System identification of the partial differential equations governing pattern-forming in biophysics

Zhenlin Wang   Xun Huan   Krishna Garikipati

Departments of Mechanical Engineering and Mathematics, University of Michigan, Ann Arbor, MI, USA.   http://www.umich.edu/ compphys/

Introduction

Pattern formation is central to the developmental biological processes of any multicellular organism. Identification of the PDEs governing pattern formation delivers insights to the biophysics of developmental dynamics.

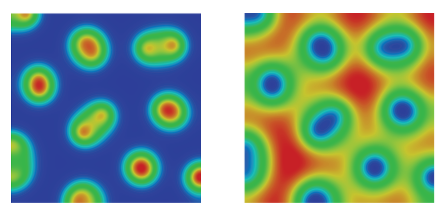
Approaches to data-driven discovery of PDEs

- Inference of the parameters in a known governing equation
- Learning the parameters that define an approximate model
- Learning the governing partial differential equations by identifying their operators
  - Symbolic regression by genetic algorithms   M. Schmidt, H. Lipson, Science, 3, 81-85, 2009
  - "Physics-informed" deep neural networks (incorporating the strong form in the loss function to learn parameters of known operators)   M. Raissi, et al., J. Comput. Phys., 378, 686-707, 2019
  - Sparse regression   S. L. Brunton, et al. Proc. Natl. Acad. Sci., 113, 3932-3937, 2016;   S. H. Rudy, et al., Sci. Adv., 3, e1602614, 2017

Parabolic PDEs that model pattern formation in biophysics

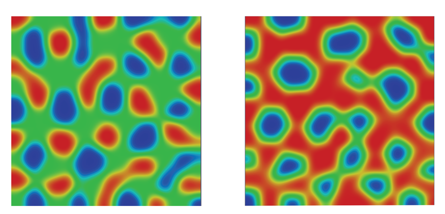
- Diffusion-reaction equations following Schnakenberg kinetics

$$\frac{\partial C_1}{\partial t} = D_1\nabla^2 C_1 + R_{10} + R_{11}C_1 + R_{13}C_1^2C_2$$
$$\frac{\partial C_2}{\partial t} = D_2\nabla^2 C_2 + R_{20} + R_{21}C_1^2C_2$$



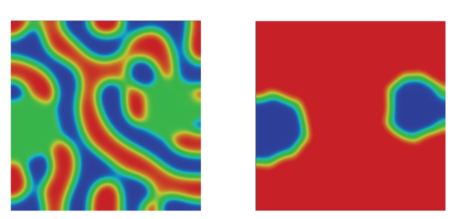
- Cahn-Hilliard equation for segregation of cell types in tissues

$$\frac{\partial C_1}{\partial t} = \nabla\cdot\left(M_1\nabla\left(\frac{\partial g}{\partial C_1} - k_1\nabla^2 C_1\right)\right)$$
$$\frac{\partial C_2}{\partial t} = \nabla\cdot\left(M_2\nabla\left(\frac{\partial g}{\partial C_2} - k_2\nabla^2 C_2\right)\right)$$



- Allen-Cahn equation for tissue patterning by nucleation and growth

$$\frac{\partial C_1}{\partial t} = \nabla\cdot\left(M_1\nabla\left(\frac{\partial f}{\partial C_1} - k_1\nabla^2 C_1\right)\right)$$
$$\frac{\partial C_2}{\partial t} = -M_2\left(\frac{\partial f}{\partial C_2} - \nabla\cdot k_2\nabla C_2\right)$$



Identification of governing parabolic PDEs in weak form

- General strong form for first-order dynamics:

$$\frac{\partial C}{\partial t} - \chi\cdot\omega = 0 \quad \chi = [1, C, C^2, ..., \nabla^2 C, ...]$$

- The Galerkin weak form (using NURBS basis functions) leads to the residual form $R_i = 0, i = 1, \ldots, N$:

$$\int_\Omega w\left(\frac{\partial C}{\partial t} - \chi\cdot\omega\right)\mathrm{d}v = 0 \Rightarrow R_i = F_i\left(\frac{\partial C^h}{\partial t}, C^h, \nabla C^h, ..., N, \nabla N...\right)$$

Aim to find ω for many operators in weak form!

- Examples of basis in the weak form

$$\Xi_i^{\dot{C}}|_n = \int_\Omega N^i \sum_{a=1}^{n_b}\frac{c_n^a - c_{n-1}^a}{\Delta t}N^a\mathrm{d}v \quad \text{time derivatives}$$
$$\Xi_i^{\nabla^2 C}|_n = -\int_\Omega \nabla N^i\cdot\sum_{a=1}^{n_b}c_n^a\nabla N^a\mathrm{d}v \quad \text{Laplace operator}$$
$$\Xi_i^{\nabla^2 C_{BC}}|_n = \int_\Gamma N^i\sum_{a=1}^{n_b}1\mathrm{d}s \quad \text{Neumann boundary: constant flux}$$

...

