

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/112306>

Please be advised that this information was generated on 2019-12-12 and may be subject to change.

# GENERALIZED RANK ANNIHILATION METHOD. I: DERIVATION OF EIGENVALUE PROBLEMS

N. M. FABER, L. M. C. BUYDENS AND G. KATEMAN

*Department of Analytical Chemistry, University of Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen,  
Netherlands*

## SUMMARY

Rank annihilation factor analysis (RAFA) is a method for multicomponent calibration using two data matrices simultaneously, one for the unknown and one for the calibration sample. In its most general form, the generalized rank annihilation method (GRAM), an eigenvalue problem has to be solved. In this first paper different formulations of GRAM are compared and a slightly different eigenvalue problem will be derived. The eigenvectors of this specific eigenvalue problem constitute the transformation matrix that rotates the abstract factors from principal component analysis (PCA) into their physical counterparts. This reformulation of GRAM facilitates a comparison with other PCA-based methods for curve resolution and calibration. Furthermore, we will discuss two characteristics common to all formulations of GRAM, i.e. the distinct possibility of a complex and degenerate solution. It will be shown that a complex solution—contrary to degeneracy—should not arise for components present in both samples for model data.

KEY WORDS RAFA GRAM Eigenvalue problem Complex solution Degenerate solution

## INTRODUCTION

Rank annihilation factor analysis (RAFA) is a method for multicomponent calibration using two data matrices simultaneously, one for the unknown and one for the calibration sample. In order to apply the technique of rank annihilation, the measured signal must be linear and additive, e.g. high-performance liquid chromatography with a diode array-UV/visible spectrophotometer as a detector (HPLC-DA-UV) or fluorescence excitation-emission spectroscopy. Data constructed in this way are called bilinear. For bilinear data the rank of a one-component data matrix is one in the absence of noise. Rank annihilation further demands that the signal for the analyte of interest be identical for both samples and finally it must be independent of the signal of the remaining substituents, i.e. the presence of the analyte of interest will raise the rank of the data matrix by one. If the data follow the assumed model, rank annihilation can be used to quantitate the analyte of interest without calibrating for the interferences.

The method was originally developed by Ho *et al.*<sup>1</sup> as an iterative procedure, but the latest developments in rank annihilation have their origin in Lorber's non-iterative reformulation of the calibration problem.<sup>2</sup> Lorber found a direct solution for the case where the calibration sample contains only one component. He derived a standard eigenvalue problem by projecting the calibration matrix on the significant principal components of the unknown data matrix.

The concentration ratio of the component was found as the only non-zero eigenvalue. Lorber's method was generalized by Sánchez and Kowalski<sup>3</sup> to the case of several components that are not necessarily present in both samples. This method, introduced as generalized rank annihilation factor analysis (GRAFA), is now known as the generalized rank annihilation method (GRAM). Wilson *et al.*<sup>4</sup> modified the procedure of Sánchez and Kowalski by projecting both matrices on a common low-dimensional subspace.

In this first paper we will compare different formulations of GRAM and present an alternative derivation that results in an eigenvalue problem for which the eigenvectors have a very simple interpretation: they form the transformation matrix that rotates the abstract factors found by a principal component analysis (PCA) to their physical counterparts. In this way a relationship can be established with other PCA-based curve resolution methods,<sup>5,6</sup> e.g. iterative target-testing factor analysis (ITTFA),<sup>7</sup> evolving factor analysis (EFA)<sup>8</sup> and the recently described window factor analysis (WFA).<sup>9</sup> Every method has its specific features, but a drawback common to all these curve resolution methods is the difficulty in constraining the (final) solution for the transformation matrix in an objective way. Finally we will discuss two characteristics of GRAM, i.e. the distinct possibility of a complex and degenerate solution. It will be shown that a complex solution should not arise for components that are present in both samples if the data follow the assumed linear additive model. However, the situation may be markedly different if the data are affected by model errors. Degeneracy constitutes a fundamental problem for ideal as well as non-ideal data. Whereas self-modeling curve resolution methods seem to work best when used for the calibration of samples that are very similar, at least part of the information may be lost if rank annihilation is applied.

We will start by introducing the relevant symbols and decompositions in the context of PCA-based curve resolution and calibration.

