LoCP-B Project Presentation

Luca Agosti Fatima Ballout Nicola Bee Lorenzo Saccaro

8 July 2022

Università degli Studi di Padova

Introduction

- Many models that we use to describe the world are based on partial differential equation (Schrödinger, Navier-Stokes, Fokker–Planck *ecc.*)
- Usually we study the data assuming we know the underlying model, while the opposite path is less travelled
- We aim to infer the equations governing a given system starting only from data, e.g. the potential V(x) in Schrodinger equation
- This is done by implementing the **S**parse **I**dentification of **N**onlinear **Dy**namics (SINDy) algorithm.

SINDy algorithm

UNIVERSITÀ DEGLI STUDI DI PADOVA

- The goal of the SINDy algorithm is to discover the equations governing the model of a dynamical system only starting from time-series data.
- The dynamics of a system could be expressed in the following form:

$$\mathbf{U}_t = \mathbf{\Theta}(\mathbf{U})\mathbf{\Xi}$$

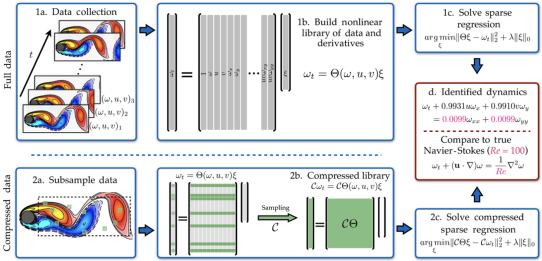


Figure: Schematic of the SINDy algorithm applied to the Navier-Stokes equations

The optimal choice for the library Θ

UNIVERSITÀ DEGLI STUDI DI PADOVA

- Our library should be as small as possible to give a generalizable solution but not so much to exclude possible solution terms
- We should exploit our prior knowledge about the dynamics of the system to add/remove terms from the library

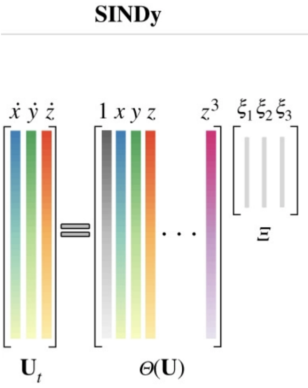


Figure: Role of the library in the SINDy algorithm

The optimizer for the matrix Ξ

- As optimizer for the matrix Ξ we used the *sequentially thresholded least squares algorithm* with Ridge regression.

- The loss function $L(\lambda)$ is given by:

$$L(\lambda) = \|\dot{\mathbf{X}} - \mathbf{\Theta}(\mathbf{X})\mathbf{\Xi}\| + \lambda\|\mathbf{\Xi}\|_0$$

Our objective is to find the sweet spot for λ in a way such that our model will be the sparsest possible but still generalizable.

The optimizer for the matrix Ξ

- if $\lambda \to 0$ we are performing a linear regression so we will obtain a model with a high complexity that is overfitting the data.

- if $\lambda \to +\infty$ we are only considering the term that values sparsity so, we will tend to the identical null solution, the sparsest one possible, with an increasingly higher error.

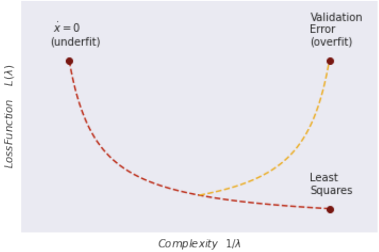


Figure: Loss function in respect to $\frac{1}{\lambda}$

Our implementation of SINDy

Università degli Studi di Padova

Now that we have defined SINDy we can use it to study a few simple examples of quantum systems. Of course the possibilities are endless but our choice landed on a few system defined by the Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\Psi(x,t) = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x,t)\right]\Psi(x,t)$$

We changed the potential $V(x,t)$ and studied how a Gaussian wave packet interacted with it.

Data gathering

- The datasets are generated from a partial differential equation with *py-pde* library.
- No real data are used given that we need to know the starting equation in order to study the goodness of SINDy results.
- For a dynamic problem the solution of the PDE produces as dataset a matrix of samples representing the evolution of the system with discrete coordinates both in space and time.

