Practical summer course:
Modeling for Systems Biology

June 29 – July 4, 2014

Centre for Genomic Regulation

Barcelona

- Outlines lectures -

- Day 1, Sunday June 29th, 2014 -

"*Basic Mathematical Concepts & Introduction to Matlab*" (Optional)"

Teachers: Kai Dierkes, Andreea Munteanu, Marie Trussart

Overview

The objective of these introductory lectures is to familiarize the participants with the main concepts, keywords and tools that will be employed throughout the week. The topics covered are extremely broad, including fundamentals of linear algebra, defining and solving ordinary differential equations. Obviously, this broad introduction will be of a graphical and intuitive nature, minimizing the algebraic content, when possible. In addition, a short presentation of Matlab's main commands will provide the starting point for the practical sessions of the week.

Lecture Linear Algebra

Abstract

Quantitative biology often aims at capturing the state of biological systems, e.g. of cells, in terms of sets of numbers, e.g. protein concentrations. Interpreting these lists of numbers as spatial coordinates, the different states of a system can be conceptualized as corresponding to distinct points within a space of suitable dimension, also referred to as state space. Linear algebra provides concepts and tools to describe and analyze the structure, the geometry, and the transformations of such abstract spaces. Within this lecture, we will introduce some of the important notions and techniques that will be used during the rest of this summer school.

Topics

State space; vector space; scalar vs vector; operations on vectors (addition, inner product); norm of a vector; orthogonality; basis and linear independence; linear transformations and matrices; multiplication of a matrix with a vector; multiplication of matrices; determinant of a matrix (2d & 3d).

Lecture Ordinary Differential Equations

Abstract

This lecture will define the main vocabulary associated to ordinary differential equations (ODEs), from their definition and necessity for modeling, to (conceptual) solving and subsequent graphical representation of the solutions. It will present an intuitive approach for transforming the information we can measure into a system of differential equations, and how to understand the output of such systems. Beyond and in contrast

to ODEs, fundamental definitions on discrete systems and probability theory will be briefly mentioned.

Topics

Differential equation, ordinary vs partial; derivative & integral; analytic vs numerical; parameters & variables (dependent & independent); linearity vs nonlinearity; steady states, phase plane (portrait); eigenvalues. Deterministic vs stochastic; probability distributions.

Lecture Introduction to Matlab

Abstract

This lecture introduces students to the scientific programming language MATLAB. Prior computer programming experience is not required.

The lectures begin with a presentation of variables within Matlab workspace like vectors, matrices, and arrays. Students are introduced to basic commands in MATLAB, as well as built-in functions that provide useful shortcuts.

Then we will focus on the differences between MATLAB scripts and MATLAB functions and describe when one method of programming organization might be preferable to the other.

The principles are illustrated through different examples and exercises that make it more interactive.

Because of its introductory nature, students who have prior programming experience may find these lectures to be essentially a review. For students without this experience, these lectures are a prerequisite to using MATLAB.

Topics

Variables in Matlab: vectors and matrices; defining functions and using m-filles; write and load data; control structures; data visualization.

- Day 2, Monday June 30th, 2014 -

"Dynamical systems theory, networks"

Teacher: Mukund Thattai

Overview: The dynamics of genetic networks

Living cells are made up of networks of interacting genes, proteins and biochemicals. Simple interactions between network components can lead to complex collective dynamics. Cells use these emergent dynamical properties to propagate signals and perform computations, to probe their surroundings and generate appropriate responses. These phenomena can only be understood at the network level, not at the level of individual components.

I will describe how to construct and analyze dynamical models of genetic networks. These lectures will begin with a brief introduction to the physical basis of gene regulation. I will then discuss how to convert cartoon models of genetic networks into differential equations. We will learn how techniques from dynamical systems theory can be used to study the behavior of these systems. Finally, I will discuss how physical processes place limits on the accuracy of biological control.

Lecture 1: The physical basis of gene regulation

Promoters and transcription factors; chemical kinetics; thermodynamic model of TF-promoter binding; binding of multiple factors; activation and repression; the Hill equation; detailed balance; out-of-equilibrium systems.

Lecture 2: Genetic networks

Derivation of the network equation; the network equation as an extension of binary threshold models; models of simple feedback systems: negative feedback and noise reduction; positive feedback and bistability; hysteretic and ring oscillators. Comparison with synthetic biology experiments.

Practical 1: Dynamical systems and bifurcation analysis I (whiteboard)

Ordinary differential equations; phase-plane analysis and vector fields; nullclines and fixed points.

