SCALABLE KERNEL METHODS

in machine learning

GEORGE BIROS
padas.ices.utexas.edu

Outline

- Discuss a very small set of machine learning methods
- Introduce kernel matrices and kernel methods
- Discuss kernel approximation (for scalability)
  - Nystrom/Nbody
- Apply to image analysis problems

Logistics

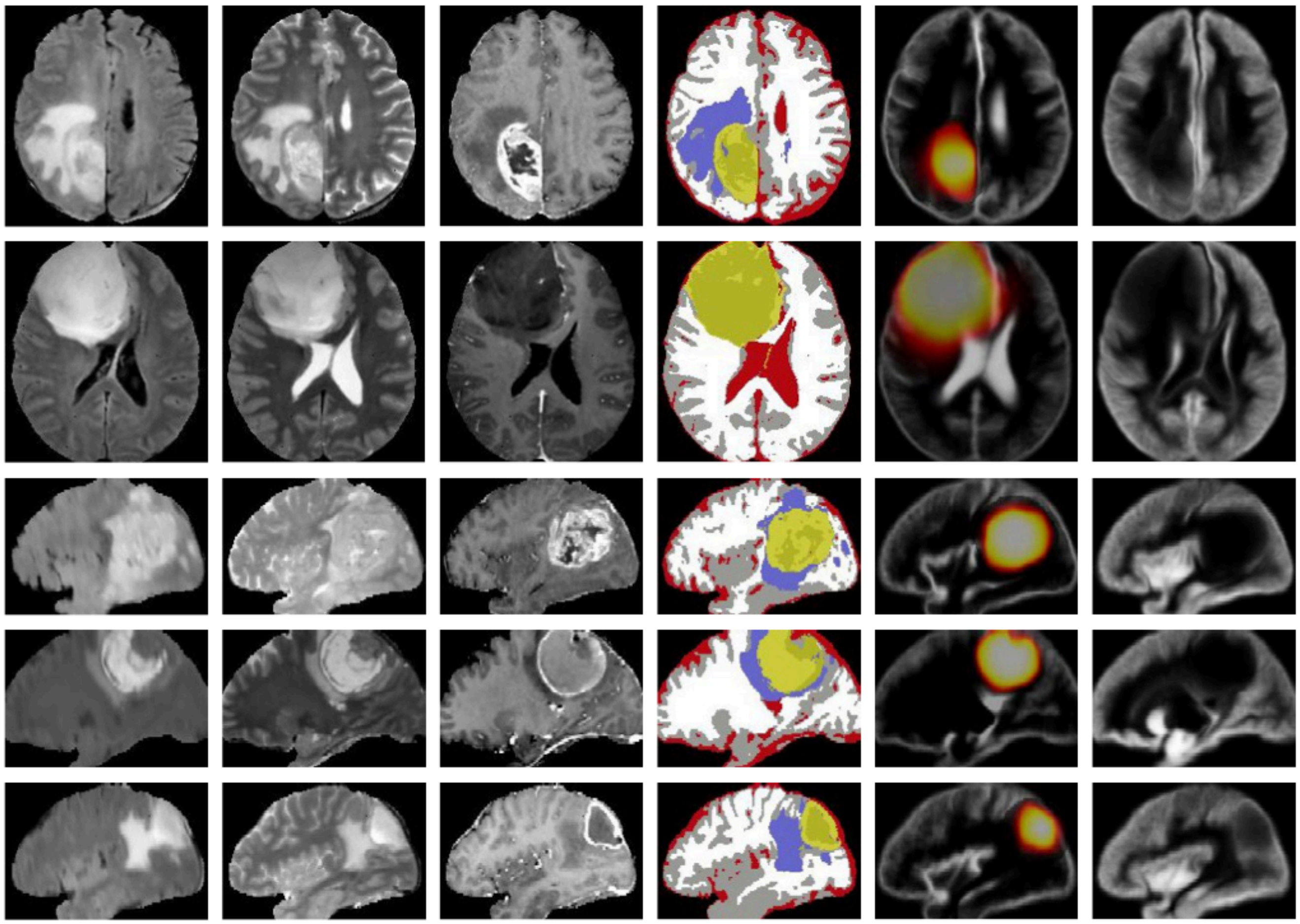
- Possible project at the end
- MATLAB, bring your laptops for practice
- https://piazza.com/tum.de/summer2016/informaticsvscientificcomputing
- Office IAS 2.006
- gbiros@gmail.com

Google's Artificial Neural Network of 16,000 CPUs Learns to Detect Cats

Posted on June 26, 2012

Digit recognition

Medical image analysis



Other examples

- Given DNA sequences
  - find genes
- Given web browsing history
  - recommend ads
- Given movie history
  - recommend new movies
- Given scientific articles
  - group them according to topics
- Given time series of pressure
  - recognize speech (words spoken)
- Given high-energy physics particle data
  - classify it as a Higgs boson
- Given MRI,CT, blood work, history profile, symptoms
  - diagnose

