Computation of the range (band boundaries) of feasible solutions and measure of the rotational ambiguity in self-modeling/multivariate curve resolution

Róbert Rajkó *

Department of Mechanical and Process Engineering, Faculty of Engineering, University of Szeged, P.O. Box 433, H-6701 Szeged, Hungary

1. Introduction

Self-modeling curve resolution (SMCR) was introduced in 1971 by Lawton and Sylvestre [1] for decomposing a bilinear two-component spectroscopic data matrix into physically interpretable profiles, i.e., concentration profiles and spectra, using the minimal constraint of nonnegativity for concentrations and absorbances. Unfortunately, the solution is not unique: without further restrictions, the method can only give feasible regions for the pure component profiles. The work that followed Lawton and Sylvestre took two directions: (1) developing and generalizing the original method to obtain feasible regions for more than two components, and (2) developing iterative (or other, e.g., Monte Carlo) methods and using constraints to get a unique solution for arbitrary number of components. Many papers were published in both areas. Hamilton and Gemperline gave a very detailed history of the early years from 1972 to 1989 [2], the next decade attempts were reviewed by Liang and Kvalheim [3], and the reader can find more recent information on the topic in Refs. [4–12] as well. We do not presume to give detailed introduction to the historical development, but guide the reader to the above-mentioned papers.

Borgen and Kowalski [13,14] have generalized analytically (i.e., without use of any approximate, Monte Carlo, etc. algorithms) the method of Lawton and Sylvestre to three-component systems, with the same minimal constraints. The concepts of Borgen were rather difficult to understand and implement. This explains to a large extent why many chemometricians turned to developing approximation methods for SMCR [2–11]. Motivated by this trend, Rajkó and Irtván [9] revisited Borgen’s method. They improved the interpretation and used tools from computational geometry to find inner and outer polygons, which form the basis for creating the analytically given feasible regions. Very recently Rajkó [15] has extended the duality concept for the minimal constrained SMCR being available a simpler algorithm.

Several years ago, Gemperline [16] proposed a novel concept of computation of the range of feasible solutions in self-modeling curve resolution algorithms. The new method for finding feasible ranges was based on a matrix of mixture spectra, $A = CP^T$, where $A$ is an $n \times m$ data matrix with spectra measured at $n$ time intervals and $m$ wavelengths; $C$ is an $n \times k$ matrix of pure composition profiles; $P$ is a $k \times m$ matrix of pure-component spectra; and $k$ is the number of components. For each pair of upper and lower boundaries a rank-one matrix, $A_q$, for the constituent of interest, $q$, was computed

$$A_q = cqp_q$$  \hspace{1cm} (1)

and the integrated signal due to the constituent was approximated by taking the sum over $n$ and $m$ as shown in Eq. (2)

$$\max \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m} (c_i p_{qj}) \right\} \quad \text{or} \quad \min \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m} (c_i p_{qj}) \right\}$$  \hspace{1cm} (2)
The upper boundary condition for the $k$th constituent was found by maximizing this sum, and the lower boundary condition was found by minimizing this sum, subject to simultaneous nonnegativity constraints for all $k$ concentration profiles and spectral profiles.

Later Tauler [17] showed that Gemperline’s concept was equivalent to minimize and maximize the function in Eq. (3)

$$f_k(T) = \frac{||c_k s_k^T||}{||CS||}$$

where $T$ is the rotation (transformation) matrix (Gemperline denoted it by $B$); $C$ is the matrix of pure composition profiles; $S$ is the matrix of pure-component spectra (Gemperline denoted it by $P$). Eqs. (1)–(3) above are copied from the original papers of Gemperline [16] and Tauler [17], thus detailed derivations and more information can be found in those papers, however the equations are shown here to introduce the basis of the problem.

The concept developed by Tauler was used for investigation of noise propagation and error estimation [18], reaction of curing epoxy resins [19], multiple runs of this reaction [20], multiple runs of gasoline blending processes [21] and two chemical and environmental simulated data sets [22].

It should be noted, that the main core of the concept described above might be based on the experiences on the simplest, two-component cases, i.e. the inner bound and the outer bound provide the minimum and the maximum of the range [1], see for example Fig. 1 in Ref. [9]. That is why both Gemperline and Tauler suggested to calculate only two boundaries, i.e. minimum and maximum.

