ELSEVIER

Chemometrics and Intelligent Laboratory Systems 63 (2002) 107–116

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

Second-order bilinear calibration: the effects of vectorising the data matrices of the calibration set

Nicolaas (Klaas) M. Faber [a,*], Joan Ferré [b], Ricard Boqué [b], John H. Kalivas [c]

[a] Department Production and Control Systems, ATO, P.O. Box 17, 6700 AA Wageningen, The Netherlands
[b] Department of Analytical and Organic Chemistry, Institute of Advanced Studies, Rovira I Virgili University, Pça. Imperial Tàrraco 1, E-43005 Tarragona, Catalonia, Spain
[c] Department of Chemistry, Idaho State University, Pocatello, ID 83209, USA

Accepted 12 March 2002

Abstract

In a groundbreaking paper, Linder and Sundberg [Chemometr. Intell. Lab. Syst. 42 (1998) 159] developed a statistical framework for the calibration of second-order bilinear data. Within this framework, they formulated three different predictor construction methods [J. Chemom. 16 (2002) 12], namely the so-called naïve method, the bilinear least squares (BLLS) method, and a refined version of the latter that takes account of the calibration uncertainty. Elsewhere [J. Chemom. 15 (2001) 743], a close relationship is established between the naïve method and the generalized rank annihilation method (GRAM) by comparing expressions for prediction variance. Here it is proved that the BLLS method can be interpreted to work with vectorised data matrices, which establishes an algebraic relationship with so-called unfold partial least squares (PLS) and unfold principal component regression (PCR). It is detailed how these results enable quantifying the effects of vectorising bilinear second-order data matrices on analytical figures of merit and variance inflation factors.

Keywords: Second-order bilinear calibration; PARAFAC; Vectorisation; Analytical figures of merit; Variance inflation factors; BLLS; PLS; PCR; GRAM

1. Introduction

Vectorisation, i.e., stringing out an *N*-way array into a vector is a routine operation in multiway data analysis [1]. For example, the so-called unfold partial least squares (PLS) and unfold principal component regression (PCR) transform the *N*-way data arrays of the calibration set into vectors that are subjected to standard PLS and PCR, respectively [2]. In contrast, the generalized rank annihilation method (GRAM) is a second-order bilinear calibration method that operates on the original data matrices [3].

The purpose of this paper is to investigate the effects of vectorising the data matrices of the calibration set under the second-order bilinear assumption. These effects are to be studied for performance characteristics such as analytical figures of merit and variance inflation factors, because these quantities are related to the standard error of prediction. Understanding the effect of vectorisation on these performance characteristics should be useful for understanding the behaviour of calibration methods. In this way, one expects to reveal essential differences between unfold-PLS and unfold-PCR on one side (vectorised matri-

* Corresponding author.
E-mail address: n.m.faber@ato.wag-ur.nl (N.M. Faber).

0169-7439/02/$ - see front matter
PII: S0169-7439(02)00018-7

ces), and GRAM on the other (original matrices). GRAM has been claimed to be inferior to alternating least squares (ALS) [4]. However, Monte Carlo simulations [5] have shown that GRAM not only compares well with ALS, but also with recently proposed alternatives to ALS, namely alternating trilinear decomposition, alternating coupled vectors resolution, alternating slice-wise diagonalization, alternating coupled matrices resolution, self-weighted alternating trilinear decomposition, and pseudo alternating least squares. These results justify including GRAM in the comparison.

Analytical figures of merit are performance characteristics of an analytical determination. They can be used to select between potentially useful methods or to evaluate or optimise a method that is already in use [6]. Analytical figures of merit such as net analyte signal (NAS), sensitivity, selectivity, signal-to-noise ratio (SNR), and limit of detection (LOD) are well defined for univariate or zeroth-order calibration. Lorber [7] has extended these figures of merit to the multivariate or first-order domain. The resulting quantities are based on a particular definition of NAS. According to Lorber, "net analyte signal for a component is equal to the part of its spectrum which is orthogonal to the spectra of the other components". This definition is based on a uniqueness concept: "Only the orthogonal part is unique to the sought-for component and, therefore, useful for quantitation". Interestingly, Morgan [8] presents equations for projections to essentially compute the NAS vector. This NAS vector is transformed into a scalar NAS by taking the Euclidean norm and the first-order figures of merit follow by inserting the scalar NAS in the formulas for the zeroth-order analogues. For example, selectivity is defined as the ratio of the NAS and the total analyte signal. It ranges between zero (complete overlap between analyte of interest and interferents) and unity (no overlap). Sensitivity is defined as the ratio of the NAS and the analyte concentration. This figure of merit is encountered in the less common classical model, where instrument responses are expressed as a function of analyte concentrations, while the 'inverse sensitivity' (size of the regression vector) is defined for the inverse model [9], where the roles of analyte concentrations and instrument responses are interchanged. Kalivas and Lang [10,11] have discussed interrelationships between Lorber's first-order analytical figures of merit and the variance inflation factors known from statistics. The latter quantities are often used to describe the stability of a system of coupled equations.

Ho et al. [12], Wang et al. [13], and Messick et al. [14] have proposed generalizations of the figures of merit to higher-order multilinear (i.e., PARAFAC structure) data arrays. These generalizations have been compared by Faber et al. [15]. Two major conclusions from that work are that (1) the generalization due to Wang et al. [13] is less useful because it does not account for the influence of interferents, and (2) the generalization due to Messick et al. [14] amounts to vectorising the data arrays. A direct consequence is that GRAM is connected with the generalization of Ho et al. [12], while the generalization proposed by Messick et al. [14] is appropriate for unfold-PLS and unfold-PCR.

