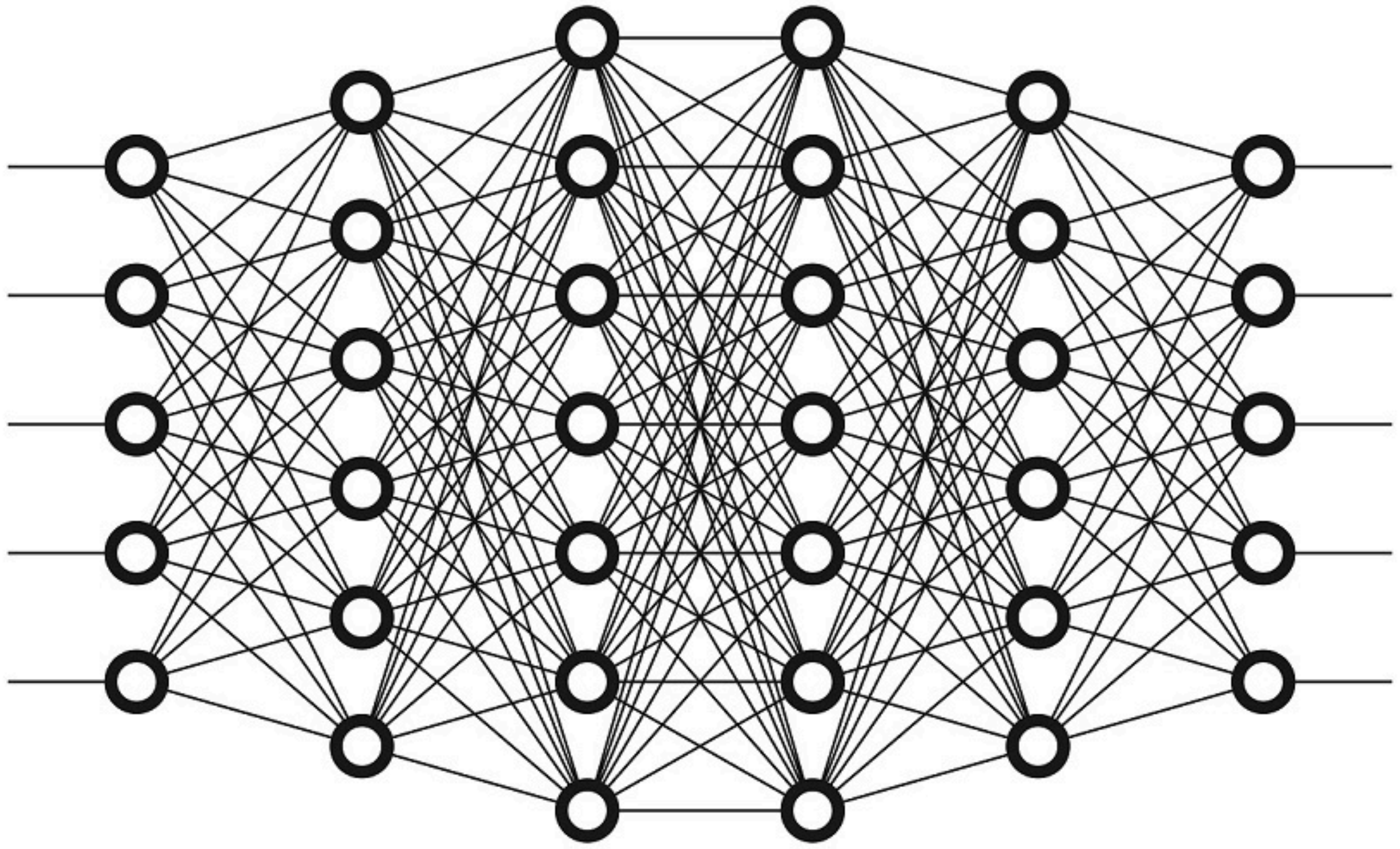
Neural Networks with Physical Inductive Biases

Maxwell Cai
maxwellcai.com

$$\frac{d\Psi}{dx} = f(x)$$



“Since Newton, mankind has come to realise that the laws of physics are always expressed in the language of differential equations.”

–Steven Strogatz

Why Differential Equations?

It is often easier to describe **changes** than absolute amount.

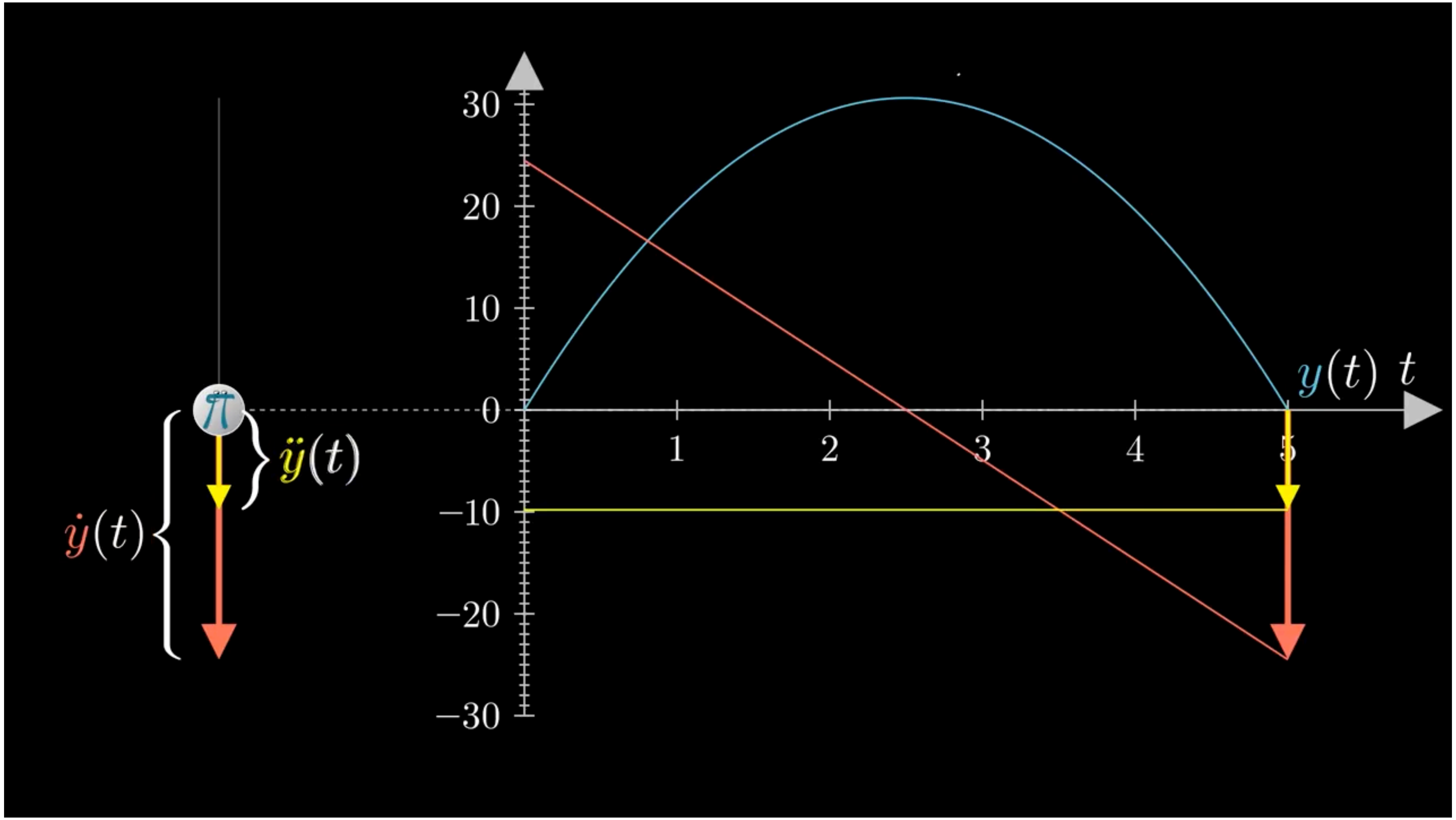
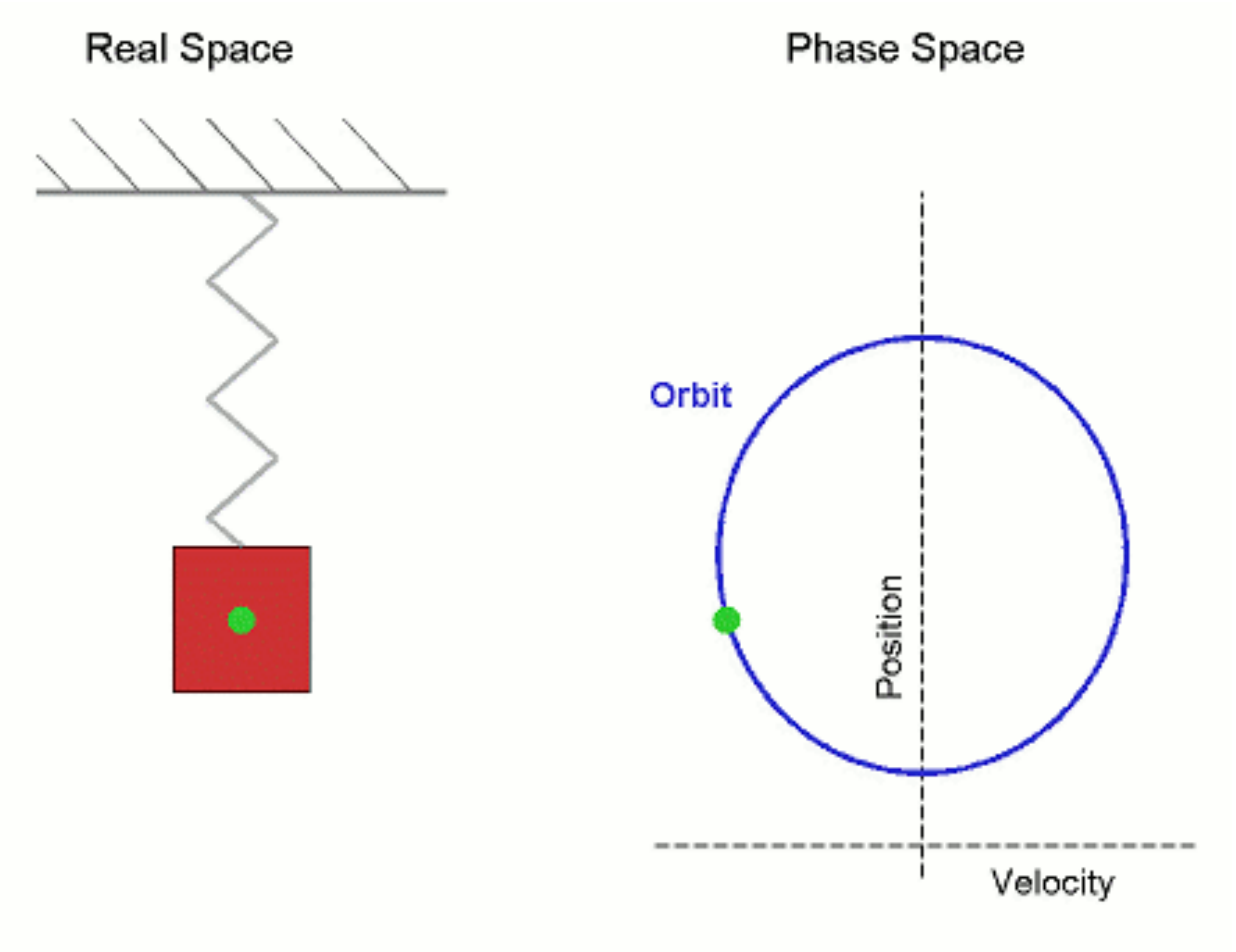


