

Track Introduction

Data Science and Scientific Computing

presented by Christoph Lampert (TrackRep)



Institute of Science and Technology

Today's Schedule

10:00-10:30	Christoph Lampert	Overview of the DSSC Track
10:30-10:45	Gasper Tkacik	Theoretical Biophysics and Neuroscience
10:45-11:00	Bingqing Cheng	Computational Materials Science
11:00-11:15	Marco Mondelli	Machine Learning at IST Austria
11:15-11:30	Matthew Robinson	Medical Genomics
11:30-11:45	Caroline Muller	Atmosphere and Ocean Dynamics
11:45-12:00	all of the above	overflow buffer and Q&A

Data Science and Scientific Computing (DSSC)

Interdisciplinary Track in the Graduate School, combining aspects of:

- data analysis
- information processing
- modelling
- numerical simulation

**Beatriz
Vicoso**



SEX-CHROMOSOME
BIOLOGY AND
EVOLUTION

**Nick
Barton**



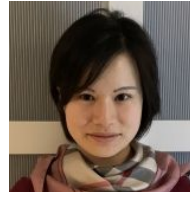
EVOLUTIONARY
GENETICS

**Bernd
Bickel**



COMPUTER
GRAPHICS AND
DIGITAL FABRICATION

**Bingqing
Cheng**



COMPUTATIONAL
MATERIALS SCIENCE

**Edouard
Hannezo**



PHYSICAL
PRINCIPLES IN
BIOLOGY

**Marco
Mondelli**



INFORMATION-
THEORETIC VIEW OF
DATA SCIENCE

**Carl
Goodrich**



COMPUTATIONAL
SOFT-MATTER
PHYSICS

**Caroline
Muller**



ATMOSPHERE AND
OCEAN DYNAMICS

**Christoph
Lampert**



MACHINE LEARNING
AND COMPUTER
VISION

**Matt
Robinson**



MEDICAL GENOMICS

**Chris
Wojtan**



COMPUTER
GRAPHICS AND
PHYSICS SIMULATION

**Gasper
Tkacik**



THEORETICAL
BIOPHYSICS AND
NEUROSCIENCE

**Sandra
Siegert**



NEURO-
IMMUNOLOGY

**future
professors**



**Beatriz
Vicoso**



SEX-CHROMOSOME
BIOLOGY AND
EVOLUTION

**Nick
Barton**



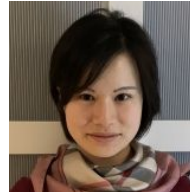
EVOLUTIONARY
GENETICS

**Bernd
Bickel**



COMPUTER
GRAPHICS AND
DIGITAL FABRICATION

**Bingqing
Cheng**



COMPUTATIONAL
MATERIALS SCIENCE

**Edouard
Hannezo**



**Deputy
TrackRep**
PRINCIPLES IN
BIOLOGY

**Marco
Mondelli**



INFORMATION-
THEORETIC VIEW OF
DATA SCIENCE

**Carl
Goodrich**



COMPUTATIONAL
SOFT-MATTER
PHYSICS

**Caroline
Muller**