Notation

TRAINING SET

$$\underline{X} = \{x_i\}_{i=1}^{N}, \quad x_i \in \mathbb{R}^d$$

$$\underline{Y} = \{y_i\}_{i=1}^{N}; \quad y_i \in \mathbb{R}$$

N : NUMBER OF POINTS

d : DIMENSION

LEARNING

Given point x

parameters (TRAINING)

$$y = f(x; w)$$

LEARNING ALGORITHM

PREDICTION

How to learn? Optimization

Solve a minimization problem
propose a loss function

$$\min_{w} \mathcal{L}[\underline{Y}, \underline{X}, w]$$

Function approximation

- Given **X** and **Y** construct f() that approximates the function
- Assumes no noise
- Does not have a unique solution
- 200 years to figure everything out

Regression

- Term used in statistics
- Unclear what is the input and output
- Even if we know **X** and **Y** are noisy
- In these lectures approximation and regression are the same

Supervised classification

- **Y** is a *categorical* variable
- **Binary classification** {-1,1}
- Multiclass {0,1,2,…}
- Trick for binary: sign(f)
- Trick for multiclass : one vs other

Unsupervised classification

- We are only give **X**
- We want to group into **k** classes
  - sometimes **k** is not known
- The algorithms for this problem don't fit the function approximation.

Digit recognition

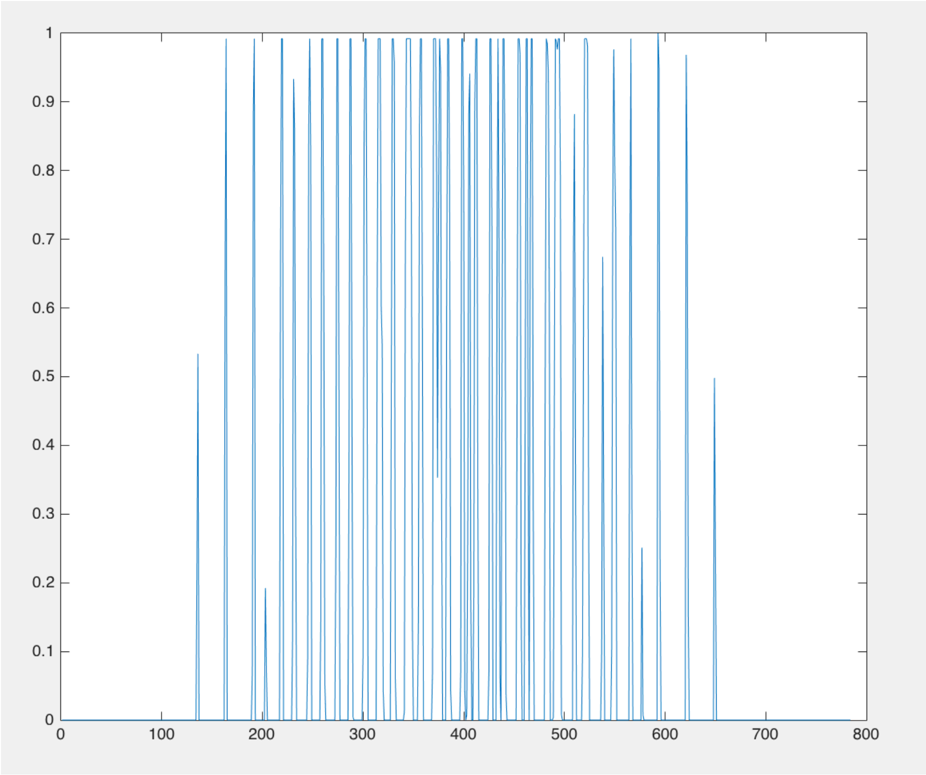
MNIST database; normalized digits from NIST institute in the US
http://yann.lecun.com/exdb/mnist/
http://www.cs.nyu.edu/~roweis/data.html
matlab format

N= 60000
D = 784

Goal given training test, construct an algorithm that classifies correctly unseen input