### PCA-BASED CURVE RESOLUTION AND CALIBRATION

The goal of curve resolution is the decomposition of a data matrix into the pure contributions of the individual components. Without loss of generality we will assume throughout this paper that the data are obtained by the spectral detection of a chromatographic separation process. Then, if Beer's law is valid, the  $S \times W$  data matrix  $\mathbf{M}$  of  $S$  mixture spectra measured at  $W$  wavelengths can be written as

$$\mathbf{M} = \mathbf{H}\mathbf{Y}^T \quad (1)$$

where  $\mathbf{H}(S \times K)$  contains the pure elution profiles of the  $K$  components and  $\mathbf{Y}(W \times K)$  contains the pure spectra. Usually the spectra in  $\mathbf{Y}$  are normalized so that the concentration dependency is absorbed in  $\mathbf{H}$ .

In curve resolution one is primarily concerned with the reconstruction of  $\mathbf{H}$  and  $\mathbf{Y}$  (qualitative solution). The problem of calibration is more difficult, because the denormalized elution profiles in  $\mathbf{H}$  have to be related to real concentration values (quantitative solution). This can be done directly if a theoretical relationship exists between the measured response and the concentrations. Otherwise an empirical relationship has to be built by estimating calibration factors from the response of a standard sample. Different calibration schemes are outlined in Reference 7. In calibration the following notation for  $\mathbf{M}$  is often preferred:

$$\mathbf{M} = \mathbf{X}\mathbf{C}_M\mathbf{Y}^T \quad (2)$$

where the columns in  $\mathbf{X}(S \times K)$  represent normalized elution profiles and  $\mathbf{C}_M$  is a  $K \times K$  diagonal matrix proportional to the concentrations.

PCA-based curve resolution proceeds in two steps.<sup>6</sup> First, PCA is applied in order to define the solution space in terms of a set of orthogonal base vectors. This results in the following decomposition of  $\mathbf{M}$ :

$$\mathbf{M} = \mathbf{A}\mathbf{B}^T \quad (3)$$

where the matrices  $\mathbf{A}(S \times W)$  and  $\mathbf{B}^T(W \times W)$  are usually denoted as score and loading matrices respectively. (It is assumed that  $S \geq W$ .) The decomposition in scores and loadings is equivalent to the singular value decomposition (SVD) of  $\mathbf{M}$ :

$$\mathbf{M} = \mathbf{U}\mathbf{\Theta}\mathbf{V}^T \quad (4)$$

where  $\mathbf{U}$  is the  $S \times W$  matrix of left singular vectors,  $\mathbf{\Theta}$  is the  $W \times W$  diagonal matrix of singular values and  $\mathbf{V}^T$  is the  $W \times W$  matrix of right singular vectors. Scores and loadings are related to the singular vectors by\*

$$\mathbf{A} = \mathbf{U}\mathbf{\Theta} \quad (5a)$$

$$\mathbf{B} = \mathbf{V} \quad (5b)$$

Next  $\mathbf{M}$  is reproduced using only the significant PCs and this decomposition is rewritten by means of a transformation matrix  $\mathbf{T}$  as

$$\bar{\mathbf{M}} = \bar{\mathbf{A}}\bar{\mathbf{B}}^T = \bar{\mathbf{A}}\mathbf{T}\mathbf{T}^{-1}\bar{\mathbf{B}}^T \quad (6)$$

The 'overbar' denotes that the corresponding decomposition (PCA or SVD) is truncated. If  $F$  is the number of PCs retained for the reproduction,  $\mathbf{T}$  is an  $F \times F$  matrix. A successful transformation yields the physical decomposition of  $\mathbf{M}$  up to a normalization constant:

$$\mathbf{H} = \bar{\mathbf{A}}\mathbf{T} \quad (7a)$$

$$\mathbf{Y}^T = \mathbf{T}^{-1}\bar{\mathbf{B}}^T \quad (7b)$$

Differences between PCA-based methods come down to differences in estimating the transformation matrix  $\mathbf{T}$ . The problem of curve resolution and subsequent calibration is therefore translated to finding the correct number of factors  $F$  in the PCA step and determining a successful transformation matrix  $\mathbf{T}$ . It is important to note that overfactoring the model will not change the PCs but will certainly affect the estimate of  $\mathbf{T}$ . However, one frequently reported advantage of rank annihilation is the relative insensitivity of the solution to the number of PCs included in the model. This fact can very well be explained by the standard errors for the eigenvalues we recently derived using the method of error propagation.<sup>10</sup>

Until now the treatment has been restricted to the analysis of a single data matrix. In the next section it will become clear how the availability of a second data matrix, obtained under identical experimental circumstances, can help in determining  $\mathbf{T}$ .