Linder and Sundberg [16,17] have recently introduced a statistical framework for second-order bilinear calibration. This framework is of considerable interest because it aims at a low standard error of prediction while taking account of the special structure of the data. (The latter property sets it somewhat aside from unfold-PLS and unfold-PCR.) It centres on three approaches, namely the so-called naïve, bilinear least squares (BLLS), and refined approach. The first one is inferior while the latter two perform similarly. In fact, the BLLS and refined predictors are identical when calibration is error-free: in that ideal scenario they are best (minimum variance) linear unbiased (BLUE) under the assumption of homoscedastic white noise in the prediction sample responses. GRAM strongly resembles the naïve approach with respect to standard error of prediction [18]. Despite its poor statistical efficiency, GRAM is a viable method because it works with a single calibration sample. Moreover, by constructing a common model for the calibration and unknown sample, it achieves the so-called second-order advantage. The second-order advantage may be essential when analysing complex samples, because it ensures correct prediction in the presence of unknown interferents. In contrast, the BLLS and refined predictors require at least as many calibration samples as constituents and do not have the second-order advantage, see Ref. [18] for further discussion. Here it is shown that the BLLS approach can be interpreted to work with vectorised data

matrices. In this way, an algebraic relationship is established with unfold-PLS and unfold-PCR. Owing to these connections with GRAM, unfold-PLS, and unfold-PCR, the framework of Linder and Sundberg provides a formal basis for discussing the effect of vectorising bilinear second-order data matrices on before mentioned performance characteristics.

This paper focuses on analytical figures of merit that have been shown to directly relate to standard error of prediction, i.e., 'inverse sensitivity' [9] and sensitivity (the reciprocal of the 'inverse sensitivity') [15]. The current analysis is restricted to the calibration of second-order bilinear data matrices. However, the generalization to *N*th-order multilinear data arrays is straightforward, since one may vectorise along each individual order, independent of the others.

2. Theory

2.1. Preliminaries

2.1.1. Vectorisation operator

The vectorisation operator converts a matrix in a vector by stacking the columns from left to right. An extremely useful relationship for performing algebraic manipulations involving the vectorisation operator is [1]:

$$\text{vec } \mathbf{FGH} = (\mathbf{H}^{\mathrm{T}} \otimes \mathbf{F})\text{vec } \mathbf{G} \tag{1}$$

where vec stands for vectorisation, **F**, **G**, and **H** are general matrices, the superscript 'T' denotes matrix transposition, and '⊗' symbolises the Kronecker product. This relationship holds whenever the product **FGH** is defined. It simplifies in straightforward fashion when scalars or vectors are involved.

2.1.2. Second-order bilinear model

The bilinear assumption implies that the errorless data matrix of a pure analyte can be written as the outer product of two vectors. Thus, the model equation for the measured $J \times K$ data matrix for an M-component mixture is

$$\mathbf{X} = \sum_{m=1}^{M} y_m \mathbf{a}_m \mathbf{b}_m^{\mathrm{T}} + \Delta\mathbf{X} = \mathbf{AYB}^{\mathrm{T}} + \Delta\mathbf{X} \tag{2}$$

where **A** ($J \times M$) and **B** ($K \times M$) are matrices that contain the column and row profiles at unit concentrations, $\{\mathbf{a}_m\}$ and $\{\mathbf{b}_m\}$, respectively, **Y** is the $M \times M$ diagonal matrix of concentrations $\{y_m\}$, and $\Delta\mathbf{X}$ is a $J \times K$ error matrix.

2.1.3. Second-order bilinear model and vectorisation

Applying Eq. (1) to **X** gives, after rearrangement

$$\mathbf{x} = \sum_{m=1}^{M} y_m(\mathbf{b}_m \otimes \mathbf{a}_m) + \Delta\mathbf{x} = \sum_{m=1}^{M} y_m \mathbf{d}_m + \Delta\mathbf{x} = \mathbf{Dy} + \Delta\mathbf{x} \tag{3}$$

where $\mathbf{x} = \text{vec } \mathbf{X}$, $\Delta\mathbf{x} = \text{vec } \Delta\mathbf{X}$, **D** contains the pure analyte 'pseudo first-order' profiles at unit concentration, $\{\mathbf{d}_m = \mathbf{b}_m \otimes \mathbf{a}_m\}$, and $\mathbf{y} = \text{diag}(\mathbf{Y})$. Eq. (3) is relevant in the context of unfold-PLS and unfold-PCR, because these methods amount to applying the standard algorithms after vectorising the data matrices of the calibration set.

2.2. Crude expression for standard error of prediction

The goal of any instrument calibration is to predict a property of interest, e.g., analyte concentration, for an unknown sample. This is achieved by constructing a model from data obtained for a calibration set. Two modes of calibration are possible, namely by constructing a classical or an inverse model. The inverse model is more flexible, because it allows one, among others, to predict individual analytes without requiring the explicit modelling of interferences. We will carry out the derivations within the inverse model formalism and clarify the connection with the classical model afterwards.