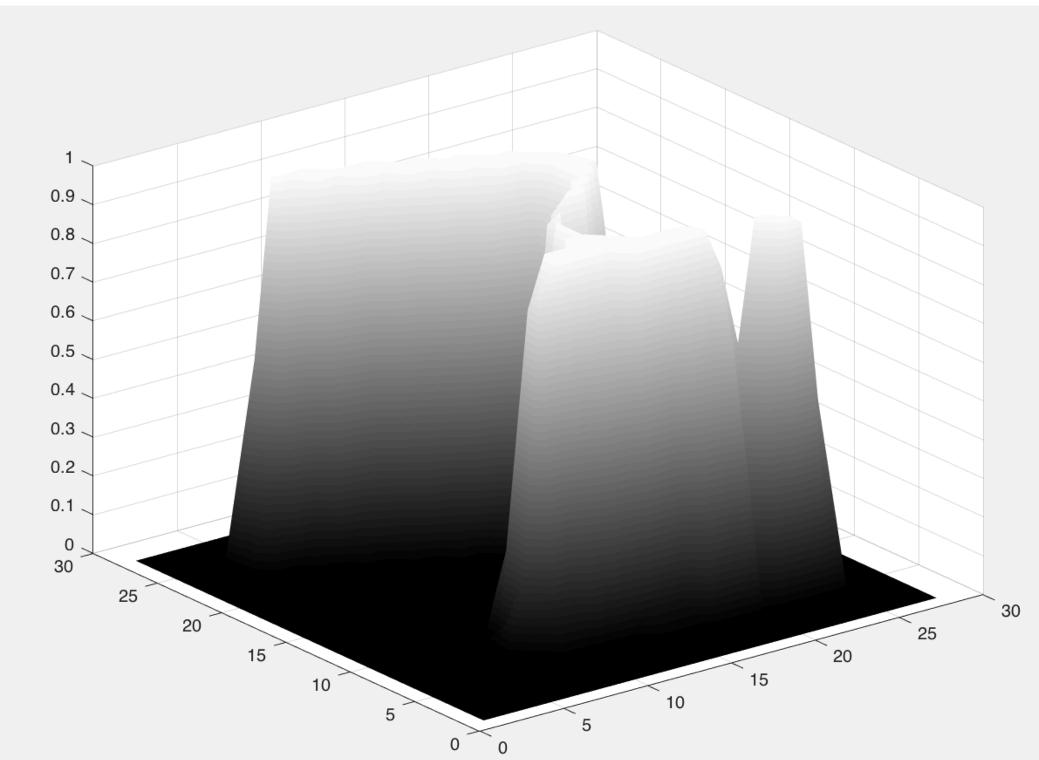
Digit recognition (mnist)

plot(x)

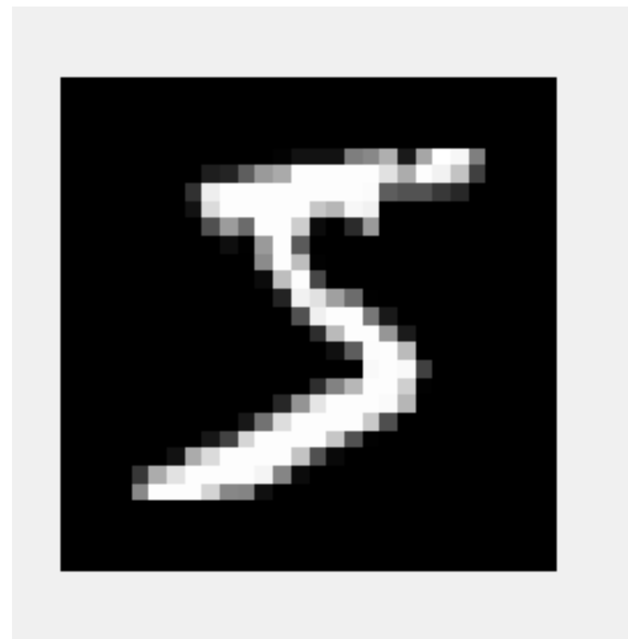


Digit recognition (mnist 60K)

surf(reshape(x,[28,28]))



imshow(reshape(x,[28,28])



- BREAK

Austin, Texas

Location in the United States of America

Coordinates: 30°15′0″N 97°45′0″W

Keep Austin Weird

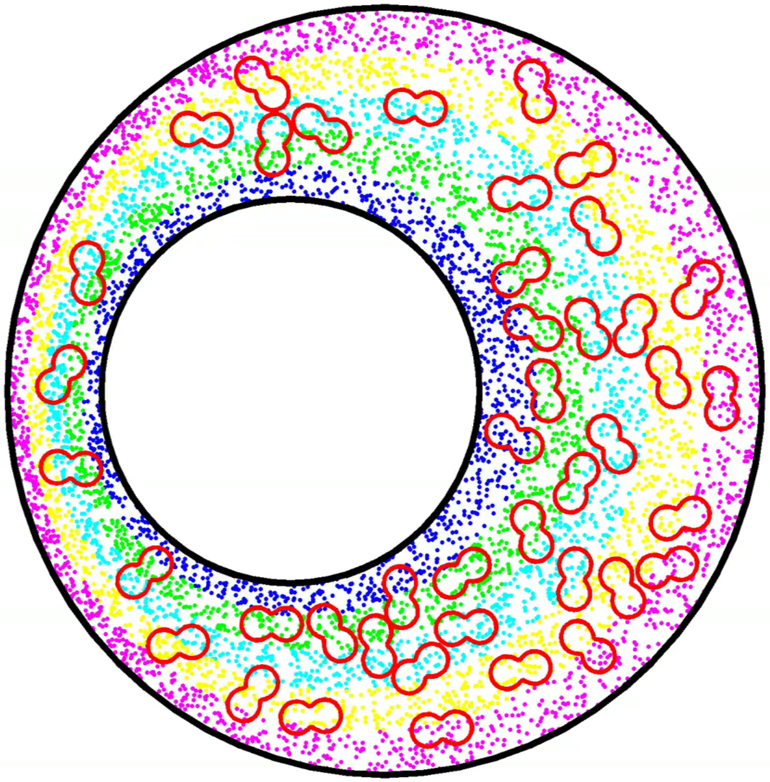
PADAS Lab: padas.ices.utexas.edu
Parallel Algorithms for Data Analysis and Simulation
PI: George Biros (ICES, CS, Mechanical Engineering)