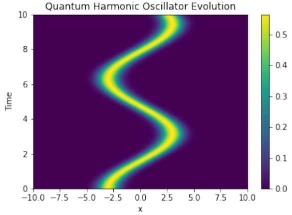


Figure: Oscillation of a wave packet under Harmonic potential using py-pde

Analyzed systems

In particular we considered the following systems:

- Harmonic Oscillator
- Double well potential
- A potential that imitates the effects of a catalyst on a chemical reaction

The reason behind this choice is that each system is incrementally more complex than the one before (more terms in the potential), allowing us to gradually study and correct our SINDy implementation.

Harmonic Oscillator (HO)

$$V(x) = \frac{1}{2}x^2$$

Symmetric double well (SDW)

$$V(x) = 0.0156x^4 - 0.25x^2$$

Asymmetric double well (ADW

$$V(x) = 0.0156x^4 + 0.0033x^3 - 0.25x^2 - 0.08x - 0.0064$$

Symmetric catalyzed (SCat)

$$
\begin{aligned}
V(x) = {} & 0.0042x^6 - 0.0013x^5 \\
& - 0.0630x^4 + 0.0210x^3 \\
& + 0.2074x^2 - 0.0838x \\
& - 0.7630
\end{aligned}
$$

Asymmetric catalyzed (ACat)

$$V(x) = 0.0036x^6 - 0.0007x^5 - 0.0520x^4 + 0.0143x^3 + 0.1467x^2 - 0.124x - 0.6807$$

In the following table we collect the results obtained after running the SINDy algorithm with optimal threshold value λ.

	Reconstructed $V(x)$	Error (%)
HO	$0.4996x^2$	0.08
SDW	$0.0156x^4 + 0.2498x^2$	$0.09, 0.08$
ADW	$0.0156x^4 + 0.0033x^3 - 0.2494x^2$ $-0.0791x - 0.0060$	$0.22, 0.8, 0.23$ $0.79, 5.44$
SCat	$0.0042x^6 - 0.0013x^5 - 0.0628x^4 + 0.0209x^3$ $+0.2067x^2 - 0.0836x - 0.7611$	$0.24, 0.22, 0.27, 0.22$ $0.33, 0.23, 0.24$
ACat	$0.0036x^6 - 0.0007x^5 - 0.0518x^4 + 0.0143x^3$ $+0.1462x^2 - 0.1237x - 0.679$	$0.24, 0.29, 0.27, 0.30$ $0.34, 0.28, 0.25$

Robustness of SINDy

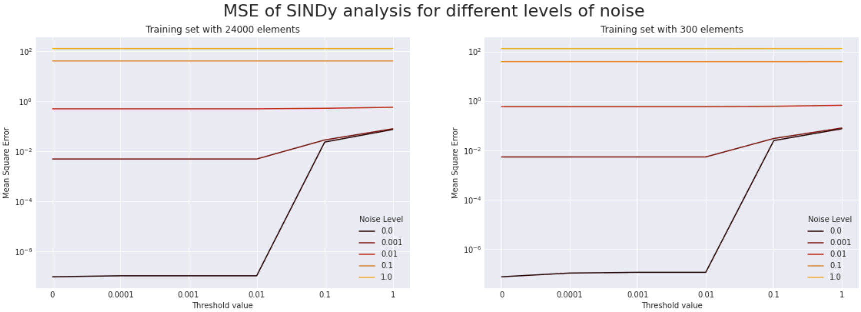
We aim to check the robustness of SINDy with our model in different cases, testing how the algorithm behave with various threshold values.

In particular:

- Progressively adding phase noise to the measure
- Changing the number of terms used for the identification
- Trying different sampling rate

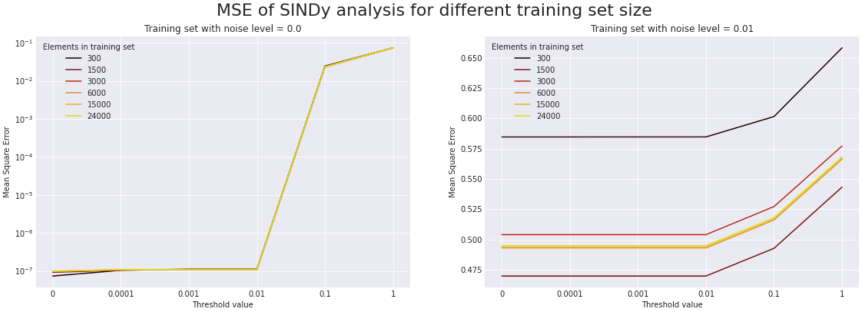
This analysis are performed with the *symmetric double well*, that we considered non-trivial but with still low complexity

