Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery

Machine Learning Physics Models for Materials Self-Assembly

Project Report

Project Dates: 14/06/2021- 20/08/2021

University of Southampton

Project Student: Aspen Fenzl, University of Sheffield

Supervised by: Nigel Clarke, University of Sheffield

Report Date: 01/10/2021

AI3SD-Intern-Series:Report-5_Fenzl

Machine Learning Physics Models for Materials Self-Assembly
AI3SD-Intern-Series:Report-5_Fenzl
Report Date: 01/10/2021
DOI: 10.5258/SOTON/AI3SD0142
Published by University of Southampton

Network: Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery
This Network+ is EPSRC Funded under Grant No: EP/S000356/1

Principal Investigator: *Professor Jeremy Frey*
Co-Investigator: *Professor Mahesan Niranjan*
Network+ Coordinator: *Dr Samantha Kanza*

Contents

1 Project Details

Title	Machine Learning Physics Models for Materials Self-Assembly
Project Reference	AI3SD-FundingCall3_006
Supervisor Institution	University of Sheffield
Project Dates	14/06/2021- 20/08/2021
Keywords	machine learning, partial differential equations, feature selection, dewetting, thin films, sparse optimization, LASSO

2 Project Team

2.1 Project Student

Name and Title	Aspen Fenzl
University Department Name	University of Sheffield, Department of Physics & Astronomy
Work Email	afenzl1@sheffield.ac.uk

2.2 Project Supervisor

Name and Title	Professor Nigel Clarke
University Department Name	University of Sheffield, Department of Physics & Astronomy
Work Email	n.clarke@sheffield.ac.uk

2.3 Researchers & Collaborators

Matthew Jones, PhD Student at the University of Sheffield Department of Physics and Astronomy, provided advice and support throughout the project.

3 Lay Summary

Material development is a very slow process, relying primarily on trial and error. By refining known theories of soft matter and its evolution, we can tailor materials to a given system using a specifically developed model based on data rather than assumptions. By combining feature selection methods with the efficiency of machine learning, we are able to determine the dominant terms of a system's partial differential equation (PDE).

In this project, we begin by simulating the process of dewetting. Height data is generated from the Thin Film Equation, the simplest model describing the time evolution of a liquid film on a surface. We then apply feature selection methods to reconstruct the PDE from the data. However, there may be physics unaccounted for when using simple models, which is why a process to learn the governing PDE from data is so valuable. The data-driven approach of machine learning will allow us to make predictions from observation alone, and expedite the process of materials discovery.

4 Aims and Objectives

We aim to develop a machine learning tool that can predict a system's PDE from data alone, using thin film dewetting as an example. The project can be broken up into the following benchmark tasks:

- Understand the physics of dewetting in thin films according to the Thin Film Equation
- Reproduce results of the dewetting process published by Sharma [1]
- Extend the method to accept inverse-space inputs
- Test against numerically generated film height data to provide proof of principle

The final output of the project is a tool for learning the structure of known and unknown PDEs from both real-space and power spectrum data.

5 Methodology

5.1 Scientific Methodology

The first part of the project focused on understanding the physics of dewetting. Liquids placed on a repellent surface can undergo dewetting, the process of a fluid retracting itself and spontaneously rearranging into smaller structures such as droplets. The key force affecting the motion of the film is the local surface tension. However, this doesn't take into effect differences in chemical interactions between the substrate and the film.

To account for this, we include a term for the effective potential between the solid and liquid as an average of local intermolecular forces, called the disjoining pressure. In very thin films, short range van der Waals interactions become increasingly dominant, which can lead to the process of dewetting [2]. The disjoining pressure $\Pi(h)$ expresses the film's tendency to change thickness in order to minimize its free energy, while the surface tension γ tries to keep the film flat [5]. The Thin Film Equation also contains a viscous term μ, which slows the growth of instabilities.

$$\frac{\partial h}{\partial t} = -\frac{1}{3\mu}\nabla \cdot (h^3 \nabla(\gamma \nabla^2 h - \Pi(h))) \tag{1}$$

Data for the height of the film was calculated on a 2D grid according to the Thin Film Equation using Python. This allowed us to visualize the formation of droplets as the system evolved with time. The spatial and temporal derivatives were calculated using finite differences. Following Schaeffer's method [3], we applied an L^1 least squares minimization to the data.