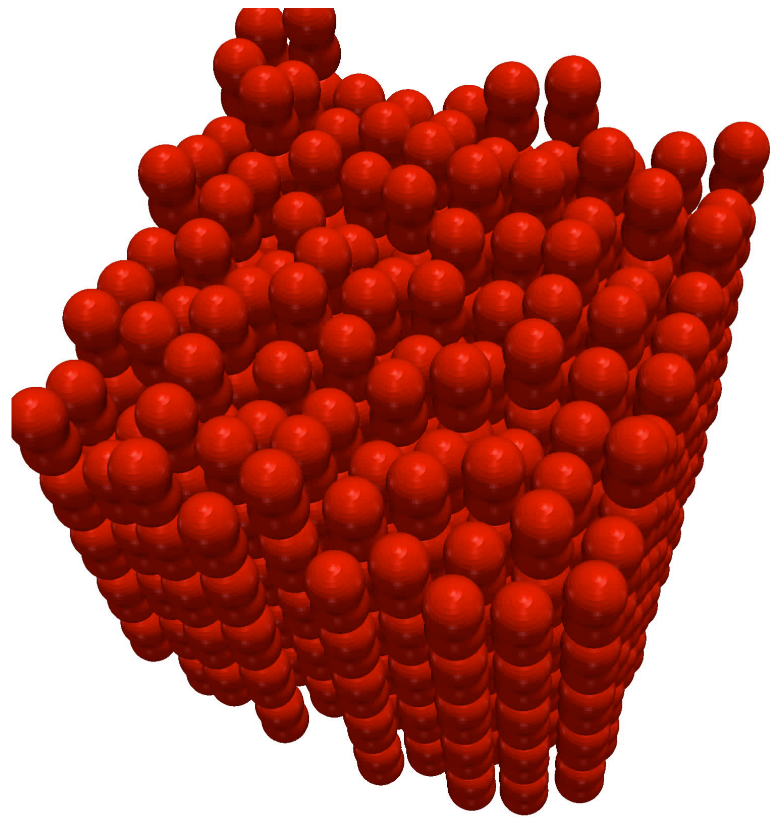
- **MISSION:** High-performance algorithms and software for core problems in numerical analysis, machine learning, and simulation; scalability on millions of heterogeneous cores

- **ALGORITHMS:** N-body algorithms, high-order methods, fast PDE solvers, fast approximation, inverse problems, regression, classification, model reduction, uncertainty quantification

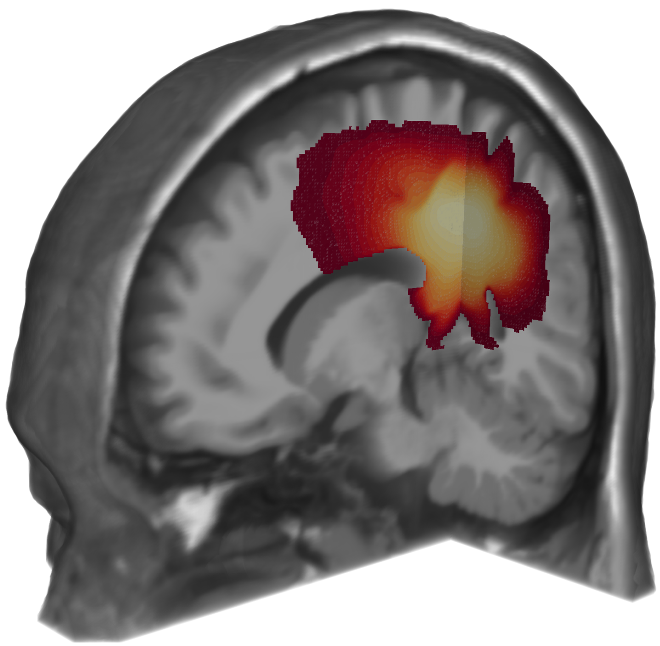
- **APPLICATIONS:** biomechanics, fluid mechanics (microfluidics, nanofluidics, complex fluids), image analysis (registration, segmentation, classification), inverse scattering, convection-reaction-diffusion systems, rare event detection, data assimilation