Image: 3Blue1Brown

Example: Harmonic Oscillator



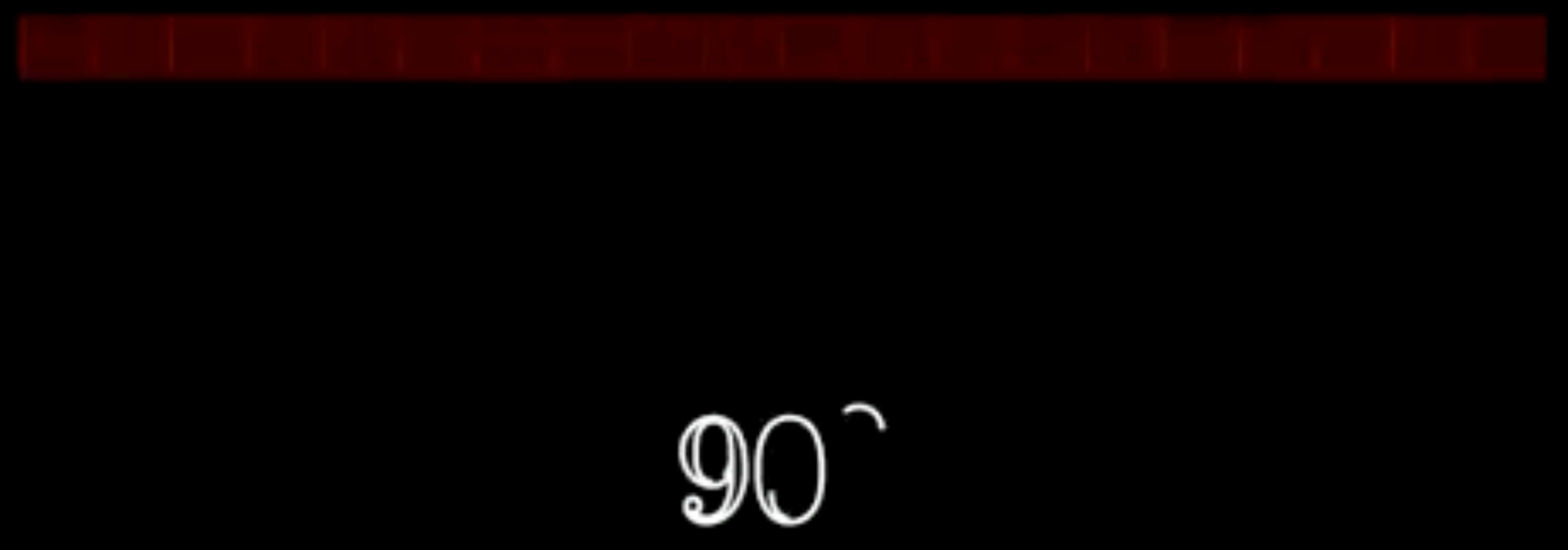
Equation of motion:

$$F = ma = m\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = m\ddot{x} = -kx$$

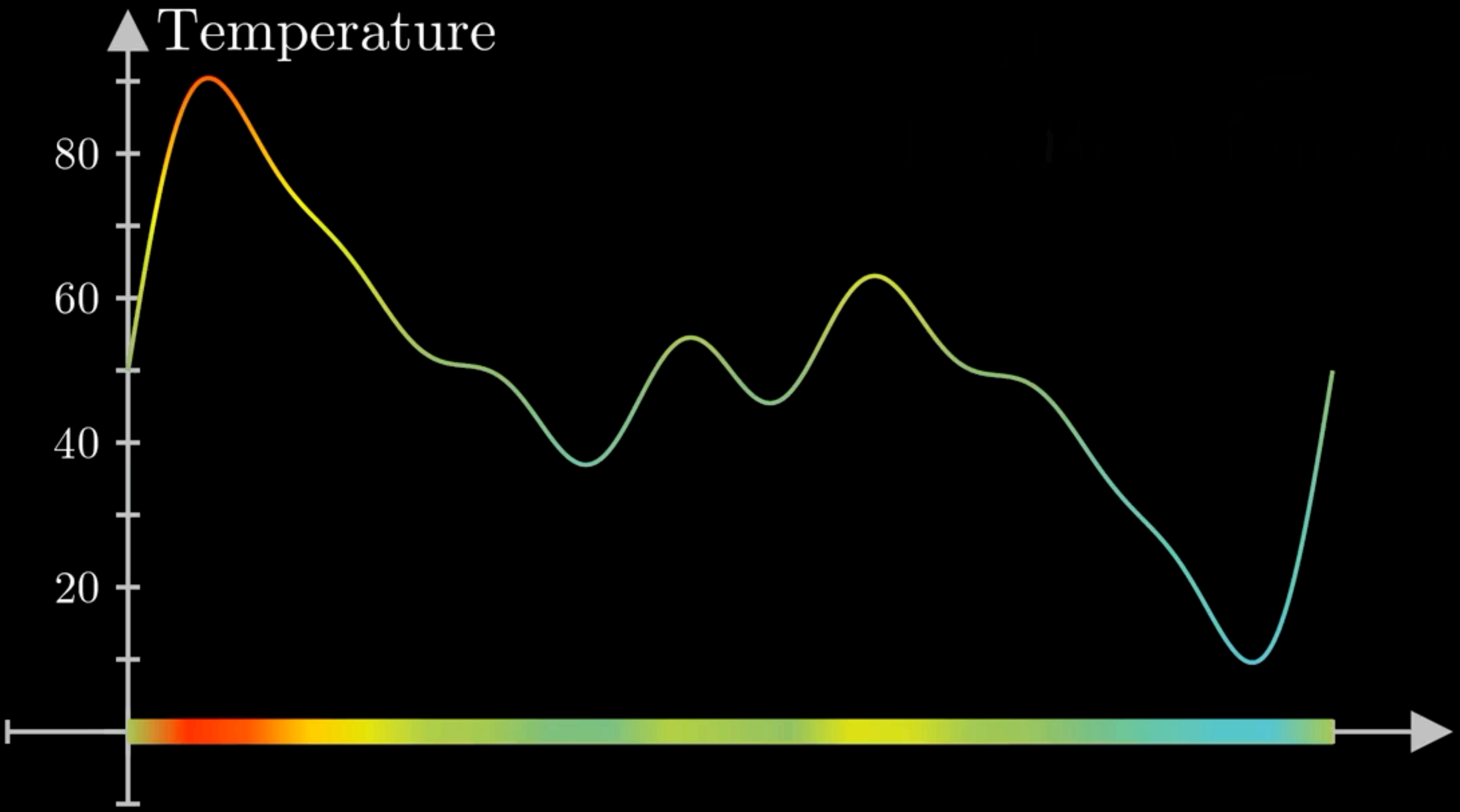
Solution:

$$x(t) = A\cos(\omega t + \phi)$$

$$\omega = \sqrt{\frac{k}{m}}$$



Movie: 3Blue1Brown



Movie: 3Blue1Brown

Physics Informed Neural Networks (PINNs)

Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations

Maziar Raissi[1], Paris Perdikaris[2], and George Em Karniadakis[1]

[1] *Division of Applied Mathematics, Brown University, Providence, RI, 02912, USA*

[2] *Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA, 19104, USA*

Raissi et al. (2017a), arXiv: 1711.10561

Physics Informed Neural Networks (PINNs)

Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations

Maziar Raissi[1], Paris Perdikaris[2], and George Em Karniadakis[1]

[1] *Division of Applied Mathematics, Brown University, Providence, RI, 02912, USA*

[2] *Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA, 19104, USA*

Raissi et al. (2017a), arXiv: 1711.10566

Example: Burger's Equation

$$u_t + uu_x - (0.01/\pi)u_{xx} = 0, \quad x \in [-1, 1], \quad t \in [0, 1],$$
$$u(0, x) = -\sin(\pi x),$$
$$u(t, -1) = u(t, 1) = 0.$$

```
def u(t, x):
    u = neural_net(tf.concat([t,x],1), weights, biases)
    return u
```

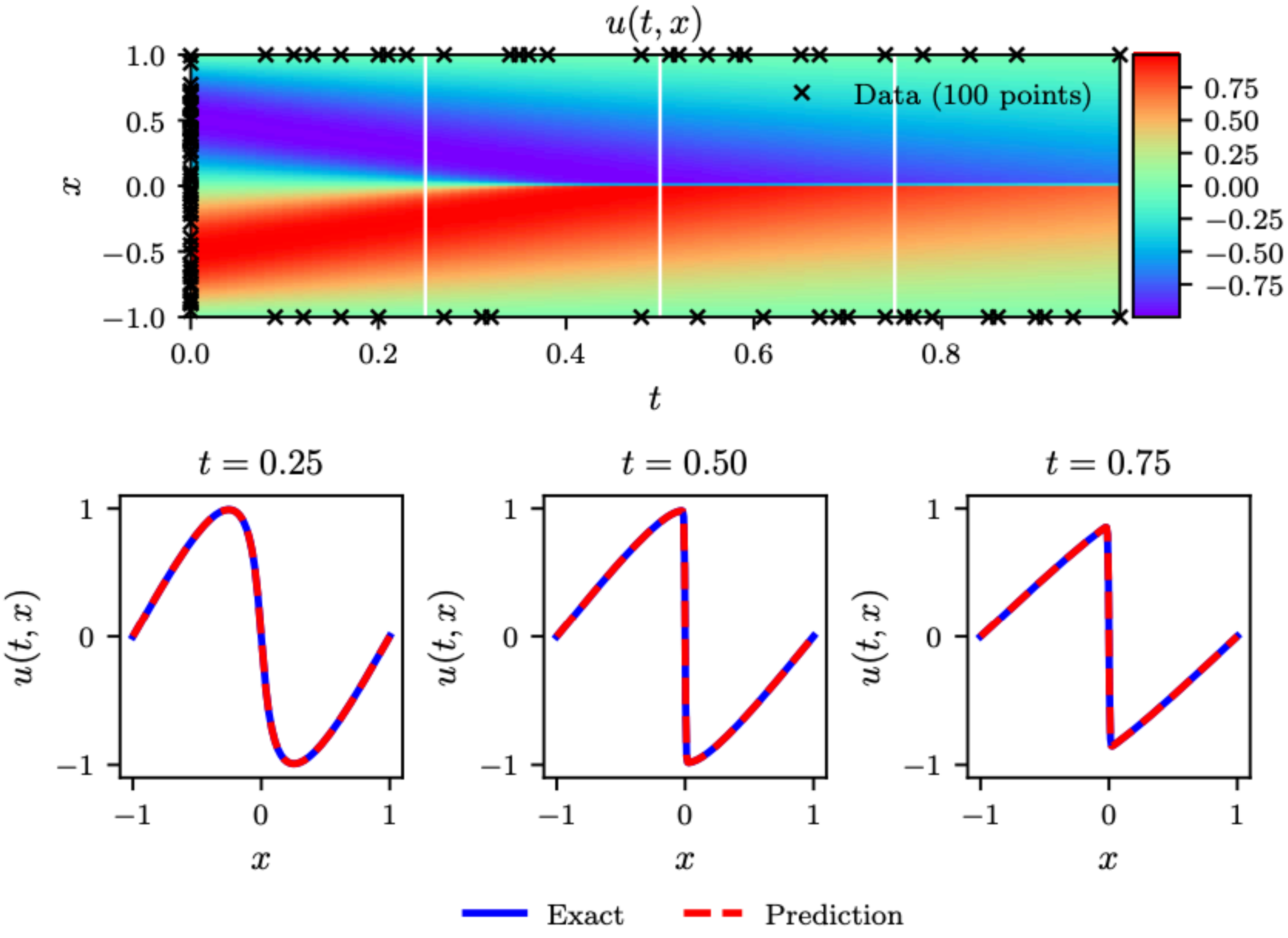
```
def f(t, x):
    u = u(t, x)
    u_t = tf.gradients(u, t)[0]
    u_x = tf.gradients(u, x)[0]
    u_xx = tf.gradients(u_x, x)[0]
    f = u_t + u*u_x - (0.01/tf.pi)*u_xx
    return f
```

