Resolving components in mass spectrometry data: parametric and non-parametric approaches

Katharine M. Mullen, Ivo H. M. van Stokkum
Department of Physics and Astronomy
Vrije Universiteit Amsterdam
E-mail: {kate|ivo}@nat.vu.nl

March 31, 2008

A fundamental problem in mass spectrometry data analysis is decomposition of a matrix of measurements D, the rows of which represent times and the columns of which represent mass-to-charge ratio, into two matrices C and S, so that $D = CS^T$ and column i of C represents a component contributing to the data with respect to time (called an *elution profile*), and column i of S represents the mass spectrum of that component. This decomposition allows the compounds in a complex sample to be identified by taking the maximum of the elution profile of a component (that is, its *retention time*) and its mass spectrum and matching these properties to those of a known compound stored in a database.

A popular nonparametric means of resolving C and S given D is multivariate curve resolution alternating least squares (MCR-ALS), which combines the alternating least squares algorithm with constraints to impose nonnegativity, unimodality, selectivity, etc. MCR-ALS also allows the resolution of components in many datasets $D_1, \ldots, D_K$ simultaneously. We present a package **ALS** to perform MCR-ALS in R. While the package can be applied to any kind of data, it includes functions to plot mass spectra in particular.

A new methodology for resolving C and S given D currently in development uses a parametric description for C (in which components are usually described by functions based on a exponentially modified Gaussian), and optimizes the resulting separable nonlinear least squares problem to improve estimates for nonlinear parameters, while treating the mass spectra S as conditionally linear parameters. Like MCR-ALS, the methodology is well-suited to resolving components in many datasets simultaneously. We present options for the package **TIMP** that implement this parametric model-based methodology, and address issues such as outliers, a baseline, and instrument saturation.