Neural Ordinary Differential Equations

Oganesyan Viktor

Chen, Tian Qi, et al. "Neural ordinary differential equations."

Talk outline

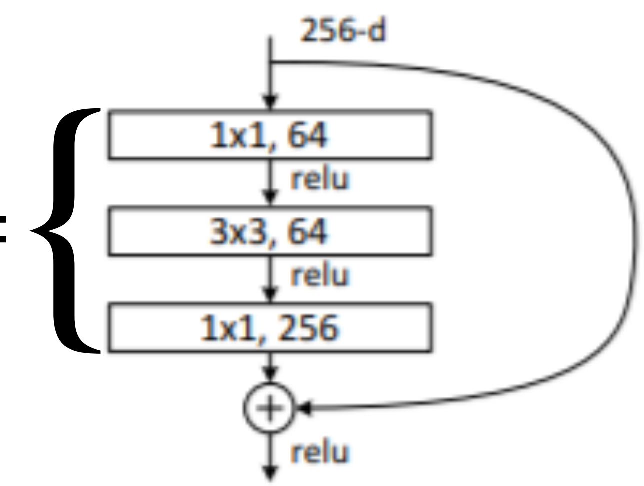
1. **Motivation (analogy with ResNet)**
2. **How to solve ODE**
3. **Training of the model**
4. **Adjoint method**
5. **Experiments**

Analogy with ResNet

$$z_{n+1} = z_n + f(z_n, \theta_n)$$

$$\frac{z_{n+1} - z_n}{(n+1) - n} = f(z_n, \theta_n)$$

$f =$



$$\frac{dz}{dt} = f(z(t), \theta(t), t)$$, where $\theta(t)$ – some function

$z(0) = z_{input}$ – input of the model

$z(T) = z_{output}$ – output of the model

How to solve ODE

$$\frac{dz}{dt} = -z^2 \qquad \left(\frac{dz}{dt}\right)_k = \frac{z_{k+1} - z_k}{h}$$

Explicit Euler method:

$$\frac{z_{k+1} - z_k}{h} = -z_k^2 \rightarrow z_{k+1} = z_k - hz_k^2$$

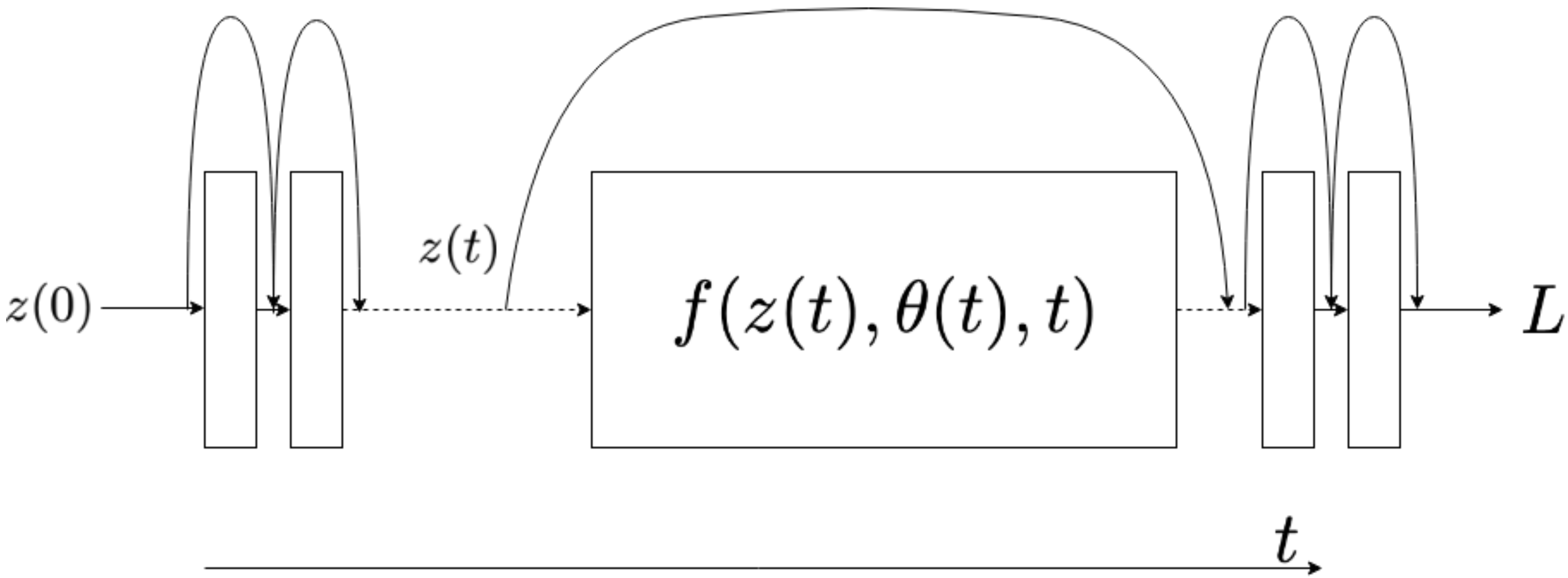
Implicit Euler method:

$$\frac{z_{k+1} - z_k}{h} = -z_{k+1}^2 \rightarrow z_{k+1} + hz_{k+1}^2 = z_k$$

Nonlinear equation

Training of the model

$$\frac{dz}{dt} = f(z(t), \theta(t), t) \qquad z(0) = z_{input} \qquad \theta(t) = \theta_0$$



$$L = L(z(T)) \qquad a_\theta(t) = \frac{\partial L}{\partial \theta(t)} = ? \qquad a_z(t) = \frac{\partial L}{\partial z(t)} = ?$$

