Sparse identification of shape variation dynamics in cardiac anatomy, and generative modeling based on sparse regression

MASTER THESIS PROJECT

TU Delft

Supervisors:

Konstantinos Kevopoulos
Dr. Alexander Heinlein
Dr. ir. Mathias Peirlinck

September 11, 2025

Delft University of Technology

1 Problem description

There is considerable natural variation in the shape of the human heart among healthy individuals. This variation is influenced by demographic and anthropometric factors such as age, sex, body mass index (BMI), and body surface area (BSA). We call these characteristics metadata. Understanding such shape differences is important since cardiac geometry directly affects heart function, and deviations from expected shape patterns may indicate early stages of disease. To study shape variations according to the aforementioned characteristics in a data-driven manner, several machine learning and statistical techniques have been proposed, such as statistical shape models, as in [2].

However, such statistical approaches aim to find correlations between cardiac anatomies and metadata, and therefore model this relationship in an implicit manner. Explicit modeling of the relationship between metadata and cardiac shapes, in terms of identifying analytical expressions that define this relationship, is a largely unexplored research direction.

2 Research objective

The main purpose of this research project is to investigate whether we can discover from data an analytical expression that describes the relationship between metadata and shape representations.

To do so, we plan to rely on the SINDy formulation [1]. A slight modification of the formulation will be needed, since we are not dealing with dynamical systems in this case. More specifically, given a shape representation $\mathbf{z}_i$, we aim to identify a relationship of the form

$$\mathbf{z} = \mathbf{\Theta}(\mathbf{c})\xi_i, \tag{1}$$

where $\mathbf{\Theta}$ denotes a dictionary of nonlinear functions, and ξ_i are the sparse coefficients to be identified via sparse regression. The vector $\mathbf{c}$ is a vector containing the metadata, and combinations of the metadata elements. For example, it could be expressed as:

$$\mathbf{c} = [1, c_{AGE}, c_{BMI}, c^2_{AGE}, c^2_{BMI}, c_{AGE}c_{BMI}, \cdots, c^d_{AGE}, c^d_{BMI}, c^d_{AGE}c^d_{BMI}, cos(c_{BMI}), cos(c_{AGE})], \tag{2}$$

for an arbitrary integer d. A visualization of the standard SINDy method, applied to dynamical systems is presented in Figure 1.

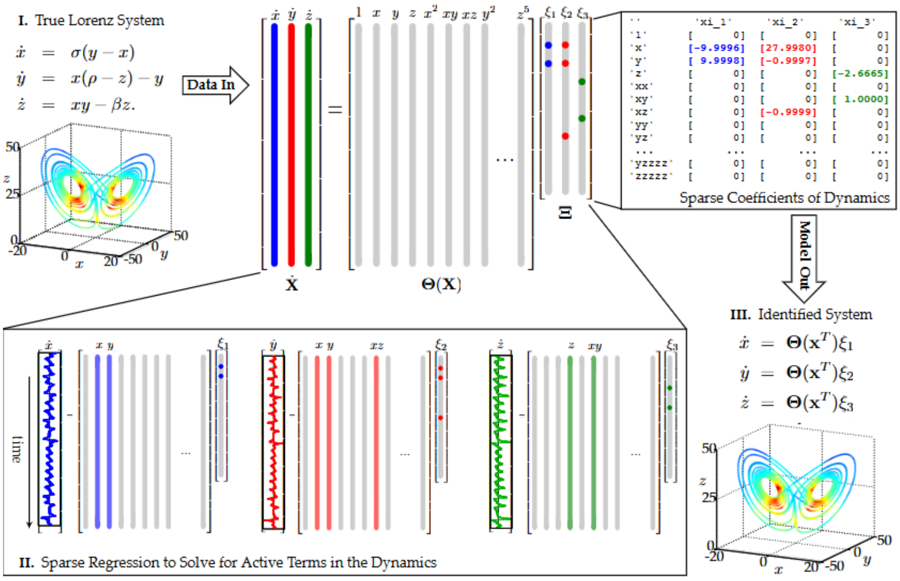

Figure 1: Visualization of the standard SINDy model, applied to discover the dynamics of the Lorenz system.

Assuming that such an analytical relationship can be identified successfully, we aim to evaluate the quality of the discovered model. In other words, we will run tests to investigate if the learned model is indeed informative of the shape representations.

The next, and most challenging step of this project will be to extend the aforementioned model to a probabilistic formulation, meaning that the learned ξ coefficients will be treated as random variables, from which we can sample to generate synthetic shape representations. To do so, we envision using a simple normalizing flow model, or any other technique explored by the student.

Overall, this project aims to investigate whether there is an analytical expression for the relationship between shapes and metadata, as well as an integrated probabilistic framework that can generate synthetic shape representations according to the equation (1), by sampling ξ. The student is expected to work on modifying the SINDY architecture to achieve the aforementioned goal, testing the interpretability of the analytical expression (if any), and potentially formulating some ideas to make the framework probabilistic.

3 Candidate requirements

- Experience in SciML and model discovery models, such as SINDy and its variants.
- Experience with developing, training, and/or using deep learning frameworks.
- Ideally, experience with developing and training flow-based generative deep learning models (normalizing flows, flow matching).

4 Tentative timeline

- Month 0-2: Literature review regarding SINDy models, model discovery methods in medical data, and potentially generative models
- Month 2-4: Develop and train the SINDy-like model, to discover an equation that models the aforementioned relationship.
- Month 4-5: Perform experiments to investigate if the learned expression can indeed model our dataset in a sufficient and interpretable manner. -7: Develop and train a small generative model, to introduce probabilistic features into the ξ coefficients. This will be explored together with the supervisors. Conduct experiments to investigate if the normalizing flow implementation is successful.
- Months 7-9: Write dissertation

References

[1] Brunton, S. L., Proctor, J. L., and Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937.

[2] Moscoloni, B., Beeche, C., Chirinos, J. A., Segers, P., and Peirlinck, M. (2025). Unveiling sex dimorphism in the healthy cardiac anatomy: fundamental differences between male and female heart shapes. *arXiv preprint arXiv:2503.00197*.