Matrix-vector form of residual equations

- Target vector: Matrix containing operators in weak form:

$$y = \begin{bmatrix} \vdots \\ \Xi^{\dot{C}}|_{n-1} \\ \Xi^{\dot{C}}|_{n} \\ \Xi^{\dot{C}}|_{n+1} \\ \vdots \end{bmatrix} \quad \Xi = \begin{bmatrix} \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \\ \Xi^{\text{cons}}|_{n-1} & \Xi^{C}|_{n} & \Xi^{C^2}|_{n-1} & \ldots & \Xi^{\nabla^2 C}|_{n-1} & \Xi^{\nabla^4 C}|_{n-1} & \Xi^{\nabla^2 C_{BC}}|_{n-1} & \ldots \\ \Xi^{\text{cons}}|_{n} & \Xi^{C}|_{n} & \Xi^{C^2}|_{n} & \ldots & \Xi^{\nabla^2 C}|_{n} & \Xi^{\nabla^4 C}|_{n} & \Xi^{\nabla^2 C_{BC}}|_{n} & \ldots \\ \Xi^{\text{cons}}|_{n+1} & \Xi^{C}|_{n+1} & \Xi^{C^2}|_{n+1} & \ldots & \Xi^{\nabla^2 C}|_{n+1} & \Xi^{\nabla^4 C}|_{n+1} & \Xi^{\nabla^2 C_{BC}}|_{n+1} & \ldots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \end{bmatrix}$$

- Residual equation:

$$R = y - \Xi\omega$$

Regression problem

- Find ω to minimize the loss function, $l = \|R\|_2$:

$$\omega = \arg\min_{\omega} l(\omega)$$

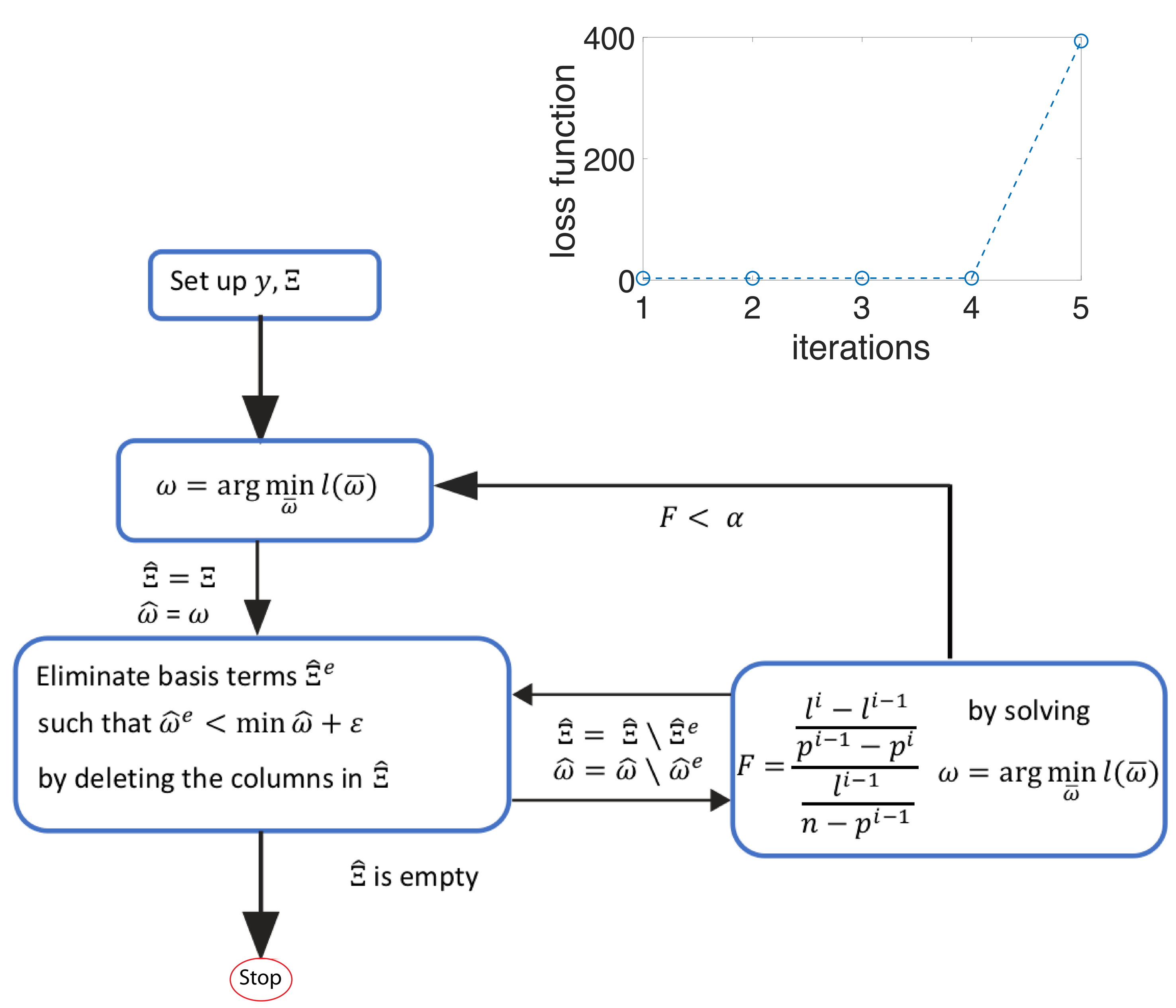
Standard regression will result in a non-parsimonious solution for ω with nonzero contributions in each component of this vector