Practical 2: Dynamical systems and bifurcation analysis II (MATLAB)

Bifurcations, multistability, limit-cycle oscillations; bifurcation analysis in 2-D; trace and determinant conditions for stability in 2-D. Application to genetic network models.

Lecture 3: Physical limits of biological control

Hopfield's kinetic proofreading mechanism. Berg/Purcell analysis of physical limits of signalling.

- Day 3, Tuesday July 1th, 2014 -

"Stochastic systems"

Teacher: Jordi Garcia Ojalvo

Overview: Stochastic systems

Dynamic mathematical models in the form of ordinary differential equations approximate reasonably well the behavior of biochemical networks under many conditions. However, the underlying kinetics of cellular processes are fundamentally stochastic, as a result of the random occurrence of biochemical reactions in the cell. This randomness manifests itself in fluctuating, or "noisy", levels of proteins and other biomolecules. Depending on the circuit of interest, this noise will affect not only quantitatively, but also qualitatively the operation of a circuit, resulting in some cases in behaviors that are not possible under a purely deterministic regime.

This part of the course focuses on the fundamental bases for describing noise in biochemical networks, starting from conceptual aspects (e.g., how to measure and detect noise) and fundamentals of chemical kinetics that allow for a quantitative description of random cellular processes, both from a continuous and a discrete perspective. Applications to prototypic circuits will reveal that stochastic noise can be detrimental –for instance, by deteriorating cellular decision processes– or advantageous –for example, by creating cellular diversity–.

Lecture 4: Noise in biochemical reactions

Noise and cellular heterogeneity; imaging techniques; intrinsic vs. extrinsic noise; elementary reaction systems and propensities; from the chemical master equation to the Fokker-Planck equation; the chemical Langevin equation.

Lecture 5: Continuous description of stochastic processes

Brief overview of stochastic processes; additive vs. multiplicative noise; stochastic interpretations; white vs. colored noise; generation of correlated noises; introduction to the simulation of stochastic differential equations.

Practical 3: Simulating the chemical Langevin equation

Simulation of stochastic effects in direct transcriptional regulation; additive vs. multiplicative noise; chemical Langevin equation of a genetic toggle switch; noise-induced bistability.

Lecture 6: Discrete stochastic simulations

The stochastic simulation algorithm (SSA); computationally efficient variants of SSA; deterministic limit of the SSA; relation between the stochastic and deterministic approaches; applications: oscillations, excitability.

Practical 4: Controlling noise in stochastic simulations

Implementation of SSA; simulation of positive feedback regulation; controlling noise levels by balancing transcription and translation; controlling noise levels by tuning cell size; noise-induced bistability.

- Day 4, Wednesday July 2th, 2014 -

"*Multivariant and multidimensional data analysis*"

Teacher: Fernando Amat

Overview

The current trend in science, and systems biology in particular, is to acquire an increasing amount of data from different sources. Oftentimes, it is impossible for a human observer to extract all the information contained in this raw data due to its size and the large number of variables involved. The objective of the lectures on day 4 is to familiarize the students with the concepts involved in visualizing, analyzing, comparing and classifying multivariate signals (e.g. microarray expression data) in order to transform the data into meaningful representations where key insights can be obtained and explained. Each lecture will have a follow-up Matlab session with exercises drawn from the topics discussed in the lectures, where the students can apply the techniques to real-world datasets.

Lecture 7: Multivariate analysis and dimensionality reduction techniques

Abstract

We will introduce the notation for multivariate analysis with different examples that will be used in throughout the day (in the lectures and the practical sessions) and discuss problems associated with high-dimensional signals such as the "curse of dimensionality" or data visualization. In this context, we will present various dimensionality reduction techniques, such as principal components analysis, multidimensional scaling and locally linear embedding, to handle the complexity of high-dimensional feature spaces.

Topics

Curse of dimensionality, data visualization, model selection, dimensionality reduction, kernel methods.

Practical 5: Dimensionality reduction techniques

Principal component analysis; multidimensional scaling; locally linear embedding applied to micro-array data.

Lecture 8: Unsupervised pattern recognition methods

Abstract

We will learn how the techniques of Lecture 7 fit into the paradigm of unsupervised learning. In particular, we will present different clustering techniques, such as k-means and hierarchical clustering, in order to recognize patterns in the multivariate data when we do not have a labeling associated with it. We will also discuss different approaches to validate clustering results and to estimate the right number of clusters.

Topics

Clustering, K-means, hierarchical clustering, GAP statistics, silhouette.

Practical 6: Clustering methods

K-means; hierarchical clustering; GAP statistics applied to micro-array data.