5.2 AI Methodology

Follwing the methodology outlined in Schaeffer's paper [3], we developed a feature selection algorithm to determine the underlying PDE of a dataset. Using related physical models as a guide, we first construct a candidate function of all possible derivative terms, similar to a Taylor expansion. In this case we use the Navier-Stokes equations for motion of viscous fluids, which are proportional to ∇h with h as the local film height. The derivative function can be written as a product of a coefficient vector $=\alpha$ and the remaining derivative terms. Each derivative is expanded into a feature vector, a column holding the term's value for each point on the lattice at a given time. Once we have calculated all the feature vectors, we normalize them to

a maximum value of 1 to prevent unintentional biasing.
The feature selection algorithm utilized is known as the least absolute shrinkage and selection operator (LASSO). LASSO yields sparse models, selecting only a subset of the initial parameters. It relies on a combination of a least-squares fit and the L^1 norm to penalize nonzero coefficients. The penalizing L^1 norm shrinks less influential terms to zero, leaving us with the dominant derivative terms.
A tuning parameter λ allows us to choose how sparse we want the resulting vector to be. If λ=0, there would be no penalty and we would get a least squares fit, but if λ was large enough, all the coefficients would equal zero. For this reason, LASSO can produce a model with any number of variables. Coefficient sparsity is the constraint for the system; we assume there are relatively few governing terms. This allows for easier interpretation of the resulting PDE, which is especially helpful in problems with a large number of variables.

6 Results

The 2D Thin Film Equation was used to generate height data on a 50×50 lattice. Constants such as those for the viscosity and surface tension were taken from a past investigation by Nigel Clarke [4]. Over the course of 800,000 timesteps, we visualized the initially uniform liquid film begin to dewet.

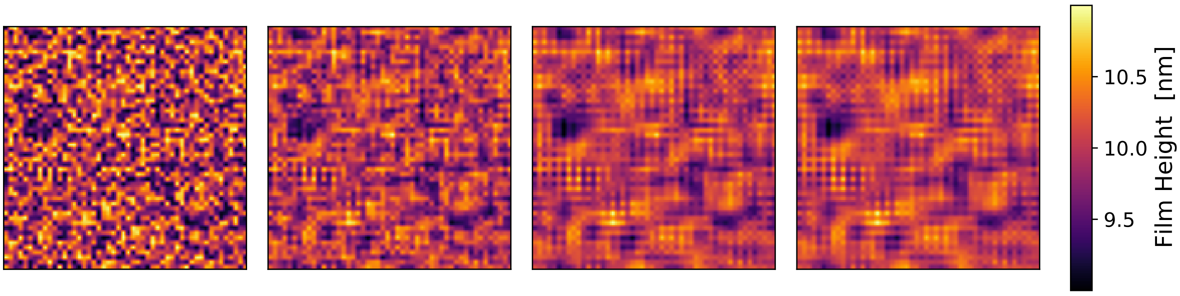

Figure 1: Time dependent evolution of the film height on a 50×50 lattice at (from left to right): T=0, 72 000, 216 000, and 288 000 seconds. A continuous colour scale shows the difference between minimum and maximum thickness of the liquid film.

The morphological pattern of the dewetting in Figure 1 follows the same as the ones in Sharma's work [1]. It begins with the development of local ridges and valleys which eventually form complete droplets given an extended simulation. Looking at the time stamps of the system in Figure 1, it is clear the dewetting process for this film lies on a fairly large time scale. In practice, the rate of evolution can be altered by changing the molecular weight or surrounding temperature of the liquid [6].

