1 Introduction

We consider noisy gradient setting of the form

$$w_{n+1} = w_n - a_n(b_n w_n + c_n M_n)$$

where a_n is learning rate, M_n X, X is some distribution, $std(X) = 1, \quad \mathbb{E}X = 0.$

$$w_{n+1} = w_n - a_n b_n (w_n + \frac{c_n}{b_n} M_n)$$

$$w_{n+1} = w_n - \alpha_n (w_n + r_n M_n)$$

$$w_{n+1} = w_n (1 - \alpha_n + \frac{\alpha_n r_n M_n}{w_n})$$

Assumption, $\quad \alpha_n(-1 + \frac{r_n M_n}{w_n}) << 1$

$$w_{n+1} \approx w_n exp(-\alpha_n + \frac{\alpha_n r_n M_n}{w_n})$$

$$w_{n+1} = w_n exp(-\alpha_n) exp(\frac{\alpha_n r_n M_n}{w_n})$$

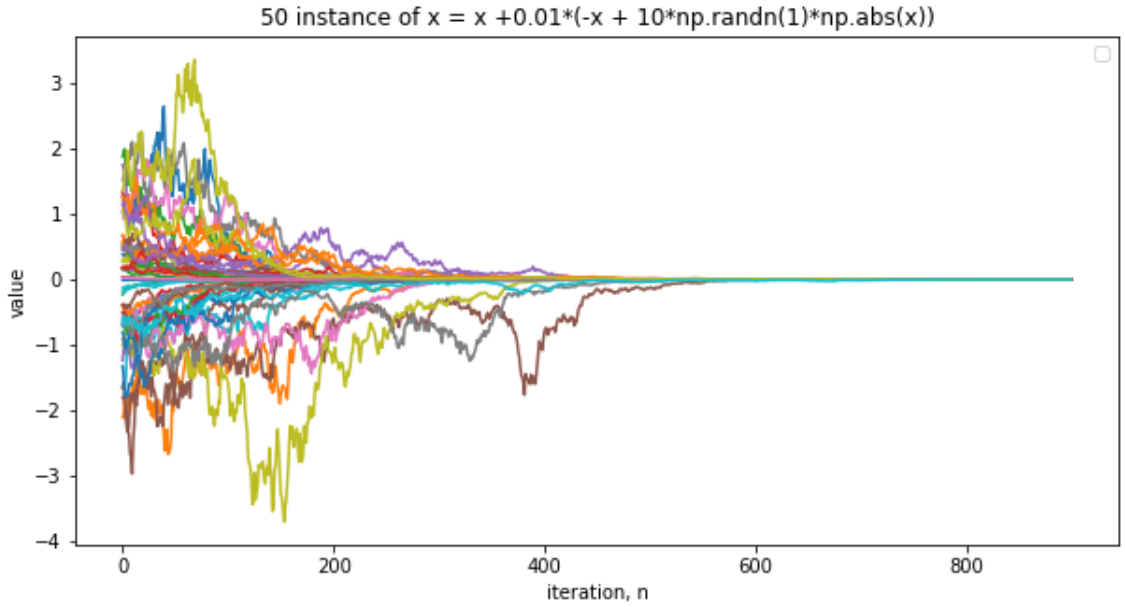
$$w_{n+1} = w_0 exp(-\Sigma_{i=0}^n \alpha_n) exp(\Sigma_{i=0}^n \frac{\alpha_n r_n M_n}{w_n})$$

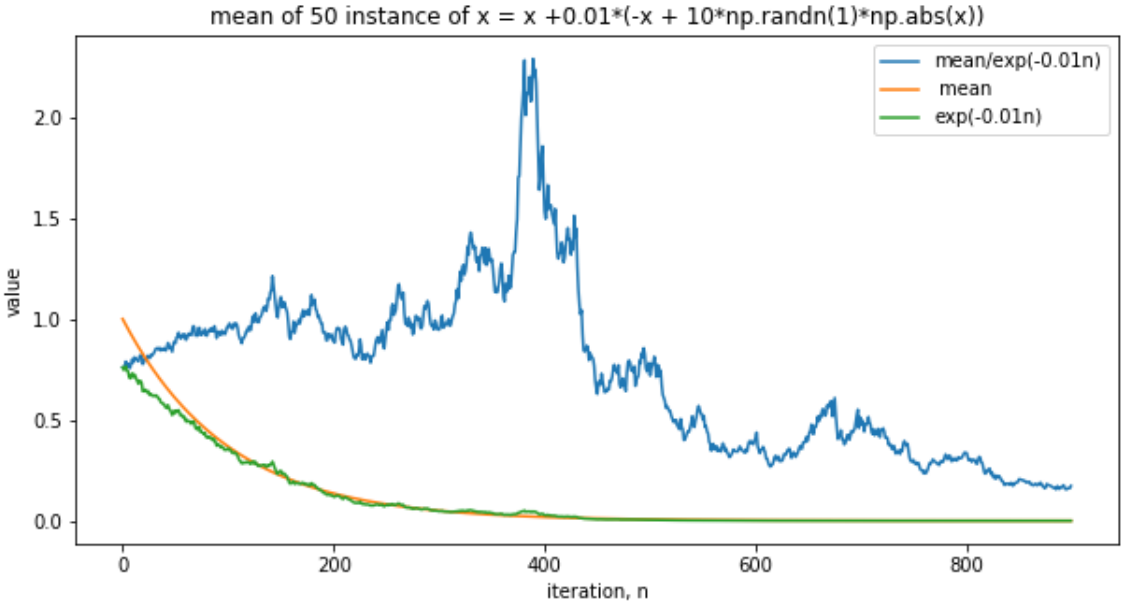
Assumption, $\quad |\frac{\alpha_n r_n}{w_n}| < b$

Hypothesis, $\quad \Sigma_{i=0}^n \frac{\alpha_n r_n M_n}{w_n} \to 0$

$$\Rightarrow w_{n+1} \to w_0 exp(-\Sigma_{i=0}^n \alpha_n)$$

2 Experiment 1





3 Experiment 2