Lecture 9: Supervised pattern recognition methods

Abstract

We will learn how the techniques of Lecture 7 fit into the paradigm of supervised learning. In particular, we will present different classification and regression techniques, such as support vector machines and boosting trees, in order to recognize patterns in the multivariate data when we do have a labeling associated with it. We will also discuss different approaches to validate the trained model in order to avoid overfitting.

Topics

Training and test set, cross-validation, linear discriminant analysis, logistic regression, support vector machines, classification and regression trees, boosting.

Practical 7: Supervised Learning Techniques

Support vector machines; classification trees; boosting applied to fly behavior data.

- Day 5, Thursday July 3rd, 2014 -

"*Parameter inference, reverse engineering*"

Teacher: Theodore Perkins

Overview

Having covered both modeling approaches (day 2 and 3) and analysis of quantitative data (day 4) we will now show how both of these can be combined to infer regulatory networks and their dynamics from observed gene expression patterns or other evidence. Such data-driven modeling is called *reverse engineering*. It is becoming increasingly powerful and popular in systems biology. We will introduce a range of methods for reverse engineering and parameter inference, for different types of models. In the practicals we will apply some of these approaches to reverse engineer a genetic network that is important in early *Drosophila* development.

Lecture 13: Inference of static network models

Abstract

In this first lecture of the day the students will be introduced to methods that allow us to infer network structure (or topology)—the interactions between different factors in a network—without considering dynamical behavior. We will focus on two broad classes of approaches, "correlation" networks and genetic interaction networks. These methods are easily scalable and especially powerful when used to analyze large-scale 'omics' data sets.

Topics

Correlation networks; Pearson and Spearman correlation; information and mutual information; Relevance Networks algorithm; ARACNE; significance determination. interaction networks; screening for synthetic lethality and other genetic interactions; epistasis analysis.

Practical 8: Practicing static network inference using the gap gene system of *Drosophila*

Lecture 14: Inference of dynamic network models

Abstract

In many cases, such as during development, genetic networks have inherently dynamical functions; cascades of gene activity guide differentiation, tissue organization, growth, etc. Thus, while static network analysis is generally an efficient

means to reconstruct relationships between genes, in some cases dynamical models are more appropriate. Dynamical model fitting is a problem arising in many scientific disciplines and is widely recognized as challenging. In this lecture, we will try to gain some insight into why dynamical model fitting is so difficult—essentially because the problem is often highly nonlinear in the model parameters—and look at various ways to ameliorate these difficulties. Solutions include alternative optimization strategies and reformulation of dynamical model fitting problems in more static form, which has its own pros and cons.

Topics

Formulation of the model fitting problem; nonlinearity; optimization strategies, including: gradient descent, simplex methods, simulated annealing; functional data analysis and related "static" reformulations of the dynamical model fitting problem; advantages and disadvantages; hybrid approaches.

Practical 9: Practicing dynamic network inference using the gap gene system of *Drosophila*

Lecture 15: Inference of stochastic network models

Abstract

As previous lectures have described, gene expression is an inherently stochastic phenomenon. It also sometimes appears stochastic due to limitations in our ability to collect data. In the final lecture of the day, we will consider the problem not just of dealing with stochastic expression data, but explicitly constructing models to capture that stochasticity. We will consider to main kinds of models, Bayesian networks and stochastic chemical kinetic models. For each, we will consider advantages and disadvantages in what they are able to represent, the feasibility of their respective model fitting problems, and how the nature of the data influences the feasibility of fitting.

Topics

Bayesian networks (in great brevity); network equivalence; model fitting for fully observable data; identifiability; stochastic chemical kinetic models; assessing the probability of the data; approximation schemes.

- Day 6, Friday July 4th, 2014 -

"*Simulating Tissue Morphogenesis and Signalling*"

Teacher: Dagmar Iber

Overview

During embryonic development tissue morphogenesis and signaling are tightly coupled. It is therefore important to simulate both tissue morphogenesis and signaling simultaneously in *in silico* models of developmental processes. The resolution of the processes depends on the questions of interest. As part of the lecture I will show how PDE-based signaling models can be solved on growing and deforming domains and I will introduce different descriptions of tissue morphogenesis. In the simplest approximation tissue is a continuous domain and tissue expansion is described according to a pre-defined function of time (and possibly space). In a slightly more advanced version the expansion speed and direction of the tissue may depend on a signaling variable that evolves on the domain. Both versions will be referred to as 'prescribed growth'. Alternatively tissue can be regarded as incompressible fluid and can be described with Navier-Stokes equations. Local cell expansion, proliferation, and death are then incorporated by a source term that may depend on local signaling. In other applications the cell boundaries may be important and cell-based models must be introduced. Finally, cells may move within the tissue, a process best described by agent-based models.