- Penalization to drive pre-factors toward zero

$$\omega = \arg\min_{\omega}\left\{l(\omega) + \lambda_2\sum_j \omega_j^2\right\} \quad \text{Ridge Regression}$$
$$\omega = \arg\min_{\omega}\left\{l(\omega) + \lambda_1\sum_j |\omega_j|\right\} \quad \text{LASSO-sparsity inducing}$$

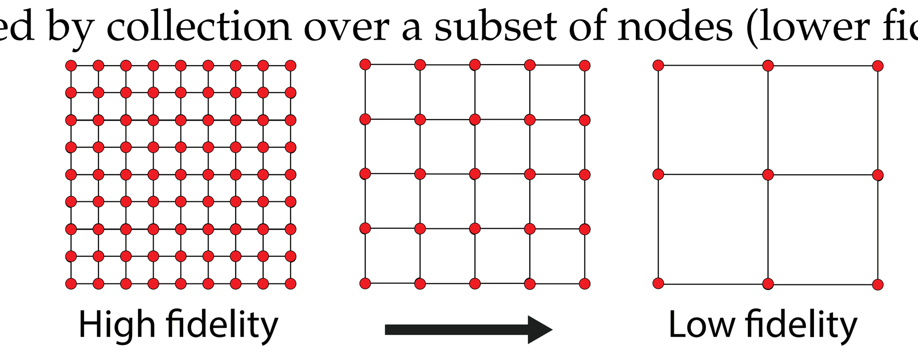
Selecting few relevant terms from large candidate set is challenging

Identification of operators via stepwise regression



Low fidelity and noisy data

- Data generated by simulation on high fidelity mesh, but may be subsampled by collection over a subset of nodes (lower fidelity mesh)



- Observed data $\widehat{C}$ may be noisy

$$\widehat{C} = C + \epsilon$$

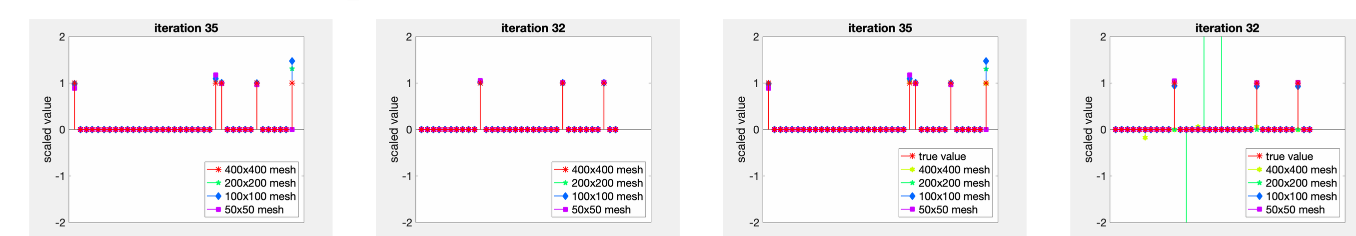
- Noise is amplified through spatial gradient and time derivative.
- Lower fidelity data is favorable to smooth out the amplified noise.
- Higher fidelity data is favorable to capture sharp variations.

