Formatting Instructions For NeurIPS 2022

David S. Hippocampus*
Department of Computer Science
Cranberry-Lemon University
Pittsburgh, PA 15213
hippo@cs.cranberry-lemon.edu

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Numerical Experiments

As the final part of our report, we implement the algorithms in the original article with `Julia` and its package `DifferentialEquations.jl` [**rackauckas2017differentialequations**]. By comparing ODE direct discretizating (DD) methods described in the article against gradient descent (GD) and Nesterov's accelerated gradient (NAG) methods, we con verify the main results in the theoretical part.

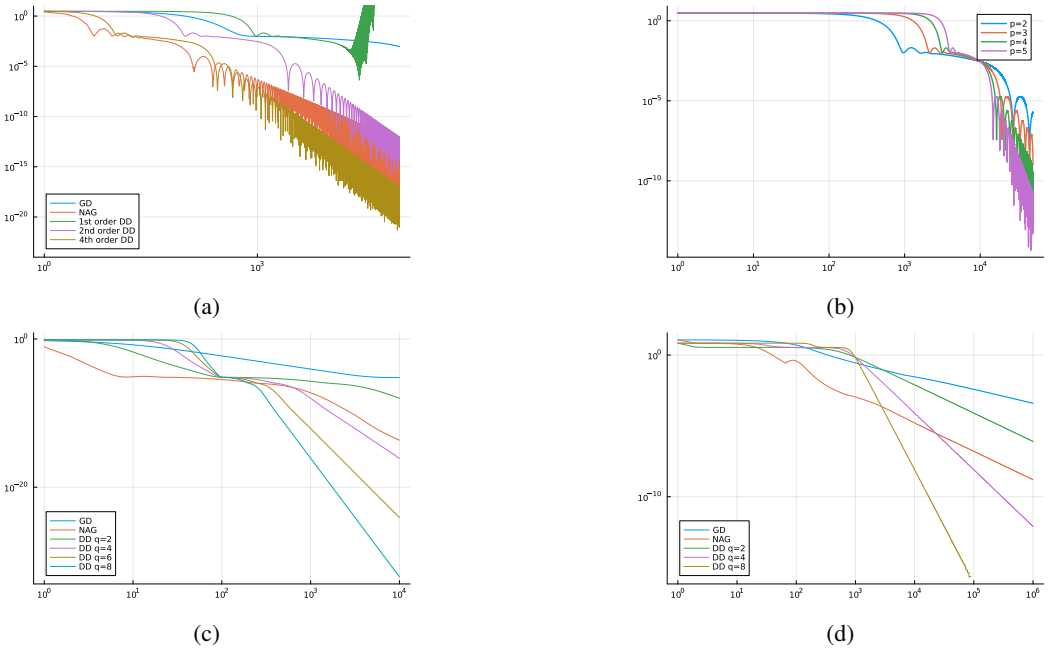
Inspired by the numerical results by **NEURIPS2019_7a2b33c6**, we generate normal distributed separable dataset and fit a linear model $Ax = b$. Then, we minimize three different kinds of loss functions:

$$\begin{aligned} f_1(x) &= \|Ax - b\|_2^2 \\ f_2(x) &= \log(1 + e^{-yw^{\mathrm{T}}x}) \\ f_3(x) &= \frac{1}{4}\|Ax - b\|_4^4 \end{aligned} \tag{1}$$

where $f_1(\cdot), f_2(\cdot), f_3(\cdot)$ are L_2 loss, logistic loss and L_4 loss, respectively. For each test case and optimization algorithm, we empirically select the learning rate as the largest step length among $\{10^{-k} | k \in \mathbb{Z}\}$ that the method remains stable during the optimization process. Main results are shown in figure **??** where all figures are on log-log scale.

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

Preprint. Under review.



(a) (b)

(c) (d)

Figure 1: (a) Convergence