$$MSE = MSE_u + MSE_f,$$

$$MSE_u = \frac{1}{N}\sum_{i=1}^{N} |u(t_u^i, x_u^i) - u^i|^2,$$

$$MSE_f = \frac{1}{N}\sum_{i=1}^{N} |f(t_u^i, x_u^i)|^2.$$

PINNs are data efficient



PDE Discovery

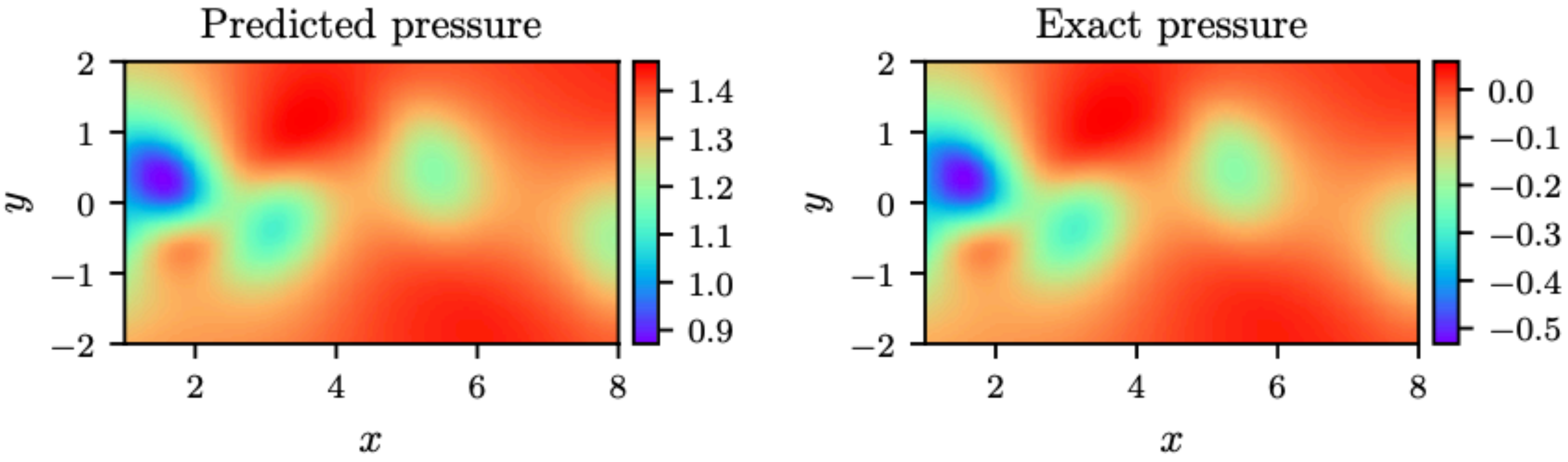
PDE general form (i.e., parameterised by λ):

$$u_t + \mathcal{N}\left[u; \lambda\right] = 0, x \in \Omega, t \in [0,T]$$

Given a small set of scattered and potentially noisy observations of the hidden state $u(t, x)$ of a system, what are the parameters λ that best describe the observed data?

Use SGD or L-BFGS to find the optimal value of λ.

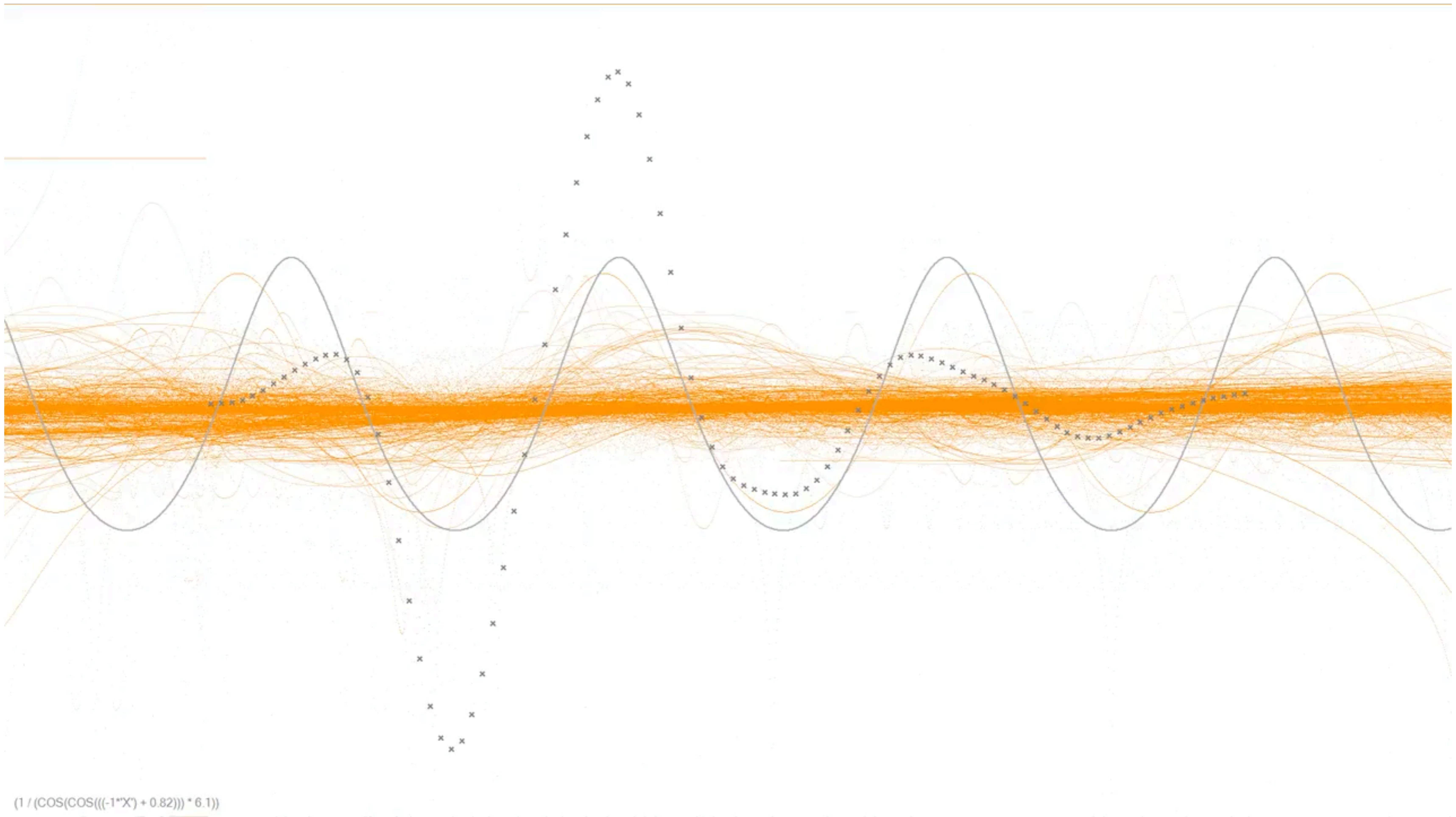
Example: Navier-Stokes Equation

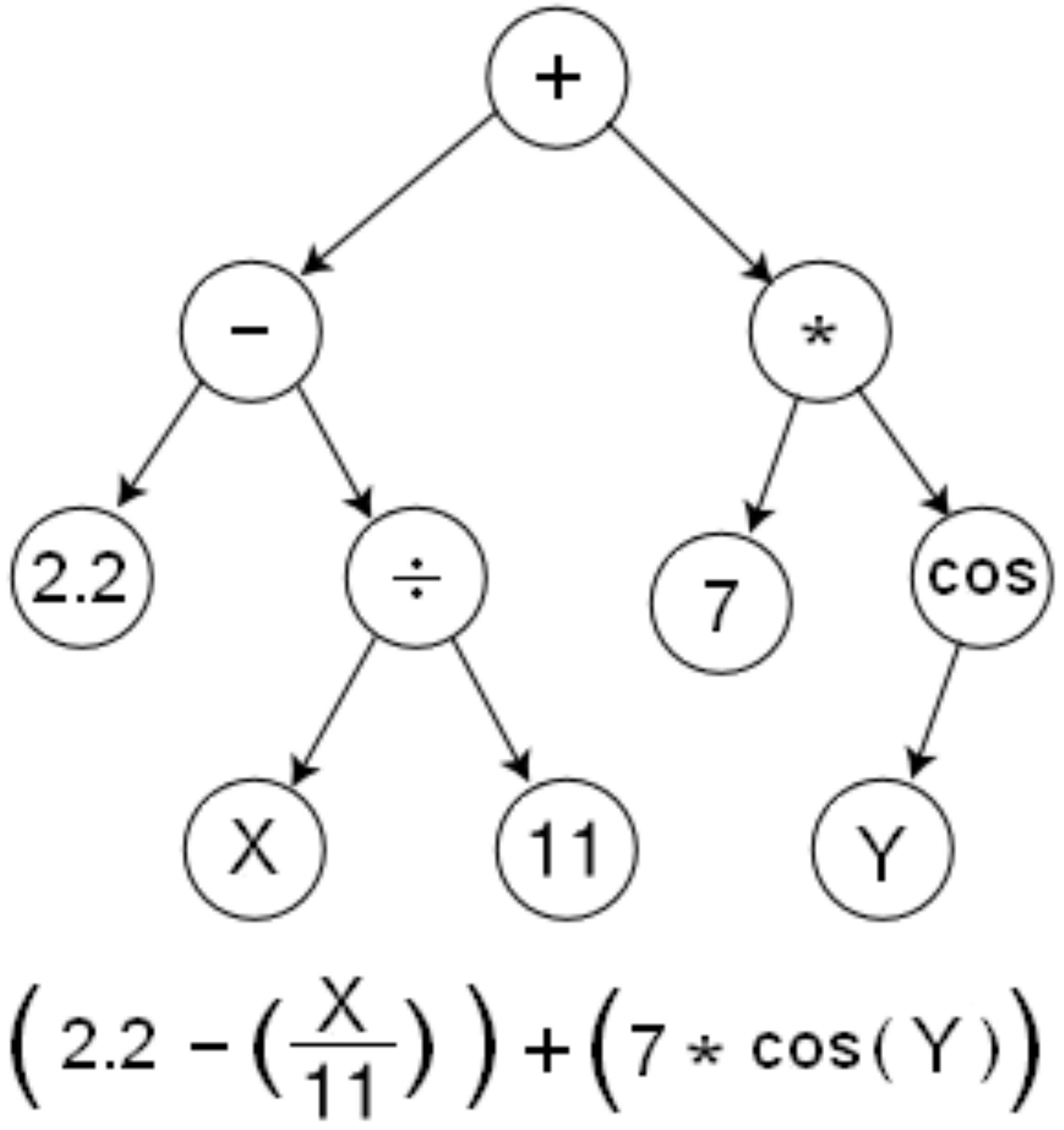


Correct PDE	$u_t + (uu_x + vu_y) = -p_x + 0.01(u_{xx} + u_{yy})$ $v_t + (uv_x + vv_y) = -p_y + 0.01(v_{xx} + v_{yy})$
Identified PDE (clean data)	$u_t + 0.999(uu_x + vu_y) = -p_x + 0.01047(u_{xx} + u_{yy})$ $v_t + 0.999(uv_x + vv_y) = -p_y + 0.01047(v_{xx} + v_{yy})$
Identified PDE (1% noise)	$u_t + 0.998(uu_x + vu_y) = -p_x + 0.01057(u_{xx} + u_{yy})$ $v_t + 0.998(uv_x + vv_y) = -p_y + 0.01057(v_{xx} + v_{yy})$