Noise Analysis



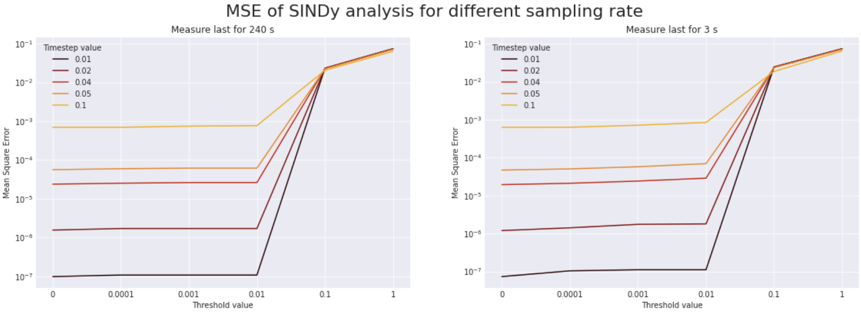
- We have found out that noise is the major cause of model misidentification
- Similar behaviour for every set size in range $[24\,000 - 300]$

Training Set Size



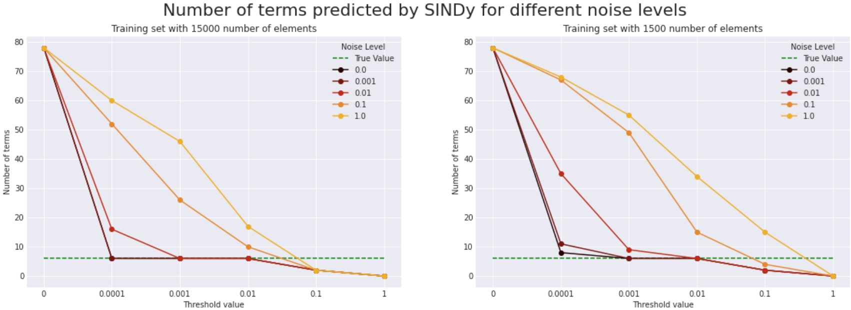
- Clean data lead to a correct identification of the model despite the training size
- The size of the dataset affect the MSE less than noise

Timestep



- The higher the sampling rate, the lower the mean square error on predicted model
- As in previous analysis, the size of the data does not significantly affect the goodness of the prediction for zero-noise datasets

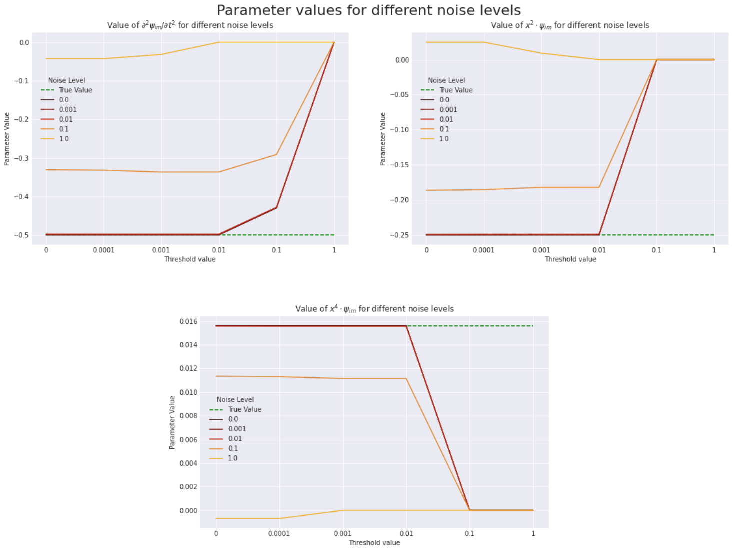
Number of Terms



- The noisiest datasets do not have a correct prediction for any threshold value.
- For less noisy dataset an optimal value exists

Parameters value

UNIVERSITÀ DEGLI STUDI DI PADOVA



The limit of our approach

The method discussed up to this point relies on measuring the complex valued $\Psi(x, t)$