The parameters of an inverse model are collected in a regression vector and focussing on the analyte of interest (index m), prediction proceeds as

$$\hat{y}_m = \hat{\beta}_m^{\mathrm{T}} \mathbf{x} \tag{4}$$

where $\hat{\beta}_m$ is the estimated regression vector.

From Eq. (4), it is clear that the standard error of prediction must have two contributions, namely one from the regression vector estimate ($\hat{\beta}_m$) and one from the prediction sample responses (**x**). Assuming that the prediction sample is not an extreme outlier, the contribution of the regression vector estimate will be

relatively small. When ignoring the contribution of the regression vector estimate, Eq. (4) yields

$$\sigma_{\hat{y}_m} \approx \| \hat{\beta}_m \| \sigma_X \tag{5}$$

where $\sigma_{\hat{y}_m}$ denotes the standard error in the prediction $\hat{y}_m$ (square root of the variance), $\| \beta \|^2 = \beta^T \beta$, and σ_X is the standard deviation of the measurement errors in the prediction sample responses. These measurement errors are assumed to be independently and identically distributed (iid).

Eq. (5) illustrates that the length of the regression vector estimate is a performance characteristic for the analytical determination. It has been termed the 'inverse sensitivity' [9], because it is the reciprocal of the sensitivity defined for the classical model. It can be interpreted as the square root of a variance inflation factor, because $\| \hat{\beta}_m \|^2$ quantifies the inflation of the input variance σ_X^2 when considering the prediction $\hat{y}_m$. Berger and Feld [19] have reported on a Raman application where Eq. (5) was an excellent approximation of the true standard error of prediction, while García et al. [20] have calculated limits of detection for beta emitter (^{14}C) activity from liquid scintillation-counting data that were consistent with Eq. (5). Although not being concerned with the calibration of second-order bilinear data, these studies demonstrate that Eq. (5) can be adequate for real applications.

It is worth mentioning that the iid requirement is not necessary for deriving expressions for standard error of prediction. In fact, the 'iid expressions' result from simplifying expressions that completely account for heteroscedastic and correlated measurement errors [9]. Moreover, Eq. (5) may still be useful for semiquantitative understanding if the measurement noise is not iid. Monte Carlo simulations can be used to investigate how the impact of heteroscedastic noise relates to simply multiplying the length of the regression vector with the average standard deviation of the measurement errors. For the GRAM analysis of the data obtained by monitoring a chemical reaction by UV–VIS spectroscopy, it was found that the impact was higher by approximately 10–30%, see last two columns in Table 5 in Ref. [21]. It is reasonable to assume that vectorising the arrays has little influence on the increased impact due to heteroscedasticity. In other words, this increase can be considered to be a constant factor in the comparison.

The preceding considerations have led to the following strategy: first, we derive the appropriate expressions for the regression vector estimates (inverse model). Next, the effect of vectorisation on the 'inverse sensitivity' (size of the regression vector estimate) is determined, from which the effect on sensitivity (classical model) is immediate. Finally, this result is translated to the effect on variance inflation factors.

2.3. Statistical framework for second-order bilinear calibration [16,17]

2.3.1. Naïve approach

An intuitive way to obtain a prediction of the concentrations $\{y_m\}$ is to pre- and post-multiply Eq. (2) with the Moore-Penrose pseudo-inverses of $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$. This procedure has been termed the naïve approach by Linder and Sundberg [16]. The reason for this terminology (i.e., naïve) is that it leads to a predicted matrix, $\hat{\mathbf{Y}}$, which unlike the (true) $\mathbf{Y}$ need not be diagonal. Clearly, the naïve approach is identical to the *unconstrained* least squares (LS) solution [22], because the latter also ignores the special form of $\mathbf{Y}$. For obtaining $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$, Linder and Sundberg [16] developed two methods, namely BLLS and the so-called singular value decomposition (SVD) estimator.

The naïve approach yields the prediction

$$\hat{y}_{\text{naïve},m} = \hat{\mathbf{a}}_m^{+\mathrm{T}} \mathbf{X} \hat{\mathbf{b}}_m^{+} = (\hat{\mathbf{b}}_m^{+} \otimes \hat{\mathbf{a}}_m^{+})^{\mathrm{T}} \mathbf{x} \tag{6}$$

where $\hat{\mathbf{a}}_m^{+}$ and $\hat{\mathbf{b}}_m^{+}$ denote the associated columns of $\hat{\mathbf{A}}^{+\mathrm{T}}$ and $\hat{\mathbf{B}}^{+\mathrm{T}}$, respectively (the superscript '+' symbolises the Moore-Penrose pseudo-inverse). The first step requires $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ to be full column rank matrices, whereas the second step in Eq. (6) follows from applying Eq. (1).

A note seems to be in order with respect to the case where some of the columns of $\mathbf{A}$ and $\mathbf{B}$ are completely overlapped with the remaining ones. This situation is sometimes referred to as rank overlap. It is easily verified that the naïve approach is still feasible if the analyte of interest is not involved in a rank overlap. One simply replaces in Eq. (2) the profiles of the components that are involved in a rank overlap by a correspondingly smaller one, and adapts $\mathbf{Y}$ accordingly. This can be done without changing the notation. (For example, M now denotes the number of components minus the number of rank overlaps.)