## DIFFERENT FORMULATIONS OF RANK ANNIHILATION

As outlined in the Introduction, RAFA comprises a number of related methods. We will restrict ourselves to the discussion of methods that can be derived from the direct solution to the one-component problem, first published by Lorber.<sup>2</sup>

\* The definition of scores and loadings mentioned before is purely conventional. A more operational definition has been given by Malinowski:<sup>5</sup> 'Attention is focused on either the row designees or the column designees. Where attention is focused is called the *scores*; the counterpart is called the *loadings*.' For the discussion of rank annihilation only the relation to the SVD is important.

### Lorber's method

The method of rank annihilation can be applied if a calibration matrix  $\mathbf{N}$  is present:

$$\mathbf{N} = \mathbf{X}\mathbf{C}_N\mathbf{Y}^T \quad (8a)$$

Equation (8a) is the preferred notation for the calibration matrix in the literature of rank annihilation. The alternative derivation presented in a later section will make use of the transcription

$$\mathbf{N} = \mathbf{H}\mathbf{\Pi}\mathbf{Y}^T \quad (8b)$$

where  $\mathbf{\Pi} = \mathbf{C}_M^{-1}\mathbf{C}_N$ . The components that are absent in the calibration sample (all but one) are indicated by a corresponding zero on the diagonal of  $\mathbf{C}_N$  and  $\mathbf{\Pi}$  respectively. Lorber<sup>2</sup> showed that combination of equations (1) and (8a) gives the generalized eigenvalue problem

$$\mathbf{N}\mathbf{Z} = \mathbf{M}\mathbf{Z}\mathbf{\Pi} \quad (9)$$

where  $\mathbf{Z} = (\mathbf{Y}^T)^+$ . The matrices  $\mathbf{N}$  and  $\mathbf{M}$  are, however, not necessarily square and consequently the common eigenvalue-problem-solving routines<sup>11</sup> cannot be used. Approximating  $\mathbf{M}$  by the truncated SVD of equation (4), i.e.  $\bar{\mathbf{M}} = \bar{\mathbf{U}}\bar{\mathbf{\Theta}}\bar{\mathbf{V}}^T$ , and making use of the orthogonality properties of  $\mathbf{U}$  and  $\mathbf{V}$  leads to the standard eigenvalue problem

$$(\bar{\mathbf{U}}^T\mathbf{N}\bar{\mathbf{V}}\bar{\mathbf{\Theta}}^{-1})\mathbf{Z}^* = \mathbf{Z}^*\mathbf{\Pi} \quad (10)$$

where  $\mathbf{Z}^* = \bar{\mathbf{\Theta}}\bar{\mathbf{V}}^T\mathbf{Z}$ . The concentration ratio of the analyte of interest is found as the only non-zero eigenvalue.

### Generalization by Sánchez and Kowalski

If the calibration sample contains several components that are also present in the unknown sample, the eigenvalues found have to be identified, i.e. the qualitative solution is needed. Sánchez and Kowalski<sup>3</sup> recognized that the pure component responses, necessary for the identification step, could be reconstructed by

$$\mathbf{H} = \bar{\mathbf{U}}\mathbf{Z}^* \quad (11a)$$

$$\mathbf{Y}^T = (\mathbf{Z}^*)^{-1}\bar{\mathbf{\Theta}}\bar{\mathbf{V}}^T \quad (11b)$$

Combination of (11a) and (11b) gives, as expected,

$$\bar{\mathbf{M}} = \bar{\mathbf{U}}\bar{\mathbf{\Theta}}\bar{\mathbf{V}}^T \quad (11c)$$

In the general case where both samples contain unique components, the solution space has to be derived from a matrix that is a combination of the unknown and calibration data matrices. Sánchez and Kowalski propose to decompose the sum matrix  $\mathbf{Q} = \mathbf{N} + \mathbf{M}$  and solve the standard eigenvalue problem with  $\mathbf{M}$  substituted for  $\mathbf{N}$ . This results in an eigenvalue matrix  $\mathbf{\Pi} = (\mathbf{C}_N + \mathbf{C}_M)^{-1}\mathbf{C}_M$  and a possible division by zero is avoided. With the necessary substitutions made, equations (10) and (11) describe the generalized rank annihilation method (GRAM).

