Neural Ordinary Differential Equations with Envolutionary Weights

No Author Given

No Institute Given

Lemma 1. *For every C^∞ curve in $\mathbb{R}^n$ without intersection, $s : [0,T] \to \mathbb{R}^n$, there is a differential equation $\frac{d\mathbf{x}}{dt} = F(\mathbf{x})$ such that:*

(1) $\mathbf{x}(t) = s(t) \quad \forall t \in [0,T]$;

(2) $F(\mathbf{x})$ *is Lipschitz continuous in* M*, where* M *is a compact set containing* s*.*

Proof. Let S denote the image of the curve s, which is a compact set in M. Since there is no intersection point in s, s is bijective. We can define a map $r : S \to \mathbb{R}^n$ such that:

$$r(\mathbf{x}) = \frac{ds}{dt}(s^{-1}(\mathbf{x})),$$

which is a C^∞ function. From the result of [4], there is a smooth funtion F defined on M, satisifing

$$r(\mathbf{x}) = F(\mathbf{x}), \quad \forall \mathbf{x} \in M.$$

So

$$F(s(t)) = r(s(t)) = \frac{ds}{dt}(s^{-1}(s(t))) = \frac{ds}{dt}.$$

As M is a compact set, the gradient of F is bounded. As a result, F is Lipschitz continuous.

Lemma 2. *Let F and $\bar{F} : M \to \mathbb{R}^n$ be Lipschitz continuous mappings and L be a Lispschitz constant of F. Suppose that for all $\mathbf{x} \in D$,*

$$|F(\mathbf{x}) - \bar{F}(\mathbf{x})| < \epsilon.$$

If $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are solutions to

$$\frac{d\mathbf{x}}{dt} = F(\mathbf{x}), \text{ and}$$

$$\frac{d\mathbf{y}}{dt} = \bar{F}(\mathbf{y}),$$

respectively, on some [0,T], such that $\mathbf{x}(0) = \mathbf{y}(0)$, then

$$|\mathbf{x}(t) - \mathbf{y}(t)| \leq \frac{\epsilon}{L}(\exp(L|t - t_0|) - 1)$$

holds, for all $t \in M$.

The proof of Lemma 2 can be found in [2].

Theorem 1. *Define a continuous function* $s : [0, T] \to \mathbb{R}^n$ *such that* $s(a) = s(b)$ *if and only if* $a = b$. *Given any* $\sigma > 0$, *there always exists a Neural ODE defined on the interval* $[0, T]$ *and its solution* $\mathbf{y}(t)$ *satisfies:* $|\mathbf{y}(t) - s(t)| \le \sigma, \quad \forall t \in [0, T]$.

Proof. The work in [1] inspires our proof roadmap. We take the neural network of Neural ODE in the following form:

$$f(\mathbf{x}) = \mathbf{A}\phi(\mathbf{B}\mathbf{x} + \boldsymbol{\theta}),$$

where $\mathbf{A}$ and $\mathbf{B}$ are weight matrices, $bm\theta$ is bias vector and ϕ is a smooth activation function.

By Stone–Weierstrass theorem, there is a smooth curve $s_1 : [0, T] \to \mathbb{R}^n$ such that:

$$|s_1(t) - s(t)| < \frac{\sigma}{2}.$$

From Lemma 1, $s_1(t)$ satisfies

$$\frac{ds_1}{dt} = F(s_1),$$

where F is Lipschitz continuous on a compact set M containing s. The Lipschitz constant of f is L. According to the approximation theorem [3]: there is an integer N and $n \times N$ matrix $\mathbf{A}$, $N \times n$ matrix $\mathbf{B}$, and $\boldsymbol{\theta}$ such that

$$|F(\mathbf{x}) - \mathbf{A}\phi(\mathbf{B}\mathbf{x} + \theta)| < \frac{\sigma L}{2(\exp(LT) - 1)}.$$

Let $\overline{F} = \mathbf{A}\phi(\mathbf{B}\mathbf{x} + \boldsymbol{\theta})$ and $\mathbf{y}(t)$ be the solution of the following equation:

$$\frac{d\mathbf{y}}{dt} = \overline{F}(\mathbf{y}),$$
$$\mathbf{y}(0) = s_1(0).$$

By Lemma 2, for any $t \in$ [0, T],

$$|s_1(t) - \mathbf{y}(t)| < \frac{\sigma L}{2(exp(LT) - 1)} \frac{\exp(Lt) - 1}{L} \le \frac{\sigma}{2},$$

So

$$|s(t) - \mathbf{y}(t)| < \sigma.$$

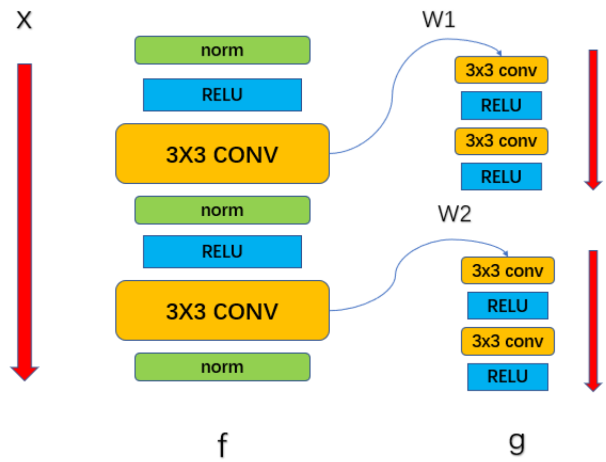

Fig. 1. NODE-EW. Norm means group normalization, f and g are both neural networks where the input of f is $\mathbf{x}$ while the input of g is the weight $\boldsymbol{\theta}$ of f. So g can be viewed as a hypernet.

References

1. Ken-ichi Funahashi and Yuichi Nakamura. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural networks*, 6(6):801–806, 1993.
2. Morris W Hirsch, Robert L Devaney, and Stephen Smale. Differential equations, dynamical systems, and linear algebra. 60, 1974.
3. Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.
4. Hassler Whitney. Analytic extensions of differentiable functions defined in closed sets. *Transactions of the American Mathematical Society*, 36(1):63–89, 1934.