By comparing Eqs. (6) and (4), the naïve regression vector is identified as

$$\hat{\beta}_{\text{naïve},m} = \hat{\mathbf{b}}_m^+ \otimes \hat{\mathbf{a}}_m^+ \tag{7}$$

It is shown elsewhere [18] that there is a close connection between the naïve approach and GRAM: both methods yield predictions by pre- and post-multiplying the data matrix of the prediction sample with the pseudo-inverses of $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$. A major difference, however, is that GRAM does not estimate $\mathbf{A}$ and $\mathbf{B}$ from the calibration set data. Instead, the required estimates are obtained by constructing a joint model for the calibration and prediction sample. One of the consequences is that the predicted matrix $\hat{\mathbf{Y}}$ is diagonal.

2.3.2. BLLS approach

As an improvement over the naïve approach, Linder and Sundberg [17] proposed the BLLS predictor. Focussing on the analyte of interest, this procedure yields

$$\hat{y}_{\text{BLLS},m} = [(\hat{\mathbf{A}}^{\mathrm{T}}\hat{\mathbf{A}} * \hat{\mathbf{B}}^{\mathrm{T}}\hat{\mathbf{B}})^{-1}\ \text{diag}(\hat{\mathbf{A}}^{\mathrm{T}}\mathbf{X}\hat{\mathbf{B}})]_m \tag{8}$$

where '*' symbolises the element-wise Hadamard product. Without errors in $\mathbf{A}$ and $\mathbf{B}$, the BLLS predictor is BLUE under the assumption that $\Delta\mathbf{X}$ contains homoscedastic white noise.

The BLLS predictor differs from the naïve approach in that it takes account of the special form of $\mathbf{Y}$. Stated differently: the BLLS approach is a so-called *constrained* LS solution [22], where the constraint is that $\mathbf{Y}$ be diagonal. It is clear from Eq. (3) that vectorising $\mathbf{X}$ effectively eliminates the zeros from $\mathbf{Y}$. Thus, unfold-PLS and unfold-PCR respect the special form of $\mathbf{Y}$ too. Moreover, Eq. (3) implies that the unfold-PLS and unfold-PCR regression vectors estimate the column of $\mathbf{D}^{+\mathrm{T}}$ that is associated with the analyte of interest (index m). Consequently, one expects that Eq. (8) can be rewritten in a form akin to that of prediction using unfold-PLS and unfold-PCR. The proof of this conjecture is straightforward. Using [1]

$$\hat{\mathbf{D}}^{\mathrm{T}}\hat{\mathbf{D}} = [\hat{\mathbf{d}}_1; \cdots, \hat{\mathbf{d}}_M]^{\mathrm{T}}[\hat{\mathbf{d}}_1; \cdots, \hat{\mathbf{d}}_M]$$

$$= [\hat{\mathbf{b}}_1 \otimes \hat{\mathbf{a}}_1; \cdots, \hat{\mathbf{b}}_M \otimes \hat{\mathbf{a}}_M]^{\mathrm{T}}[\hat{\mathbf{b}}_1 \otimes \hat{\mathbf{a}}_1; \cdots, \hat{\mathbf{b}}_M \otimes \hat{\mathbf{a}}_M]$$

$$= \begin{pmatrix} \hat{\mathbf{b}}_1^{\mathrm{T}}\hat{\mathbf{b}}_1 \cdot \hat{\mathbf{a}}_1^{\mathrm{T}}\hat{\mathbf{a}}_1 & \cdots & \hat{\mathbf{b}}_1^{\mathrm{T}}\hat{\mathbf{b}}_M \cdot \hat{\mathbf{a}}_1^{\mathrm{T}}\hat{\mathbf{a}}_M \\ \vdots & \ddots & \vdots \\ \hat{\mathbf{b}}_M^{\mathrm{T}}\mathbf{b}_1 \cdot \hat{\mathbf{a}}_M^{\mathrm{T}}\hat{\mathbf{a}}_1 & \cdots & \hat{\mathbf{b}}_M^{\mathrm{T}}\hat{\mathbf{b}}_M \cdot \hat{\mathbf{a}}_M^{\mathrm{T}}\hat{\mathbf{a}}_M \end{pmatrix}$$

$$= \hat{\mathbf{A}}^{\mathrm{T}}\hat{\mathbf{A}} * \hat{\mathbf{B}}^{\mathrm{T}}\hat{\mathbf{B}} \tag{9}$$

and

$$\hat{\mathbf{a}}_m^{\mathrm{T}}\mathbf{X}\hat{\mathbf{b}}_m = (\hat{\mathbf{b}}_m^{\mathrm{T}} \otimes \hat{\mathbf{a}}_m^{\mathrm{T}})\mathbf{x} = \hat{\mathbf{d}}_m^{\mathrm{T}}\mathbf{x} \tag{10}$$