### Generalization by Wilson *et al.*

Wilson *et al.*<sup>4</sup> devised a different algorithm for carrying out the rank annihilation. Equation (9) is converted to the usual generalized eigenvalue problem by approximating  $\mathbf{M}$  using

orthogonal bases  $\mathbf{F}$  and  $\mathbf{G}$  for the column and row space as follows:

$$\bar{\mathbf{M}} = \bar{\mathbf{F}}\bar{\mathbf{M}}_{\text{FG}}\bar{\mathbf{G}}^T \quad (12a)$$

$$\bar{\mathbf{N}} = \bar{\mathbf{F}}\bar{\mathbf{N}}_{\text{FG}}\bar{\mathbf{G}}^T \quad (12b)$$

If  $\mathbf{M}$  spans the space of  $\mathbf{N}$ , then  $\mathbf{F}$  and  $\mathbf{G}$  can be calculated from  $\mathbf{M}$ . In the general case it is recommended to calculate  $\mathbf{F}$  and  $\mathbf{G}$  from the column and row augmented matrices  $(\mathbf{N} | \mathbf{M})$  and  $(\bar{\mathbf{N}})$  respectively. Introducing  $\mathbf{Z}_G = \bar{\mathbf{G}}\mathbf{Z}$  leads to

$$\bar{\mathbf{N}}_{\text{FG}}\mathbf{Z}_G = \bar{\mathbf{M}}_{\text{FG}}\mathbf{Z}_G\mathbf{\Pi} \quad (13)$$

The eigenvalue matrix  $\mathbf{\Pi}$  is obtained from the QZ algorithm<sup>12,11</sup> in the form  $\pi_k = \alpha_k/\beta_k$ , where  $\alpha_k$  and  $\beta_k$  are scalars, possibly zero or near zero. The divisions in  $\pi_k = \alpha_k/\beta_k$  become the responsibility of the program's user.<sup>12</sup> The pure component responses are now reconstructed by

$$\mathbf{H} = \bar{\mathbf{F}}\bar{\mathbf{M}}_{\text{FG}}\mathbf{Z}_G \quad (14a)$$

$$\mathbf{Y}^T = (\mathbf{Z}_G)^{-1}\bar{\mathbf{G}}^T \quad (14b)$$

Combination of (14a) and (14b) gives equation (12a). The authors claim a better stability for their algorithm, since only unitary transformations are involved in solving equation (13).

### Alternative derivation

Assuming that the components present in the calibration sample are a subset of those present in the unknown, a standard eigenvalue problem similar to Lorber's eigenvalue problem is derived. (The generalization is obtained by replacing  $\mathbf{Q} = \mathbf{N} + \mathbf{M}$  for  $\mathbf{M}$  and  $\mathbf{M}$  for  $\mathbf{N}$ .) The derivation has as its main advantage that it is short and leads to simple reconstruction expressions for the pure component responses.

Premultiplying  $\mathbf{N}$  of equation (8b) by the pseudoinverse of  $\mathbf{H}$ , postmultiplying by the pseudoinverse of  $\mathbf{Y}^T$  and introducing  $\mathbf{H}^+ = (\bar{\mathbf{A}}\mathbf{T})^+ = \mathbf{T}^{-1}\bar{\mathbf{A}}^+$  and  $(\mathbf{Y}^T)^+ = (\mathbf{T}^{-1}\bar{\mathbf{B}}^T)^+ = \bar{\mathbf{B}}\mathbf{T}$  from equation (7) immediately leads to the standard eigenvalue problem\* in scores and loadings

$$\mathbf{H}^+\mathbf{N}(\mathbf{Y}^T)^+ = \mathbf{T}^{-1}(\bar{\mathbf{A}}^+\mathbf{N}\bar{\mathbf{B}})\mathbf{T} = \mathbf{\Pi} \quad (15)$$

that has already been derived by Öhman *et al.*<sup>13</sup> by manipulations similar to those employed by Lorber.<sup>2</sup> (It is interesting to note that Kubista<sup>14</sup> arrives at the transpose problem by correlating the data matrices by means of a Procrustes rotation.)