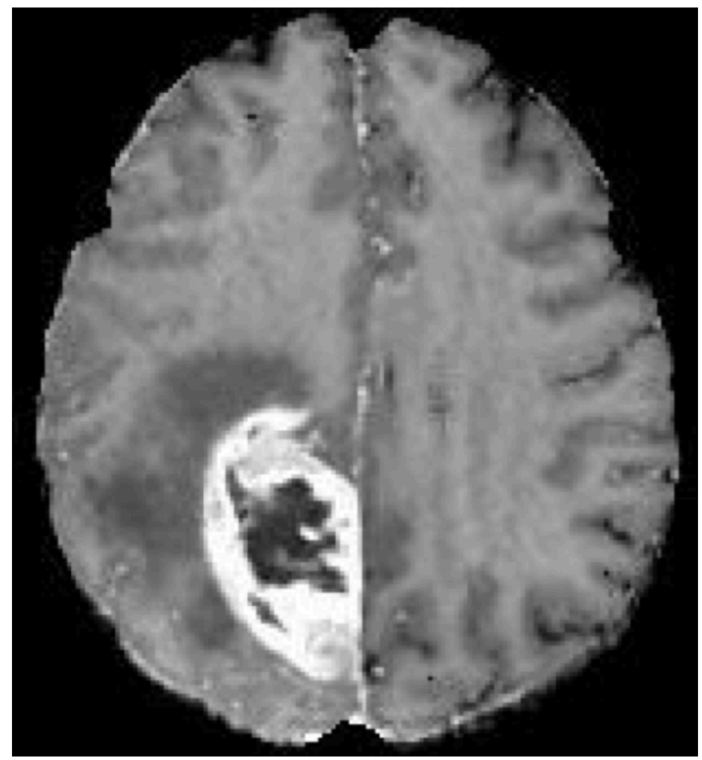
- **SOFTWARE:** padas.ices.utexas.edu/software



