On (non)parametric estimation of low and high-dimensional data

Low and high-dimensional data is increasingly available in data-driven applications, and the question of studying the geometrical aspects of such data remains a primary objective. In this talk, I will cover three lines of research: (i) low-dimensional parametric approximation; (ii) high-dimensional nonparametric estimation; and (iii) explore the power of studying the data geometry.

For one-dimensional data, even after decades of research, there were classes of functions that a parametric approximation for them was unknown. In this talk, we will present a method that, under certain assumptions on the data structure, constructs a reflecto-multiscale function. Moreover, we rigorously show that the limit function of the reflecto-multiscale refinement process is Holder continuous, and has Holder continuity of the highest-order well-defined derivatives.

The problem is further enhanced in case the data lies in a high dimension and is contaminated with noise and outliers. In such cases, a common assumption is that high-dimensional data is an embedding of a low-dimensional manifold. Unfortunately, existing solutions for surface reconstruction become impractical in high dimensions due to the curse of dimensionality. Moreover, most of the existing approaches for high-dimensional data are challenged by the presence of noise. I will present a method designed for denoising and reconstructing a low-dimensional manifold in a high-dimensional space from scattered data. A theoretical analysis demonstrates that the proposed optimization solution converges to a quasi-uniform reconstruction of the manifold within a bounded time. This nonparametric approach can then be extended to address various approximation tasks in high dimensions, such as function approximation or recovering missing information in the data. The presentation concludes with delving into the exploration of data geometry using manifold learning analysis. I will showcase that a unified line of research persists across various data-driven applications, highlighting the potential that lies in the (non) parametric estimation.