

P4.4 Analysis and implementation of regularization techniques for pure component decompositions

Topic and overall goal. Multivariate curve resolution (MCR) methods aim at extracting the concentration profiles of the pure components and their spectra from mixed spectroscopic data. MCR methods require that the measured spectral data comply with the bilinear Lambert-Beer law. A first step is to determine the number of chemical species which is followed by the recovery of the pure component profiles.¹ MCR methods can be applied to (time-)series of UV-vis, IR or Raman spectra and others as, in the case of spectroelectrochemistry, to series of spectra depending on the electric potential. For this purpose, we develop improved mathematical methods and numerical algorithms based on our previous work and apply the developed methods in the project areas of the RTG. The work focuses on reducing factor ambiguities by the application of constraint-based optimisation techniques with constraints as unimodal concentration profiles or their consistency with a kinetic model and many others. These methods have the advantage to produce reliable pure component information. However, they suffer from their complexity which sometimes makes them difficult to use for non-experienced users.

Specific aims and work plan. The core algorithms for computing MCR-based decompositions can be categorized into three classes, namely those based on multiplicative update formula (first introduced by Lee and Seung²), alternating least-squares iterations (by Kim and Park³), and penalized solvers for optimization problems working with a singular value decomposition of the given spectral data matrix⁴. For this PhD project we focus on the third approach. The truncated singular value decomposition provides a proper basis for reconstructing the measured data and for representing the pure component profiles. Nonnegativity of these profiles is a necessary requirement, but is not sufficient in order to find a unique pure component factorization. Instead, additional constraints are used in order to regularize the optimization problem. Such constraints can be used to favour solutions with special properties as sharp peaks, smooth profiles, selectivity in the concentration profiles and others. The constraint functions and the more heavily weighted penalty functions (as nonnegativity) are used to form an objective function for the numerical optimization. Typically, the objective function is nonlinear and its minimization is difficult. Good initial starting factors are very important. Within this project certain initial factors are to be analyzed which are related to simplexes of a maximal volume in the so-called AFS space. Such volume maximizing factors are well-known to have a profile pattern that can be anticipated for the true factors. Such an approach can reduce the fraction and the magnitude of negative entries in the dual factor of the concentration profiles. A further research topic of this PhD project is to exploit the pre-knowledge of quantum mechanically simulated pure component spectra (for instance by DFT calculations) for regularizing the pure component factorization process. Pattern recognition techniques are to be developed and translated into a constraint function. Such a spectral pattern recognition in application to MCR computations is not known in the chemometric literature so far, but can be expected to be a valuable tool of steering the pure component recovery.

Within this PhD project the student will receive training in state-of-the-art chemometric methods for solving multivariate curve resolution methods. The student will find ample fields of application of these methods in the RTG and can support experimentally working project partners by chemometric analyses of (time-)series of spectra. The PhD student will learn the mathematical background about nonnegative matrix factorizations and for solving blind source separation

-
- 1 M. Maeder, Y. M. Neuhold, *Data Handling Sci. Technol.* **2007** 26; b) E. R. Malinowski, *Factor analysis in chemistry* **2002**, Wiley c) E. Widjaja, C. Li, M. Garland, *Organometallics* **2002**, 21, 1991-1997; d) H. Schröder, C. Ruckebusch, A. Brächer, M. Sawall, D. Meinhardt, C. Kubis, S. Mostafapour, A. Börner, R. Franke, K. Neymeyr, *Anal. Chim. Acta* **2020**, 1137, 170-180; e) K. Lindenau, N. Jannsen, M. Rippke, H. Al Hamwi, C. Selle, H. J. Drexler, A. Spannenberg, M. Sawall, K. Neymeyr, D. Heller, F. Reiss, T. Beweries, *Catal. Sci. Technol.* **2021**, 11, 4034-4050.
 - 2 D.D. Lee, H.S. Seung, *Advances in neural information processing systems* **2001**, 12, 556-562.
 - 3 H. Kim and H. Park, *SIAM J. Matrix Anal. Appl.* **2008**, 30, 713-730.
 - 4 M. Sawall, H. Schröder, D. Meinhardt, K. Neymeyr, In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*; Brown, S., Tauler, R., Walczak, B., Eds., Elsevier, 2020; pp 199-231.

problems. The developed new mathematical theory and the corresponding numerical algorithms can be tested and evaluated for many problems within the framework of the graduate school.

Connection within the RTG. This PhD project will be jointly supervised by Klaus Neymeyr and Robert Francke. The work will be done in close collaboration with Angelika Brückner, Torsten Beweries and Ralf Ludwig.