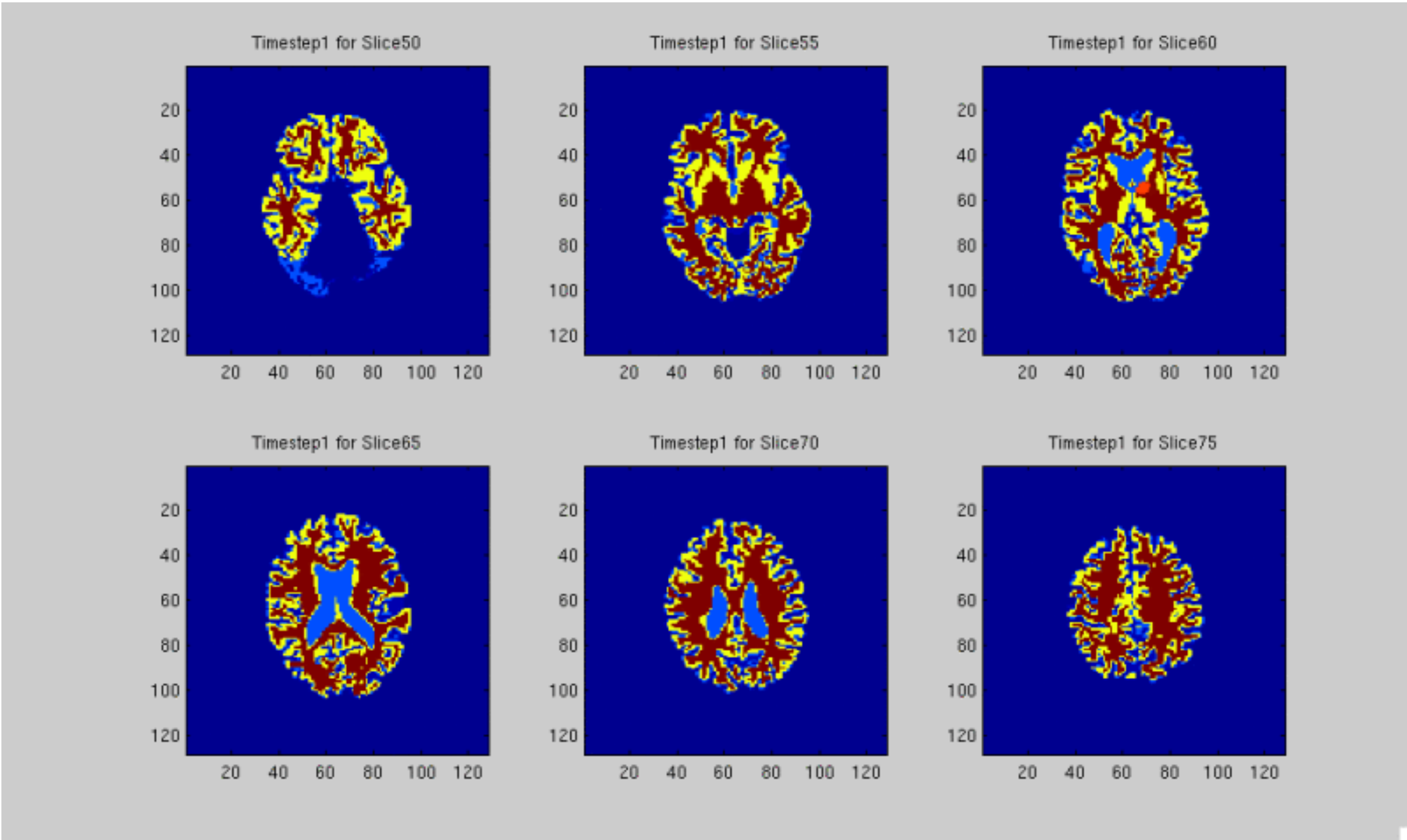


Biophysical models and image analysis

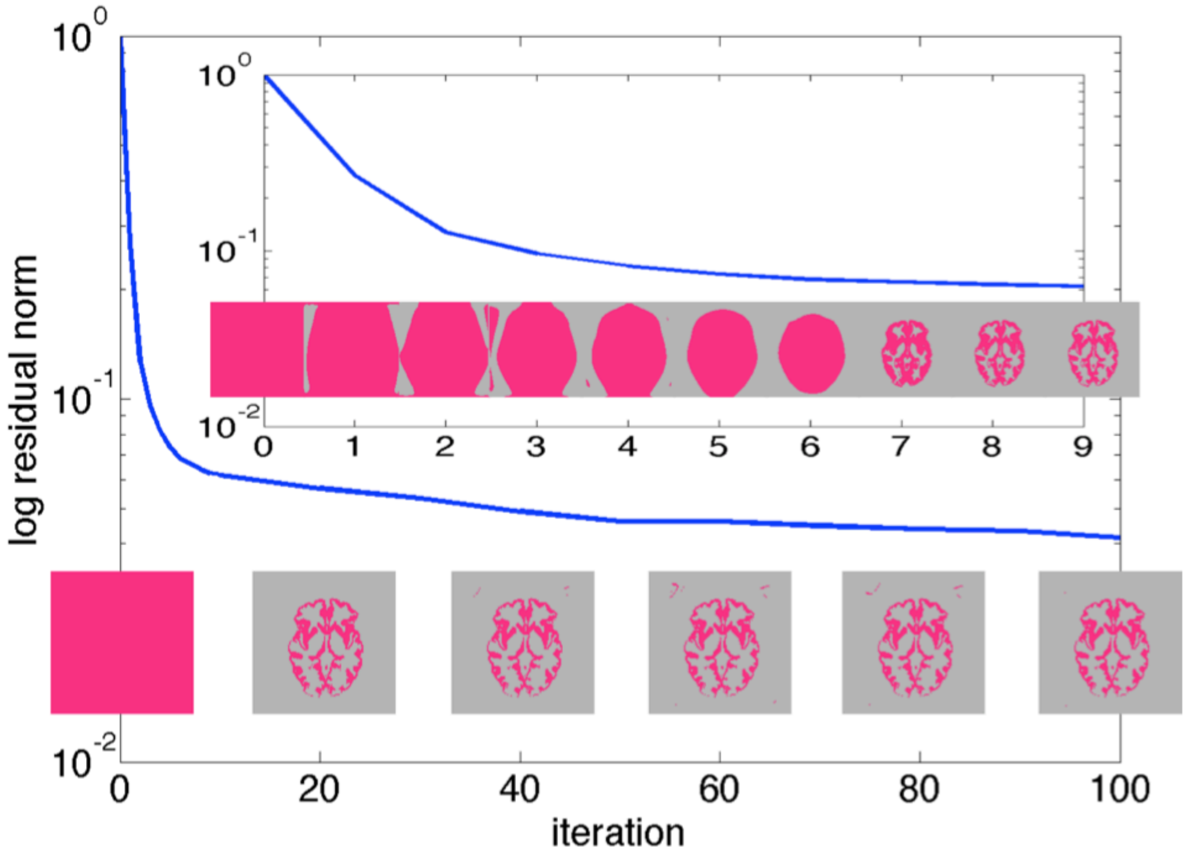




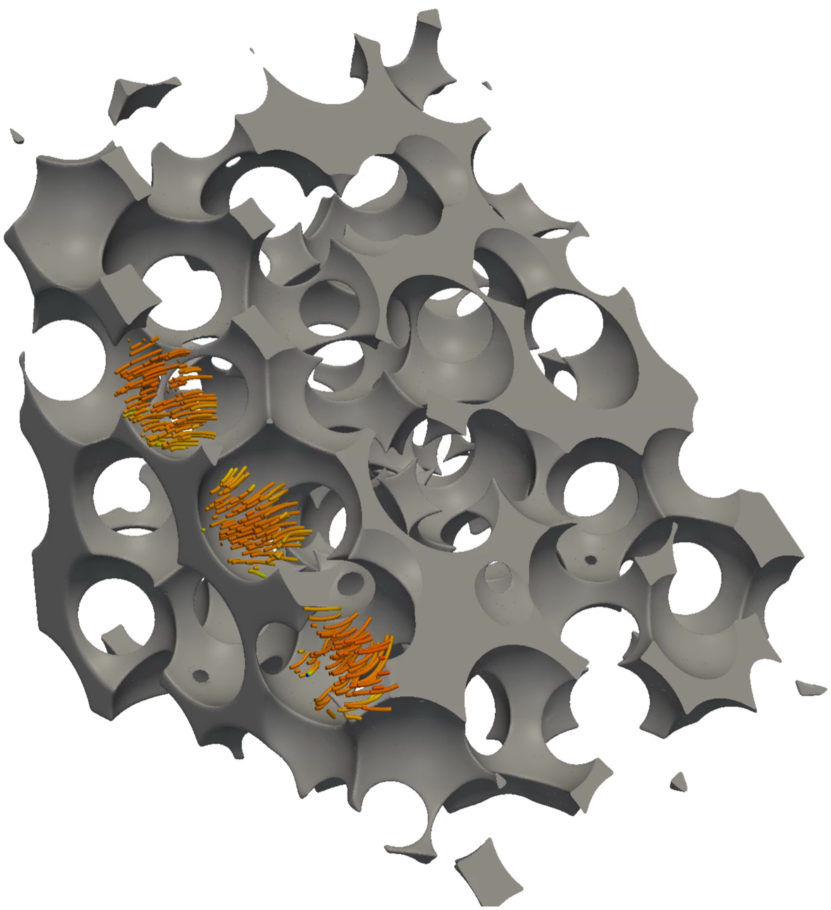
Tumor simulations

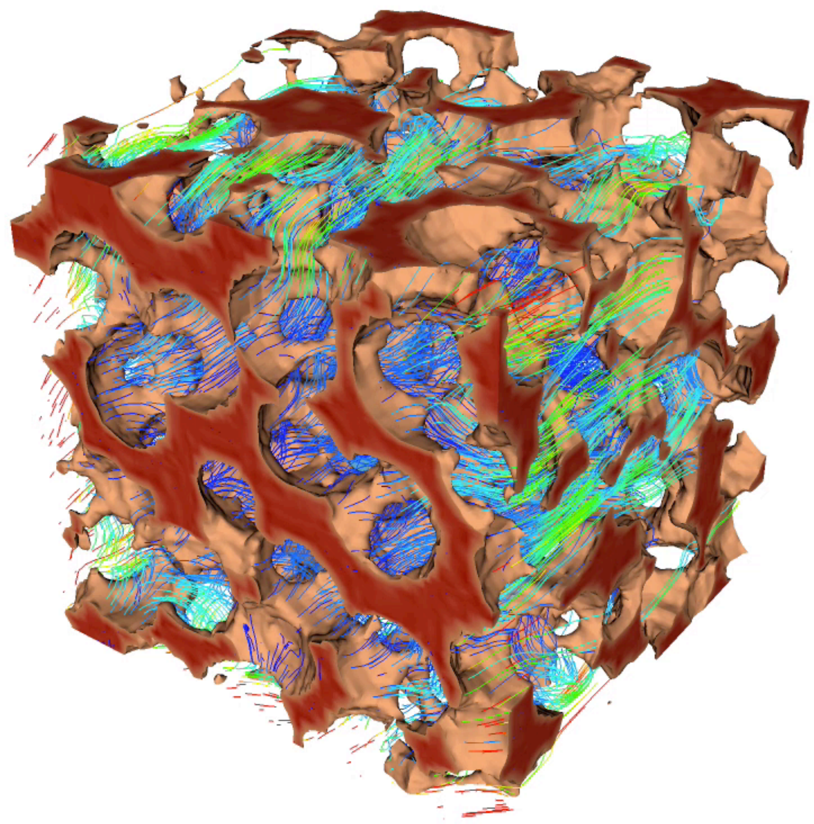