Eq. (8) simplifies to

$$\hat{y}_{\text{BLLS},m} = \hat{\mathbf{d}}_m^{+\mathrm{T}}\mathbf{x} \tag{11}$$

where $\hat{\mathbf{d}}_m^+$ is the associated column of $\hat{\boldsymbol{D}}^{+\mathrm{T}}$. By comparing Eqs. (11) and (4), the BLLS regression vector is identified as

$$\hat{\beta}_{\text{BLLS},m} = \hat{\mathbf{d}}_m^+ \tag{12}$$

which establishes an algebraic relationship with unfold-PLS and unfold-PCR. The setup of the data only (matrix or vector form) determines this relationship. Clearly, differences arise because of the criterion used for estimation of $\hat{\beta}_m$. Unfold-PCR generates scores that maximise the explained variance of the calibration set response matrices. Subsequently, these scores are regressed onto the analyte concentration using ordinary least squares (OLS) to generate a model. Similarly, unfold-PLS scores have maximum covariance with the concentration vector subject to being orthogonal to each other. In contrast, the BLLS parameters result from fitting the calibration set data matrices in the LS sense. It is important to note that this LS fit is different from the one obtained using the common alternating least squares (ALS) algorithm, because the analyte concentrations are assumed to be known. The BLLS method may have an advantage over unfold-PLS and unfold-PCR, because the latter do not utilise the bilinear structure of the instrument data. (For example, unfold-PLS and unfold-PCR do not yield estimates of the profile matrices $\mathbf{A}$ and $\mathbf{B}$.) This conjecture is an interesting subject for further study.

2.4. Effects of vectorisation

2.4.1. Effect on analytical figures of merit

For convenience, the effect of vectorisation is derived in a relative form. Using $\|\hat{\beta}_{\text{naïve},m}\| = \left(\left(\hat{\mathbf{b}}_m^+ \otimes \hat{\mathbf{a}}_m^+\right)^{\mathrm{T}} \left(\hat{\mathbf{b}}_m^+ \otimes \hat{\mathbf{a}}_m^+\right)\right)^{1/2} = \left(\hat{\mathbf{b}}_m^{+\mathrm{T}} \hat{\mathbf{b}}_m^+ \cdot \hat{\mathbf{a}}_m^{+\mathrm{T}} \hat{\mathbf{a}}_m^+\right)^{1/2} = \|\hat{\mathbf{a}}_m^+\| \cdot \|\hat{\mathbf{b}}_m^+\|$ gives

$$\frac{\|\hat{\beta}_{\text{BLLS},m}\|}{\|\hat{\beta}_{\text{naïve},m}\|} = \frac{\|\hat{\mathbf{d}}_m^+\|}{\|\hat{\mathbf{a}}_m^+\| \cdot \|\hat{\mathbf{b}}_m^+\|} \tag{13}$$

which is rewritten by introducing the scalar NAS as

$$\frac{\|\hat{\beta}_{\text{BLLS},m}\|}{\|\hat{\beta}_{\text{naïve},m}\|} = \frac{\|\hat{\mathbf{a}}_m^*\| \cdot \|\hat{\mathbf{b}}_m^*\|}{\|\hat{\mathbf{d}}_m^*\|} \tag{14}$$

where $\hat{\mathbf{a}}_m^*$, $\hat{\mathbf{b}}_m^*$, and $\hat{\mathbf{d}}_m^*$ are the net analyte contributions to $\hat{\mathbf{a}}_m$, $\hat{\mathbf{b}}_m$, and $\hat{\mathbf{d}}_m$, respectively. Lorber [7] has already given their calculation from the pseudo-inverse matrices, e.g., $\|\hat{\mathbf{a}}_m^*\| = \|\hat{\mathbf{a}}_m^+\|^{-1}$. Dividing the numerator and denominator in Eq. (14) by $\|\hat{\mathbf{d}}_m\| = \|\hat{\mathbf{a}}_m\| \cdot \|\hat{\mathbf{b}}_m\|$ yields that

$$\begin{aligned}\frac{\|\hat{\beta}_{\text{BLLS},m}\|}{\|\hat{\beta}_{\text{naïve},m}\|} &= \frac{(\|\hat{\mathbf{a}}_m^*\| / \|\hat{\mathbf{a}}_m\|)(\|\hat{\mathbf{b}}_m^*\| / \|\hat{\mathbf{b}}_m\|)}{\|\hat{\mathbf{d}}_m^*\| / \|\hat{\mathbf{d}}_m\|} \\ &= \frac{\hat{\xi}_{\text{A},m} \cdot \hat{\xi}_{\text{B},m}}{\hat{\xi}_{\text{D},m}}\end{aligned} \tag{15}$$

where $\xi \leq 1$ denotes Lorber's selectivity [7] and the subscript refers to the associated profile.

Eq. (15) enables an interpretation of the relative size in terms of the selectivities. The higher efficiency of the BLLS approach, i.e., the smaller amount of error propagation according to Eq. (5), is proved as follows. Messick et al. [14] have shown that

$$\hat{\xi}_{\text{D},m} \geq \max(\hat{\xi}_{\text{A},m}, \hat{\xi}_{\text{B},m}) \tag{16}$$

with equality holding if and only if *at least one* of the first-order selectivities ($\hat{\xi}_{\text{A},m}$ or $\hat{\xi}_{\text{B},m}$) equals unity. Inserting inequality (16) in Eq. (15) results in

$$\begin{aligned}\frac{\|\hat{\beta}_{\text{BLLS},m}\|}{\|\hat{\beta}_{\text{naïve},m}\|} &\leq \frac{\hat{\xi}_{\text{A},m} \cdot \hat{\xi}_{\text{B},m}}{\max(\hat{\xi}_{\text{A},m}, \hat{\xi}_{\text{B},m})} \\ &= \min(\hat{\xi}_{\text{A},m}, \hat{\xi}_{\text{B},m})\end{aligned} \tag{17}$$

which, using $\hat{\xi} \leq 1$, implies that

$$\|\hat{\beta}_{\text{BLLS},m}\| \leq \|\hat{\beta}_{\text{naïve},m}\| \tag{18}$$

with equality holding if and only if *both* first-order selectivities equal unity.

