Modelling of Complex Real-World Systems
Introduction and overview

Ton Coolen
Department of Biophysics, Radboud University
module NWI-NM127, January 2021

What are complex real-world systems?

Simple systems

processes described by mathematical equations
that are easy to solve analytically, e.g.

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x} = \mathbf{A}\mathbf{x} \quad \text{or} \quad \mathbf{x}(t{+}1) = \mathbf{x}(t) + \mathbf{A}\mathbf{x}(t), \qquad \text{with} \;\; \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \;\; \mathbf{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

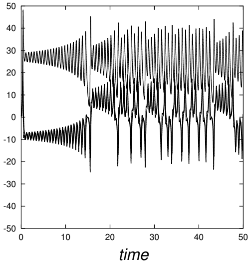
- linear eqns, hence solvable (exponential functions of time), stationary states $\mathbf{x}$ solved from $\mathbf{A}\mathbf{x} = \mathbf{0}$
- deterministic eqns, so $\mathbf{x}(t)$ exactly defined at all times
- small nr of parameters (here: four), so one can classify *all* types of behaviour

simple = linear, deterministic, small nr of parameters
what happens if we lose those features?

- linear → nonlinear eqns

  – no longer solvable analytically
  – possibly chaotic dynamics

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}x_1 &= \sigma(x_2 - x_1)\\ \frac{\mathrm{d}}{\mathrm{d}t}x_2 &= x_1(\rho - x_3) - x_2\\ \frac{\mathrm{d}}{\mathrm{d}t}x_3 &= x_1 x_2 - \beta x_3\end{aligned}$$



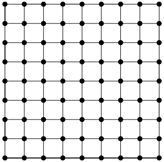
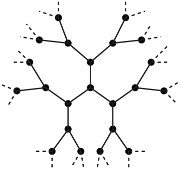
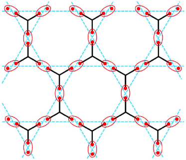
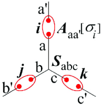
- deterministic → stochastic eqns

  – dynamical laws for probability density $p_t(\mathbf{x})$
  – can find equations only for certain types of noise distributions
  – problem effectively becomes infinite-dimensional

- small $\rightarrow$ large nr of parameters

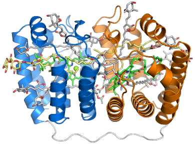
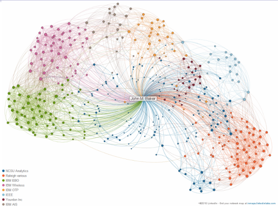
  e.g. $\mathbf{x} \in \mathbb{R}^N$ with $N \sim 10^{24}$ ...

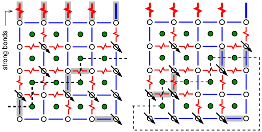
  common strategy: choose many of the parameters to be identical
  (i.e. identical forces, periodic structures, ...)

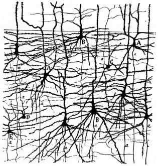
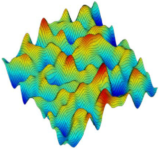
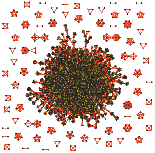





– choose uniform ‘all-to-all’ interactions (e.g. mean-field models)
– or use periodicity of the system (e.g. transfer matrices)
– only works for homogeneous and regular systems

unfortunately:
systems in the real world tend to be large, messy, noisy, and nonlinear




Complex systems

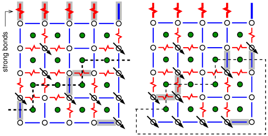
- many interacting variables
- non-linear and non-deterministic (i.e. ‘noisy’) dynamics
- heterogeneity/irregularity in interactions or forces

Examples from different disciplines

Physics

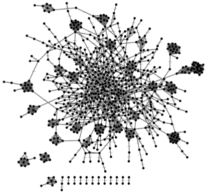
- *spin glasses*

  magnetic materials with (pseudo-) random interactions between $\sim 10^{24}$ spins



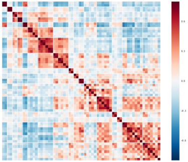
- *spin systems on random topologies*

  models of social interactions between $\sim 10^{6}$ (or more) individuals



- *eigenvalue distribution of random matrices*

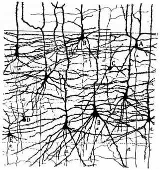
  random matrix elements in symmetric $10^4 \times 10^4$ matrices (e.g. energy levels of large atoms)



