Development of statistical and algorithmic methods in mass spectrometry

One of the main applications of mass spectrometry (MS) is in the field of diagnostics in molecular medicine. The MS instruments produce spectra of input samples, such as for example of blood plasma, extracting data which may be used to identify chemical compounds present in the sample - and then, to identify deviations from norm, and to diagnose the patient.

One of the particularly difficult problem is the identification of proteins based on their spectra. The observation of the way a protein fragments inside of a tandem-MS instrument allows insights into its structure, as well as the discovery of so-called post-translational modifications. This is especially useful for diagnostics, as well as scientific endeavours, especially in the field of proteomics. Unfortunately, the computer identification of proteins based on their spectra remains a difficult endeavour, with existing algorithms frequently proving inadequate, forcing a manual analysis of the spectrum - which is a very time-consuming task.

The goal of the presented project is the improvement of the aforementioned situation, using (mostly ignored by previous approaches) data about isotopic structure of the spectra, combined with efficient algorithms for performing various computations on them.

It is worth mentioning that the applications of the statistical and algorithmic methods using the property of measure concentration, developed over the course of this project, will be useful in other fields of scientific endeavours, particularly when an efficient algorithm for sampling from any discrete distribution is needed, as well as in stochastic simulations.