ofir Manor, 316084623, [ofir.manor@campus.technion.ac.il](mailto:ofir.manor@campus.technion.ac.il)

**Q1.**

**Q2.**

****

Diagram

Description automatically generated**Q3.**

We see here that from the most part (excluding the higher the learning rate, the faster the descent of the loss is. This makes sense because small learning rates create a slow gradient descent, so that we arrive at the minimum of the loss only after a great amount of iterations. As for we can see that the loss is divergent after a few iterations, the large “jumps” that occur do not allow the gradient descent to arrive at a minimal loss.

Our best learning rate (the one that achieved the minimal validation loss) is , and it does not make sense to increase the number of gradient steps as it seems to achieve this minimum before step 1500 and slightly diverge around the minimum.