Substitution of  $\bar{\mathbf{A}}^+ = (\bar{\mathbf{U}}\bar{\mathbf{\Theta}})^+ = \bar{\mathbf{\Theta}}^{-1}\bar{\mathbf{U}}^T$  and  $\bar{\mathbf{B}} = \bar{\mathbf{V}}$  and premultiplication by  $\mathbf{T}$  yields

$$(\bar{\mathbf{\Theta}}^{-1}\bar{\mathbf{U}}^T\mathbf{N}\bar{\mathbf{V}})\mathbf{T} = \mathbf{\Pi}\mathbf{\Pi} \quad (16)$$

The matrices  $\mathbf{H}$  and  $\mathbf{Y}^T$  are now calculated from equation (7).

There is a remarkable difference between equation (16) and equation (10) with respect to the position of the matrix  $\bar{\mathbf{\Theta}}^{-1}$ . The origin of this difference lies in the attribution of the singular values to the score matrix according to equation (5), which is merely a conventional choice

\* In a subsequent paper we will show that the eigenvalues found by GRAM are biased estimates of the concentration ratios  $\mathbf{\Pi}$  in equation (8). Deriving an expression for the bias obviates the introduction of a new symbol for the estimated concentration ratio, e.g.  $\hat{\mathbf{\Pi}}$ . Making a distinction between the concentration ratio actually present,  $\mathbf{\Pi}$ , and the concentration ratio estimated from the response matrices,  $\hat{\mathbf{\Pi}}$ , is not necessary for the present discussion.

from the point of view of this paper. It is, however, well known from matrix algebra that premultiplication of a matrix by a diagonal matrix followed by postmultiplication by the inverse matrix constitutes a similarity transformation and therefore leaves the eigenvalues unchanged.<sup>15</sup> The eigenvectors  $\mathbf{T}$ ,  $\mathbf{Z}^*$  and  $\mathbf{Z}_G$  are related according to

$$\mathbf{T} = \bar{\Theta}^{-1} \mathbf{Z}^* = (\bar{\mathbf{G}}^T \bar{\mathbf{V}})^{-1} \mathbf{Z}_G \quad (17)$$

The matrix  $\mathbf{T}$  is found by multiplying the rows of  $\mathbf{Z}^*$  by the inverse singular values of  $\mathbf{M}$ .  $\mathbf{Z}_G$  reduces to  $\mathbf{T}$  if for the orthogonal bases  $\mathbf{F}$  and  $\mathbf{G}$  the singular vectors  $\mathbf{U}$  and  $\mathbf{V}$  of  $\mathbf{M}$  are taken. This is immediate after premultiplication of equation (9) by  $\bar{\Theta}^{-1} \bar{\mathbf{U}}^T$  and insertion of  $\mathbf{Z} = \bar{\mathbf{V}} \mathbf{Z}_G = \bar{\mathbf{V}} \mathbf{Z}_V$ . The manipulations are still very straightforward but the eigenvector matrix  $\mathbf{Z}_V$  is identified as the desired transformation matrix  $\mathbf{T}$  *a posteriori*. For the derivation of (16) the transformation matrix was the starting point.

### CHARACTERISTICS OF RANK ANNIHILATION

We have seen that when applying the method of rank annihilation, an eigenvalue problem has to be solved. This leads to difficulties that are characteristic for this method. It would be nice if these difficulties already become apparent from the formulation of the model. Stated differently: rather than deriving an eigenvalue problem and discussing the properties of this eigenvalue problem, it should become clear why the formulation of the calibration problem leads to an eigenvalue problem with all its inherent properties.

The eigenvalue matrix  $\mathbf{\Pi}$  enters the derivation by the transcription of equation (8a). It is seen that both (8a) and (8b) are misleading with respect to the information displayed. For the components not present in the calibration sample we have zeros on the diagonal of  $\mathbf{C}_N$  or  $\mathbf{\Pi}$  and the corresponding profiles in  $\mathbf{X}$ ,  $\mathbf{H}$  and  $\mathbf{Y}$  are in no way restricted to correlate with the profiles that reproduce  $\mathbf{M}$ . They may even be complex, since they do not contribute to the data anyway. Furthermore, it is apparent from equation (8b) that under certain circumstances some (non-zero) diagonal elements of  $\mathbf{\Pi}$  may be identical and in that case the corresponding profiles may be linear combinations. Therefore the characteristic difficulties of rank annihilation already show up from the alternative formulation of the model. How these difficulties must be encountered follows from the study of the eigenvalue problem.

