Equation-learning for Cancer Models
Pamela Burrage
(Queensland University of Technology)
H. Weerasinghe, K. Burrage

There has been an explosion recently (that is at the heart of Machine Learning) in how to discover dynamical systems models from data. The sparse identification of the nonlinear dynamics algorithm (SINDy [1]) uses the fact that many dynamical systems $x' = f(x)$ have dynamics encoded in f that can be approximated by a simple library of nonlinear functions $\Theta(X)$ constructed from the data matrix $X = [x(t_1), \cdots, x(t_m)]^\top$. A similar matrix of derivatives is also formed $X' = [x'(t_1), \cdots, x'(t_m)]^\top$. Note if x^d is in the library then X^d is a matrix with column vectors given by all possible time series of dth degree polynomials in state x. This then leads to the data system

$$X' = \Theta(X)\, A,$$

with as few terms in A as possible. These terms in A are found from a sparse regression

$$\xi_k = \arg\min_{\hat{\xi}_k} ||X'_k - \Theta(X)\, \hat{\xi}_k||_2 + \lambda ||\hat{\xi}_k||_1,$$

where X'_k is the kth column of X'. This finally leads to the dynamical system

$$x'_k = \Theta(x)\, \xi_k$$

(see the excellent exposition in [2]).

In this talk we will illustrate these ideas for an example that takes data from an agent-based model of tumour growth (under cellular stress) in the tumour microenvironment, and derives a corresponding 2D ODE system. The agent-based model has two cell types (x for tumour cells and y for healthy cells), and 500 simulations of the model are averaged to determine cell counts at a sequence of m time points. The possible library functions $\Theta(x, y)$ to represent the model are

$$\Theta(x, y) = [x, x^2, y, y^2, xy].$$

These are evaluated at the m time points, and then the SINDy approach is followed to construct the ODE system. Information about the equilibrium points and stability is used to choose appropriate ODE parameters, and numerical results demonstrate the effectiveness of this approach.

[1] Brunton, Proctor, Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, PNAS 113(15): 3932-3937, 2016.

[2] Brunton and Kutz, Data-Driven Science and Engineering, Machine Learning, Dynamical Systems, and Control, 2nd edition, 2022.