Ordinary Differential Equation and Gradient Descent

October 1, 2024 → October 4, 2024

Abstract

This blog explains the relationship between Ordinary Differential Equation (ODE) and Gradient Descent (GD).

Contents

1 Introduction

This blog explains the relationship between Ordinary Differential Equation (ODE) and Gradient Descent (GD). Specifically, this blog explains the contents in Li et al. (2019), which is shown as follows.

> The stating motivation is the observation that GD iterations is a (Euler) discretization of the continuous-time, ordinary differential equation
>
> $$\frac{dx}{dt} = -\nabla f(x), \tag{1}$$
>
> and studying Eq. (1) can give us important insights to the dynamics of the discrete-time algorithm for small enough learning rates.

2 Preliminaries

I will try to use the definitions in the Wiki (for simplicity and convenience), and I will also provide some references for the formal definitions.

Definition (Ordinary Differential Equation)**.** In mathematics, an ordinary differential equation (ODE) is a differential equation dependent on only a single independent variable. (Ordinary differential equation)

The definition in Link (Euler method) is ok, yet I think the definition in Thomas et al. (2014) is more straightforward.

Definition (Euler's Method)**.** Given a differential equation $\frac{dy}{dx} = f(x, y)$ and an initial condition $y(x_0) = y_0$, we can approximate the solution $y = y(x)$ by its linearization

$$y_{n+1} = y(x_n) + y'(x_n)(x_{n+1} - x_n) \tag{2}$$
$$= y_n + f(x_n, y_n)(x_{n+1} - x_n) \tag{3}$$

where the second equation is due to the differential equation $\frac{dy}{dx} = f(x, y)$. (Thomas et al., 2014, Section 9.1 Solutions, Slope Fields, and Euler's Method)

3 Ordinary Differential Equation and Gradient Descent

Let us to get the numerical solution of the ODE $\frac{dx}{dt} = -\nabla f(x)$ (see Eq. (1)) by the Euler's method.

$$
\begin{aligned}
x_{n+1} &= x_n + \frac{dx_n}{dt}(t_{n+1} - t_n) \\
&= x_n - \nabla f(x_n) \underbrace{(t_{n+1} - t_n)}_{\text{step size: } \alpha} && (\because \text{Eq. (1)}) \\
&= x_n - \alpha \nabla f(x_n)
\end{aligned}
$$

where x_n is the numerical approximate solution. Here we let $t_{n+1} - t_n = \alpha$, which is the step size of the Euler's method. Note that the last equation is the standard process of GD.

References

Qianxiao Li, Cheng Tai, and E Weinan. Stochastic modified equations and dynamics of stochastic gradient algorithms i: Mathematical foundations. *Journal of Machine Learning Research*, 20(40):1–47, 2019.

George Brinton Thomas, Ross L Finney, Maurice D Weir, and Frank R Giordano. *Thomas' calculus*. 13 edition, 2014.