Lecture 13: Introduction to Reaction-Diffusion Models

Practical 10: Reaction Diffusion Models

Simple examples, which can be solved both by hand and using MATLAB

Lecture 14: Patterning dynamics on growing domains

Practical 11: Reaction Diffusion Models on growing domains using MATLAB

Lecture 15: Simulating Tissue Morphogenesis

- Prescribed Growth: Function-based or Image-based Modelling
- Fluid-based Growth Models: Navier-Stokes Description
- Cell-based Models: Viscoelastic Cell Models, Cellular Potts Models, Agent-based Simulations

Practical 12: Image Segmentation & Calculation of Displacement Fields using MATLAB

- Teachers & Coordinators -

Fernando Amat, PhD
Janelia Farm Research Farm
19700 Helix Drive
Ashburn, VA 20147
Phone: +1-571-209-4000
Email: famat81@gmail.com
Website: http://www.fernandoamat.com/

Rut Carbonell Garcia, coordinator
Systems Analysis of Development
EMBL/CRG Research Unit in Systems Biology
CRG - Centre de Regulació Genòmica
Dr. Aiguader 88,
08003 Barcelona, Spain
Phone: +34 93 316 00 98
Email: rut.carbonell@crg.eu

Kai Dierkes, PhD
Biomechanics of Morphogenesis
CRG - Centre de Regulació Genòmica
Dr. Aiguader 88,
08003 Barcelona, Spain
Phone: +34 93 316 03 65
Email: kai.dierkes@crg.eu
Website: http://www.crg.eu/jerome_solon

Johannes Jaeger, PhD, scientific coordinator
Comparative Analysis of Developmental Systems
EMBL/CRG Research Unit in Systems Biology
CRG - Centre de Regulació Genòmica
Dr. Aiguader 88
08003 Barcelona, Spain
Phone: +34 93 316 02 85
Email: yogi.jaeger@crg.eu
Website: http://www.crg.es/johannes_jaeger

Hilde Janssens, PhD, coordinator
Comparative Analysis of Developmental Systems
EMBL/CRG Research Unit in Systems Biology
CRG - Centre de Regulació Genòmica
Dr. Aiguader 88
08003 Barcelona, Spain
Phone: +34 93 316 02 90
Email: hilde.janssens@crg.eu
Website: http://www.crg.es/johannes_jaeger

Matthieu Louis, PhD, scientific coordinator
Sensory Systems and Behaviour
EMBL/CRG Research Unit in Systems Biology
CRG - Centre de Regulació Genòmica
Dr. Aiguader 88
08003 Barcelona, Spain
Phone: +34 93 316 02 18
Email: matthieu.louis@crg.eu
Website: http://www.crg.es/matthieu_louis

Andreea Munteanu, PhD
Systems Analysis of Development
EMBL/CRG Research Unit in Systems Biology
CRG - Centre de Regulació Genòmica
Dr. Aiguader 88,
08003 Barcelona, Spain
Phone: +34 93 316 02 46
Email: andreea.munteanu@crg.eu
Website: http://www.crg.eu/james_sharpe

Theodore J. Perkins, PhD
- Scientist, Sprott Center for Stem Cell Research
Regenerative Medicine Program
Ottawa Hospital Research Institute
- Assistant Professor, Department of Biochemistry,
Microbiology and Immunology
University of Ottawa
- Cross-Appointed
School of Information Technology and Engineering
University of Ottawa
Phone: +1 613 737 8899 ext. 79795
Email: tperkins@ohri.ca
Website: http://www.perkinslab.ca

James Sharpe, PhD, scientific coordinator
Systems Analysis of Development
EMBL/CRG Research Unit in Systems Biology
CRG - Centre de Regulació Genòmica
Dr. Aiguader 88,
08003 Barcelona, Spain
Phone: +34 93 316 00 98
Email: james.sharpe@crg.eu
Website: http://www.crg.eu/james_sharpe

Mukund Thattai, PhD
Computational Cell Biology
National Centre for Biological Sciences
Tata Institute of Fundamental Research
GKVK, Bellary Road,
Bangalore 560065, India
Phone: 91 80 23666001
Email: thattai@ncbs.res.in
Website: http://www.ncbs.res.in/thattai

Marie Jeanne Trussart, PhD
Design of Biological Systems
EMBL/CRG Research Unit in Systems Biology
CRG - Centre de Regulació Genòmica
Dr. Aiguader 88,
08003 Barcelona, Spain
Phone: +34 93 316 01 86
Email: marie.trussart@crg.eu
Website: http://www.crg.eu/luis_serrano