Analytica Chimica Acta 827 (2014) 1–14

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Newer developments on self-modeling curve resolution implementing equality and unimodality constraints

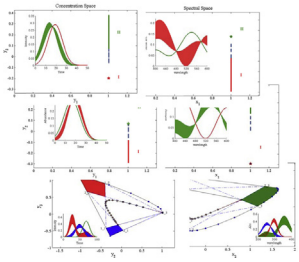
Samira Beyramysoltan [a,b], Hamid Abdollahi [b], Róbert Rajkó [a,*]

[a] Institute of Process Engineering, Faculty of Engineering, University of Szeged, Moszkvai krt. 5-7, Szeged H-6725, Hungary
[b] Department of Chemistry, Institute for Advanced Studies in Basic Sciences, P.O. Box 45195-1159, Zanjan, Iran

HIGHLIGHTS

- In this study, analytical SMCR methods are revised and described using simple concepts.
- Equality and unimodality constraints are successfully implemented in the Lawton–Sylvestre method.
- Procedure is proposed to impose equality constraint in Borgen plots.
- Detailed descriptions and explanations are given based on the obtained abstract spaces.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 20 November 2013
Received in revised form 12 March 2014
Accepted 15 March 2014
Available online 20 March 2014

Keywords:
Self-modeling curve resolution (SMCR)
Lawton–Sylvestre method
Borgen plot
Equality constraint
Unimodality constraint

ABSTRACT

Analytical self-modeling curve resolution (SMCR) methods resolve data sets to a range of feasible solutions using only non-negative constraints. The Lawton–Sylvestre method was the first direct method to analyze a two-component system. It was generalized as a Borgen plot for determining the feasible regions in three-component systems. It seems that a geometrical view is required for considering curve resolution methods, because the complicated (only algebraic) conceptions caused a stop in the general study of Borgen's work for 20 years. Rajkó and István revised and elucidated the principles of existing theory in SMCR methods and subsequently introduced computational geometry tools for developing an algorithm to draw Borgen plots in three-component systems. These developments are theoretical inventions and the formulations are not always able to be given in close form or regularized formalism, especially for geometric descriptions, that is why several algorithms should have been developed and provided for even the theoretical deductions and determinations. In this study, analytical SMCR methods are revised and described using simple concepts. The details of a drawing algorithm for a developmental type of Borgen plot are given. Additionally, for the first time in the literature, equality and unimodality constraints are successfully implemented in the Lawton–Sylvestre method. To this end, a new state-of-the-art procedure is proposed to impose equality constraint in Borgen plots. Two- and three-component HPLC-DAD data set were simulated and analyzed by the new analytical curve resolution methods with and without additional constraints. Detailed descriptions and explanations are given based on the obtained abstract spaces.

1. Introduction

The curve resolution (CR) methods [1,2] magnetize substantial research efforts aimed to discover knowledge of multicomponent

* Corresponding author. Tel.: +36 62546030.
E-mail address: rajko@sol.cc.u-szeged.hu (Ró. Rajkó).

http://dx.doi.org/10.1016/j.aca.2014.03.019
0003-2670/

systems. The soft methods establish a specified part in CR methods, as they allow going into insight the data without prior knowledge of the studied system. Using soft methods, the outcome of the analysis of one system is not usually the true one due to rotational ambiguity. The rotational ambiguity leads to a range of possible solutions that fulfill constraints and represent the measured data correctly. However, the accuracy of the result of soft methods was abolished by systematic error in case of quantitative analysis [3].

Consequently, taking into consideration of all feasible solutions can provide useful information about the system and the process under study when there is no unique solution for the system. Additionally, the calculation of feasible bands is dramatically advantageous not only for quantitative and qualitative application of the resolved profiles but also for decision about numbers and the types of further constraints applied to the model. SMCR methods undertake the task of estimation of the area of feasible solutions (AFS) for every component according to the definition stated by Lawton and Sylvestre [1]. Whereas, there are many efforts on multivariate curve resolution (MCR) methods in the literature (these methods can provide only one solution, but not necessarily a true one, but they can work on noisy data as well); this paper focuses only to SMCR methods.