ATMOSPHERE AND
OCEAN DYNAMICS

**Christoph
Lampert**



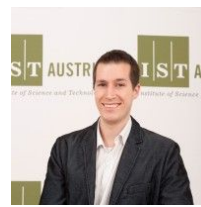
TrackRep
LEARNING
AND COMPUTER
VISION

**Matt
Robinson**



MEDICAL GENOMICS

**Chris
Wojtan**



COMPUTER
GRAPHICS AND
PHYSICS SIMULATION

**Gasper
Tkacik**



THEORETICAL
BIOPHYSICS AND
NEUROSCIENCE

**Sandra
Siegert**



NEURO-
IMMUNOLOGY

**future
professors**



**Beatriz
Vicoso**



SEX-CHROMOSOME
BIOLOGY AND
EVOLUTION

**Nick
Barton**



EVOLUTIONARY
GENETICS

**Bernd
Bickel**



COMPUTER
GRAPHICS AND
DIGITAL FABRICATION

**Bingqing
Cheng**



COMPUTATIONAL
MATERIALS SCIENCE

**Edouard
Hannezo**



PHYSICAL
PRINCIPLES IN
BIOLOGY

**Marco
Mondelli**



INFORMATION-
THEORETIC VIEW OF
DATA SCIENCE

**Carl
Goodrich**



COMPUTATIONAL
SOFT-MATTER
PHYSICS

**Caroline
Muller**



ATMOSPHERE AND
OCEAN DYNAMICS

**Christoph
Lampert**



MACHINE LEARNING
AND COMPUTER
VISION

**Matt
Robinson**



MEDICAL GENOMICS

**Chris
Wojtan**



COMPUTER
GRAPHICS AND
PHYSICS SIMULATION

**Gasper
Tkacik**



THEORETICAL
BIOPHYSICS AND
NEUROSCIENCE

**Sandra
Siegert**



NEURO-
IMMUNOLOGY

**future
professors**





Beatriz Vicoso

Sex-Chromosome Biology and Evolution

How do sex chromosomes evolve?

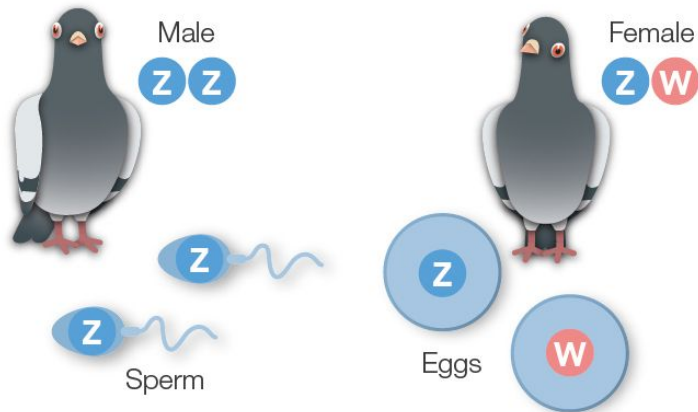


Image: Wikipedia



Nick Barton

Evolutionary Genetics

Hybrid zones:

- study selection, gene flow, random fluctuations

Genetics of complex traits:

- theory, experimental evolution, data analysis



Making sense of DNA sequence:

- how can we infer population history, and detect selection?

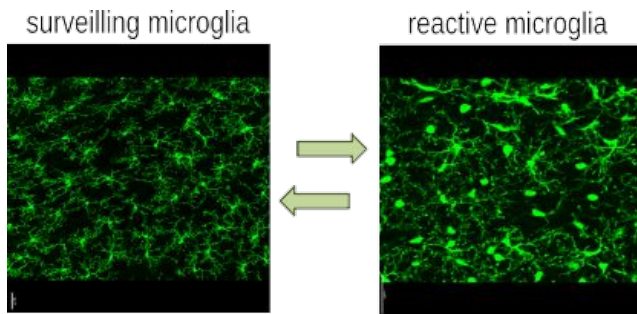




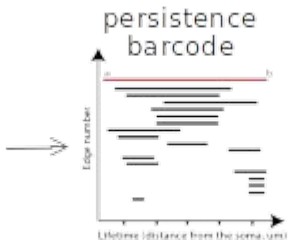
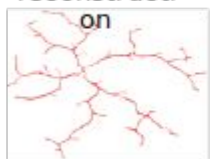
Sandra Siegert

Neuroimmunology in Health and Disease

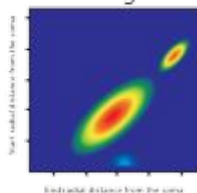
(Topological) data analysis to classify immune cells' morphology and function.



microglia 3D reconstruct



persistence image



Edouard Hannezo

Physical Principles in Biology

Example: how do cells “know” how to make the right decisions?