The connection with the classical model is established by observing that the 'inverse sensitivity' is the reciprocal of the sensitivity. It is immediate that

$$s_{\text{BLLS},m} \geq s_{\text{naïve},m} \tag{19}$$

Thus, the BLLS approach is more sensitive than the naïve one. It is emphasised that sensitivity with respect to changes in analyte concentration is meant here, not sensitivity of the model parameters and predictions with respect to predictor noise. In contrast, the opposite is inferred from Eq. (5). For the other figures of merit, namely NAS, SNR, and LOD, similar inequalities can be derived. Clearly, a higher sensitivity implies a higher NAS as well as SNR and a lower LOD. It is interesting to note that the early work of Appellof and Davidson [23] already implies that vectorisation improves the analytical figures of merit. However, they do not give a proof of this property. Finally, an implication of the current work is that recent advances with respect to calculation of first-order figures of merit [24] have wide applicability to the multiway domain.

2.4.2. Effect on variance inflation factors

Linder and Sundberg [17] have proved that the BLLS predictor is more efficient than its naïve counterpart. In other words, the variance inflation factors should decrease upon vectorising the data matrices of the calibration set. However, the amount by which the variance inflation factors decrease is not clear. The purpose of the current derivations is to take the work of Linder and Sundberg one step further by quantifying this decrease in terms of characteristics that are easily determined from the data.

Using previously defined quantities, it is straightforward to rewrite the two sides of inequality (18) in terms of variance inflation factors. For the naïve approach,

$$\begin{aligned}\hat{\mathrm{VIF}}_{\text{naïve},m} &\equiv \|\hat{\beta}_{\text{naïve},m}\|^2 = \|\hat{\mathbf{a}}_m^+\|^2 \|\hat{\mathbf{b}}_m^+\|^2 \\ &= (\hat{\mathbf{A}}^{\mathrm{T}}\hat{\mathbf{A}})^{-1}_{mm} (\hat{\mathbf{B}}^{\mathrm{T}}\hat{\mathbf{B}})^{-1}_{mm}\end{aligned} \tag{20}$$

Similarly, using Eq. (9),

$$\widehat{\mathrm{VIF}}_{\mathrm{BLLS},m} \equiv \| \hat{\beta}_{\mathrm{BLLS},m} \|^2 = \| \hat{\mathbf{d}}_m^+ \|^2$$

$$= (\hat{\mathbf{D}}^{\mathrm{T}}\hat{\mathbf{D}})^{-1}_{mm} = (\hat{\mathbf{A}}^{\mathrm{T}}\hat{\mathbf{A}} * \hat{\mathbf{B}}^{\mathrm{T}}\hat{\mathbf{B}})^{-1}_{mm} \qquad (21)$$

Combining Eqs. (18), (20), and (21) yields that

$$\widehat{\mathrm{VIF}}_{\mathrm{BLLS},m} \leq \widehat{\mathrm{VIF}}_{\mathrm{naïve},m} \qquad (22)$$

The quantities in inequality (22) can be seen as the extensions of the variance inflation factors defined for the first-order calibration model. In particular, the variance inflation factor for the naïve approach is the product of the variance inflation factors associated with the individual modes. This result should be contrasted with the form obtained for the BLLS approach: now the invertibility of $\hat{\mathbf{D}}^{\mathrm{T}}\hat{\mathbf{D}} = \hat{\mathbf{A}}^{\mathrm{T}}\hat{\mathbf{A}} * \hat{\mathbf{B}}^{\mathrm{T}}\hat{\mathbf{B}}$ counts, rather than the invertibility of $\hat{\mathbf{A}}^{\mathrm{T}}\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}^{\mathrm{T}}\hat{\mathbf{B}}$ separately.

Finally, the generalization of inequality (22) to *N*th-order multilinear data is obtained by replacing the single product by a multiple product (in which the order of the factors is immaterial).

3. Illustrative examples

We have conducted a small simulation study with noiseless predictor arrays to illustrate the potential effect of vectorisation. This effect is quantified in terms of the relative efficiency of the naïve approach (η), which is obtained by inserting the true column and row profiles in Eq. (15).

3.1. Extreme overlap in one of the modes

Inspection of Eq. (15) shows that a *low* relative efficiency of the naïve approach results if one of the first-order selectivities is small. The reason for this is that the signal in a highly overlapping mode does not contain additional information, hence it will hardly contribute to the overall second-order selectivity ξ_{D}. Consequently, ξ_{D} must approach the first-order selectivity of the other mode from above. Assuming almost 'degenerate' B-mode profiles yields that $\eta \approx \xi_{\mathrm{D}}$. With these considerations in mind, a three-component system is simulated by multiplying moderately overlapped Gaussian elution profiles by the extremely overlapped ultraviolet (UV) spectra of the proteins myoglobin, α-chymotrypsin and carbon anhydrase [25], see Fig. 1 ($J = K = 50$).