The first algorithm for determining the AFS in two-component systems was established by Lawton and Sylvestre with supposing the non-negative spectral and concentration values (minimal information about the system) [1]. As the generalization of Lawton and Sylvestre (LS) method, Borgen et al. [4,5] developed an analytical solution of the CR problem for three-component data using the tangent and the simplex rotation algorithms. The presented technique was found mathematically hard to understand and implement; thus Borgen's study could not fascinate the researcher's interest and remains an unknown question in chemometrics literature for 20 years. Rajkó and István [6] revised Borgen's study and could elucidate the concepts of SMCR methods based on the geometry of the abstract space. They used computational geometry tools for developing an algorithm to draw Borgen plots of any three-component systems. The Borgen plot (BP) method cannot resolve the systems with immoderate noise, because it is based on SVD/PCA calculations and BP inherits the sensitivity of SVD/PCA to the noise. Henry and Kim proposed a procedure [7,8] to discover the feasible ranges for an N-component system breaking the problem down into a number of separate linear programming tasks. These methods require non-negativity and other physical constraints to get feasible solutions as points and not regions. The point estimation could be dramatically biased, if some of the restrictions failed. Wentzell et al. [9] used a simplex optimization procedure to develop a new technique for computing the feasible bands for true profiles of components. Later, Leger and Wentzell [10] developed a fast dynamic Monte Carlo SMCR method. The algorithm applies the usual non-negativity assumptions, but adapts to problems arising from measurement noise by accepting some solutions as valid, which may exist slightly outside the principal component subspace. Unfortunately, if the data are noise-free, the method tends to give profiles with negative parts, too. Gemperline [11] and then Tauler [12] presented a method, later coined as MCR-BANDs method, to estimate the band boundaries of feasible solutions based on fast maximization and minimization of the relative signal contribution of each component. MCR-BANDs method can be used for data sets with more than three components. The method indicates the extent of rotational ambiguity for each component in the system but does not provide explicitly full range of feasible solutions [13]. The systematic grid search method was introduced to approximate the AFS numerically in two-component systems, and subsequently it extended for three-component systems [14,15]. The methods explore all feasible solutions that minimize the sum of square of residual obtained from the difference between real data set and reconstructed data set. Nevertheless, the concept of grid search methods is so simple, but it is time-consuming. Recently, Sawall et al. [16] suggested a fast and accurate algorithm to find AFS in two- and three-component systems using the inflation of polygons. This procedure starts with an initial triangle located in a topologically connected subset of the AFS, and an automatic extrusion algorithm is used to form a sequence of growing polygons that approximate the AFS from the interior.

Developing the fast and robust calculation of AFS allows faster discoveries of information from the system. Although the analytical SMCR (LS and BP) methods are able to resolve systems with no or only moderate noise [17,18], they are of interest in this study due to their speed and concepts, i.e. the theoretical exploration of the abstract space of the data. The authors hope that the research's attention could be attracted by these methods after illustrating their abilities. The algorithm of analytical SMCR methods is based on the bilinear model and the non-negativity property of data set. To implement the other constraints in these methods is not a trivial or easy task, and it still needs additional investigation. Thus, the aim of this study is to apply the constraints in SMCR methods. The proposed procedures, for the first time, are suggested to impose the equality and unimodality constraints in LS method and to implement the equality constraint in BP method.

As BP is a fast procedure to discover the geometry of the abstract space of the data set, the main structure of the paper emphasizes on the Borgen plot and the detailed description will be given about its theory. The developed algorithm is the improved type of the BP existing in the study of Rajkó et al. [6]; the Borgen plot can be drawn using the duality concept, the convex hull and tangent methods for all three-component systems. These developments are theoretical inventions and the formulations cannot be given in close form; thus the theoretical deduction means to give an algorithm to reveal all steps of the formalism. The algorithm for the analytical derivation of the feasible regions with the imposed constraints is given illustrating the steps with figures.

2. On the analytical SMCR methods

2.1. SMCR problem

For any spectral chemical data set (**R**), the Bouguer–Lambert–Beer law allows expressing the relation between spectral (**S**) and concentration (**C**) profiles as follows:

$$\mathbf{R}_{I,J} = \mathbf{C}_{I,n}(\mathbf{S}^{\mathrm{T}})_{n,J} \tag{1}$$

where I and J introduce the number of rows and columns, respectively; n is the number of components that cause the variance in the data set. For the column and row space of the data set, orthonormal basis vectors (**U** and **V**, respectively) can be given by performing the singular value decomposition on the data matrix according to Eq. (1):

$$\mathbf{R}_{I,J} = \mathbf{U}_{I,n}\mathbf{D}_{n,n}(\mathbf{V}^{\mathrm{T}})_{n,J} = \mathbf{X}_{I,n}(\mathbf{V}^{\mathrm{T}})_{n,J} = \mathbf{U}_{I,n}(\mathbf{Y}^{\mathrm{T}})_{n,J} \tag{2}$$

D is a diagonal matrix including singular values; the **X** and **Y** outline the coordinates of row and column vectors of the data set in a row and column spaces, respectively. The problem of the SMCR technique is to acquire the correct basis profiles **C** and **S** for a given data set. The **X** and **V** can be transformed by any invertible transformation matrix like matrix **T** to physically and chemically interpretable matrix, **C** and **S** based on Eq. (2):

$$\mathbf{R}_{I,J} = (\mathbf{X}_{I,n}(\mathbf{T}^{-1})_{n,n})(\mathbf{T}_{n,n}(\mathbf{V}^{\mathrm{T}})_{n,J}) = (\mathbf{U}_{I,n}(\mathbf{T}_{*}^{-1})_{n,n})(\mathbf{T}_{*n,n}(\mathbf{Y}^{\mathrm{T}})_{n,J}) = \mathbf{C}_{I,n}(\mathbf{S}^{\mathrm{T}})_{n,J} \tag{3}$$

Note that the transformation matrices for the two spaces have one-to-one relationships: $\mathrm{D}_{n,n}(\mathrm{T}^{-1})_{n,n} = (\mathrm{T}_{*}^{-1})_{n,n}$ and $\mathbf{T}_{n,n} = \mathbf{T}_{*n,n}\mathbf{D}_{n,n}$.