Very recently, reviewing multivariate curve resolution (MCR) methods, de Juan and Tauler [11] have expressed their suspicion: "Some of the latest tendencies opt for determining, one at a time for each component, the minimum and maximum boundaries of the feasible solution bands for the dyad of resolved profiles. The general idea is finding the ‘boundary’ dyads of profiles that provide the maximum and the minimum contribution of the component signal to the overall signal measured. . . ." Other methodologies question whether the minimum and maximum boundaries really enclose all possible feasible profiles and, whether defined as maximum and minimum of signal contributions, they really represent the ‘extreme’ solutions."

As we will show it below, their doubting was right. Actually, for two-component noisy systems, Vosough et al. [23] realized that not all profiles within a Gemperline–Tauler (GT)'s band of feasible solutions are possible, only those that are represented as a point in T-space. Using a simple three-component example we show that the minimum–maximum concept of Gemperline and Tauler for calculation of the range of feasible regions is not correct in general. We explain why, and we suggest a correct concept, which is a Borgen plot founded one, based on the recently published reasoning [24]. We also elucidate why the so called “relative component contribution” concept introduced by Tauler cannot be used for measuring the rotational ambiguity.

From time to time the chemometric methods should be reviewed and revalued, otherwise the scientific community can draw wrong theoretical and/or practical considerations accepting and using incorrect developments. For corresponding notes, the readers can refer the papers recently published in a special issue of TriC titled as “Use and abuse of chemometrics” [25], a short communication in which we proved that a recently developed SMCR method is not the generalization of a previous one [26], and another paper in which we revealed some surprising properties of MCR-ALS algorithms [27].

2. Comparing old and new concepts for computations of range of feasible solutions and measure of the rotational ambiguity in S/MCR: an illustrative example

The data to illustrate deeper discrepancy for three-component noiseless systems was generated using figure A1 and A2 of the paper of Borgen and Kowalski [13]. The Borgen plots of the noise-free data with known concentration and spectral profiles are plotted in Fig. 1. The three well-separated feasible regions shaded by blue can be seen on both panels. The green crossed circles depict the pure component profiles (they are connected with dashed green lines resulting in a triangle, i.e. a two-dimensional simplex). The cyan diamonds depict the minimum and the maximum bounds computed by Tauler’s algorithm [22] called MCR-BANDS. If the two-point bounds are proper, the segment line (cyan dotted) between the two cyan diamonds should contain the pure profile points. If the feasible region is approximated by two points (as GT’s band boundaries concept does), then the true solution should be in the region (i.e. on the line segment defined by the two points).

In the Borgen plot, i.e. the abstract space of the principal components, the ‘mixture’ of two any points should be on the straight line between the two points. It is because the basic function represented by Eq. (1) is linear. As was mentioned above, the pure components are on the vertices of the green triangle. The mixture of the three pure components should be in or on the triangle. If only two pure components are mixed, the point representing the mixture is on the corresponding side of the triangle. And this is true in general, because the mixture of any of two points should be on the straight line segment connecting to two points.

As it can be seen, there are feasible regions in each panel of Fig. 1 where the pure component is not on the segment line. It means that in these cases the boundaries computed based on the GT’s concept are theoretically wrong, thus it cannot be used in general: a feasible region, which is a complicated geometrical form defined by several vertices, straight and curved lines in case of three-component systems, cannot be represented by only two points (i.e. with a line segment).

Fig. 2A and B shows the feasible regions and boundaries transformed into the more interpretable profiles (indeed every point in Borgen plots can be transformed to a concentration profile or spectrum, respectively). Because 1-norm normalization was used to create every Borgen plot in this paper, every profile can be compared with each other; of course other norms can be used for normalization, such study has been recently published [24]. The brightly shaded bands are calculated using GT’s approach. The true profiles plotted by dashed black lines are sometimes outside the shaded bands (thick black arrows show the outsider points) indicating that the bands do not cover all possible feasible solutions not even the true profiles. The faintly shaded bands are calculated using the extreme points of the feasible regions on the Borgen plots. The true profiles and the GT’s bands are covered by the Borgen-Rajkó (BR) bands as well. Because Fig. 2A and B may contain too much information, Fig. 2C was generated showing the retransformed bands for only one feasible region of component 3 in concentration space. The dimmed band is calculated using the extreme points of the feasible regions on the Borgen plots (see them in Fig. 3A). The brightly shaded band is calculated using GT’s approach. Now it is evident that the GT’s band is much narrower than BR’s band.