3.2. Moderate overlap in both modes

This example is intended to yield a more favourable situation for methods such as GRAM, because the overlap in the separate modes is less extreme. Fig. 2 shows the excitation and emission fluorescence spectra of the amino acids tryptophan, tyrosine and phenylalanine ($J = 61$ and $K = 201$). These spectra were obtained from a three-factor PARAFAC model applied to the five samples contained in the data set

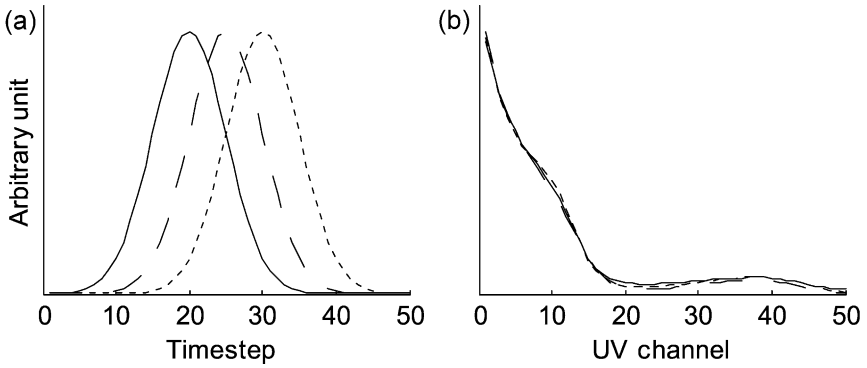

Fig. 1. (a) Simulated elution profiles and (b) experimental UV spectra for myoglobin (—), α-chymotrypsin (- - -), and carbon anhydrase (. . .).

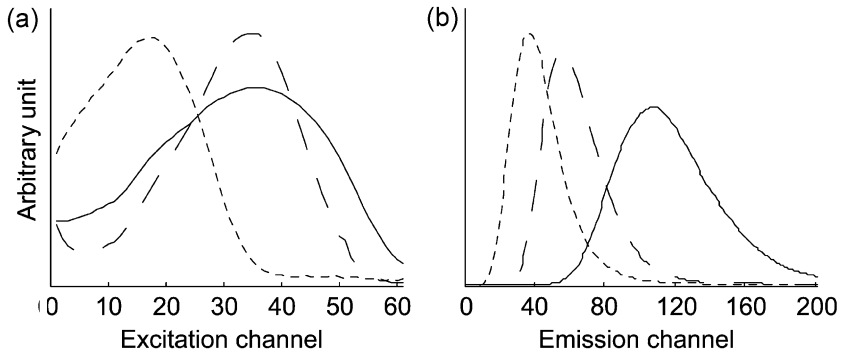

Fig. 2. (a) Excitation and (b) emission spectra for tryptophan (—), tyrosine (- - -), and phenylalanine (. . .).

claus.mat, which is available in the *N*-way Toolbox for Matlab [26].

4. Results and discussion

4.1. Extreme overlap in one of the modes

The characterisation of the elution profiles (columns of **A**) and spectra (columns of **B**) as 'moderately' and 'extremely' overlapped can be motivated from their first-order selectivities (ξ_A and ξ_B, see Table 1). For example, a value of 0.0282 for the UV spectrum of myoglobin is equivalent to 2.82% of the analyte spectrum being orthogonal to the space spanned by the spectra of α-chymotrypsin and carbon anhydrase [7]. Calibration of the resulting data matrices using GRAM may be troublesome, because the results depend on the overlap in the individual modes. The true parameter vector is constrained not to pick up signal from the interferences, since this would lead to bias. As a result, it must be orthogonal to the space spanned by the interferences. It follows that only 2.82% of the UV spectrum of myoglobin contributes to the GRAM model. It depends on the amount of predictor noise whether this amount of useful signal suffices. One might argue that the selectivity values for the spectra are unusually small. However, similar values have been reported for applications of (first-order) near-infrared spectroscopy [27,28]. For the elution profiles, the effective loss of signal due to overlap is less dramatic (ξ_A is substantially larger than ξ_B). It is observed that the second-order selectivities for the BLLS approach (ξ_D) are slightly larger than the first-order selectivities associated with the elution profiles (ξ_A), which is the expected behaviour. The relative efficiency of the naïve approach is poor; it is 'best' for α-chymotrypsin, which is explained from the higher value for ξ_B.

The selectivities can be used to customise an analytical procedure to the specific needs of an application, which is claimed to be one of the major assets of analytical figures of merit. We will focus on the effect of increasing the chromatographic resolution on the efficiency of the LS approach. (Kalivas [29] has described how to improve the spectral selectivity in high-performance liquid chromatography (HPLC) with UV detection.) The ξ_D values imply that an analytical determination using the BLLS approach is based on 34–50% of the total signal. Increasing the chromatographic resolution to baseline separation would yield ξ_A values close to unity, which directly translates to ξ_D values close to unity. It follows that the maximum reduction of standard error of prediction is approximately a factor of two for

Table 1
Selectivities and relative efficiency of naïve approach for protein data

Analyte	Selectivity				Relative efficiency
	ξ_A	ξ_B	$\xi_A \cdot \xi_B$	ξ_D	
Myoglobin	0.499	0.028	0.014	0.502	0.028
α-Chymotrypsin	0.336	0.047	0.016	0.339	0.046
Carbon anhydrase	0.499	0.028	0.014	0.502	0.028

The symbols are explained in the text.

myoglobin and carbon anhydrase and a factor of three for α-chymotrypsin. Intermediate gain is obtained if full baseline separation is not achieved. Fig. 5 in Ref. [15] describes the relationship between chromatographic resolution and first-order selectivity: the latter increases almost linearly to approximately 0.9 when a resolution of 0.5 is reached, after which it levels off. It follows that only 10% of the maximum efficiency is lost when settling for a rather poor resolution of 0.5. This is believed to be useful information from the practical point of view.

