The dynamic development of modern medicine, biology and chemistry is based on complex measurement techniques such as mass spectrometry and magnetic resonance spectroscopy. Both techniques can reveal the composition of complex substances such as drugs, body fluids, human tissues or synthetic polymers.

Although the physical phenomena used by these techniques are different, it turns out that the obtained measurement data can be analyzed using similar methods. The development of algorithms for the analysis of spectrometric data is extremely important as they are high-throughput techniques, which means that we are dealing with gigabytes of data for a single experiment. Hence the only way to analyze and interpret spectra is by means of automatic procedure.

As an example, consider mass spectrometryimaging, in which we examine the spatial composition of a tissue, i.e. each pixel contains information about thousands of metabolites and peptides in a given place. Another example is the nuclear magnetic resonance spectrum for a specific drug that contains several or a dozen substances (both active substances and excipients). Also one obtains very complex spectra for synthetic polymers because of chains of different lengths and different side groups present in the sample.

In these examples, the main task for the computer scientist is to identify the substances that are in the sample and estimate their proportions. In the first case, we do not know what molecules are in the tissue, so we have to use a whole range of potential substances. In the case of a drug, we know its composition, but the proportions of the substance can varied and spectral analysis should detect this, as well as possible contamination.

Let us assume that the spectrum of the analyzed substance is a bunch of peaks on the positive number axis. In the case of the mass spectrum, the position of the peak corresponds to the molecular mass of the substance, and the height (treated as the mass of the peak) to the signal intensity. Our algorithms for the comparison of two such spectra use the concept of optimal transport, which can be considered a generalization of the sorting problem. Optimal transport is a scenario that shifts peaks from one spectrum to obtain the other, minimizing the work done, i.e. the product of the mass to be moved and the distance to be covered.

Optimal transport allows the distance between spectra to be found. It turns out that such a distance has many good properties and it adequately compares the spectra. Using this distance, we can solve the problem of decomposition of a complex spectrum into components by means of linear regression. We intend to solve this problem for thousands of metabolites in spectrometric imaging, for complex drug spectra and for synthetic polymer spectra that are difficult to analyze due to overlapping signals.