Adjoint functions

$$a_\theta(t) = \frac{\partial L}{\partial \theta(t)} = ? \qquad a_z(t) = \frac{\partial L}{\partial z(t)} = ?$$

$$\frac{da_\theta}{dt} = -a_z \frac{\partial f(z(t), \theta(t), t)}{\partial \theta} \qquad a_\theta(T) = 0$$

$$\frac{da_z}{dt} = -a_z \frac{\partial f(z(t), \theta(t), t)}{\partial z} \qquad a_z(T) = \frac{\partial L}{\partial z(T)}$$

$$\frac{da_\theta}{dt} = \lim_{\epsilon \to 0} \frac{a_\theta(t+\epsilon) - a_\theta(t)}{\epsilon} \qquad \frac{da_z}{dt} = \lim_{\epsilon \to 0} \frac{a_z(t+\epsilon) - a_z(t)}{\epsilon}$$

Adjoint method

$$a_z(t) = \frac{\partial L}{\partial z(t)} = \frac{\partial L}{\partial z(t+\epsilon)}\frac{\partial z(t+\epsilon)}{\partial z(t)} + \frac{\partial L}{\partial \theta(t+\epsilon)}\frac{\partial \theta(t+\epsilon)}{\partial z(t)}$$

$$a_\theta(t) = \frac{\partial L}{\partial \theta(t)} = \frac{\partial L}{\partial z(t+\epsilon)}\frac{\partial z(t+\epsilon)}{\partial \theta(t)} + \frac{\partial L}{\partial \theta(t+\epsilon)}\frac{\partial \theta(t+\epsilon)}{\partial \theta(t)}$$

Adjoint method

$$a_z(t) = \frac{\partial L}{\partial z(t)} = \frac{\partial L}{\partial z(t+\epsilon)} \frac{\partial z(t+\epsilon)}{\partial z(t)} + \frac{\partial L}{\partial \theta(t+\epsilon)} \frac{\partial \theta(t+\epsilon)}{\partial z(t)}$$

$$a_\theta(t) = \frac{\partial L}{\partial \theta(t)} = \frac{\partial L}{\partial z(t+\epsilon)} \frac{\partial z(t+\epsilon)}{\partial \theta(t)} + \frac{\partial L}{\partial \theta(t+\epsilon)} \frac{\partial \theta(t+\epsilon)}{\partial \theta(t)}$$

$$z(t+\epsilon) = z(t) + \epsilon \frac{dz}{dt} + O(\epsilon^2) = z(t) + \epsilon f(z(t), \theta(t), t) + O(\epsilon^2) \rightarrow$$

$$\frac{\partial z(t+\epsilon)}{\partial z(t)} = 1 + \epsilon \frac{\partial f(z(t), \theta(t), t)}{\partial z} + O(\epsilon^2) \qquad \frac{\partial z(t+\epsilon)}{\partial \theta(t)} = \epsilon \frac{\partial f(z(t), \theta(t), t)}{\partial \theta} + O(\epsilon^2)$$

Adjoint method

$$a_z(t) = \frac{\partial L}{\partial z(t)} = \frac{\partial L}{\partial z(t+\epsilon)} \frac{\partial z(t+\epsilon)}{\partial z(t)} + \frac{\partial L}{\partial \theta(t+\epsilon)} \frac{\partial \theta(t+\epsilon)}{\partial z(t)}$$

$$a_\theta(t) = \frac{\partial L}{\partial \theta(t)} = \frac{\partial L}{\partial z(t+\epsilon)} \frac{\partial z(t+\epsilon)}{\partial \theta(t)} + \frac{\partial L}{\partial \theta(t+\epsilon)} \frac{\partial \theta(t+\epsilon)}{\partial \theta(t)}$$

$$z(t+\epsilon) = z(t) + \epsilon \frac{dz}{dt} + O(\epsilon^2) = z(t) + \epsilon f(z(t), \theta(t), t) + O(\epsilon^2) \rightarrow$$

$$\frac{\partial z(t+\epsilon)}{\partial z(t)} = 1 + \epsilon \frac{\partial f(z(t), \theta(t), t)}{\partial z} + O(\epsilon^2) \qquad \frac{\partial z(t+\epsilon)}{\partial \theta(t)} = \epsilon \frac{\partial f(z(t), \theta(t), t)}{\partial \theta} + O(\epsilon^2)$$

$$\theta(t+\epsilon) = \theta(t) \rightarrow \qquad \frac{\partial \theta(t+\epsilon)}{\partial z(t)} = 0 \qquad \frac{\partial \theta(t+\epsilon)}{\partial \theta(t)} = 1$$