Convergence



Multiphysics

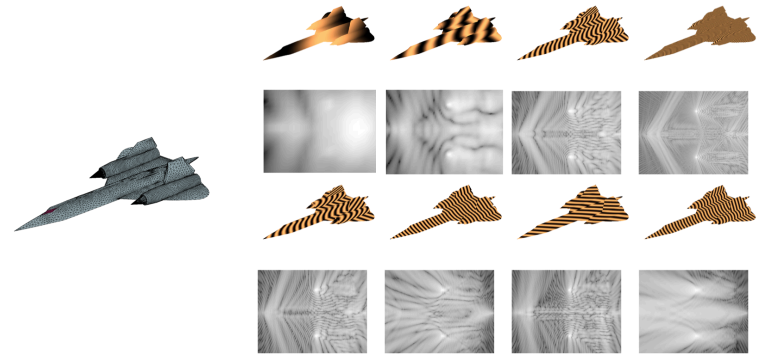




FaIMS: $\mathcal{O}(N_\omega N)$

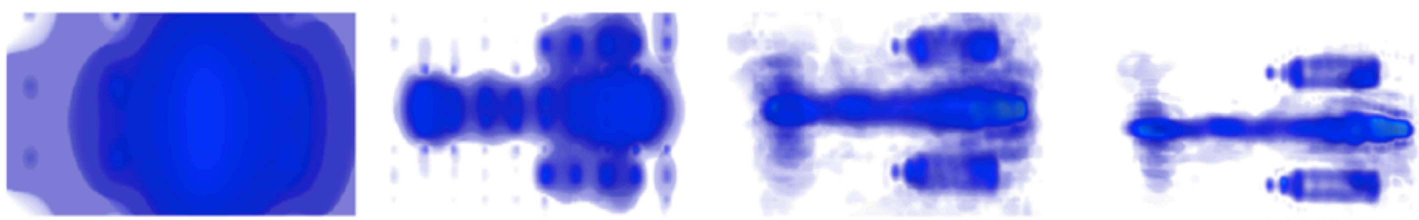
SVD: $\mathcal{O}(N^2 N_s N_d N_\omega)$