Results

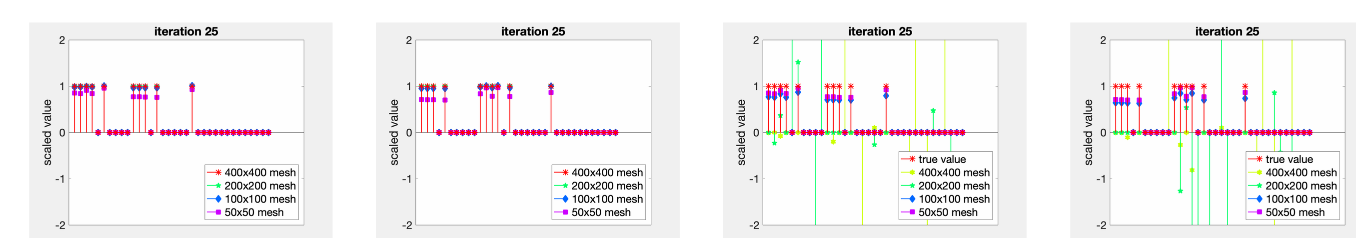
- Candidate operators

Strong form operators	Weak form operators
y	$\int_\Omega w\frac{\partial C_1}{\partial t}\mathrm{d}v$ $\int_\Omega w\frac{\partial C_2}{\partial t}\mathrm{d}v$
$\nabla(*\nabla C_1)$	$\int_\Omega \nabla w\nabla C_1\mathrm{d}v$ $\int_\Omega \nabla wC_1\nabla C_1\mathrm{d}v$ $\int_\Omega \nabla w\nabla C_2\mathrm{d}v$ $\int_\Omega \nabla wC_1^2\nabla C_1\mathrm{d}v$ $\int_\Omega \nabla wC_1C_2\nabla C_1\mathrm{d}v$ $\int_\Omega \nabla wC_2^2\nabla C_1\mathrm{d}v$ $\int_\Omega \nabla wC_1^3\nabla C_1\mathrm{d}v$ $\int_\Omega \nabla wC_1^2C_2\nabla C_1\mathrm{d}v$ $\int_\Omega \nabla wC_1C_2^2\nabla C_1\mathrm{d}v$ $\int_\Omega \nabla wC_2^3\nabla C_1\mathrm{d}v$
$\nabla(*\nabla C_2)$	$\int_\Omega \nabla w\nabla C_2\mathrm{d}v$ $\int_\Omega \nabla wC_1\nabla C_2\mathrm{d}v$ $\int_\Omega \nabla w\nabla C_2\mathrm{d}v$ $\int_\Omega \nabla wC_1^2\nabla C_2\mathrm{d}v$ $\int_\Omega \nabla wC_1C_2\nabla C_2\mathrm{d}v$ $\int_\Omega \nabla wC_2^2\nabla C_2\mathrm{d}v$ $\int_\Omega \nabla wC_1^3\nabla C_2\mathrm{d}v$ $\int_\Omega \nabla wC_1^2C_2\nabla C_2\mathrm{d}v$ $\int_\Omega \nabla wC_1C_2^2\nabla C_2\mathrm{d}v$ $\int_\Omega \nabla wC_2^3\nabla C_2\mathrm{d}v$
$\nabla^2(*\nabla^2 C)$	$\int_\Omega \nabla^2 w\nabla^2 C_1\mathrm{d}v$ $\int_\Omega \nabla^2 wC_1\nabla^2 C_1\mathrm{d}v$ $\int_\Omega \nabla^2 w\nabla^2 C_2\mathrm{d}v$ $\int_\Omega \nabla^2 wC_2\nabla^2 C_2\mathrm{d}v$
non-gradient	$-\int_\Omega w1\mathrm{d}v$ $-\int_\Omega wC_1\mathrm{d}v$ $-\int_\Omega wC_2\mathrm{d}v$ $-\int_\Omega wC_1^2\mathrm{d}v$ $-\int_\Omega wC_1C_2\mathrm{d}v$ $-\int_\Omega wC_2^2\mathrm{d}v$ $-\int_\Omega wC_1^3\mathrm{d}v$ $-\int_\Omega wC_1^2C_2\mathrm{d}v$ $-\int_\Omega wC_1C_2^2\mathrm{d}v$ $-\int_\Omega wC_2^3\mathrm{d}v$
boundary condition	$-\int_{\Gamma_1} w1\mathrm{d}s$ $-\int_{\Gamma_4} w1\mathrm{d}s$ $-\int_{\Gamma_3} w1\mathrm{d}s$ $-\int_{\Gamma_4} w1\mathrm{d}s$

- Stem-and-leaf plots show active operators selected by stepwise regression out of a set of 38 possible choices



Diffusion-reaction operators for two fields C_1 and C_2 using noise-free and noisy data.



Cahn-Hilliard operators for two fields C_1 and C_2 using noise-free and noisy data.

Key takeaways

- The weak form regularizes higher-order derivatives.
- The variational approach naturally delineates and identifies boundary conditions.

We are working on identifying governing equations using real experimental data that is:

- Sparse, being available at very coarse time steps.
- Incomplete, being only available over subdomains of the full field that are uncorrelated with respect to time.

Reference:

Z.Wang, X. Huan, K. Garikipati, Variational system identification of the partial differential equations governing pattern-forming physics: Inference under varying fidelity and noise, *Computer Methods in Applied Mechanics and Engineering*, 356, 44-74, 2019.
doi.org/10.1016/j.cma.2019.07.007