Symbolic Regression

Objective: Find the model that best describes the data





Expression trees

Movie: HeuristicLab
Image: Wikipedia

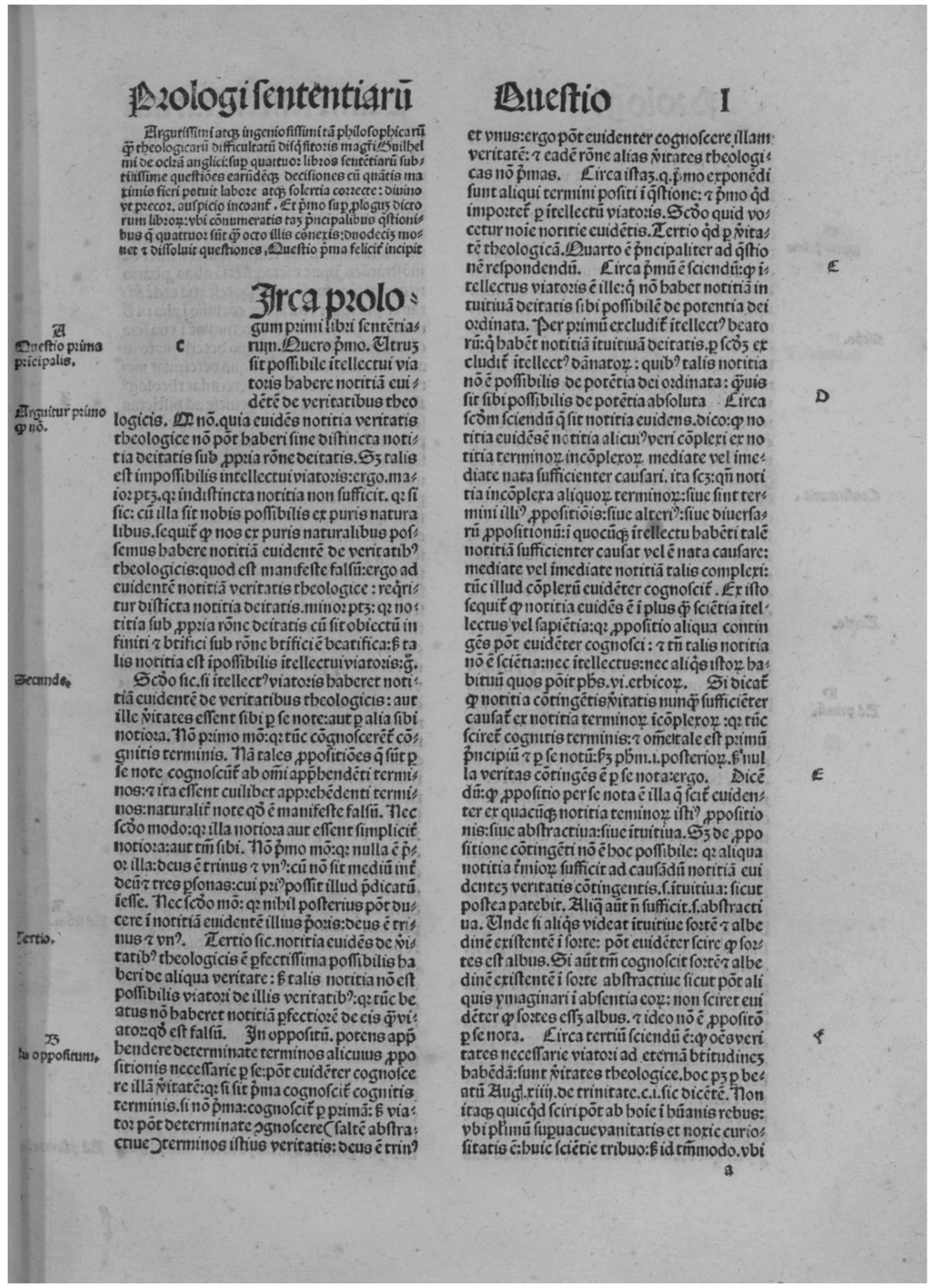
大道至簡

Laozi (aka. Lao-Tze)
ca. 6 century B.C.

Occam’s Razor (*novacula Occami*)

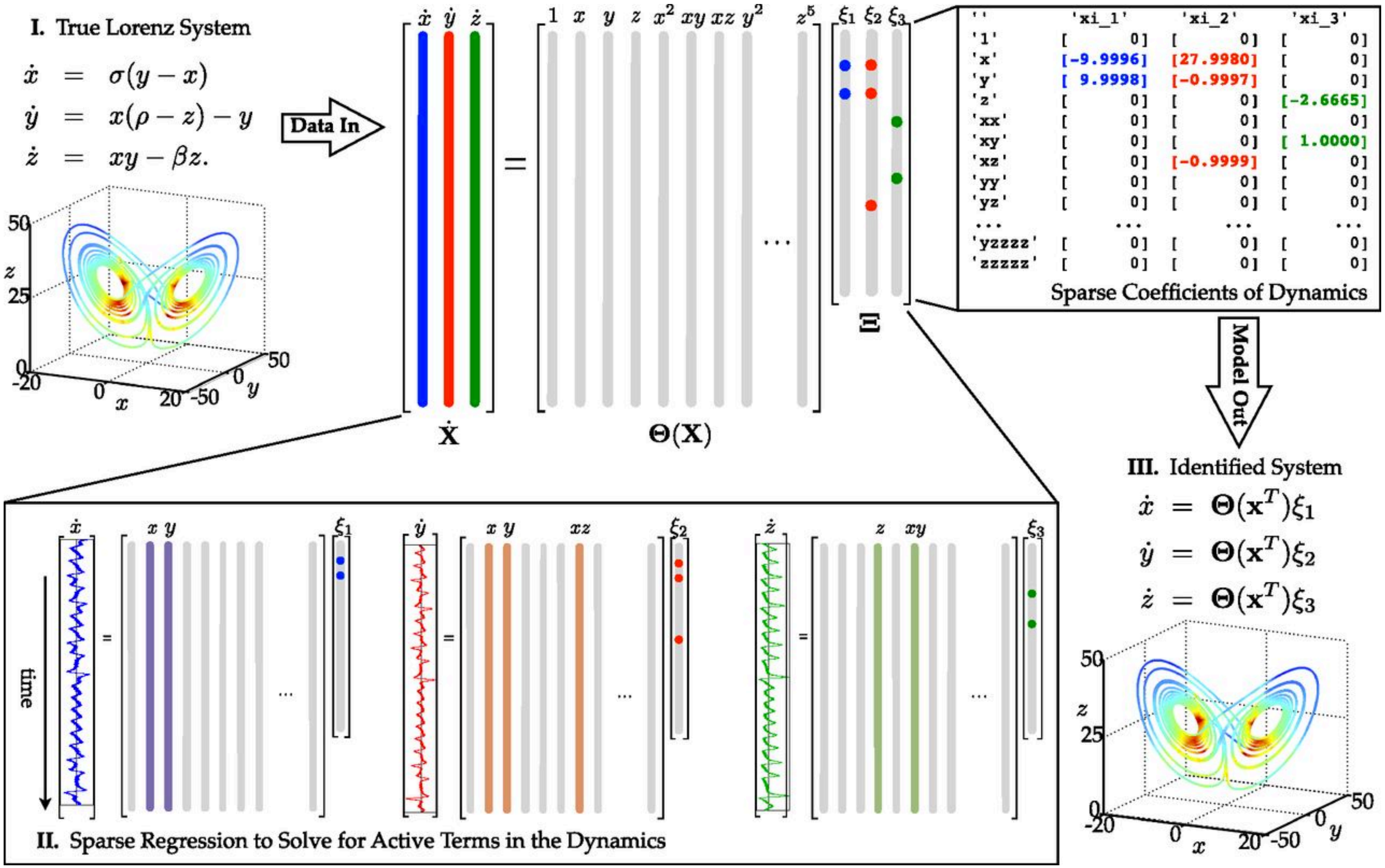
“Entities are not to be multiplied beyond necessity”

(*Entia non sunt multiplicanda praeter necessitatem)*


Prologi sententiarū — Questio I

William of Ockham (c. 1287 — 1347)

Sparse Identification of Nonlinear Dynamics (SINDy)



Brunton et al. (2016, PNAS)

Hamiltonian Mechanics

Hamiltonian Formulation

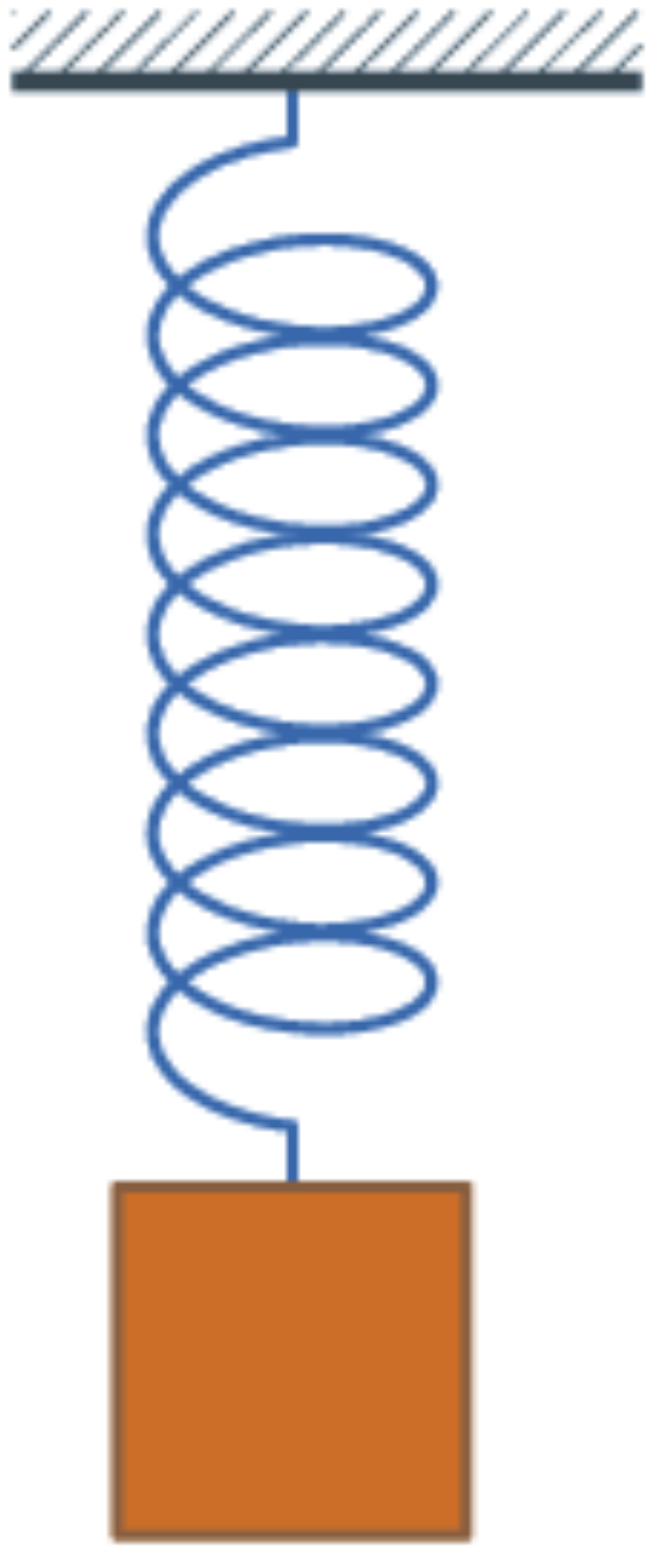
$$\mathscr{H} = T + V$$

$$\dot{\mathbf{q}} \equiv \frac{d\mathbf{q}}{dt} = +\frac{\partial \mathscr{H}}{\partial \mathbf{p}}, \qquad \dot{\mathbf{p}} \equiv \frac{d\mathbf{p}}{dt} = -\frac{\partial \mathscr{H}}{\partial \mathbf{q}}.$$