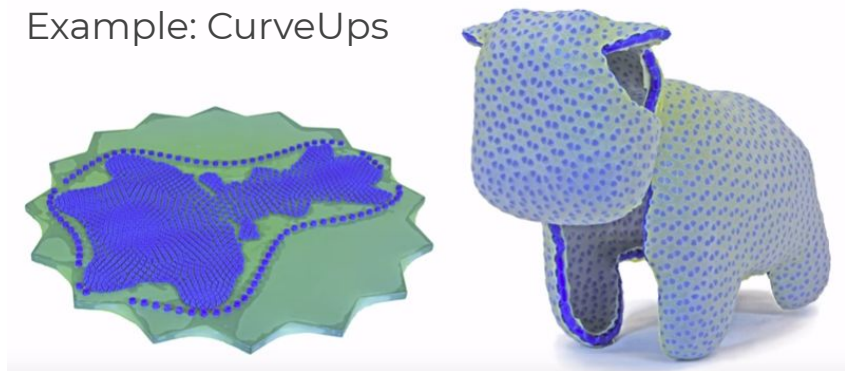


Bernd Bickel

Computer Graphics and Digital Fabrication

Methods for modeling, simulating and optimizing (printable) 3D objects

Example: CurveUps

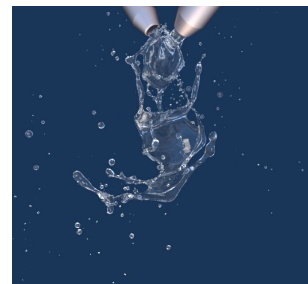


Chris Wojtan

Physics Simulation and Geometry Processing



Numerical algorithms for solving differential equations



Efficient and robust methods for animating physics



Create tools for manipulating shapes

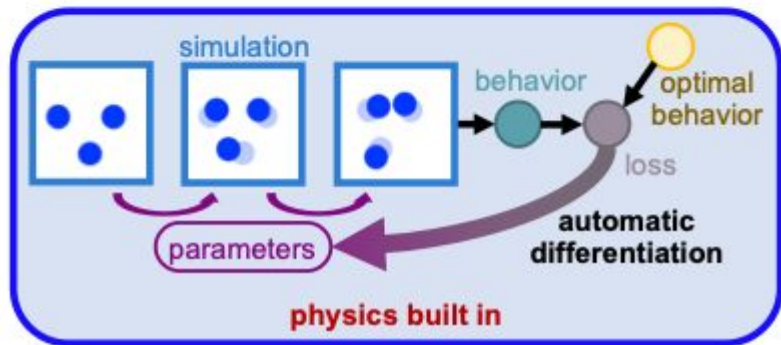


Carl Goodrich

Theoretical and Computational Soft Matter

Discovering basic soft matter principles using computational and theoretical tools, such as artificial neural networks

B Differentiable Statistical Physics Calculations

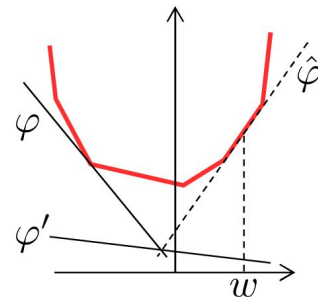


Christoph Lampert

Machine Learning and Computer Vision

Statistical machine learning:

- transfer learning,
- continual learning,
- trustworthy learning,
- theory of deep learning

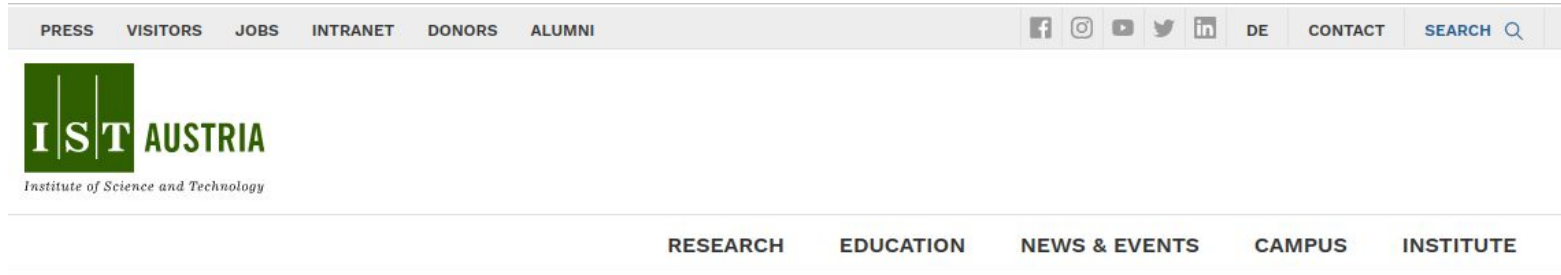


Applications in Computer Vision:

- scene understanding
- generative models of dynamic scenes

future professor(s)

... keep checking the IST Austria homepage ...



Five new professors join IST Austria

IST Austria President Thomas Henzinger presents successful young researchers | Appointments in neuroscience, physics, mathematics, and computer science



DSSC Courses 2021/22

Applications of Stochastic Processes

(Nick Barton) - Fall II

The course will cover basic stochastic processes, emphasizing examples from a range of fields. This will include Markov chains, branching processes, and the diffusion approximation.

Mathematical rigour will be avoided.

