

# Homework 2

January 2019

## 1 Problem 1:

This exercise makes use of Jupyter, a "computational notebook" application which allows the user to embed Python code (or any other language) in a document, which can be reset and run, with output stored in the document. The document can also be edited to contain illustrative text, using a simple mark-down language. To gain familiarity with Jupyter, please take a look at this tutorial ([http://bi1.caltech.edu/code/t0b\\_jupyter\\_notebooks.html](http://bi1.caltech.edu/code/t0b_jupyter_notebooks.html)). Recall the SignSGD Algorithm from the class,

---

**Algorithm 1** SignSGD Algorithm

---

- 1: **Input:** Learning rate  $\delta$ , current point  $x_k$
  - 2:  $\tilde{g}_k \leftarrow \text{stochasticGradient}(x_k)$
  - 3:  $x_{k+1} \leftarrow x_k - \delta \cdot \text{Sign}(\tilde{g}_k)$ .
- 

Please refer to [1] for more information about the algorithm and its analysis. The goal of this exercise is for you to come up with a simple example of an optimization, for which the Sign SGD diverges or fails to achieve a minimum. Please justify your example by simulating your proposed optimization using Jupyter Notebook and include the code and the plots in your submission.

## 2 Problem 2

Show that SGA with some step size  $\eta$  fails to converge to the stationary point  $(0, 0)$  when run on the function  $U(\theta, \omega) = \theta\omega$  from any initial point  $\theta_0, \omega_0$  where both  $\theta_0, \omega_0$  are non-zero.

## References

- [1] Jeremy Bernstein, Yu-Xiang Wang, Kamyar Azizzadenesheli, and Anima Anandkumar. signsgd: compressed optimisation for non-convex problems. *arXiv preprint arXiv:1802.04434*, 2018.