Mathematical modelling in biology and data science

Lisa Maria Kreusser

University of Bath

PiWORKS Seminar
October 26, 2021

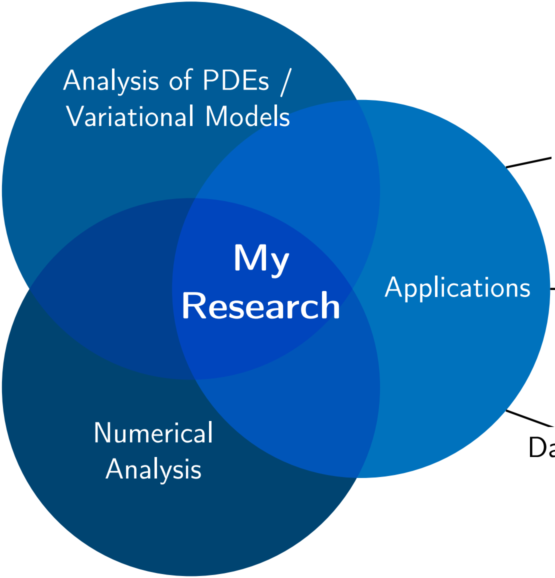
Who am I?

I am an applied mathematician.

- **My position:** Lecturer (Assistant Professor) in the Department of Mathematical Sciences, University of Bath

- **My research:** Mathematical methods in partial differential equations, mathematical and numerical analysis with applications in biology and data science

My Research



Biology →

fingerprints, networks

Industry →

textile production

Data Science →

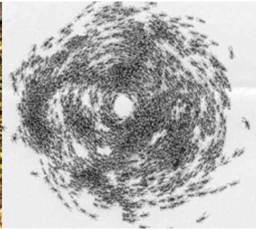
image analysis, data classification

CV in short

- 10/2010–07/2013: **B.Sc. in Mathematics (Minor: Physics)**, *University of Kaiserslautern, Germany.*
- 08/2013–09/2015: **M.Sc. in Mathematics**, *University of Kaiserslautern, Germany.*
  - **Study abroad** (08/2013–12/2013): *National University of Singapore, Singapore.*
  - **Research internship** (01/2014–04/2014): *Imperial College London, UK.*
- 10/2015-09/2019: **PhD in Mathematics**, *University of Cambridge, UK.*
- 10/2019-07/2021: **Research Fellow in Mathematics**, *Magdalene College, University of Cambridge, UK.*
- since 08/2021: **Lecturer (Assistant Professor)**, Department of Mathematical Sciences, University of Bath, UK.

Collective behaviour in nature[1]

(a) Marching locusts

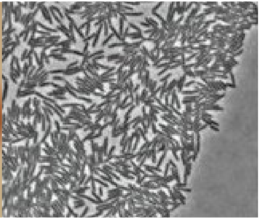


(b) Colony of army ants

(c) School of fish

(d) Swarms of birds

(e) Fingerprints



(f) Swarming of E. coli

[1] C. Liu et al. Controllable swarming and assembly of micro/nanomachines. Micromachines, 9(1), 2018.

General setting for isotropic pattern formation

- Large system of interacting particles
- **Interaction forces**
  - Short-range repulsive
  - Long-range attractive
  - Radially symmetric
  - Alignment along the distance vector

$\Rightarrow F(r) = f(|r|)r = -\nabla W(r)$

Isotropic interaction model[2]

- **Microscopic model**:

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = \frac{1}{N} \sum_{\substack{k=1 \\ k \neq j}}^{N} F(x_j - x_k),$$

$$F(r) = -\nabla W(r) = f(|r|)r$$

- **Associated continuum model**:

$$\rho_t + \nabla \cdot (\rho u) = 0, \qquad u = -\nabla W * \rho$$

- Minimization of **interaction energy**

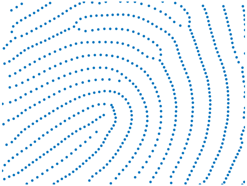
$$\mathcal{W}[\rho] = \frac{1}{2} \int_{\mathbb{R}^2} (W * \rho)(x) \rho(\mathrm{d}x)$$

[2] Albi, Balagué, Bertozzi, Burger, Carrillo, Di Francesco, Fellner, Figalli, Fornasier, James, Kolokolnikov, Laurent, Mellet, Raoul, von Brecht, Tadmor, Toscani, Uminsky, . . .