Schaeffer's regularized least-squares method was able to successfully identify the key features of the dewetting dataset, which match those in Equation 1. The sparsity of the resultant coefficients depended on the tuning parameter λ, as shown in Table 1. Smaller values such as λ=1 gave a larger number of terms, making it difficult to determine the key features. λ=50 was the perfect spot, giving the three most dominant features of the dataset. Increasing λ from here resulted in too sparse of a model with zeros for all of the derivative terms. Being able to adjust the tuning parameter is very useful, allowing us to modify the feature selection process as we see fit.

derivative terms	coefficients, λ=1	coefficients, λ=10	coefficients, λ=50
Δh	-23.67682341501	-19.95427479121	-3.40949155255
Δh^2	2.36612471188	1.99370659228	0
Δh^3	-0.07881969781	-0.06640836085	-0.01124698840
Δh_x	0	0	0
Δh_y	0	0	0
Δh_{xx}	0	0	0
Δh_{xy}	0.00012990081	0	0
Δh_{yy}	0.00041191849	0.00017939352	0
$\Delta^2 h$	0	0	0
$\Delta^2 h_x$	0.00000651966	0	0
$\Delta^2 h_y$	0	0	0
$\Delta^2 h_{xx}$	0	0	0
$\Delta^2 h_{xy}$	-0.00992948769	-0.00990987686	-0.00978232115
$\Delta^2 h_{yy}$	-0.00993628664	-0.00988517293	0

Table 1: Coefficients for candidate derivative terms generated using the LASSO shrinkage method. The data from Figure 1 was applied to Schaeffer's method for feature selection. Different values of λ correspond to how sparse the results are, which can be adjusted such as here.

Schaeffer's method successfully resolved the key features of the dewetting dataset, identifying Δh, Δh^3, and $\Delta^2 h_{xy}$. Although the nature of these terms isn't intrinsically clear, simply finding them allows us to learn more about a system's underlying dynamics. This method would work especially well as a foundation for further study, giving hints about properties and dynamics.

7 Conclusions & Future Work

In this project we have explored the process and physics behind dewetting in thin liquid films. We have also developed and tested a machine learning model to determine the underlying PDE of a dataset. We have demonstrated the use of a regularized least-squares minimization able to correctly identify dominant terms in a system's PDE. Although we have chosen to investigate the Thin Film Equation in this project, the methodology outlined in this project can be applied to any dynamic system in the same manner. This project has allowed us to explore some of the many uses for AI outside of computer science, granting us a new tool to integrate into present and future research.

Our group is continuing to work applying Schaeffer's method to simulated data from the Cahn-Hilliard Equation for phase separation in fluids, a process similar yet more complex than dewetting. We aim to continue the investigation of statistical and machine learning methods applied to soft matter as soon as the computing and lab resources become available.

Future improvements to the current work include increasing the time and length scales of the simulations. With increased computing power, this would allow us to observe how the results of our feature selection model vary with the size of the data. The next step in this ongoing investigation is to apply the algorithm to experimental data, and see how the results compare when the data is noisier. The long-term goal of this work is to apply the current methodology to ongoing research in materials development and nanotechnology to aid and expedite new discoveries.

8 Outputs, Data & Software Links

- A tested tool to learn the governing terms of a system's PDE from a dataset
- Thin film dewetting data and an algorithm for creating more
- Presentation given at the AI3SD Skills4Scientists Posters & Careers Symposium, September 2021 (https://www.ai3sd.org/ai3sd-event/01-02-09-21-skills4scientists-poster-symposium/)

Programs and libraries used in this work:

- Dewetting and LASSO data can be found in this repository
- MATLAB
- Matplotlib
- NumPy
- scikit-learn

References

[1] A. Sharma and R. Khanna, *J. Chem. Phys.*, 1999, **110**, 4929-4936.

[2] D. Bonn et al., *Rev. Mod. Phys.*, 2009, **81**, 739-805.

[3] H. Schaeffer, *Proc. R. Soc. A.*, 2017, **473**, 1-20.

[4] N. Clarke, *Macromolecules*, 2005, **38**, 6775-6778.

[5] S. Singh, *Adv. Condens. Matter Phys.*, 2011, **2011**, 1-14.

[6] A. Higgins et al., *Eur. Phys. J. E*, 2002, **8**, 137-143.