Lanczos: $\mathcal{O}(N_i N N_s N_\omega)$

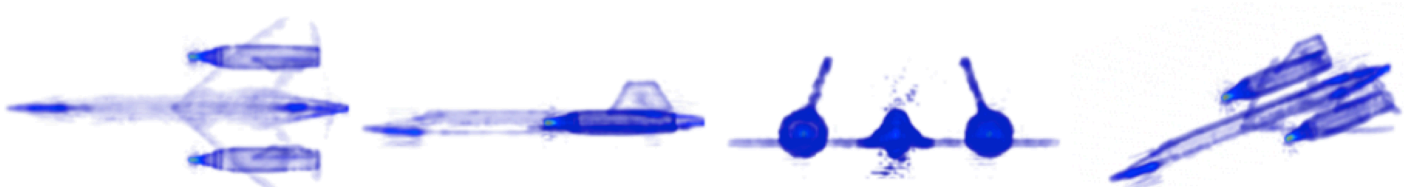


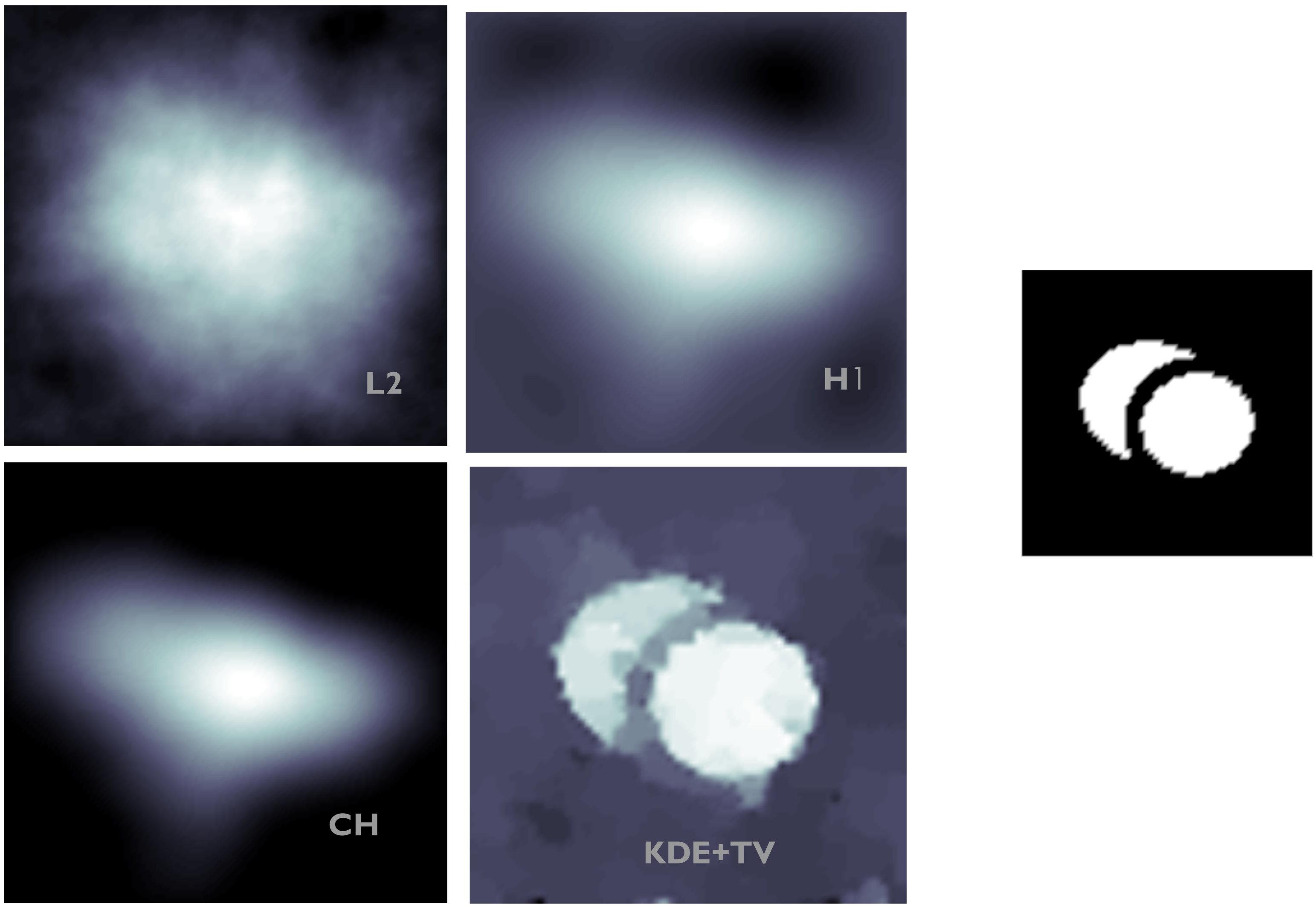
Low frequency, small number of events

0.5L/1 event | 1L/8 events | 4L/8 events | 4L/32 events



8L/512 events




L2
H1
CH
KDE+TV