Analytica Chimica Acta 891 (2015) 101–112

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

A multiresolution approach for the convergence acceleration of multivariate curve resolution methods

Mathias Sawall [a], Christoph Kubis [b], Armin Börner [b], Detlef Selent [b], Klaus Neymeyr [a, b, *]

[a] Universität Rostock, Institut für Mathematik, Ulmenstrasse 69, 18057 Rostock, Germany
[b] Leibniz-Institut für Katalyse e.V. an der Universität Rostock, Albert-Einstein-Strasse 29a, 18059 Rostock, Germany

HIGHLIGHTS

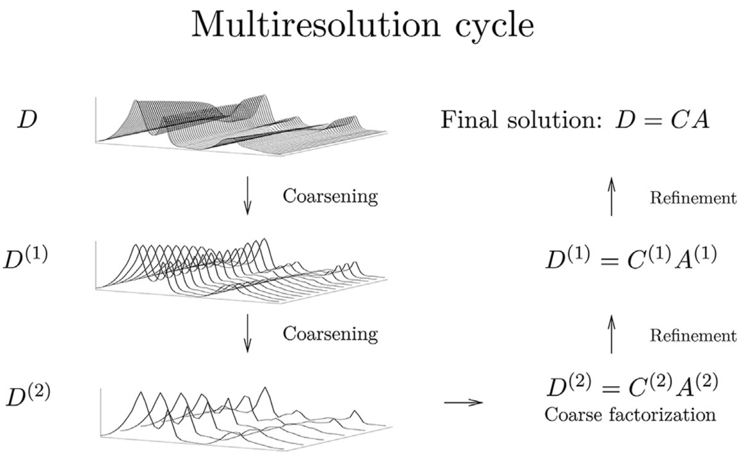
- The multivariate curve resolution problem is solved on a sequence of coarsened levels of resolution.
- Proper restriction and prolongation operations allow to transfer the factorization results between the levels of resolution.
- The new algorithm allows to reduce the computational costs by nearly the factor 10.
- The multiresolution MCR method is tested for model and for experimental data for different coarsening strategies.
- The multiresolution approach can also be applied to similar optimization problems in chemometrics.

ARTICLE INFO

Article history:
Received 27 May 2015
Received in revised form
23 July 2015
Accepted 27 July 2015
Available online 14 August 2015

Keywords:
Chemometrics
Factor analysis
Pure component decomposition
Non-negative matrix factorization
Multiresolution methods

GRAPHICAL ABSTRACT



ABSTRACT

Modern computerized spectroscopic instrumentation can result in high volumes of spectroscopic data. Such accurate measurements rise special computational challenges for multivariate curve resolution techniques since pure component factorizations are often solved via constrained minimization problems. The computational costs for these calculations rapidly grow with an increased time or frequency resolution of the spectral measurements.

The key idea of this paper is to define for the given high-dimensional spectroscopic data a sequence of coarsened subproblems with reduced resolutions. The multiresolution algorithm first computes a pure component factorization for the coarsest problem with the lowest resolution. Then the factorization results are used as initial values for the next problem with a higher resolution. Good initial values result in a fast solution on the next refined level. This procedure is repeated and finally a factorization is determined for the highest level of resolution. The described multiresolution approach allows a considerable convergence acceleration. The computational procedure is analyzed and is tested for experimental spectroscopic data from the rhodium-catalyzed hydroformylation together with various soft and hard models.

---

* Corresponding author. Universität Rostock, Institut für Mathematik, Ulmenstrasse 69, 18057 Rostock, Germany.
E-mail address: klaus.neymeyr@uni-rostock.de (K. Neymeyr).

http://dx.doi.org/10.1016/j.aca.2015.07.043

1. Introduction

The Lambert–Beer law determines the absorption $d(t,\nu)$ for an s-component system with time-dependent concentration profiles $c_i(t)$, $i = 1,\dots,s$, and frequency-dependent pure component spectra $a_i(\nu)$ in the form

$$d(t,\nu) = \sum_{i=1}^{s} c_i(t) a_i(\nu) + e \tag{1}$$

with small error terms e. The continuous-time-frequency model is approximated in practical spectroscopic measurements if spectroscopic data is recorded on a discrete time-frequency grid. For k separate spectra which include a number of n spectral channels the measurements can be recorded in a k-times-n matrix D.