Example: Ideal mass spring (harmonic oscillator)

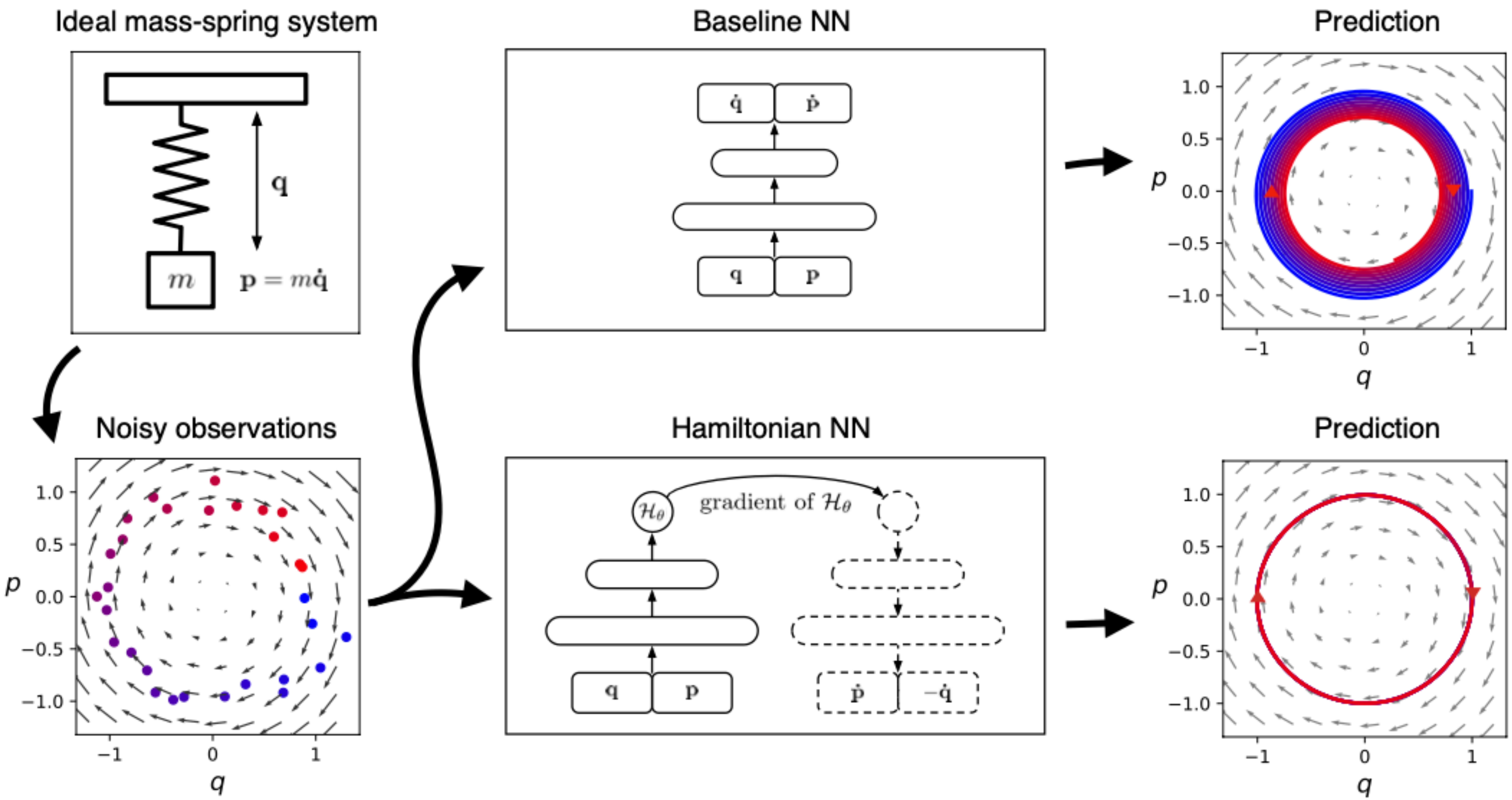
$$\mathscr{H} = T + V = \frac{p^2}{2m} + \frac{k}{2}q^2$$

$$\frac{\partial \mathscr{H}}{\partial q} = kq \qquad \frac{\partial \mathscr{H}}{\partial p} = \frac{p}{m}$$



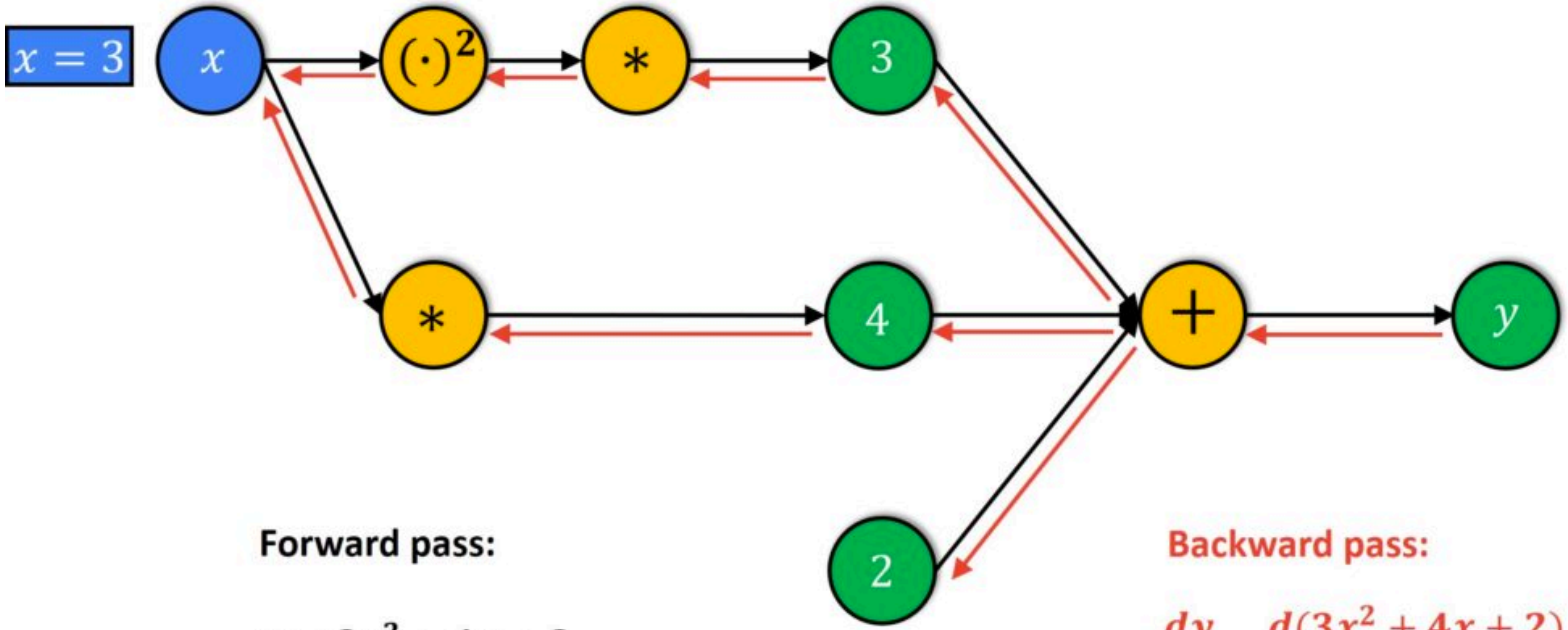
Sir William Hamilton (1805–1865)

Application: Hamiltonian Neural Networks



Greydanus et al. (2019)

Autograd



Forward pass:

$$y = 3x^2 + 4x + 2$$

$$y = 3 \cdot 9 + 4 \cdot 3 + 2$$

$$y = 41$$

Backward pass:

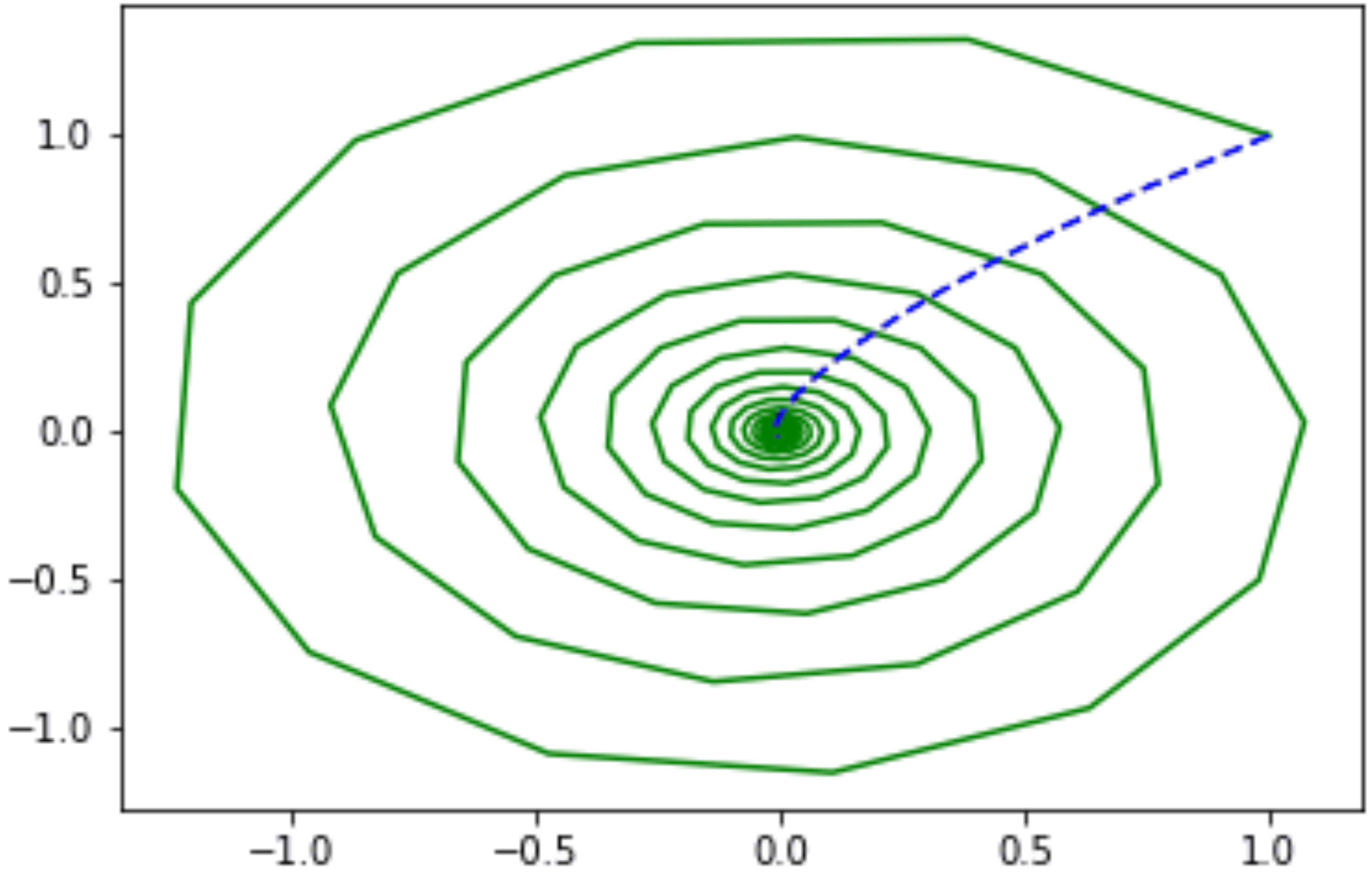
$$\frac{dy}{dx} = \frac{d(3x^2 + 4x + 2)}{dx}$$