#### Complex eigensolution

If the data follow the assumed linear additive model, i.e. equations (1) and (8) hold, the eigenvalues and eigenvectors corresponding to the calibrated components should be real. This is easily verified, since only the  $n$ th row and column of  $\mathbf{H}^+$  and  $(\mathbf{Y}^T)^+$  respectively are involved in the estimation of the  $n$ th eigenvalue in equation (15). This row and column are orthogonal to the remaining columns and rows in  $\mathbf{H}$  and  $\mathbf{Y}^T$  respectively, leading through equation (7) to a transformation vector  $\mathbf{t}_n$  that is real. Therefore the eigenvector matrix  $\mathbf{T}$  can be partitioned into a real and (possibly) complex part as  $(\mathbf{T}_{\text{real}} | \mathbf{T}_{\text{complex}})$  or, using the terminology of Liang *et al.*,<sup>16</sup> into a white and black part as  $(\mathbf{T}_{\text{white}} | \mathbf{T}_{\text{black}})$ , since the unknown background remains unresolved.

It is in fact logical that the calibration matrix can only provide information on the components that are present in that sample ( $c_{N,n} > 0$ ). The opposite must also be expected to hold: the components that are not present in the calibration sample (irrelevant part of the solution) cannot *misinform* about the components that are present in both samples (relevant part). For perfectly bilinear data the solution for the analytes of interest should always be real.

Two small problems remain. First, the *calculated* eigenvectors may be arbitrary to the extent of a complex multiplier of modulus one depending on the normalization used.<sup>15</sup> This problem is eliminated by normalizing the vectors in such a way that the largest component becomes one. It is easily verified that this solution amounts to the first similarity transformation of Li *et al.*<sup>17</sup> Second, for the reconstruction of **Y** the inverse of **T** must be calculated. It is easily verified that performing the calculations in complex arithmetic<sup>18</sup> will provide a real solution for the calibrated components. Two alternatives have been proposed in the chemometrical literature. The first alternative is to calculate **Y**<sup>T</sup> from the transpose problem (so replacing **N** and **M** by **N**<sup>T</sup> and **M**<sup>T</sup>) instead of using the inverse eigenvector matrix.<sup>13</sup> The second alternative is to convert the partially complex eigenvalue and eigenvector matrices to real matrices. This procedure amounts to the second similarity transformation of Li *et al.*<sup>17</sup> The three procedures should all lead to identical results for the analytes of interest.

Until now the discussion has been restricted to the analysis of model data. Model errors may cause the transformation vectors for the calibrated components to be complex as well.<sup>13</sup> Li and co-workers<sup>17,19</sup> have shown that in that case acceptable results can only be obtained if GRAM is modified by the second similarity transformation, thereby increasing the applicability of GRAM to very difficult practical situations.

### Degenerate eigensolution

If some of the eigenvalues are identical, the direction of the corresponding eigenvectors is not fixed. In fact, every linear combination of vectors belonging to the eigenspace of the degenerate eigenvalues forms a valid eigenvector and the eigenvector actually calculated by the computer will even depend on the algorithm used. As a result, part of the qualitative solution of GRAM is not unique. In this case it is useful to have more than one calibration sample in order to obtain the complete qualitative solution.<sup>20</sup> It is also possible to remove the degeneracy by imposing extra noise on the data matrices. We have good experience with this procedure for simulated as well as real data. Although essentially correct and easily implemented, the calibration by rank annihilation now becomes a matter of trial and error that we find difficult to recommend for routine purposes. Moreover, using different calibration samples has the additional advantage that the assumed linear model can be validated.<sup>21</sup> Finally, it is possible to subtract the resolved components and use a self-modeling curve resolution method on the residual matrices.