Simulation of fingerprints

Motivation

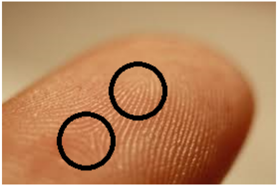
- Large databases required for **fingerprint identification algorithms**
  - Difficulty: collecting databases of real fingerprints
  - Solution: Create synthetic fingerprint images
- Modeling **formation of fingerprints**
- **Challenging generalization** of popular class of mathematical models

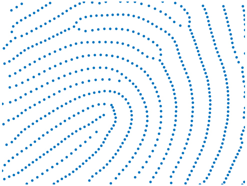


Simulation of fingerprints

Motivation

- Large databases required for **fingerprint identification algorithms**
  - Difficulty: collecting databases of real fingerprints
  - Solution: Create synthetic fingerprint images
- Modeling **formation of fingerprints**
- **Challenging generalization** of popular class of mathematical models

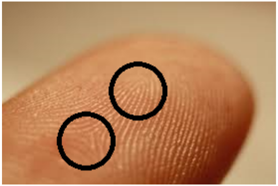


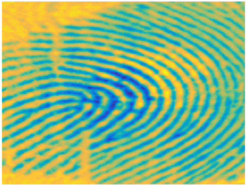


Simulation of fingerprints

Motivation

- Large databases required for **fingerprint identification algorithms**
  - Difficulty: collecting databases of real fingerprints
  - Solution: Create synthetic fingerprint images
- Modeling **formation of fingerprints**
- **Challenging generalization** of popular class of mathematical models

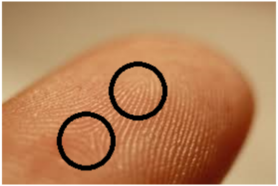


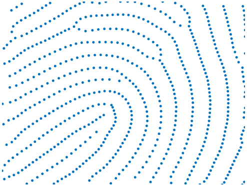


Simulation of fingerprints

Motivation

- Large databases required for **fingerprint identification algorithms**
  - Difficulty: collecting databases of real fingerprints
  - Solution: Create synthetic fingerprint images
- Modeling **formation of fingerprints**
- **Challenging generalization** of popular class of mathematical models

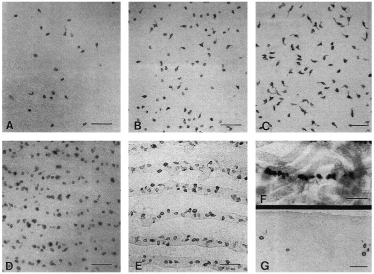




Biological background for fingerprint development[3]

Three phases of fingerprint development

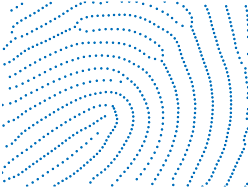
1. Growth forces in epidermis and shrinkage of volar pad create **compressive stress** (modeled by Kücken, Newell: 2004, 2005)
2. **Rearrangement of Merkel cells** from a random configuration into parallel ridges along the lines of smallest compressive stress
3. **Primary ridges** are induced by the Merkel cells.



[3]D.-K. Kim, K. Holbrook, *The appearance, density, and distribution of Merkel cells in human embryonic and fetal skin*, J. Invest. Dermat., 1995

Model assumptions:

- Extension of the isotropic microscopic model resulting in an **anisotropic interaction model**
- Large number of **interacting cells**
- **Interaction forces** F
  - short-range repulsive
  - long-range attractive
- **Underlying stress field** T influencing forces



Mathematical formulation:

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = \frac{1}{N} \sum_{\substack{k=1 \\ k \neq j}}^{N} F(x_j - x_k, T(x_j)), \quad j = 1, \dots, N$$

[4] M. Kücken, C. Champod, *Merkel cells and the individuality of friction ridge skin*, Journal of Theoretical Biology, 317 (2013), pp. 37-72

Contributions: Synthetic Fingerprint Images [5] [6] [7]

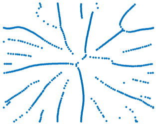
Mathematical and numerical analysis

- **Analysis** of interaction model **without simplifications** requiring **new mathematical methods**
- **Rigorous limit**
- **Pattern formation** (form, stability)
- Patterns and parameter dependence

