

Tim Laux: The large-data limit of the MBO scheme for data clustering

mardi 20 juin 2023 16:20 (50 minutes)

The MBO scheme is an efficient algorithm for data clustering, the task of partitioning a given dataset into several meaningful clusters. In this talk, I will present the first rigorous analysis of this scheme in the large-data limit.

The starting point for the first part of the talk is that each iteration of the MBO scheme corresponds to one step of implicit gradient descent for the thresholding energy on the similarity graph of the dataset. It is then natural to think that outcomes of the MBO scheme are (local) minimizers of this energy. We prove that the algorithm is consistent, in the sense that these (local) minimizers converge to (local) minimizers of a suitably weighted optimal partition problem.

To study the dynamics of the scheme, we use the theory of viscosity solutions. The main ingredients are (i) a new abstract convergence result based on quantitative estimates for heat operators and (ii) the derivation of these estimates in the setting of random geometric graphs.

To implement the scheme in practice, two important parameters are the number of eigenvalues for computing the heat operator and the step size of the scheme. Our results give a theoretical justification for the choice of these parameters in relation to sample size and interaction width.

This is joint work with Jona Lelmi (U Bonn).