For clearer illustration of the role of the extreme points of the feasible regions on the Borgen plot, Fig. 3 was created. Fig. 3A shows one of the magnified feasible regions belonging to the concentration space. The numbers on yellow background indicate the extreme points of the feasible region. It can be seen that the feasible region is not a proper convex polygon, because two sides of it are curves (the limiting curve, as was called it by Borgen, can be plotted using $x_i$ and $y_i$ coordinates calculated by Eq. (44) of Ref. [9]). The
letter ‘A’ and ‘B’ on cyan background indicate the boundary points of the GT’s band. The letter ‘C’ on red background indicates the point belonging to the pure component. The retransformed profiles are plotted in Fig. 3B. It can be seen that all 10 profiles should be used for depicting the proper band. Note that profile 10 coincides with profile ‘A’.

The bias of GT’s approach can be reduced if the bounded region for each individual component is compiled by taking the union of the bounded regions computed by finding the minimum and maximum signal of each individual component. Table 1 contains the maximum and minimum values of the measure defined by Eq. (3) for all three components. The six profiles belonging to the measures in every column should be unified. However as it can be seen in Fig. 3, six profiles are still not enough for making the proper band, despite the unquestionable improvement in the shown example.

Tauler insists on only two profiles, i.e. ones which have the real maximum and minimum function values according to Eq. (3). He calls these measures as the maximum and minimum relative component contributions (MaRCC and MiRCC, respectively). The two profiles defined by MaRCC and MiRCC will be the band boundaries.

<table>
<thead>
<tr>
<th>Comp. 1</th>
<th>Comp. 2</th>
<th>Comp. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>0.753</td>
<td>0.244</td>
</tr>
<tr>
<td>Min</td>
<td>0.282</td>
<td>0.662</td>
</tr>
<tr>
<td>Max</td>
<td>0.285</td>
<td>0.724</td>
</tr>
<tr>
<td>Min</td>
<td>0.672</td>
<td>0.241</td>
</tr>
<tr>
<td>Max</td>
<td>0.384</td>
<td>0.406</td>
</tr>
<tr>
<td>Min</td>
<td>0.367</td>
<td>0.665</td>
</tr>
</tbody>
</table>
Fig. 2. Transformed profiles. The brightly shaded areas are the bands computed by the Gemperline–Tauler’s concept and the faintly shaded bands are calculated using the extreme points of the feasible regions on the Borgen plots. Black lines with black dots are the pure component profiles, and thick arrows show the outlier points. (A) Spectral profiles. (B) Concentration profiles. (C) Concentration bands for only component 3.
Fig. 3. Illustration of the role of the extreme points of the feasible regions on the Borgen plot. (A) Enlarged feasible region from concentration abstract space. The six cyan diamonds represented the six profiles should be unified for generating the improved band. (B) Retransformed 10 profiles ('1'–'10') for obtaining the proper band, Gemperline–Tauler’s band boundaries ('A' and 'B'), and the true profile ('C'). The faintly shaded red band was calculated by unifying the six profiles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Tauler stated in Ref. [17] that “It may happen, however, as in Fig. 6E, that feasible solutions appear outside the boundary plots. In this case the feasible solutions are still numerically within the boundaries, but when they are plotted as in Fig. 6E, they may appear to be outside the boundary plots. The reason for this is that the boundaries give the maximum and minimum of the optimization function defined in terms of signal contribution, but these numerical limits may not coincide exactly with the physical limits in the graphical representation of feasible solutions in two-dimensional plots such as those shown in Fig. 6E. For instance, although the boundaries define the maximum and minimum species signal contributions and all the feasible solutions should be within these boundaries, the graphical representation of the boundaries for a particular species in a two-dimensional plot could apparently appear to be outside the boundaries. This is a problem related only to the graphical representation of the boundaries and not to their intrinsic meaning and interpretation.”

He strengthened it in Ref. [22]: “In all cases, the theoretical values and the MCR solutions are within the boundaries estimated by MCR-BANDS.”

Evidently Tauler believes that the maximum and minimum band boundaries should cover the true signal and all other feasible solutions. In this paper we reveal that this conception can be used only in limited cases. We can postulate that notion of the band boundaries and notion of the relative component contributions should not be confused. Band boundaries should cover all feasible solution and they can be represented graphically. Relative component contributions measured by the function values of Eq. (3) are just numerical values, and it is not recommended to draw the corresponding profiles as band boundaries.