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = \frac{1}{N}\sum_{k=1}^{N} F(x_j - x_k, T(x_j)), \quad k = 1, \ldots, N$$

$$\downarrow$$

$$\partial_t \rho + \nabla_x \cdot [\rho \, (F(\cdot, T(x)) * \rho)] = 0$$



[5] Burger, Düring, LMK, Markowich, Schönlieb, M3AS
[6] Düring, Gottschlich, Huckemann, LMK, Schönlieb, J. Math. Bio.
[7] Carrillo, Düring, LMK, Schönlieb, SIAM J. Appl. Dyn. Sys.

Contributions: Synthetic Fingerprint Images [5] [6] [7]

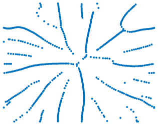
Mathematical and numerical analysis

- **Analysis** of interaction model **without simplifications** requiring **new mathematical methods**
- **Rigorous limit**
- **Pattern formation** (form, stability)
- Patterns and parameter dependence

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = \frac{1}{N}\sum_{k=1}^{N} F(x_j - x_k, T(x_j)), \quad k = 1, \ldots, N$$

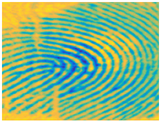
$$\downarrow$$

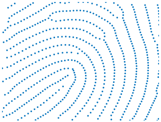
$$\partial_t \rho + \nabla_x \cdot [\rho\,(F\,(\cdot, T(x)) * \rho)] = 0$$



Applications

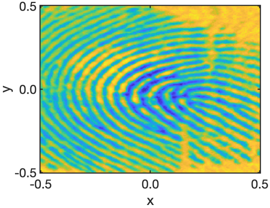
- Proposed a **bio-inspired model**
- **Realistic fingerprint simulations**
- Real-world phenomena



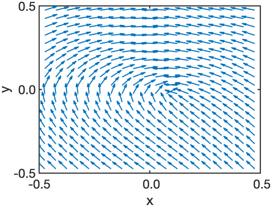


[5] Burger, Düring, LMK, Markowich, Schönlieb, M3AS
[6] Düring, Gottschlich, Huckemann, LMK, Schönlieb, J. Math. Bio.
[7] Carrillo, Düring, LMK, Schönlieb, SIAM J. Appl. Dyn. Sys.

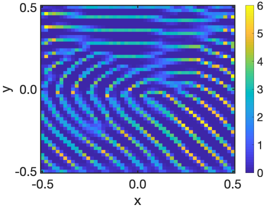
Fingerprint simulations for the macroscopic model



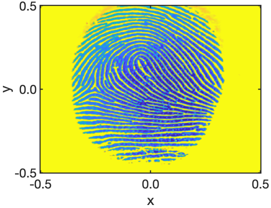
(a) Original



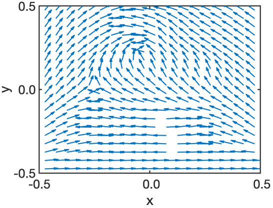
(b) s



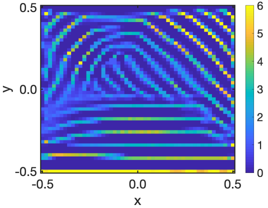
(c) Stationary solution



(d) Original



(e) s



(f) Stationary solution

Characteristics of models for collective behaviour[8]

Biological models

- **Large number** of interacting particles
- **Interactions** over short distances
- **Complex patterns** and stationary states such as flocks or clusters

From biology to label propagation

- Apply ideas from **collective dynamics** in the context of labeling and classification problems
- **Semi-supervised learning:** large number of data points, some of them being already correctly matched to labels

[8] Albi, Balagué, Bertozzi, Burger, Carrillo, Di Francesco, Fellner, Figalli, Fornasier, James, Kolokolnikov, Laurent, Mellet, Raoul, von Brecht, Tadmor, Toscani, Uminsky, . . .