Adjoint method

$$\frac{da_z}{dt} = \lim_{\epsilon \to 0} \frac{a_z(t+\epsilon) - a_z(t)}{\epsilon} \qquad \frac{da_\theta}{dt} = \lim_{\epsilon \to 0} \frac{a_\theta(t+\epsilon) - a_\theta(t)}{\epsilon}$$

$$a_z(t) = a_z(t+\epsilon)(1 + \epsilon \frac{\partial f(z(t), \theta(t), t)}{\partial z} + O(\epsilon^2))$$

$$a_\theta(t) = a_z(t+\epsilon)(\epsilon \frac{\partial f(z(t), \theta(t), t)}{\partial \theta} + O(\epsilon^2)) + a_\theta(t+\epsilon)$$

$$\frac{da_z}{dt} = -a_z(t) \frac{\partial f(z(t), \theta(t), t)}{\partial z} \qquad \frac{da_\theta}{dt} = -a_z(t) \frac{\partial f(z(t), \theta(t), t)}{\partial \theta}$$

Final algorithm

Forward:

$$z(T) = z_{input} + \int_0^T f(z(\tau), \theta_0, \tau) d\tau$$

$$\int_0^T \left(\ldots \right) d\tau =$$ **ODE solver**

Back:

$$\frac{\partial L}{\partial z(t)} = \frac{\partial L}{\partial z(T)} - \int_T^t a_z(\tau) \frac{\partial f(z(\tau), \theta_0, \tau)}{\partial z} d\tau$$

$$\frac{\partial L}{\partial \theta(t)} = - \int_T^t a_z(\tau) \frac{\partial f(z(\tau), \theta_0, \tau)}{\partial \theta} d\tau$$

Final algorithm

Forward:

$$z(T) = z_{input} + \int_0^T f(z(\tau), \theta_0, \tau) d\tau$$

$$\int_0^T \Big(\dots \Big) d\tau =$$ **ODE solver**

Back:

$$\frac{\partial L}{\partial z(t)} = \frac{\partial L}{\partial z(T)} - \int_T^t a_z(\tau) \frac{\partial f(z(\tau), \theta_0, \tau)}{\partial z} d\tau$$

$$\frac{\partial L}{\partial \theta(t)} = - \int_T^t a_z(\tau) \frac{\partial f(z(\tau), \theta_0, \tau)}{\partial \theta} d\tau$$

$$z(t) = z(T) + \int_T^t f(z(\tau), \theta_0, \tau) d\tau$$

Final algorithm

Forward:

$$z(T) = z_{input} + \int_0^T f(z(\tau), \theta_0, \tau) d\tau$$

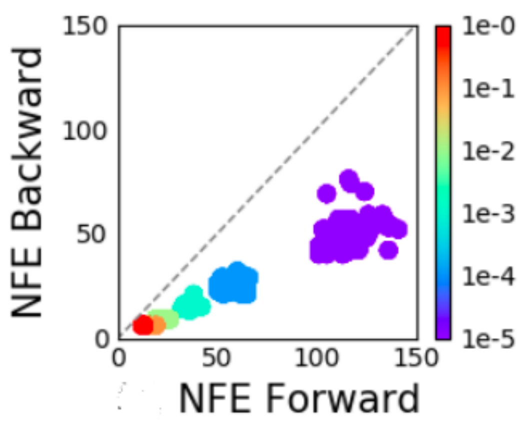
$$\int_0^T \Big(\dots \Big) d\tau =$$ **ODE solver**

Back:

$$\frac{\partial L}{\partial z(t)} = \frac{\partial L}{\partial z(T)} - \int_T^t a_z(\tau) \frac{\partial f(z(\tau), \theta_0, \tau)}{\partial z} d\tau$$

$$\frac{\partial L}{\partial \theta(t)} = - \int_T^t a_z(\tau) \frac{\partial f(z(\tau), \theta_0, \tau)}{\partial \theta} d\tau$$

$$z(t) = z(T) + \int_T^t f(z(\tau), \theta_0, \tau) d\tau$$



Experiments

MNIST:

	Test error	# Params	Memory	Time
ResNet(6)	0.41 %	0.60 M	$O(L)$	$O(L)$
ODE-Net	0.42 %	0.22 M	$O(1)$	$O(\hat{L})$
ResNet(1)	0.42 %	0.22 M	$O(L)$	$O(L)$

L- number of layers, $\hat{L}$ - number of function evaluations (implicit number of layers)

Experiments

CIFAR10 (160 epoch):

	Test accuracy	# Params	Time, sec
ResNet(6)	85.43 %	0.60 M	3698
ODE-Net	83.90 %	0.22 M	10763
ResNet(1)	83.62 %	0.22 M	1498
ODE-Net (no time)	82.37 %	0.22 M	10643