Possible Improvements to Multivariate Curve Resolution with Particle Swarm Optimisation: Estimation of Rotation Ambiguity

Skvortsov A.N.

St-Petersburg State Polytechnical University. Polyteknicheskaja, 29. Biophysics dept. 195251 St-Petersburg, Russia

Multivariate curve resolution (MCR) employs a set of various techniques widely used in modern chemometrics. They include target factor analysis (TFA), resolving factor analysis (RFA), alternating least squares (ALS), and many other. These techniques aim to resolve data matrices into spectra and concentrations of individual chemical components, given a set of physicochemical constraints. Well known issue of these methods is the rotation ambiguity (RA). If the model is not sufficiently constrained, e.g. in the case of non-negativity constrains only; or the data matrix is rank-deficient, the solution may be not unique (ambiguity set). Except for the simple case of 2 components, RA is hard to predict *a priori*. However, in case of RA, a single MCR solution has very little practical value. So good MCR techniques are to numerically evaluate the extent of the ambiguity and report it to the researcher. There are many works devoted to RA in the literature, and several numeric approaches to its solution do exist (MCR-BANDS, MCR-FMIN etc). Unfortunately these approaches lose the elegance and simplicity of the initial methods: they base on complicated nonlinear algorithms, or produce huge amount of data. The latter data are hard to visualize, and they themselves need non-trivial analysis.

One of the possible options, which gains popularity, is to combine MCR with global search algorithms or evolution algorithms. The power of combining MCR with particle swarm optimization (PSO) has been recently shown (MCR-PSO; *e.g.* Parastar et al., Anal Chim Acta. 2013. 772:16). PSO builds a set of solutions (particles) which are updated by some simple search algorithm, typically including stochastic terms. This inner-level algorithm does not need to have good convergence *per se*. The convergence of PSO is provided by interaction of the particles (swarm intelligence), which is typically some simple potential function. In common PSO, interaction is optimized to find the minimum of generic target function (which is sum of squares of residues in case of MCR). However, we believe that the inner-level algorithm and interaction in PSO may be tuned specifically for solving MCR problems.

The present work focused on the ability of PSO-like algorithms to find ambiguity set for MCR. PSO was modified by adding small penalty function for similarity of solutions (repulsion potential, as it corresponds to repulsion between the particles). When several particles are in the ambiguity set and have the same values of target function, the repulsion potential is the sole force that controls the particles. It repels the particles from each other towards the borders of the ambiguity set. When the procedure finishes, it should in theory produce the set of solutions, which are the least similar to each other (in terms of the selected repulsion potential). The solutions are biased, but the bias can be made small by carefully selecting the repulsion potential.

Various types of repulsion potentials and weight coefficients were tested on common optimization tasks, simulated multivariate data, and real MCR problems (fluorescence spectra). Different inner-level algorithms were tested (stochastic search, ALS, RFA). The algorithm had to account for permutation and scale ambiguities of MCR, which incorrectly reduced penalty for similarity. For two-component data matrices the results were compared to the analytical boundaries of the ambiguity set. The results of the work have shown a fairly good ability of PSO with repulsion potential to span the ambiguity set of MCR solutions. The significant improvement of convergence of PSO over the respective parent method was confirmed. The developed approach is compatible with all types of constrains found in MCR problems including model-based MCR. It also retains the simplicity of the parent algorithms. The only cost for it is the computational expense for optimizing multiple particles, which is partially ameliorated by easy parallelization of PSO.