Graph-based classification

Given $n + m$ data points $V = \{X_1, \dots, X_{n+m}\}$:

- Determine **similarity measure** $w_{i,j}$ between data points X_i and X_j
- **Graph construction** based on similarity measure
- **Partition of the graph** using biological models

(a) Data set

(b) Graph $G = (V, w)$

Graph Laplacian methods

Given graph $G = (V, w)$:

- **Graph Laplacian**

$$\Delta_G u_i = \sum_{j \in V} w_{ij}(u_i - u_j), \quad i \in V$$

- **Generalised continuum Laplacian**

$$\mathcal{L}(u) = \frac{1}{\rho^p} \nabla \cdot \left(\rho^q \nabla \left(\frac{u}{\rho^r} \right) \right)$$

  for distribution of points ρ with fixed parameters $p,\ q,\ r \in \mathbb{R}$ with particular choice $(p, q, r) = (1, 2, 0)$

- **Analysis** of eigenvalues and eigenfunctions in suitable scaling limits: Garcia Trillos and Slepcev, Hoffmann et al., ...

Mathematical formulation of the biological model

- n **data points** with $m \ll n$ correctly labelled
- **Information propagation** of correct labels to unlabelled points
- **Characteristic** $u_i = u_i(t)$ of agent i for $i = 1, \dots, n$
- **Interaction radius** $\epsilon > 0$
- **Influence** $w_{ij} = \eta(x_i, x_j, \epsilon)$ that agent i has on j
- **Interacting system:**

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \sum_{j \neq i} w_{ij}(u_j - u_i) \text{ for } i = 1, \dots, n$$

- Includes **well-known models**, e.g. Cucker-Smale model and Krause's opinion formation model

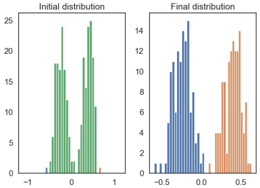
Properties of the continuum model[9]

We provide rigorous proofs for

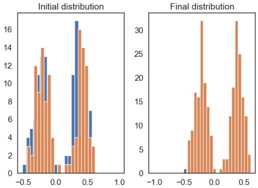
- **Derivation of the continuum model** (limit as $n \to \infty$)
- **Existence and uniqueness** of solutions
- **Consistency of labelling:** Stationary solution u to macroscopic model has a similar structure as the given probability measure
- **Maximum principle:** Solutions can attain their maximum and minimum on the parabolic boundary only
- **Dependence on edge weights and regularisation:**
  - Edge weights w_{ij} and parameter γ correspond to rescaling in time
  - Γ-convergence result as regularisation parameter $\kappa \to \infty$

[9] LMK, M.-T. Wolfram. *On anisotropic diffusion equations for label propagation*, arXiv:2007.12516

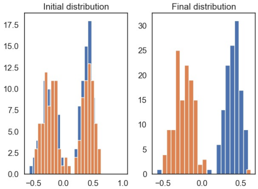
Dependence on initial data



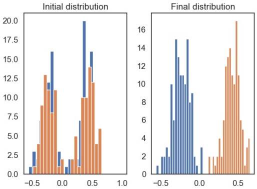
(a) Zero initial data



(b) Normally distributed initial data

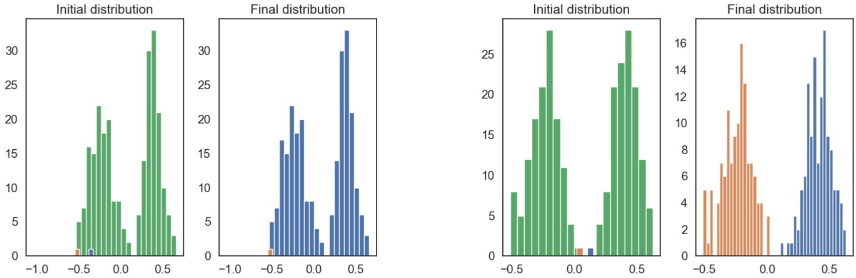


(c) Uniformly distributed initial data



(d) Uniformly distributed initial data

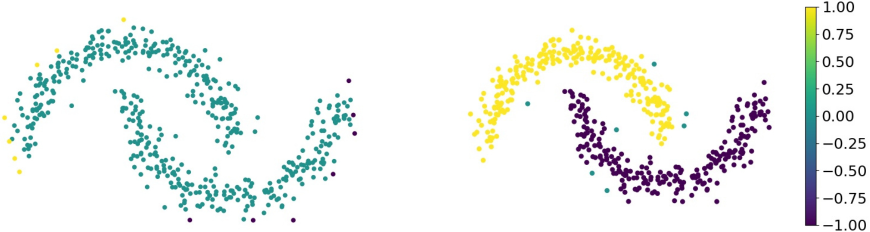
Dependence on label location



(e) Test case 1 (f) Test case 2

Figure: Dependence on label location for hom. initial distribution and their stationary solution for the microscopic discretisation.