Applied Algorithms and Datastructures

(Tobias Meggendorfer) -- Fall I+II

This course aims to teach the concepts of efficient algorithms through a practical, hands-on format. Each week treats a particular class of problems, beginning with basic data structures and simple algorithms, and continuing with more advanced topics, e.g. shortest path or dynamic programming.

Computational Physics (Chris Wojtan) -- Fall I

This course surveys some moderate/advanced topics for solving problems in computational physics and computer animation. The course will be structured as a seminar, with students primarily presenting material and discussing the details of various approaches.

Information Theory (for Data Science)

(Marco Mondelli) -- Fall I

The goal of the course is to present fundamental concepts in Information Theory and describe their relevance to emerging problems in Data Science and Machine Learning.

DSSC Courses 2021/22

Introduction to Python Programming for Data Science (Eder Miguel Villalba) -- Fall I

The goal of the course is to provide a deeper understanding of programming fundamentals in Python, develop programming skills through small programming projects and learn how to use some of the most common Python libraries for scientific computing and data analysis.

Methods of Data Analysis (Gasper Tkacik) -- Fall II

This course introduces five topics in data analysis and simulation methods. The focus is on sampling and inferring probabilistic models.

Statistical Machine Learning (Christoph Lampert) -- Fall I (inverted classroom format)

Introduction to modern statistical machine learning, in particular probabilistic models. The focus is on concepts, not applications.

Computational Bayesian Statistics (Matthew Robinson) -- Spring II

This course aims to start from the basic fundamentals of Bayesian inference and slowly evolves to cover more recent advances. The focus is on methods and algorithms closer to the application level, but with the hope of providing a solid theoretical foundation.

Data Science and Scientific Computing Track Core Course

Track Core Course -- Spring I+II

- introduction to data analysis / predictive models
- introduction to numerical simulation / optimization
- individual projects that combine both aspects



Prerequisites

- programming skills (preferably Python)
- strong mathematical skills (linear algebra, calculus)
- good understanding of statistics / probabilities



Introduction to Data Analysis / Predictive Models

data
 input: $\mathcal{D} = \{\vec{x}_1, \dots, \vec{x}_T\}$
 $\vec{x} \in \mathbb{R}^d$

K-MEANS CLUSTERING

Partition data into k disjoint sets $S_i, i=1, \dots, k$
 to minimize cluster distortion

$$D = \sum_{i=1}^k \sum_{\vec{x} \in S_i} \|\vec{x}_t - \vec{\mu}_i\|^2$$

where $\vec{\mu}_i = \frac{1}{|S_i|} \sum_{\vec{x} \in S_i} \vec{x}$

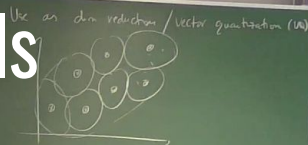
Use as clustering tool

$$P(\vec{x}) = \sum_i w_i P_i(\|\vec{x} - \vec{\mu}_i\|)$$

Mixture of Gaussians

$$P(\vec{x}) = \sum_{i=1}^k w_i \frac{1}{\sigma_i^d} \mathcal{N}(\vec{x}; \vec{\mu}_i, \sigma_i^2)$$

d dimensional and dd



output: \circ cluster assignment

$$\vec{x}_t \rightarrow i, k$$

\circ center of mass, "centroids"
 $\vec{\mu}_i$



SPARSE CODING / ICA

$\mathcal{D} = \{\vec{x}_1, \dots, \vec{x}_T\}$

WHITENING

mean $\langle \vec{x} \rangle_T = \frac{1}{T} \sum_{t=1}^T \vec{x}_t$

Covariance

$$C = \langle \vec{x} \vec{x}^T \rangle_T$$

$$C = V D V^T$$

\parallel diagonal
 orthogonal
 $V V^T = I$

$$\vec{x}_t \leftarrow \vec{x}_t - \langle \vec{x} \rangle_T$$

$$\vec{x}_t \leftarrow V D^{-1/2} V^T \vec{x}_t$$

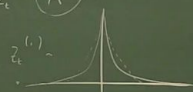
after whitening: $\langle \vec{x} \vec{x}^T \rangle = I$

