IsotopeFit – data extraction and processing software for cluster mass spectra

Michal Ďurian* [1], **Stefan Ralser**†, **Johannes Postler**† **and Paul Scheier**† [2]¶

* Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava 84248, Slovak Republic
† Institut für Ionenphysik und Angewandte Physik, Universität Innsbruck, A-6020 Innsbruck, Austria

Synopsis We are developing a software capable of aiding in evaluation of complex cluster mass spectra. Theoretical knowledge of characteristic isotopic pattern of each fragment is used to create a matrix of all fragments contained in the mass spectrum. Least squares routine is employed to find abundances of these fragments. Currently, efforts are concentrated on making the software efficient enough to run on common computers with reasonable performance and execution times.

Mass spectroscopy of clusters often produces fairly complex spectra, complicating the extraction of useful data, such as magic numbers, cluster sizes and species identities. To alleviate encountered difficulties, a dedicated software tool – IsotopeFit – has been developed [1] in MATLAB. Its core ability is finding the abundancy of each fragment that contributes to the mass spectrum. Currently, our main goal is to make calculations more resource efficient to tackle even more complex spectra and most importantly, to make even a regular laptop capable of running the software without limiting the user's productivity. Therefore, a new performance-optimized version is being developed in C/C++, which will be published under one of the open-source licenses.

Abundance (A_i) extraction procedure begins, after background subtraction and mass axis calibration, by calculating isotopic patterns $p_i(m)$ for all fragments specified by user. Afterwards, a convolution core $\kappa(m)$, typically a Gaussian, is applied, thus creating a characteristic spectral fingerprint of the fragment. These subsequently constitute the design matrix $\bar{\bar{S}}$ of the calculated mass spectrum $s_{calc}(m)$:

$$\bar{s}_{calc}(m) = \sum_i A_i \,.\, p_i(m) * \kappa(m) = \bar{\bar{S}}.\bar{A}$$

Finally, a non-negative linear least squares fitting routine yields the abundance vector $\bar{A}$ by minimizing the residuum between $s_{calc}(m)$ and the experimental spectrum being evaluated. More detailed description of underlying concepts can be found in [1].

From a technical point of view, the first non-trivial routine is the creation of the design matrix $\bar{\bar{S}}_{M\times N}$. Typical dimensions range between $10^5 - 10^6 \times 10^3 - 10^4$. Due to the fact, that columns, representing individual fragment fingerprints, are being calculated from lighter to heavier ones, we obtain a tall-and-skinny sparse matrix with non-zero values dispersed roughly around the diagonal. Sparseness allows for lower memory requirements.

Second, and the most important technical point is the preparation of input data for the least squares fitting routine, which absolutely requires optimizations due to the dimensions and properties of the design matrix. We take advantage of the fact that calculating QR-decomposition of the design matrix and then using the upper triangular matrix R for the least squares routine does not change the resulting abundance vector $\bar{A}$, while also having a matrix with dimensions of only $N \times N$. Moreover, the QR-decomposition can be calculated by blocks, which, with the use of Givens rotations implementation, allows for parallelization [2].

In its current state, the program is capable of running all necessary subroutines, although with many optimizations and checks still not in place, most notably in the least squares fitting routine. At present, a standard least squares algorithm is used, which is to be replaced with non-negative least squares routine [3], which provides a speed boost and better numerical stability for our case.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 692335.

References

[1] S. Ralser et al. 2015 *Int. J. Mass Spectrom.* **379 pp. 194-199**
[2] S. Hammarling, C. Lucas 2008 *MIMS* *ISSN 1749-9097*
[3] R. Bro, S. D. Jong 1997 *J. Chemom.* **11 pp. 393-401**

[1] E-mail: durian@fmph.uniba.sk
[2] E-mail: Paul.Scheier@uibk.ac.at