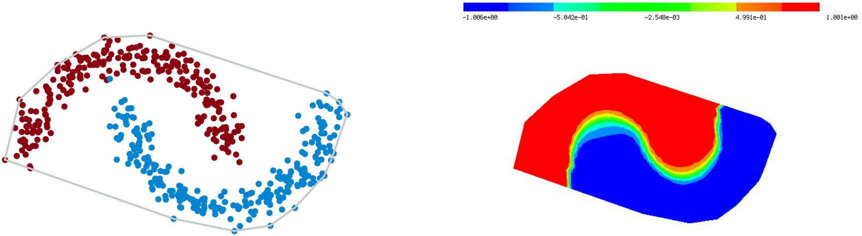
Two moons: Initial and final distribution



(a) Initial distribution (b) Final distribution

Figure: Initial and final distribution at $T = 25$ using the parameters $\gamma = 1.0$ and $\kappa = 10$. We assume that all points defining the convex hull are correctly labelled.

Two moons: micro- and macroscopic



(a) Microscopic dataset (b) Final macroscopic distribution

Figure: Final distribution at $T = 25$ using the parameters $\gamma = 1.0$ and $\kappa = 10$ for the microscopic and macroscopic models. We assume that all points defining the convex hull are correctly labelled.

Multi-label and more complex labelling problems

- MNIST dataset:
  1797 samples of digital digits
- Labels $L = \{0, \ldots, 9\}$
- Weights $w_{ij} = \mathbb{1}_{d_{\mathcal{W}_2(x_i,x_j)} \leqslant \bar{c}} d^{-1}_{\mathcal{W}_2(x_i,x_j)}$
- Choose 320 samples, determine weight matrix, assume that the first 40 digits are correctly labelled and apply to

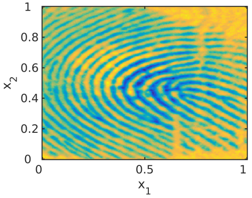
$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \gamma \sum_{j \neq i} w_{ij}(u_j - u_i) - \kappa W'(u_i),$$

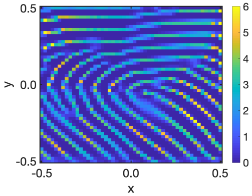
$$i = 1, \ldots, n$$

- 84.285% of the labels assigned correctly
- Performance improvable by fine-tuning the parameters

Conclusion

1. **Model development**
2. **Continuum models:**
   Derivation of model in the limit $n \to \infty$.
3. **Quantitative behaviour: Analytic results** to characterize the behaviour of solutions to the continuum problem.
4. **Computational experiments** illustrating and exemplifying the structure of solutions to the micro- as well as macroscopic equation





PhD opportunities in Bath

EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath

SAMBa

- 10-15 fully-funded PhD studentships available annually for four year programme
- Fusion of applied mathematics, numerical analysis, probability, statistics
- Distil industrial and interdisciplinary problems into mathematical ones, and solve them
- Choose and shape your own research direction according to your interests
- Diversity of cohort in an inclusive environment an essential component

You'll need:
- Good degree with high mathematical content
- Some research or professional experience

Research themes include:
- Inverse problems and compressed sensing
- Efficient numerical algorithms and scientific computing
- Data assimilation and uncertainty quantification
- Hybrid applied and stochastic modelling
- Spatial statistics and Bayesian networks
- Probability, statistical physics and applied analysis
- Mathematical machine learning

Application and industrial areas include:
- Medical imaging
- Environmental modelling
- Drug safety and development
- Advanced materials
- Remote sensing
- Nuclear safety
- Fluid dynamics

University of Bath

UKRI Engineering and Physical Sciences Research Council

More information: www.bath.ac.uk/samba

Thank you very much for your attention!

More information:
https://people.bath.ac.uk/lmk54/
Email: lmk54@bath.ac.uk