Biology

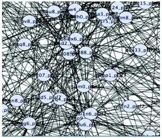
- *recurrent neural networks*

  network of $\sim 10^{12}$ information processing neurons with modifiable excitatory and inhibitory interactions (heterogeneity: stored information)



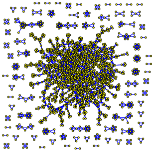
- *gene regulation networks*

  control of cell states via switching on/off of $\sim 10^4$ interacting genes (heterogeneity: transcription factors)



- *adaptive immune system*

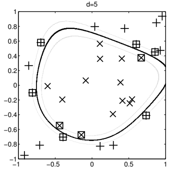
  signalling between $\sim 10^8$ T-cell and B-cell clones (heterogeneity: encountered viruses/bacteria/...)



Data science and statistics

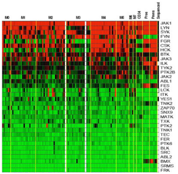
- *binary classifiers*

  inference from data of algorithm
  for prediction of binary outcome labels
  (heterogeneity: realisation of the data)



- *parameter inference*

  inference of model parameters in
  medical outcome prediction from genomic data
  (heterogeneity: realisation of the data)



Economics

- *models of financial markets*

  understanding instabilities of financial markets, resulting from (partly irrational) decision-making of interacting agents

(heterogeneity: trading strategies of individuals)

Aims and structure of the module

Aim

explain mathematical tools for modelling complex systems

- many interacting variables → *statistical approach*
- non-linear and non-deterministic dynamics → *stochastic processes*
- heterogeneity in interactions or forces → *replica method, GFA*

Outline

- methods
- applications in physics
- applications in biology
- applications in data science and economics
- advanced topics – if time permits

Practicalities

- weekly lectures: Monday 13:30-15:15, online (initially)
- weekly exercise classes: Friday 10:30-12:15, in HG01.028
- final examination, second half of june
- new module, so contents may evolve

A. General Methods

A.1 *Mathematical preliminaries.* Gaussian integrals, the delta function, steepest descent integration, exponential distributions and generating functions.

A.2 *Stochastic processes.* Langevin, Fokker-Planck and master equations. Detailed balance, convergence to equilibrium and the H-theorem. Correlation and response functions.

A.3 *Networks and graphs.* Definitions of standard concepts (nodes, links, degrees, degree distributions). Random graph ensembles.

B. Methods for Heterogeneous Systems

B.1 *Analysis of homogeneous systems.* Homogeneous systems. Sequential master equation. Parallel Markov chain. Fokker-Planck equation.

B.2 *The replica method.* Standard definition via the logarithm identity. Application to generating functions and to optimization algorithms. Alternative forms and derivations.

B.3 *Generating functional analysis.* Generating functionals for dynamical observables. Illustration on simple toy models. The use of causality to eliminate non-physical solutions.

C. Applications in physics

C.1 *Spin glasses.* The SK model. Replica analysis of the SK model. Interpretation of order parameters. Ergodicity and replica symmetry (RS). Replica symmetry breaking (RSB).

C.2 *Spin systems on graphs - equilibrium.* Equilibrium replica analysis of spin or voter models on Erdös-Renyi graphs. Phase diagrams. Extension to other random graphs.

C.3 *Spin systems on graphs – dynamics.* Generating functional analysis for parallel discrete time dynamics. Derivation of closed macroscopic laws. Impact of interaction symmetry.

C.4 *Spectra of random matrices.* Derivation of the Edwards-Jones formula. Derivation of Wigner's semi-circular law and the Marchenko-Pastur distribution, using the replica method.

D. Applications in biology

- D.1 *Neural networks.* The Hopfield model of associative memory. Solution away from saturation. Derivation of phase diagram using the replica method. The RSB transition.
- D.2 *Neural network dynamics.* Generating functional analysis for parallel discrete time dynamics. Derivation of macroscopic laws. Impact of interaction symmetry and dilution.
- D.3 *Coupled oscillators on graphs.* Kuramoto oscillators. Replica analysis of randomly interacting Kuramoto oscillators on Erdös-Renyi graphs. Guzai expansion and phase diagram.

D. Applications in data science and economics

E.1 *Binary classifiers.* Discriminant analysis and version space for binary classifiers. Gardner's approach to storage capacity. Replica analysis and derivation of the critical capacity.

E.2 *Agent-based models of financial markets.* El Farol problem and minority games. Replica analysis of minority games (MG). Dynamical solution via generating functional analysis.

E.3 *Parameter inference.* Maximum a posteriori probability (MAP) inference. Replica analysis of overfitting in the high-dimensional (big data) regime. Application to logistic regression.