An Application of the Sparse Identification of Nonlinear Dynamics (SINDy) Algorithm

Han Zhang

University of California Santa Barbara

Intro to SINDy

Accurately modeling the nonlinear dynamics of a system from measurement data is a challenging yet vital topic. **Sparse Identification of Nonlinear Dynamics (SINDy)** is a data-driven method used for uncovering the underlying dynamics of a nonlinear system from observational data. The key idea behind SINDy is to express the dynamics of the system as a **sparse linear combination of a library of candidate functions or basis functions**. These basis functions can be chosen based on prior knowledge or intuition about the system, such as polynomials, trigonometric functions, or exponential functions.
SINDy has found applications in various fields, including **physics, biology, engineering, and finance**. It enables the discovery of simplified mathematical models that capture the essential dynamics of complex systems, even in the absence of complete knowledge about the system's governing equations.

The SINDy Method

Consider the following initial value problem

$$\frac{dx}{dt} = f(x,t), \qquad x(t_0) = x_0 \in \mathbb{R}^n$$

where f is a Lipschitz continuous function in x. SINDy addresses the problem of inferring the function f from data and takes advantage of the fact that many of these systems have dynamics with only **a few active terms** in the space of potential functions f. This avoids the intractable combinatorial search across all possible model structures. SINDy approximates f by a generalized linear model

$$f(x) \approx \Sigma_{k=1}^{p} \theta_k(x)\xi_k = \Theta(x)\xi$$

with the fewest nonzero terms in ξ as possible. In the formula above, $\theta_k(x)$ represents the candidate functions we fit to the data and ξ_k is the corresponding coefficients of these functions that demonstrate the weight of these functions on the overall dynamics. $\Theta(x)$ is a library of candidate nonlinear functions and may be constructed from **X**, e.g.,

$$\Theta(X) = [1\; X\; X^2 \cdots X^d \cdots sin(X)\; e^X \cdots]$$

Then, it is possible to use **sparse regression** to solve for the relevant terms that are active in the dynamics.The dynamical system can then be represented as

$$\dot{X} = \Theta(X)\Xi$$

where Ξ contains a column vector ξ_k that represents the coefficients determining the active terms in the k^{th} row. A parsimonious model will provide an accurate model in fit with as few terms as possible in Ξ. Such a model may be identified using a **convex ℓ_1-regularized sparse regression**:

$$\xi_k = argmin_{\xi_k'} ||\dot{X}_k - \Theta(X)\xi_k'||_2 + \lambda||\xi_k'||_1$$

Here $\dot{X}_k$ is the k^{th} column of $\dot{X}$, and λ is a sparsity-promoting knob.
The sparse vector ξ_k may be synthesized into a dynamical system:

$$\dot{x}_k = \Theta(x)\xi_k$$

Note that x_k is the k^{th} element of **x** and Θ(x) is a row vector of symbolic functions of **x**.

Applying SINDy to the Duffing Equation

The Duffing equation is a mathematical model that describes the motion of a **damped, driven oscillator**. It is named after the German engineer **Georg Duffing**, who first introduced it in the early 20th century. The simplified version of this equation we explore takes the form:

$$\ddot{x} + \gamma\dot{x} + \beta x + \epsilon x^3 = 0.$$

In this equation, x represents the displacement of the oscillator, t is time, overhead dot represents differentiation. γ controls the amount of **damping**, β controls the linear **stiffness** and ϵ controls the amount of **non-linearity** in the restoring force. We consider three dynamical systems generated by the equation. Our goal is to test **how well SINDy performs** in three different situations.

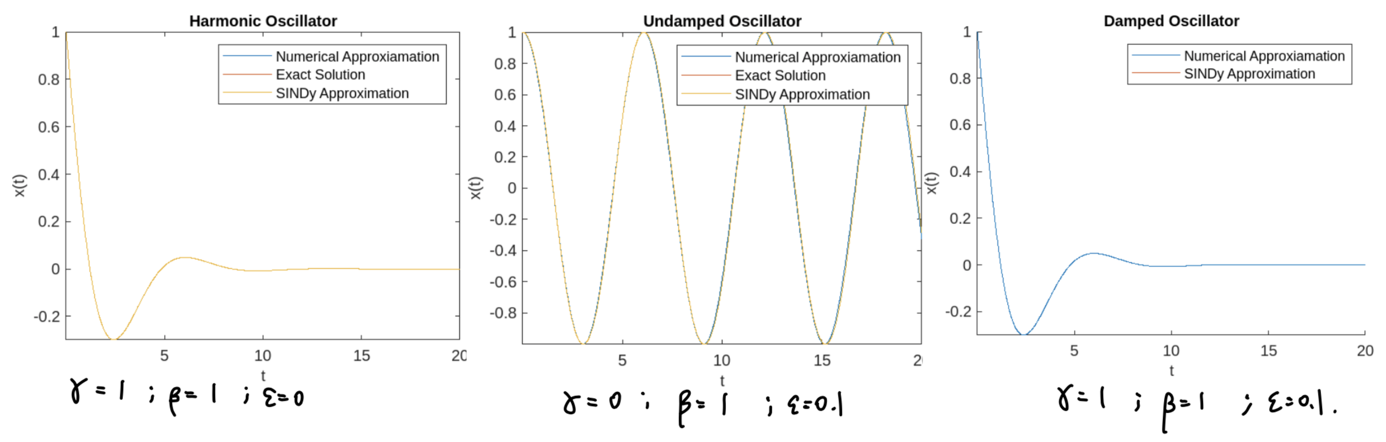
Case I: Damped Linear Oscillator(ϵ=0)
A harmonic oscillator is a fundamental concept that refers to a system exhibiting simple harmonic motion. The mathematical description of this system is:

$$\ddot{x} + \gamma\dot{x} + \beta x = 0$$

This ordinary differential equation can be solved algebraically. By assuming γ>0, β>0, $\gamma^2 - 4\beta$<0, and setting the initial conditions to be x(0)=1 and $\dot{x}$=0, the exact solution for this ODE is:

$$x(t) = e^{-\frac{\gamma t}{2}}\left(\cos\omega t + \frac{\gamma}{4\omega}\sin\omega t\right)$$

where $\omega^2 = \gamma^2 - 4\beta$. By comparing the $x-t$ graph of the **exact solution** to the $x-t$ graph generated by **SINDy algorithm** and the $x-t$ graph approximated by the **Runge-Kutta Method(ode45)** with the same initial conditions, we conclude that SINDy **is efficient** in identifying the dynamics of harmonic oscillators.



Case II: Undamped Nonlinear Oscillator(γ=0)
An undamped oscillator is a system that exhibits oscillatory motion without any dissipation or damping effects. The mathematical description of this system is:

$$\ddot{x} + \beta x + \epsilon x^3 = 0$$

This equation cannot be solved algebraically due to the involvement of the nonlinearity ϵx^3. Considering ϵ to be a small parameter, we apply **multiple-scale analysis** to construct a uniformly valid approximation to the solution of the undamped nonlinear oscillator:

$$x(t) = \cos\left(1 + \frac{3\epsilon}{8}\right)t + O(\epsilon) \qquad (1)$$

By comparing the $x-t$ graph of the **leading-order approximation** to the undamped nonlinear oscillator with the $x-t$ graph obtained by **SINDy** and numerics, we conclude that SINDy performs well in the undamped case.

Case III: Duffing oscillator($\gamma \neq 0,\ \epsilon \neq 0$)
The last case is more general with both γ and ϵ not equal to 0. Having built confidence in our methodology in the previous two cases, we rely solely on numerical simulations of the differential equation to test the SINDy method. We then compare the graph generated by **SINDy** with the **numerical approximation**. The two graphs suggest that SINDy yet again performs well.

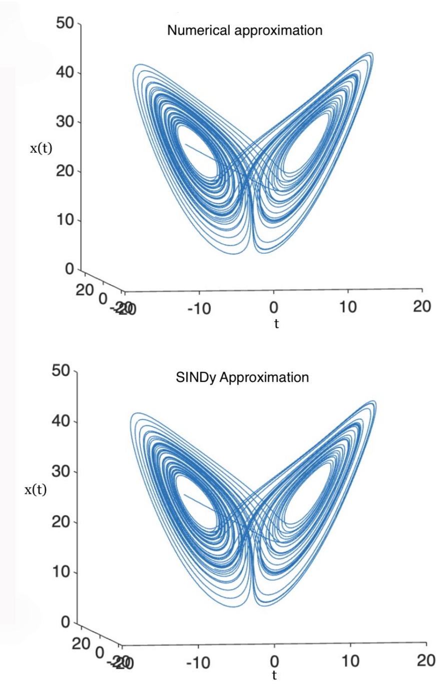
Applying SINDy to the Lorenz system

Lorenz system
The Lorenz system refers to a set of three differential equations that were discovered by the mathematician and meteorologist Edward Lorenz in 1963. These equations are used to describe a simplified model of atmospheric convection, which is the process by which heat is transferred through the motion of a fluid. It is a well-known model in chaos theory.
The system is defined by the following equations:

$$x' = \sigma(y - x),$$
$$y' = x(\rho - z) - y,$$
$$z' = xy - \beta z.$$

In these equations, x, y, and z represent the variables that describe the state of the system over time, and t represents time itself. σ, ρ, and β are parameters that determine the behavior of the system. We simulate the system with **ode45**, and show the resulting chaotic trajectory of the system upright corner of the figure below.
We then apply SINDy to the data generated by the numerics. We choose all polynomials up to 3^{rd} order as our library of functions. The output of the SINDy algorithm is a sparse matrix of coefficients Ξ. We show the coefficient matrix Ξ found by the SINDy algorithm below:

''	'xdot'	'ydot'	'zdot'
'1'	[0]	[0]	[0]
'x'	[-10.0000]	[28.0000]	[0]
'y'	[10.0000]	[-1.0000]	[0]
'z'	[0]	[0]	[-2.6667]
'xx'	[0]	[0]	[0]
'xy'	[0]	[0]	[1.0000]
'xz'	[0]	[-1.0000]	[0]
'yy'	[0]	[0]	[0]
'yz'	[0]	[0]	[0]
'zz'	[0]	[0]	[0]
'xxx'	[0]	[0]	[0]
'xxy'	[0]	[0]	[0]
'xxz'	[0]	[0]	[0]
'xyy'	[0]	[0]	[0]
'xyz'	[0]	[0]	[0]
'xzz'	[0]	[0]	[0]
'yyy'	[0]	[0]	[0]
'yyz'	[0]	[0]	[0]
'yzz'	[0]	[0]	[0]
'zzz'	[0]	[0]	[0]



Comparing the two graphs, we conclude that SINDy works efficiently to identify the dominant terms that account for the observed behaviors of the Lorenz system.

Acknowledgements

I express my gratitude to my mentor Jimmie Adriazola for his invaluable guidance and unwavering support throughout the project as well as the UCSB Directed Reading Program for the opportunity to work on this project.

References

1. S.L. Brunton, J.L. Proctor, J.N. Kutz(2016), *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, National Acad Sciences.
2. S.L. Brunton J.N. Kutz (2019), *Data-driven Science and engineering: Machine Learning, Dynamical Systems, and Control*, Control-Cambridge University Press, NY.