A special form of degeneracy is known as a defect eigensystem:<sup>15</sup>

$$\lim_{\beta \rightarrow \alpha} \begin{pmatrix} \alpha & 1 \\ 0 & \beta \end{pmatrix} = \begin{pmatrix} \alpha & 1 \\ 0 & \alpha \end{pmatrix}$$

In the limit of  $\beta \rightarrow \alpha$  the two eigenvalues are  $\lambda_1 = \lambda_2 = \alpha$  and the corresponding eigenvectors coincide:  $\mathbf{z}_1^T = \mathbf{z}_2^T = (1, 0)$ . The eigenspace is one-dimensional and the matrix cannot be reduced to diagonal form, since the eigenvector matrix is not invertible. Now the direction of the eigenvectors is fixed and the pure component responses can be recovered. (The numerical consequences of eigenvalues belonging to so-called quadratic elementary divisors are discussed in Reference 12.) This situation, which could be expected to result from severe collinearity in the data, will, however not occur. If components are undistinguishable within the noise level, this will result in a reduction of the number of factors found from PCA and the reconstructed component will simply be an average of the components that were too similar. The selectivity of the experiment must be improved in that case in order to obtain the complete quantitative solution (e.g. by employing bimodal data<sup>22</sup>). This situation applies to all PCA-based techniques and is therefore not typical for rank annihilation.

## CONCLUSIONS

We have presented a simple derivation of the standard eigenvalue problem that arises if the method of rank annihilation is used for calibration. The availability of two data matrices, obtained under identical circumstances, enables the direct evaluation of the matrix that transforms the abstract decomposition of a matrix into the desired physical one. In this way a relationship is established with other PCA-based curve resolution and calibration methods. The transformation matrix found by rank annihilation may consist of a real and (possibly) complex part if some of the components in the unknown sample are not calibrated. However, in the absence of model errors the relevant part of the transformation matrix is not affected by the irrelevant part. In the presence of model errors (e.g. irreproducibility of chromatographic data) the relevant part of the eigensolution may also become complex.<sup>13,17,19</sup> In that case it is mandatory to follow up the solution of the eigenvalue problem with unitary transformations as recommended in References 17 and 19. It remains an important issue as to how the accuracy and precision of the transformed solution must be estimated.

## REFERENCES

1. C.-N. Ho, G. D. Christian and E. R. Davidson, *Anal. Chem.* **50**, 1108 (1978).
2. A. Lorber, *Anal. Chim. Acta*, **164**, 293 (1984).
3. E. Sánchez and B. R. Kowalski, *Anal. Chem.* **58**, 496 (1986).
4. B. E. Wilson, E. Sánchez and B. R. Kowalski, *J. Chemometrics*, **3**, 493 (1989).
5. E. R. Malinowski and D. G. Howery, *Factor Analysis in Chemistry*, Wiley, New York (1991).
6. W. Windig, *Chemometrics Intell. Lab. Syst.* **16**, 1 (1992).
7. B. G. M. Vandeginste, F. Leyten, M. Gerritsen, J. W. Noor and G. Kateman, *J. Chemometrics*, **1**, 57 (1987).
8. H. Gampp, M. Maeder, C. J. Meyer and A. D. Zuberbuhler, *Talanta*, **32**, 1133 (1985).
9. E. R. Malinowski, *J. Chemometrics*, **6**, 29 (1992).
10. N. M. Faber, L. M. C. Buydens and G. Kateman, *J. Chemometrics*, **7**, 495 (1993).
11. IMSL, International Mathematical and Statistical Libraries, Inc., 7500 Bellaire Blvd., Houston, TX 77036, U.S.A.
12. C. B. Moler and G. W. Stewart, *SIAM J. Numer. Anal.* **10**, 241 (1973).
13. J. Öhman, P. Geladi and S. Wold, *J. Chemometrics*, **4**, 135 (1990).
14. M. Kubista, *Chemometrics Intell. Lab. Syst.* **7**, 273 (1990).
15. J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon, Oxford (1965).
16. Y.-Z. Liang, O. M. Kvalheim and R. Manne, *Chemometrics Intell. Lab. Syst.* **18**, 235 (1993).
17. S. Li, J. C. Hamilton and P. J. Gemperline, *Anal. Chem.* **64**, 599 (1992).
18. J. R. Westlake, *A Handbook of Numerical Matrix Inversion and Solution of Linear Equations*, Wiley, New York (1968).
19. S. Li and P. J. Gemperline, *J. Chemometrics*, **7**, 77 (1993).
20. E. Sánchez and B. R. Kowalski, *J. Chemometrics*, **2**, 265 (1988).
21. A. Lorber, *Anal. Chem.* **57**, 2395 (1985).
22. L. S. Ramos, E. Sánchez and B. R. Kowalski, *J. Chromatogr.* **385**, 165 (1987).