Analytica Chimica Acta 772 (2013) 16–25

Contents lists available at SciVerse ScienceDirect

Analytica Chimica Acta

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

Multivariate curve resolution-particle swarm optimization: A high-throughput approach to exploit pure information from multi-component hyphenated chromatographic signals

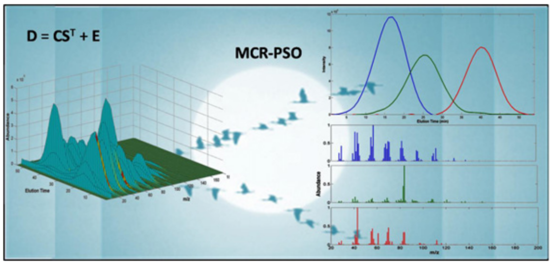
Hadi Parastar[a,*], Heshmatollah Ebrahimi-Najafabadi[b], Mehdi Jalali-Heravi[b]

[a] *Department of Chemistry, University of Isfahan, Isfahan 81746-73441, Iran*
[b] *Department of Chemistry, Sharif University of Technology, Tehran, Iran*

HIGHLIGHTS

- A new MCR method based on particle swarm optimization (PSO) is developed.
- Multi-component simulated GC–MS and HPLC–DAD data are successfully resolved.
- Performance of MCR-PSO algorithm is compared with MCR-ALS and MCR-FMIN.
- MCR-PSO is used for high-throughput analysis of real chromatographic data.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 12 September 2012
Received in revised form 4 December 2012
Accepted 25 February 2013
Available online 6 March 2013

Keywords:
Chemometrics
Multivariate curve resolution
Particle swarm optimization
Hyphenated chromatography
Rotational ambiguity

ABSTRACT

Multivariate curve resolution-particle swarm optimization (MCR-PSO) algorithm is proposed to exploit pure chromatographic and spectroscopic information from multi-component hyphenated chromatographic signals. This new MCR method is based on rotation of mathematically unique PCA solutions into the chemically meaningful MCR solutions. To obtain a proper rotation matrix, an objective function based on non-fulfillment of constraints is defined and is optimized using particle swarm optimization (PSO) algorithm. Initial values of rotation matrix are calculated using local rank analysis and heuristic evolving latent projection (HELP) method. The ability of MCR-PSO in resolving the chromatographic data is evaluated using simulated gas chromatography–mass spectrometry (GC–MS) and high-performance liquid chromatography–diode array detection (HPLC–DAD) data. To present a comprehensive study, different number of components and various levels of noise under proper constraints of non-negativity, unimodality and spectral normalization are considered. Calculation of the extent of rotational ambiguity in MCR solutions for different chromatographic systems using MCR-BANDS method showed that MCR-PSO solutions are always in the range of feasible solutions like true solutions. In addition, the performance of MCR-PSO is compared with other popular MCR methods of multivariate curve resolution-objective function minimization (MCR-FMIN) and multivariate curve resolution-alternating least squares (MCR-ALS). The results showed that MCR-PSO solutions are rather similar or better (in some cases) than other MCR methods in terms of statistical parameters. Finally MCR-PSO is successfully applied in the resolution of real GC–MS data. It should be pointed out that in addition to multivariate resolution of hyphenated chromatographic signals, MCR-PSO algorithm can be straightforwardly applied to other types of separation, spectroscopic and electrochemical data.

---

* Corresponding author. Tel.: +98 311 7932737; fax: +98 311 6689732.
E-mail addresses: h.parastar@sci.ui.ac.ir, h.parastar@gmail.com (H. Parastar).

0003-2670/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.aca.2013.02.042

1. Introduction

In recent decades, hyphenated chromatographic techniques, such as gas chromatography–mass spectrometry (GC–MS), high-performance liquid chromatography–diode array detection (HPLC–DAD) and liquid chromatography–mass spectrometry (LC–MS) have been frequently used in the analysis of complex samples [1–10]. Analysis of complex samples has been a great challenge in analytical chemistry. Though many analytical methods have been developed, most of them are not feasible to be used for analysis of real samples without tedious separation and concentration operations. Indeed, hyphenated chromatographic techniques have been revolutionized the routines in analytical chemistry for the detection, quantification and identification of chemical constituents [5,7,9]. The chromatographic profiles describe the chemical composition of a sample and provide a unique identification tool. With the hyphenated instruments, a huge amount of data can be produced. However, the effective separation of the components in a sample is the prerequisite for qualitative and quantitative analysis. For most of the cases, to obtain a satisfactory separation, optimized experimental conditions should be employed which is usually laborious and time-consuming. Therefore, in spite of advance hyphenated chromatographic systems, the incomplete separation issues still exist in the analysis of complex samples [11–13]. The incomplete separation directly reflects the lack of selectivity in the chromatographic systems. Therefore, exploiting pure chromatographic and spectroscopic information of target compounds in the presence of interferences is very difficult. As a consequence, reliable qualitative and quantitative information cannot be obtained.

Great efforts, therefore, have been made by chemometricians to extract the chemical information from hyphenated chromatographic signals [2,11–17]. Fortunately, in recent decade, different chemometric resolution methods have been proposed by different research groups to compensate the lack of selectivity problem in chromatographic systems and to obtain pure chromatographic and spectroscopic profiles of target compounds in the presence of interferences (which is called second-order advantage) [18–20].