The limit of our approach

The method discussed up to this point relies on measuring the complex valued $\Psi(x,t)$

- Not practical to measure directly the wavefunction $\Psi(x,t)$ or its phase

The limit of our approach

The method discussed up to this point relies on measuring the complex valued $\Psi(x, t)$

- Not practical to measure directly the wavefunction $\Psi(x, t)$ or its phase
- One has usually access only to $|\Psi(x, t)|^2$

The limit of our approach

The method discussed up to this point relies on measuring the complex valued $\Psi(x,t)$

- Not practical to measure directly the wavefunction $\Psi(x,t)$ or its phase
- One has usually access only to $|\Psi(x,t)|^2$
- Move from a quantum to a semi-classical problem

Quantum Smoluchowski equation

Università degli Studi di Padova

$$\gamma m \frac{\partial}{\partial t}\rho(x,t) = \frac{\partial}{\partial x} V'_{\text{eff}}(x)\ \rho(x,t) + \frac{\partial^2}{\partial x^2} D_{\text{eff}}(x)\ \rho(x,t)$$

Where

$$\begin{cases} V_{\text{eff}}(x) = V(x) + \frac{1}{2}\lambda V''(x) \\ D_{\text{eff}}(x) = \frac{1}{\beta}\left[1 + \lambda\beta V''(x)\right] \\ \lambda = \frac{\hbar}{\pi m \gamma} \ln\left(\frac{\hbar\beta\gamma}{2\pi}\right)\ ;\ \beta = \frac{1}{k_B T} \end{cases}$$

Validity regime:

$$\lambda \gg \frac{1}{m\beta\gamma^2} \qquad\qquad k_B T \ll \hbar\gamma$$

Simulations

SINDy results

Possible issues

Few factors can be responsible :

Possible issues

Few factors can be responsible :

- Stability issue with the integrator

Possible issues

Few factors can be responsible :

- Stability issue with the integrator
- We are dealing with very small coefficient ($\gamma \approx 100$), noisy terms are introduced

Possible issues

Few factors can be responsible :

- Stability issue with the integrator
- We are dealing with very small coefficient ($\gamma \approx 100$), noisy terms are introduced
- Try different optimizers and more advanced techniques (model ensembling, multiple trajectories, weak-form etc...)

Final remarks

Challenges and shortcomings of our approach:

Final remarks

Challenges and shortcomings of our approach:

- Library Θ grows exponentially when adding new terms:

Final remarks

Challenges and shortcomings of our approach:

- Library Θ grows exponentially when adding new terms:
  - Can easily run out of system memory

Final remarks

Challenges and shortcomings of our approach:

- Library Θ grows exponentially when adding new terms:
  - Can easily run out of system memory
  - The optimization problem becomes ill-conditioned

Final remarks

Challenges and shortcomings of our approach:

- Library Θ grows exponentially when adding new terms:
  - Can easily run out of system memory
  - The optimization problem becomes ill-conditioned
- Can only have linear combination of terms: it is not possible to learn function parameters such as mean and standard deviation of a gaussian term $e^{\frac{(x-\mu)^2}{2\sigma^2}}$

Future works

Both physics and coding developments (pull requests) can be made:

Future works

Both physics and coding developments (pull requests) can be made:

- Explore more on the transition between quantum and classical mechanics

Future works

Both physics and coding developments (pull requests) can be made:

- Explore more on the transition between quantum and classical mechanics
- Lots of SINDy features for ODEs are not working out of the box for PDEs

Future works

Both physics and coding developments (pull requests) can be made:

- Explore more on the transition between quantum and classical mechanics
- Lots of SINDy features for ODEs are not working out of the box for PDEs
- Complex number support

Acknowledgements and references

Steven L. Brunton: https://www.youtube.com/c/Eigensteve
PDE-FIND paper: https://www.science.org/doi/10.1126/sciadv.1602614
Smoluchowski equation: https://doi.org/10.1016/j.physa.2004.12.007
pysindy docs: https://pysindy.readthedocs.io/en/latest/index.html
py-pde docs: https://py-pde.readthedocs.io/en/latest/index.html
source code: https://github.com/lorenzo-saccaro/LoCP_B_project