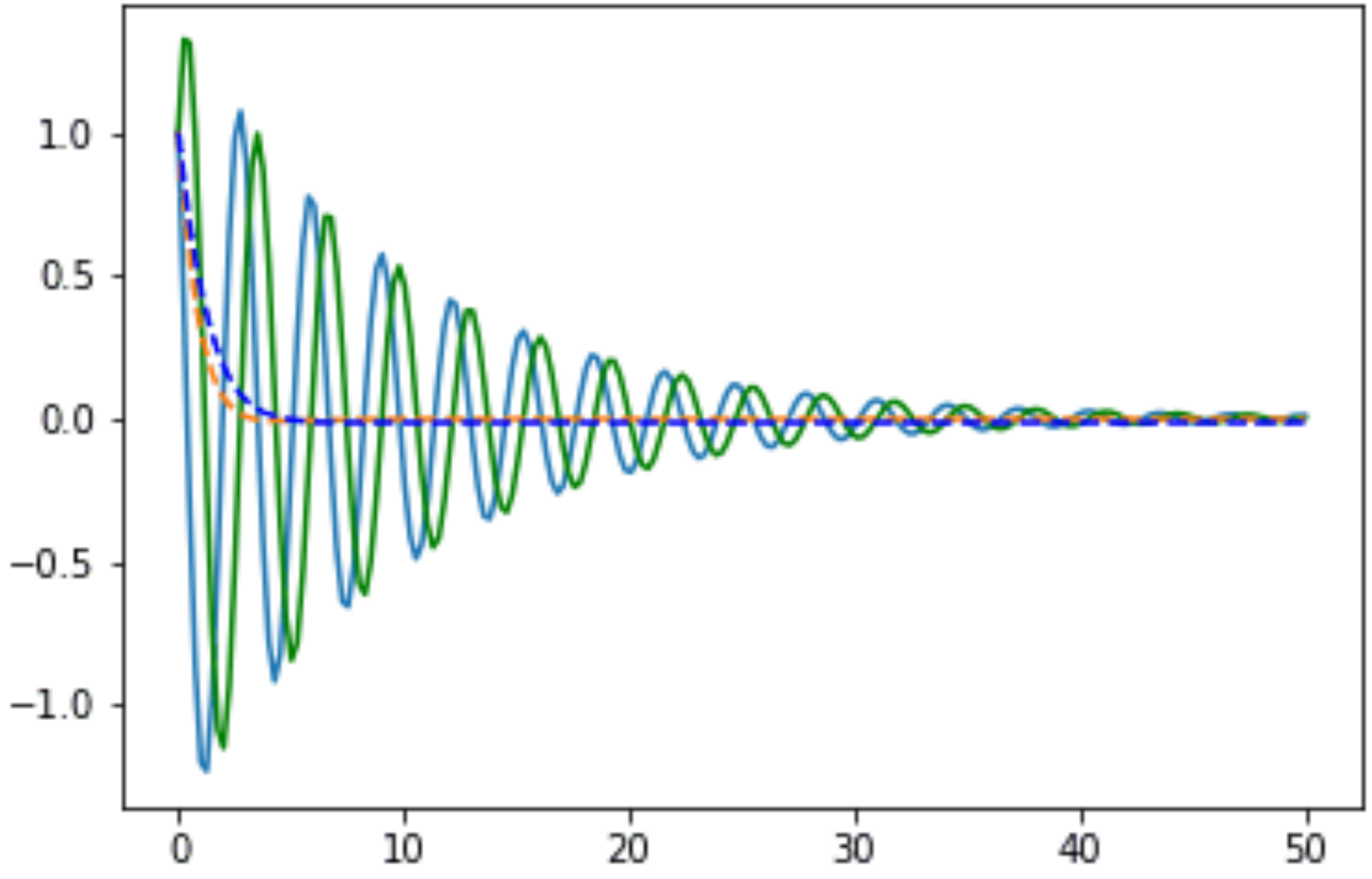
$$\frac{dy}{dx} = 2 \cdot 3x + 4 = 6x + 4$$

$$\frac{dy}{dx} = 18 + 4 = 22$$

Image: DataHacker.rs

Neural ODE Solvers



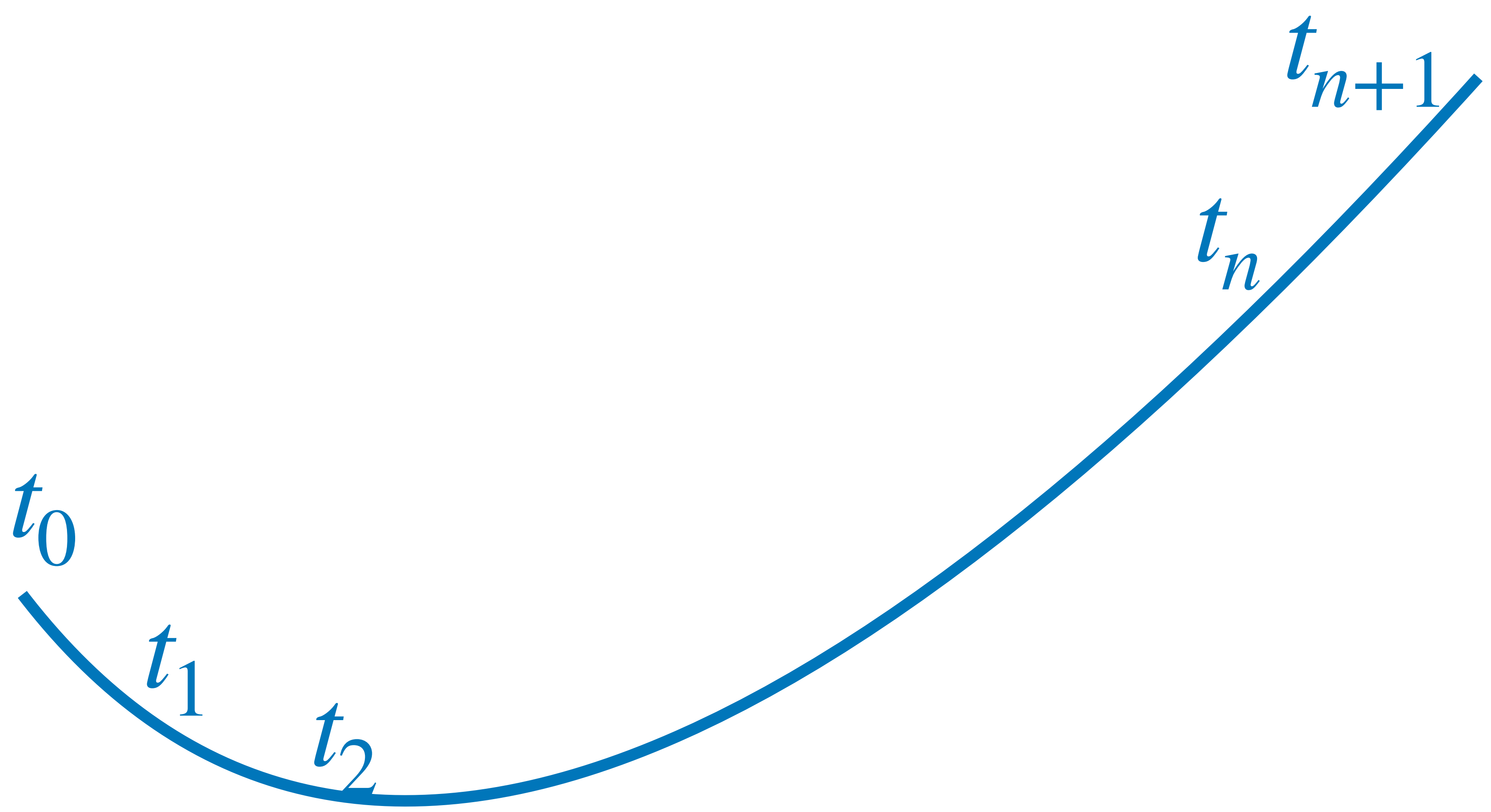


Chen et al. (2018)

From PINNs to NDEs

Neural Differential Equations

$$\frac{d\mathbf{h}(t)}{dt} = f\left(\mathbf{h}(t), t, \theta\right)$$



Initial value problem (IVP)

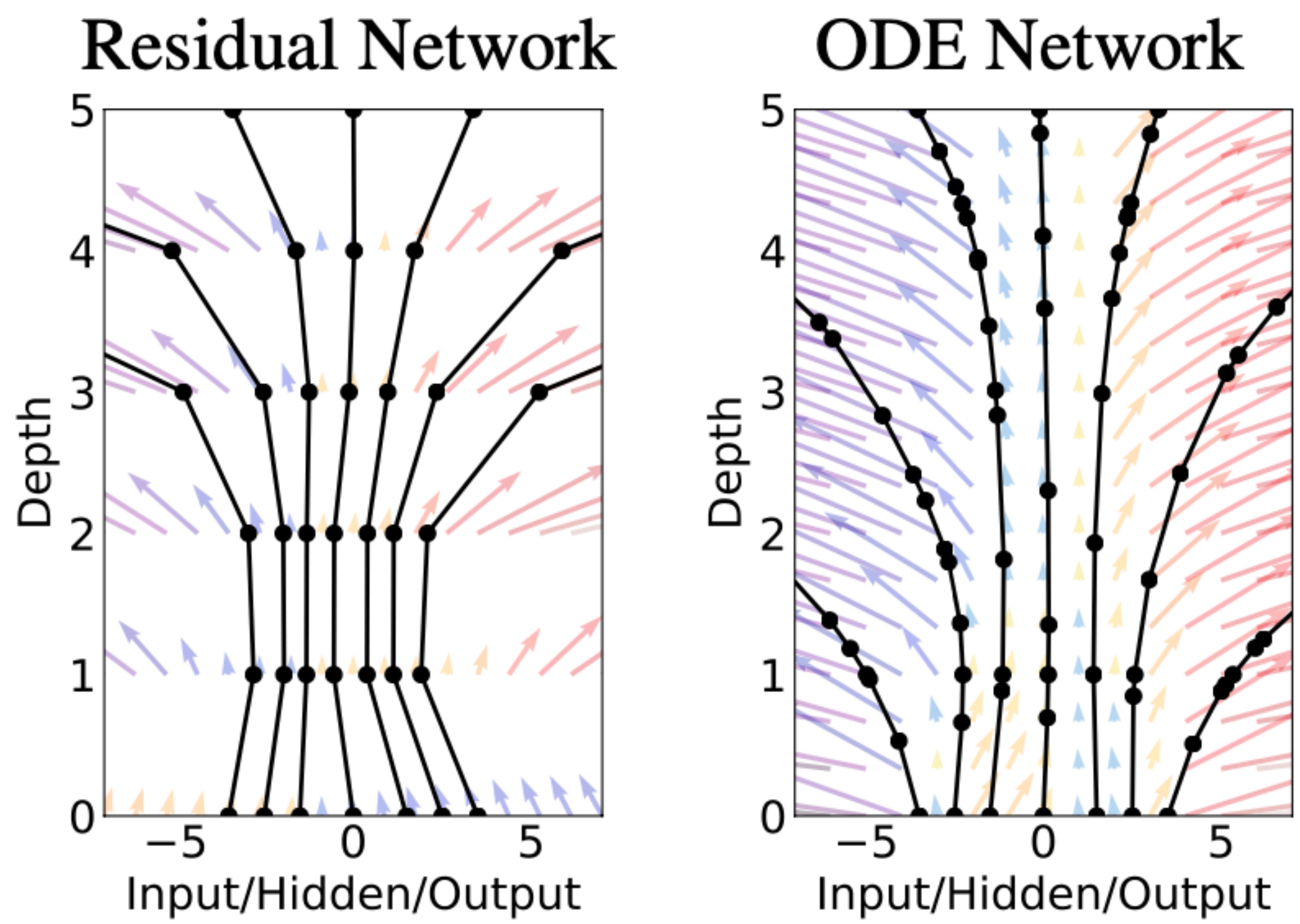
$$\frac{dy}{dt} = f(t, y) \qquad y(t_0) = y_0$$

