

Prof. Dr. Andrej Tarlecki
Scientific Councils Office
University of Warsaw
Żwirki i Wigury 93

02-089 Warsaw

University Medical Center Mainz
Institute for Immunology
Proteomics Core Facility
Univ.- Prof. Dr. Stefan Tenzer
Geb. 308A Zi.3.102
Langenbeckstr. 1
55131 Mainz
Telefon:+49 (0) 6131 17-6199
Telefax: +49 (0) 6131 17-6202
E-Mail: tenzer@uni-mainz.de
www. immunologie-mainz.de

Mainz, 06.01.2023

Thesis evaluation report

“Algorithms for computational mass spectrometry based on the optimal transport theory”

By

Michał Aleksander Ciach

Mass spectrometry is one of the major analytical technologies for the analysis of biological samples. Recent generation instruments are capable of highly sensitive and high throughput analyses, enabling the analysis of complete proteomes even at single cell level. To facilitate the analysis of the ever-increasingly complex dataset, computational methods are of utmost importance. The recent improvement of both instruments and computational hardware allow researchers to carry out increasingly complex calculations on increasingly accurate data, resulting in the continuous development of computational mass spectrometry methods.

One of the challenges of the modern computational mass spectrometry is the task of separating of isotopic envelopes, i.e. deconvolution of mass spectra, which is also referred to as regression of mass spectra. The respective methods are tailored to explain experimentally measured mass spectra of mixtures of chemical by a linear combination of theoretically predicted spectra of the mixture's components.

In the present dissertation, Mr. Ciach has developed an approach to computational spectrometry and spectroscopy, which is theoretically based on the optimal transport of signal between spectra to determine their similarity. Toward this purpose, Mr. Ciach has harnessed a mathematical tool, the Wasserstein distance, and uses it to quantify the difference between two spectra as the minimal distance in the m/z axis needed to match their signals.

Based on the Wasserstein Distance, Mr. Ciach developed a regression-denoising algorithm for fitting a linear combination of reference spectra to an experimental spectrum of a mixture of chemical compounds and implemented it successfully as a Python 3 package called

masserstein.

The work presented is fascinating to me as a proteomics researcher in multiple aspects. One of its main conceptual contributions is the treatment of spectra with probabilistic measures to encompass both profile and centroid spectra in a single mathematical formalism, and to use the notion of optimal transport to compare different spectra to each other. This approach enabled Mr Ciach to develop a novel method which is fitting a linear combination of discrete theoretical spectra to a continuous experimental one.


Notably, the presented thesis is a highly successful example of an interdisciplinary collaboration between mathematicians, biologists, statisticians, chemists, and computer scientists. Throughout his work, Mr. Ciach has successfully demonstrated that he is able to identify and overcome practical difficulties encountered in experimental research with both mass spectrometry and NMR spectroscopy, to formalize these challenges as mathematical problems and solve them using optimization algorithms. In addition, Mr. Ciach has shown that his approaches can be successfully generalized to other types of data analysis and lead to significant improvements of the biological accuracy of the result, as exemplified by the demonstration of the practical applicability of his approach by improving the methods of segmentation of mass spectrometric images.

In the thesis, Mr. Ciach has thus demonstrated an excellent level of understanding and also of critical assessment of the current literature in the field. His expertise is also confirmed by his publications and an very comprehensive bibliography. The literary quality and general presentation of the thesis meets highest international standards. The results of the various algorithm development and validation steps are presented very comprehensively and in detail, enabling other researchers to easily reproduce the findings. The results and findings have been evaluated and discussed in detail, taking both own results and the published literature into account.

Overall, the thesis by Mr. Ciach constitutes an independent and highly valuable scientific contribution to the field and provides a substantial improvement for the mass spectrometry community. The algorithms developed as a part of this thesis are available in an open-source Python Package “masserstein”, which enables easy adoption by other laboratories worldwide.

I therefore highly recommend to the faculty to accept this excellent work as a dissertation with honorary distinction.

With best regards,



Univ.-Prof. Dr. Stefan Tenzer