Matrix Factorization for the near real time analysis of XRF imaging data

Matthias Alfeld[1,2], Mirwaes Wahabzada[3], Christian Bauckhage[4], Kristian Kersting[5], Gerd Wellenreuther[6], Gerald Falkenberg[1]

(1) DESY, Hamburg, Germany
(2) Current affiliation: Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 8220, Laboratoire d'archéologie moléculaire et structurale (LAMS), Paris, France
(3) INRES-Phytomedicine, University of Bonn, Germany
(4) Fraunhofer IAIS, Sankt Augustin, Germany
(5) Computer Science Department, TU Darmstadt, Germany
(6) European XFEL GmbH, Hamburg, Germany

matthias.alfeld@upmc.fr

Technological progress allows for an ever accelerating acquisition of raw data, creating the thus labeled BIG DATA. To fully exploit its potential is an ongoing challenge in computer science. Spectroscopic imaging is facing a similar challenge of faster acquisition and growing data volumes, and, albeit spectroscopic imaging data is more structured and in principle well understood, benefits from the developments in computer science.

The challenge of interpretation of spectroscopic imaging data is largest in situations where samples with limited *a-priori* knowledge are investigated with severe time constrains, e.g. during a beamtime at a synchrotron radiation source or during an *in-situ* measurement of cultural heritage objects with mobile instruments.

The core problem is to reduce the complexity of data sets to make them easier interpretable. This is commonly achieved by reducing redundancies and enhancing correlations. In practice it not only requires a mathematical correct, but a well understandable representation.

We explored the potential of matrix factorization in the form of Non-Negative Matrix Factorization [1,2] and Archetypal Analysis/Simplex Volume Maximization [3] for near real time data analysis of XRF imaging data. We will give a critical review of the potential and limitations of these approaches and discuss in how far these methods allow to replace full spectral fitting of the data, as this allows to avoid human error during the definition of fitting models. Further, we will discuss how far they are applicable for *needle-in-a-haystack* problems.

[1] M. Alfeld, M. Wahabzada, C. Bauckhage, K. Kersting, G. Wellenreuther, G. Falkenberg, J. Phys. Conf. Ser. 499 (2014) 12013.

[2] M. Alfeld, M. Wahabzada, C. Bauckhage, K. Kersting, G. Wellenreuther, P. Barriobero-Vila, G. Requena, U. Boesenberg, G. Falkenberg, J. Synchrotron Radiat. 23 (2016) 579–589.

[3] M. Alfeld, M. Wahabzada, C. Bauckhage, K. Kersting, G. van der Snickt, P. Noble, K. Janssens, G. Wellenreuther, G. Falkenberg, Microchem. J. 132 (2017) 179–184.