Tauler reported [22] that the maximum and the minimum band boundaries calculated for the solutions of MCR-ALS and MCR-FMIN (and of course the measures defined by Eq. (3) belonging to these boundaries) were different (see Fig. 1 in Ref. [22]). How-
ever, it is impossible, because the rotational ambiguity belongs to the data matrix and is independent of the algorithms of the curve resolution. MCR-BANDS is seeking for a suitable T rotational matrix getting the maximum (or the minimum) value of the function in Eq. (3). Eq. (4) shows the relationship among the relevant matrices:

\[ C_{\text{MCR-ALS}} T_{\text{MCR-ALS}, \text{max}} = C_{\text{MCR-ALS}} \]
\[ C_{\text{MCR-FMIN}} T_{\text{MCR-FMIN}, \text{max}} = C_{\text{MCR-FMIN}} \text{max} \]  

(4)

Suppose that the measure of the kth component is higher for \( C_{\text{MCR-FMIN}, \text{max}} \) than for \( C_{\text{MCR-ALS}, \text{max}} \). With a suitable rotation matrix \( T \),

\[ C_{\text{MCR-ALS}} T^{-1} = C_{\text{MCR-FMIN}} \text{ max} \]
\[ C_{\text{MCR-ALS}} T^{-1} T_{\text{MCR-FMIN}, \text{max}} = C_{\text{MCR-FMIN}} \text{max} \]  

(5)

\( T^{-1} T_{\text{MCR-FMIN}, \text{max}} \) can be considered as a new rotational matrix, but it was not found for MCR-ALS solution in Ref. [22].

The complex character of the problem can be illustrated by recent work of Rajkó [24], in which he demonstrated that the rotational ambiguity (defined as the area of the corresponding band normalized to maximum 1) of minimal constrained SMCR method is Borgen norm-dependent.

However, a slightly better definition of the rotational ambiguity is the area ratio of the band transformed from the feasible region of the observed component to the band transformed from the outer polygon regarding the suitable Borgen plot. In Table 2 the values given by using this new definition are compared with ones based on the relative component contribution. These latter are calculated simply as the difference between MaRCC and MiRCC.

Studying Table 2, the most important statement is that the relative component contribution cannot make any difference for the concentration and spectral spaces. Only one indicator can be calculated for every component. For the explanation why this property is critical, let us imagine that one of the spectra is known. In this way the spectral rotational ambiguity of the component with known spectrum will be zero. But the rotational ambiguity for concentration is not necessarily zero in general. Thus the measure for the rotational ambiguity calculated based on the relative component contribution must be wrong: if its value is zero then it cannot handle the concentration space suitably; if its value is not zero then it cannot work on the spectral space properly.

However, the rotational ambiguity can be measured correctly by using Borgen plot method.

3. Conclusions

As a concluding remark, we have to mention that the analytical solution, i.e. the Borgen plot method can be used only for three-component noise-free systems presently (however the first paper according to the application of Borgen plot for real (noisy) measurements has been already published [28], and another one is under reviewing [29]). Thus there is a real need to develop approximation methods for more than three components and noisy data to calculate the feasible band boundaries. It turned out that GPF's approach is biased. How this bias can be altered by the correlation and/or shape property of the pure component profiles is a very exciting research question and it should be focused on near future. It would be also very fruiting to study the effects of the supplementary (unimodality, closure, etc.) on the feasible regions on Borgen plot. Thus self-modeling curve resolution is far not a closed chemometric area, there are a lot of unanswered questions fertilizing corresponding chemometric researches.

Acknowledgements

Roma Tauler is acknowledged for providing the Matlab code with which the minimum and maximum boundaries could be calculated. The anonymous reviewers of the manuscript are thanked for their helpful suggestions and comments. Last but not least, the author highly appreciates Paul Gemperline and Roma Tauler's excellent contributions to developing, using and teaching curve resolution methods in chemometrics.

References


Table 2

Measures of the rotational ambiguity: comparing the new definition with one based on the relative component contribution.

<table>
<thead>
<tr>
<th></th>
<th>Comp. 1</th>
<th>Comp. 2</th>
<th>Comp. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration</td>
<td>0.2912</td>
<td>0.5158</td>
<td>0.4722</td>
</tr>
<tr>
<td>Spectrum</td>
<td>0.2675</td>
<td>0.6309</td>
<td>0.3654</td>
</tr>
<tr>
<td>Max-Min</td>
<td>0.471</td>
<td>0.483</td>
<td>0.300</td>
</tr>
</tbody>
</table>