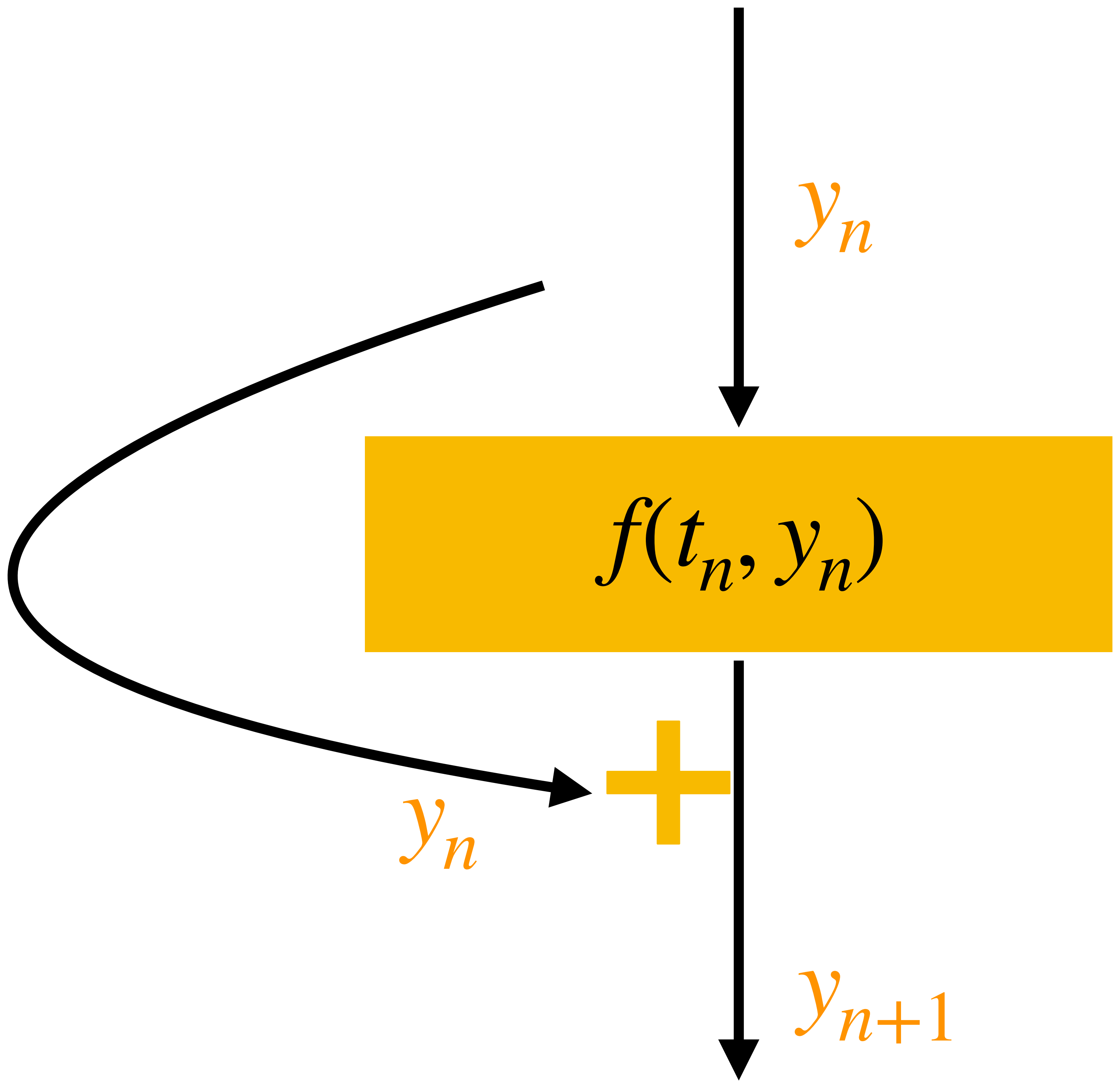
Integrate (Euler)

$$y_{n+1} = y_n + f\left(t_n, y_n\right)\Delta t \qquad \Delta t \equiv t_{n+1} - t_n$$

Beyond NDE…

$$y_{n+1} = y_n + f\left(t_n, y_n\right) \Delta t$$



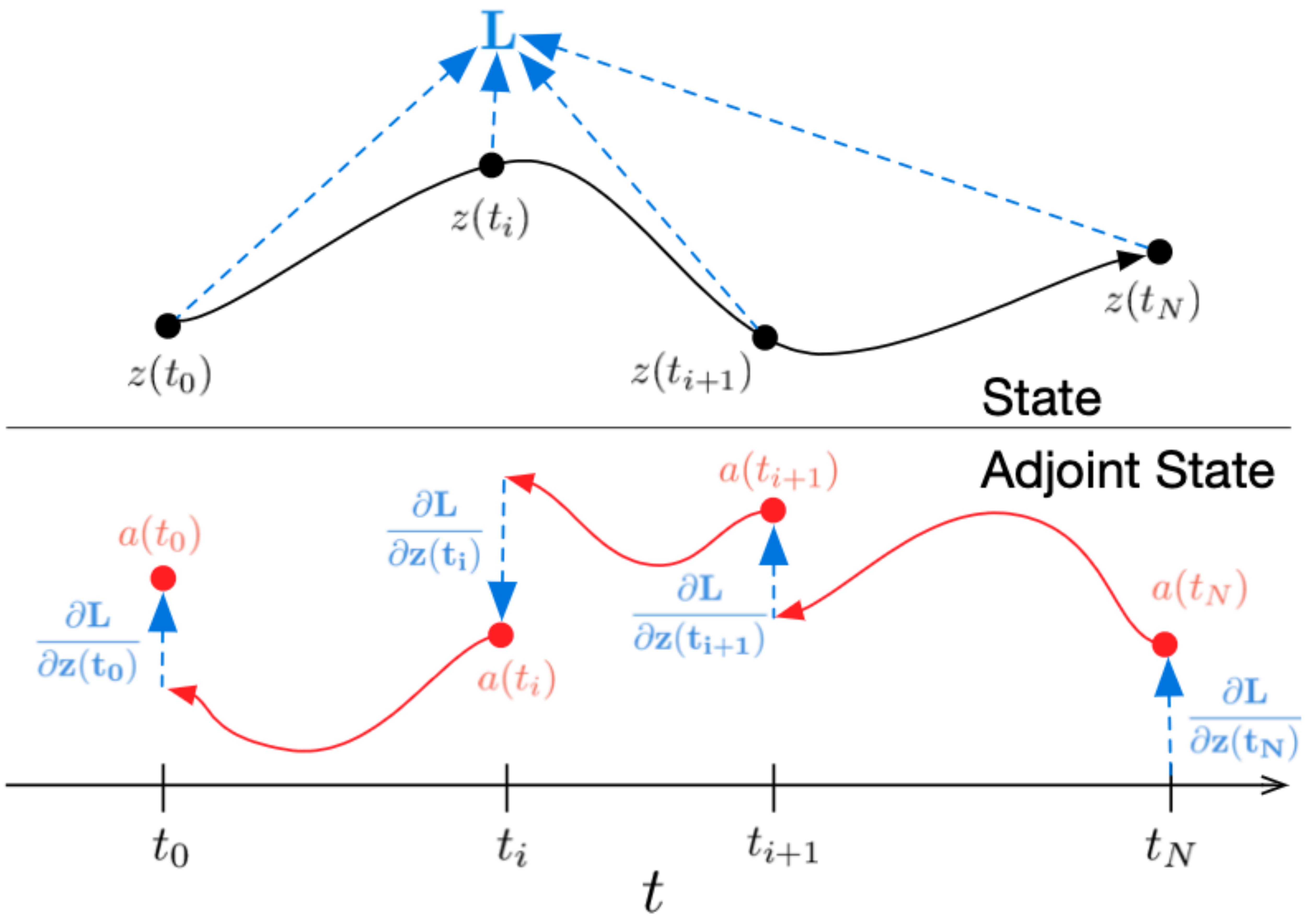


ResNet is essentially an Euler propagator!

Chen et al. (2018)

A flow from input to output

$$L(\mathbf{z}(t_1)) = L\left(\mathbf{z}(t_0) + \int_{t_0}^{t_1} f(\mathbf{z}(t), t, \theta)dt\right) = L\left(\text{ODESolve}(\mathbf{z}(t_0), f, t_0, t_1, \theta)\right)$$



Adjoint $\mathbf{a}(t) = \frac{\partial L}{\partial \mathbf{z}(t)}$

$$\frac{d\mathbf{a}(t)}{dt} = -\mathbf{a}(t)^{\mathsf{T}} \frac{\partial f(\mathbf{z}(t), t, \theta)}{\partial \mathbf{z}}$$

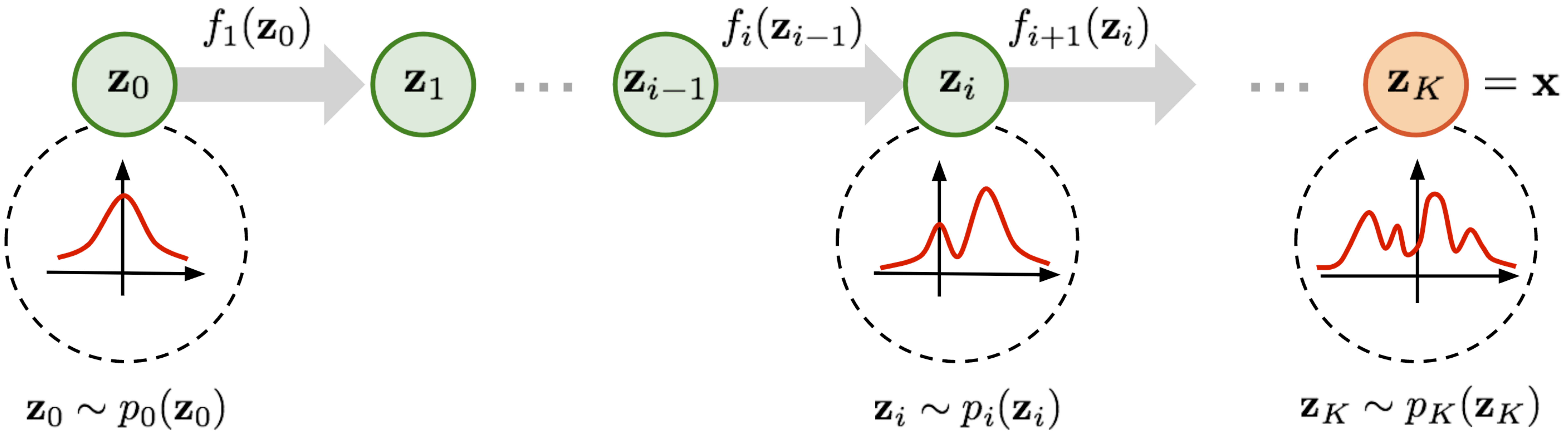
$$\frac{dL}{d\theta} = -\int_{t_1}^{t_0} \mathbf{a}(t)^{\mathsf{T}} \frac{\partial f(\mathbf{z}(t), t, \theta)}{\partial \theta} dt$$