Multivariate curve resolution methods aim at a factorization of this $k \times n$ matrix D in a nonnegative matrix $C \in \mathbb{R}^{k\times s}$ of concentration profiles and a nonnegative matrix $A \in \mathbb{R}^{s\times n}$ of pure component spectra. If a coarse time-frequency grid is selected, i.e. the number kn is small, then the computational costs for the determination of a feasible factorization CA are relatively small. But the resulting small matrices constitute only a poor approximation of the continuous model. In contrast to this, a high time-frequency resolution with potentially oversampled data can yield accurate results at the cost of time-consuming computations. Typically the number k of spectra and the number n of channels are determined by the experimental setup and the spectrometer. The key point of this paper is to develop a computational strategy which uses a sequence of submatrices

$$D^{(1)}, D^{(2)}, \dots, D^{(L)}$$

of the spectral data matrix $D \in \mathbb{R}^{k\times n}$ in order to accelerate the pure component factorization. These submatrices $D^{(i)}$ are representations of the initial matrix $D = D^{(0)}$ with lower resolutions. The nonnegative factorization problem is solved in a way that first the matrix $D^{(L)}$ with the lowest resolution, which is the smallest submatrix, is factored. Then the factorization with respect to the current grid is used as the starting point for the iterative factorization procedure on the next finer time-frequency grid. The resulting iterative procedure is much faster compared to a direct computation of the factorization of the initial high-dimensional matrix $D = D^{(0)}$.

Such a successive approximation of the solution of a general optimization problem (not necessarily related to chemometrics) with respect to the finest grid by means of a sequence of relaxed subproblems, which are cheaper or easier to solve, is a well-known iterative technique for high-dimensional problems. For some classes of problems the sequence of coarsened grids can be used in order to construct very effective solvers for the problem. This is especially the case for the famous multigrid or multilevel methods for the solution of boundary value and eigenvalue problems for elliptic partial differential operators by means of a finite element method [10]. For these problems one has to solve a minimization problem for the elliptic energy functional or for the Rayleigh quotient [4].

The present chemometric matrix factorization problem, which is essentially a multicomponent decomposition, can also be formulated as a minimization problem. For high-dimensional data the solution of such minimization problems can be extremely time-consuming. A severe obstacle to a fast numerical solution of the nonnegative matrix factorization is the non-uniqueness of its solutions. This fact is paraphrased by the *rotational ambiguity* of the solution [1,2,18,25]. A possible approach to single out specific important solutions from the continuum of feasible nonnegative solutions is the usage of hard or soft models [5,12,17]. Finally, a constrained minimization problem is to be solved and the computational costs for the minimization of the target function depend on the dimension of D and on the number of necessary iterations. The number of iterations decreases if the quality of the initial approximation increases.

1.1. Central idea

The aim of this paper is to introduce a multiresolution method for the convergence acceleration of a multivariate curve resolution method. The key idea is to utilize a sequence of coarsened factorization problems in order to compute an associated sequence of gradually refined approximations of the solution. The coarsest problem can be solved with relatively low computational costs and provides good starting values for the factorization problem for the next refined resolution level. These two steps of a correction of the solution with respect to a given resolution level together with the subsequent refinement form a “correction-refinement cycle”. This cycle is applied on the sequence of refined grids until a nonnegative matrix factorization of the initial spectral data matrix D is computed, see Fig. 1.

1.2. Organization of the paper

The paper is organized as follows: In Section 2 a short introduction to multivariate curve resolution techniques is given which includes the principles of soft- and hard-modeling. The central multiresolution approach is introduced in Section 3. Its application to model data and to experimental data from the rhodium-catalyzed hydroformylation process is presented in Section 4. Different strategies for the refinement steps are analyzed.

2. Multivariate curve resolution methods

Multivariate curve resolution methods are powerful tools to extract pure component information from spectroscopic data of chemical mixtures. The spectroscopic measurements are recorded in a $k \times n$ absorption matrix D with k points in time of measurement along the time axis and n spectral channels along the frequency axis. Whereas the continuous form of the Lambert–Beer law is given in (1) its discrete matrix form reads

$$D \approx CA + E.$$

The small error term E collects all measurement errors and nonlinearities. The matrix $C \in \mathbb{R}^{k\times s}$ of concentration profiles and the matrix $A \in \mathbb{R}^{s\times n}$ of the spectra contain columnwise or rowwise the information on the s pure components. The factors C and A and the spectral data matrix D are componentwise nonnegative matrices. The mathematical problem is to compute a chemically meaningful nonnegative matrix factorization CA from a given D. The most common way to compute this factorization is to start with a singular value decomposition (SVD) of D with the form $D = U\Sigma V^T$, [8]. If D has the rank s, then the matrix can also be represented by a truncated SVD. This truncated SVD uses only the first s columns of U and V. Then Σ is an $s \times s$ diagonal matrix containing the s largest singular values on its diagonal. With these matrices the desired factors C and A can be constructed with a regular matrix $T \in \mathbb{R}^{s\times s}$ as follows