Multivariate curve resolution (MCR) [16,19–21] techniques are a family of chemometric methods based on bilinear model assumption which attempt doing the multivariate resolution of mixed chromatographic signals into the contribution of pure components by means of mathematical tools but without their physical separation. The main assumption of MCR bilinear model is that multivariate chromatographic signals are a linear sum of the pure individual contribution of the different chemical components in the system [14,21].

Despite of the theoretical ability of MCR bilinear models, MCR methods have sometimes experienced difficulties in yielding sufficient resolution performances due to the presence of rotational ambiguity [22–24]. For example multivariate curve resolution-alternating least squares (MCR-ALS) [25–27] as one of the most popular MCR methods, sometimes reaches the suboptimal local minimum due to its linear objective function [28–31]. One potential solution for this problem is to use more powerful optimization algorithms with a more robust objective function, which directly reflects the constraints fulfillment [32,33]. Particle swarm optimization (PSO) [34] is one of these methods.

PSO is a population-based evolutionary technique which searches for optima by updating generations [35]. A remarkable advantage of PSO compared to other evolutionary techniques is that its algorithm is conceptually much simpler and demands lower computational costs. Additionally, it contains few parameters to be adjusted. In the literature, there is only one report on the use of particle swarm optimization (PSO) algorithm for the multivariate resolution purposes [28]. However, in this article, PSO was used as a method to estimate the initial values of concentration or spectral profiles for resolution of complex near infrared (NIR) data and the potential of PSO to obtain meaningful components profiles has not been used in this work.

In the present contribution, MCR-PSO algorithm is developed using a newly developed objective function by Tauler [32] based on non-fulfillment of constraints. The performance of this novel algorithm is tested using simulated GC–MS and HPLC–DAD data with different number of components and different levels of noise. In addition, the performance of this algorithm was compared with MCR-ALS and multivariate curve resolution-objective function minimization (MCR-FMIN) which is based on a similar objective function but it uses local search optimization algorithms such as Quasi-Newton (QN) and Nelder–Mead simplex (NM) algorithms [32,33]. Finally, to test the reliability of MCR-PSO in real cases, a real GC–MS data set was analyzed by this algorithm.

The resolution results were satisfactory; therefore, MCR-PSO can be applied for high-throughput analysis of complex multi-component hyphenated chromatographic signals. As a side point, with this method, multi-component samples may be analyzed in a very fast elution way without considering the separation effect of the components.

2. Theory

2.1. MCR bilinear model

The data obtained from hyphenated chromatographic measurements such as GC–MS, LC–DAD and LC–MS can be arranged in a two-way data matrix (**D**) with elution time points as rows and mass-to-charge ratio (m/z) as columns for GC–MS/LC–MS and wavelengths (λ) for LC–DAD. The two-way data matrix **D** can be decomposed using the PCA (Eq. (1)) and MCR (Eq. (2)) bilinear models:

$$\mathbf{D} = \mathbf{U}\mathbf{V}^{\mathrm{T}} + \mathbf{E}_{\mathrm{PCA}} = \mathbf{D}_{\mathrm{PCA}} + \mathbf{E}_{\mathrm{PCA}} \quad (1)$$

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E}_{\mathrm{MCR}} = \mathbf{D}_{\mathrm{MCR}} + \mathbf{E}_{\mathrm{MCR}} \quad (2)$$

where $\mathbf{D}_{\mathrm{PCA}}$ and $\mathbf{D}_{\mathrm{MCR}}$ are the reconstructed data matrices using PCA and MCR, respectively. **U** and **C** are the row vector matrices and $\mathbf{V}^{\mathbf{T}}$ and $\mathbf{S}^{\mathbf{T}}$ are the column vector matrices obtained by PCA and MCR, respectively. $\mathbf{E}_{\mathrm{PCA}}$ and $\mathbf{E}_{\mathrm{MCR}}$ are the residual matrices with the data variance unexplained by $\mathbf{U}\mathbf{V}^{\mathrm{T}}$ and $\mathbf{C}\mathbf{S}^{\mathrm{T}}$, respectively. The dimensions of matrices are: **D** ($I \times J$), **C** and **U** ($I \times N$), $\mathbf{S}^{\mathbf{T}}$ and $\mathbf{V}^{\mathrm{T}}$ ($N \times J$), and **E** ($I \times J$); I is the number of elution time points in chromatographic direction of the GC–MS/LC–DAD/LC–MS data, J is the number of spectral variables (e.g. m/z or λ), and N is the number of chemical components.

Due to the very restrictive constraints applied during the determination of the PCA scores and loadings, PCA decomposition is unique [36]. On the other side, MCR decomposition is not unique because the constraints imposed during the determination of factor matrices are less restrictive and allow for an undetermined number of equivalent solutions with the same fitting values. Nevertheless, MCR solutions have physical meaning because they are based on natural constraints fulfilled by the true sources of data variation but, PCA solutions lack this physical meaning [21]. Therefore, the good mathematical PCA solutions can be transformed to good physically MCR ones using an appropriate rotation matrix **R** under application of proper constraints. This may be stated by the following equation [21]:

$$\mathbf{D}_{\mathrm{PCA}} = \mathbf{U}\mathbf{V}^{\mathrm{T}} = \mathbf{U}\mathbf{R}\mathbf{R}^{-1}\mathbf{V}^{\mathrm{T}} \Rightarrow \mathbf{C}\mathbf{S}^{\mathrm{T}} = \mathbf{D}_{\mathrm{MCR}} \quad (3)$$

It is important to note that due to experimental noise, going from PCA to MCR solutions is not always an easy trivial task. This problem may incorporate some noise into PCA and MCR solutions