Chen et al. (2018)

Normalising Flows

$$y_{n+1} = y_n + f\left(t_n, y_n\right)\,\Delta t$$

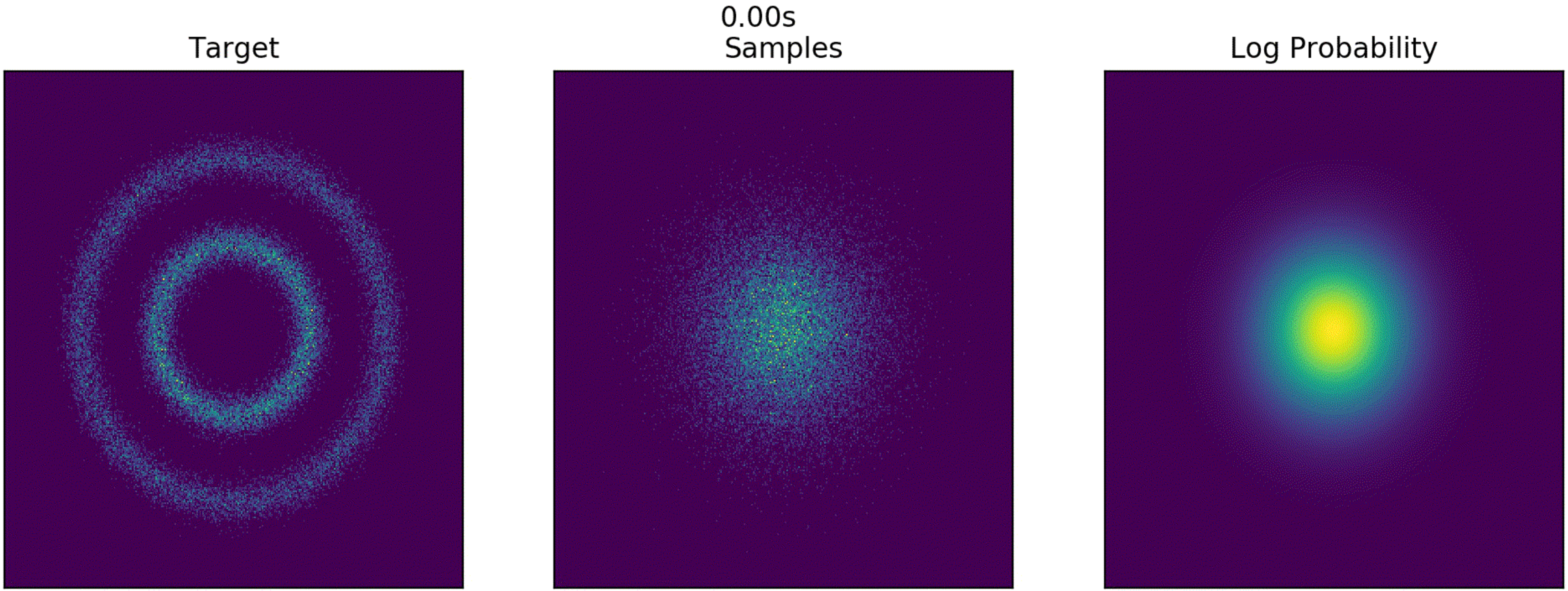
(Discrete) normalising flows



Continuous Normalising Flows

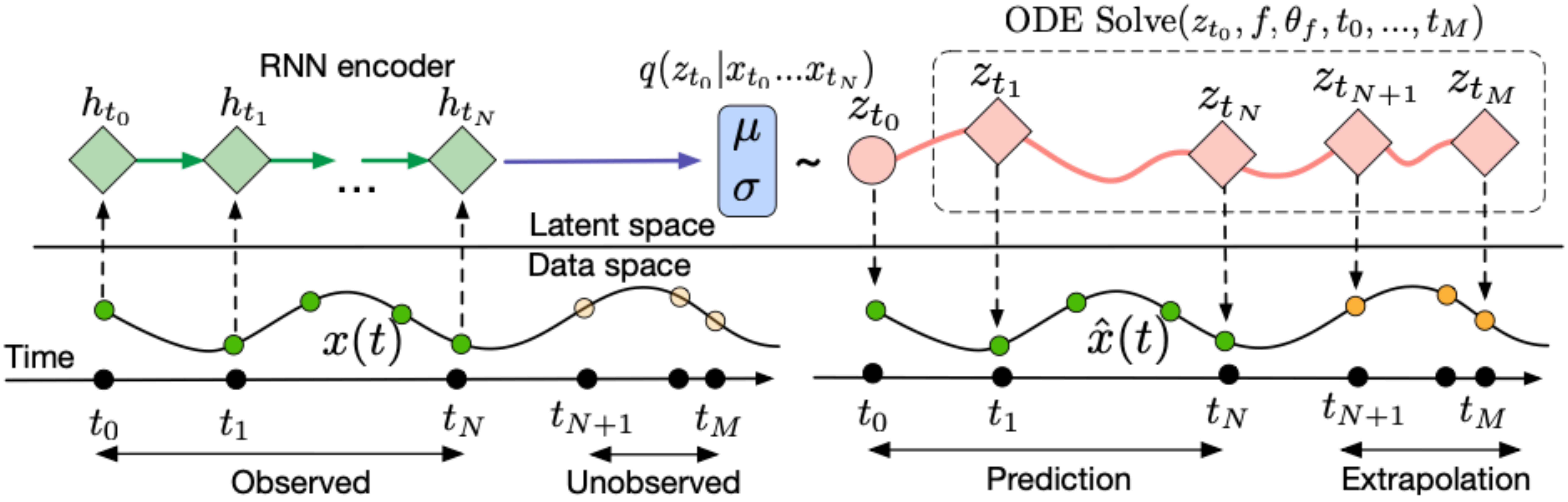
$$\frac{d\mathbf{h}(t)}{dt} = f\left(\mathbf{h}(t), t, \theta\right)$$

Integrate w.r.t. t



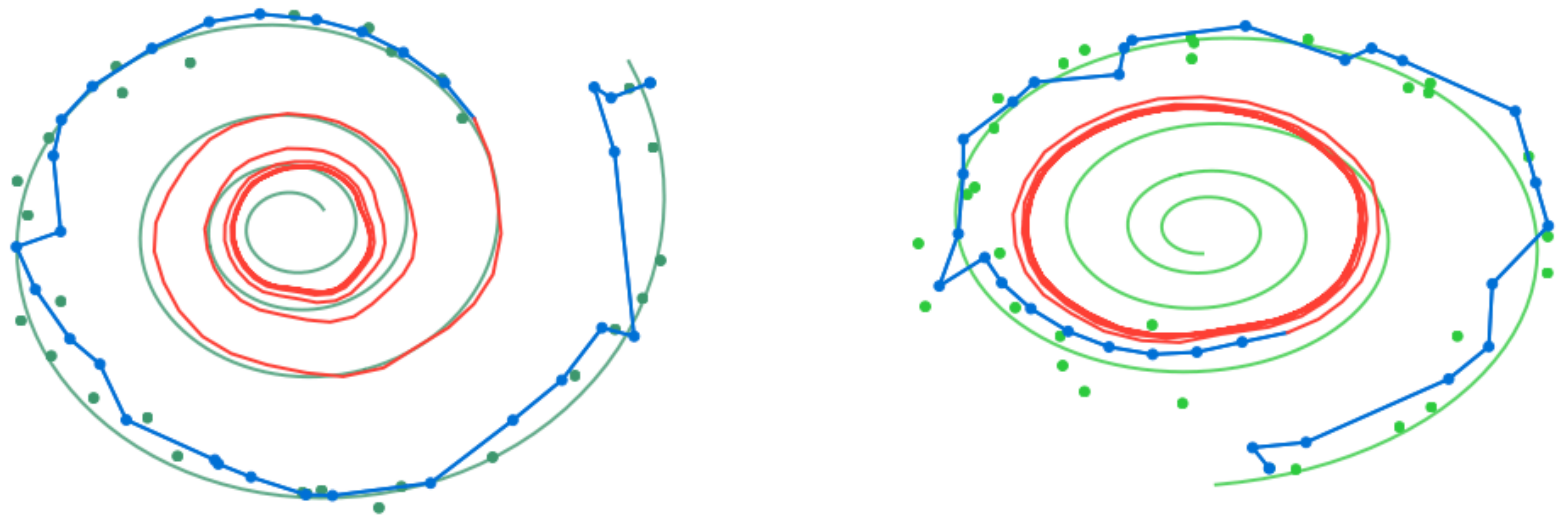
Chen et al. (2018)

Time Series VAE

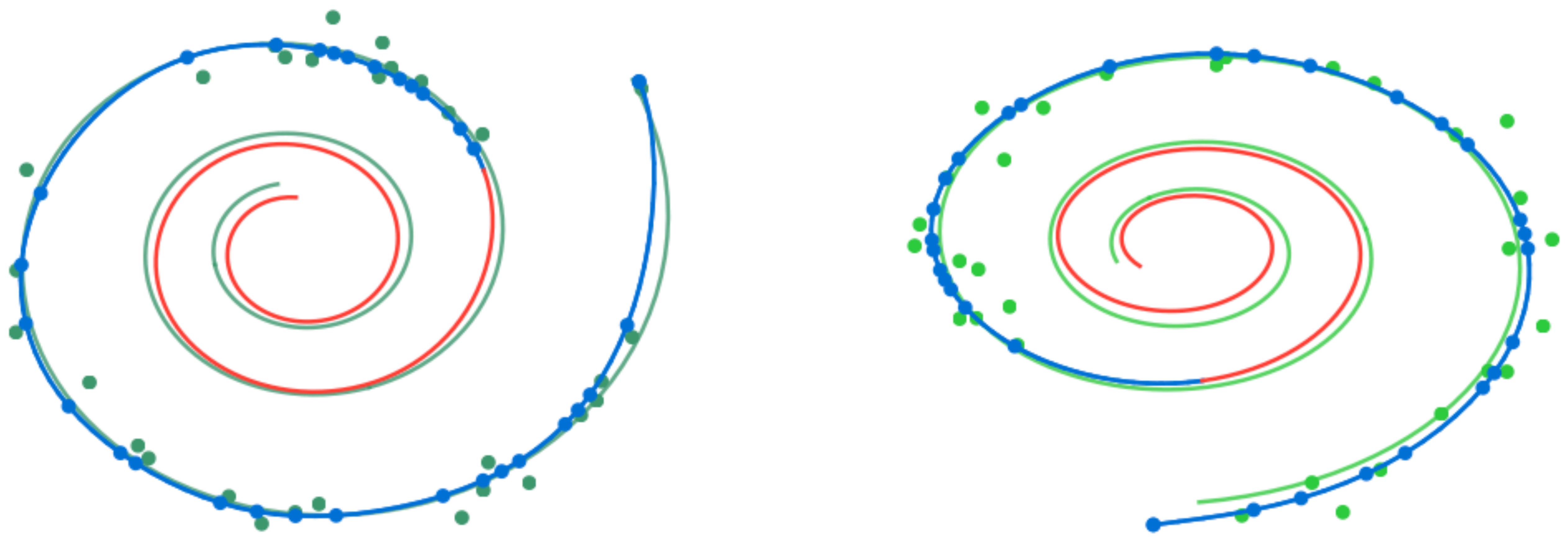


Chen et al. (2018)

Time Series VAE



(a) Recurrent Neural Network



Chen et al. (2018)