$$z_i = \vec{w}_i^T \vec{x}_t \quad P_f(z)$$

$\frac{d_{max}}{\sigma}$

$\vec{x}_1, \dots, \vec{x}_T$

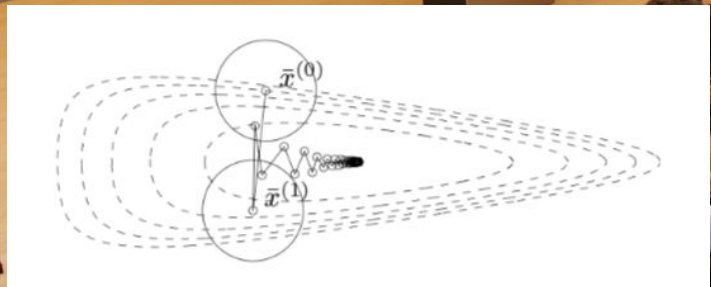
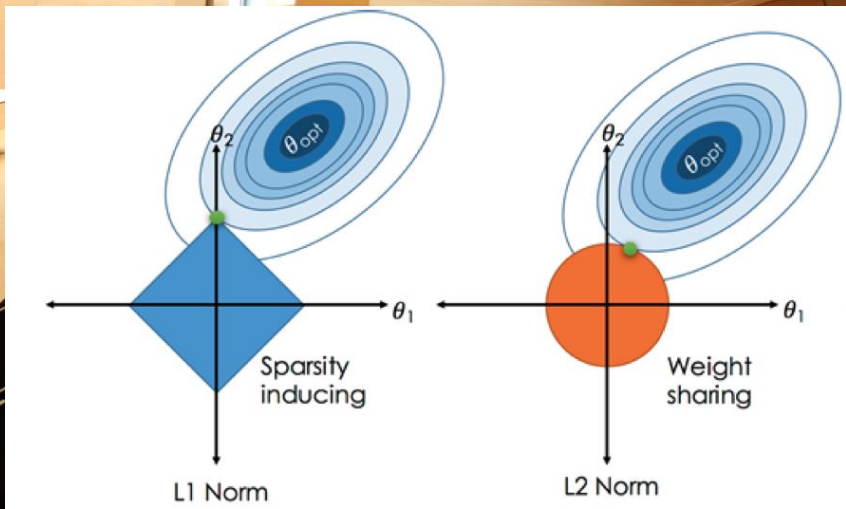
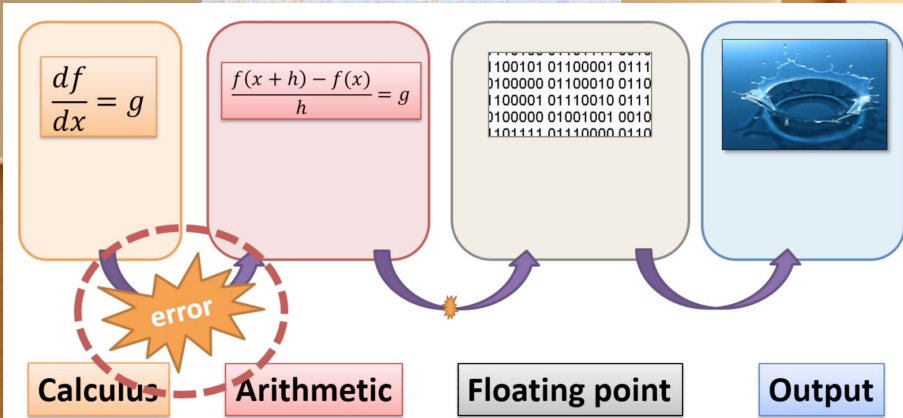
$$\vec{x}_t = \sum_i z_i^{(t)} \vec{A}^{(i)}$$



After whitening, how do we find the "sparse" non-Gaussian source?




Introduction to Numerics and Optimization



Individual Projects

- simulating neuron firing together and developing patterns
- simulating molecular dynamics of chiral proteins to learn group behaviors
- simulating/animating ant colonies
- N-body simulator to study evaporation of star clusters
- study of pattern formation in reaction-diffusion equations
- simulating a "turbidostat" - a lab tool for growing bacteria and studying mutations
- study of data set compression for machine learning models
- game theory simulation to find stable population behaviors
- study of stable/unstable balances between predator/prey interactions on a graph



QUESTIONS

ANSWERS

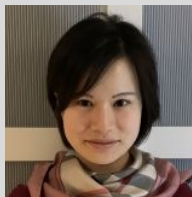
Following: Individual Presentations

**Gaspar
Tkacik**



Theoretical
Biophysics and
Neuroscience

**Bingqing
Cheng**



Computational
Materials Science

**Marco
Mondelli**



Machine Learning
at IST Austria

**Matthew
Robinson**



Medical
Genomics

**Caroline
Muller**



Atmosphere and
Ocean Dynamics