4.2. Moderate overlap in both modes

As expected, the relative efficiency of the naïve approach takes much higher values for the amino acid data set (Table 2). The smallest values are obtained for tryptophan (0.26) and tyrosine (0.24), because of the relatively poor selectivity of their excitation spectra. It is worth mentioning that the ξ_D values approach the ideal limit (unity). Although the BLLS approach is certainly preferable from a precision point of view, GRAM can often be a reasonable choice in practice. This can be understood as follows. Unlike conventional calibration methods, which construct a model from calibration set data, GRAM constructs a joint 'model' for the calibration and prediction sample data matrices [3]. In this way, GRAM ensures that the analyte can be determined in the presence of unknown interferences. This so-called second-order advantage is already obtained using a single calibration sample. In contrast, the other methods require that there be at least as many calibration samples as the number of independently contributing constituents. The need for a calibration set that adequately spans the variation expected in the prediction samples introduces many practical problems. It is not uncommon that the instrument responses as well as the samples may change over time. These circumstances seriously hamper utilising a model for prediction. However, when using GRAM, measuring both samples (calibration and prediction) with short time intervals automatically solves these kinds of problems. It is emphasised that correction procedures have been developed to perform conventional calibration while keeping the second-order advantage. Öhman et al. [30] have proposed the so-called residual bilinearization (RBL) to be used in combination with unfold-PLS. The resulting method, RBL–PLS, was compared to GRAM by Öhman et al. [31] and Gerritsen et al. [32] for quantitation from HPLC–UV data. No clear winner emerged from these studies. Linder and Sundberg [17] developed a procedure similar to RBL and noted that it is unknown how this correction will affect the precision of prediction.

Table 2
Selectivities and relative efficiency of naïve approach for amino acid data

Analyte	Selectivity				Relative efficiency
	ξ_A	ξ_B	$\xi_A \cdot \xi_B$	ξ_D	
Tryptophan	0.260	0.961	0.250	0.970	0.26
Tyrosine	0.285	0.761	0.217	0.918	0.24
Phenylalanine	0.735	0.785	0.577	0.945	0.61

The symbols are explained in the text.

5. Conclusions and outlook

Performance characteristics have been compared for the BLLS approach, which is related to unfold-PLS and unfold-PCR, and the naïve approach, which is similar to GRAM. The BLLS prediction approach has a clear advantage in terms of precision of prediction. In principle, GRAM seems to be promising only if the overlap among the separate modes is moderate and the second-order advantage is strictly required (i.e., getting a representative calibration data set is not feasible). The latter requirement seems to restrict correct application of GRAM to the analysis of complex samples (food industry, environmental applications, biological samples). Of course, the relevance of the current discussions depends on the adequacy of the approximations leading to Eq. (5).

An interesting subject for future research is to investigate how other calibration methods relate to the framework of Linder and Sundberg. For example, Bro [33] has introduced multilinear PLS as a powerful alternative to unfold-PLS. Bro gives an example where the coefficient estimates are larger for multilinear PLS by approximately a factor of two (see Fig. 7 in Ref. [33]). However, the coefficient estimates are much noisier for unfold-PLS, which violates one of the assumptions behind Eq. (5). This explains why for that particular example the prediction results are better

for multilinear PLS. More extensive simulations corroborate these observations [34]. An interesting result is that in the absence of calibration error the regression vector obtained using multilinear PLS is different from the BLLS vector. We have only been able to determine that some coefficients are expanded, whereas others are shrunk. Consequently, at this stage, we do not know how multilinear PLS relates to the framework of Linder and Sundberg.

Acknowledgements

The reviewers are thanked for pointing out some errors and unclear passages.

References

[1] H.A.L. Kiers, J. Chemom. 14 (2000) 105.
[2] S. Wold, P. Geladi, K. Esbensen, J. Øhman, J. Chemom. 1 (1987) 41.
[3] E. Sánchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496.
[4] H.A.L. Kiers, A.K. Smilde, J. Chemom. 9 (1995) 179.
[5] N.M. Faber, Anal. Bioanal. Chem. 372 (2002) 683.
[6] D.L. Massart, A. Dijkstra, L. Kaufman, Evaluation and Optimization of Laboratory Methods and Analytical Procedures, Elsevier, Amsterdam, 1978.
[7] A. Lorber, Anal. Chem. 58 (1986) 1167.
[8] D.R. Morgan, Appl. Spectrosc. 31 (1977) 404.
[9] K. Faber, A. Lorber, B.R. Kowalski, J. Chemom. 11 (1997) 181.
[10] J.H. Kalivas, P.M. Lang, Mathematical Analysis of Spectral Orthogonality, Marcel Dekker, New York, 1994.
[11] J.H. Kalivas, P.M. Lang, Chemometr. Intell. Lab. Syst. 32 (1996) 135.
[12] C.N. Ho, G.D. Christian, E.R. Davidson, Anal. Chem. 52 (1980) 1071.
[13] Y. Wang, O.S. Borgen, B.R. Kowalski, M. Gu, F. Turecek, J. Chemom. 7 (1993) 117.
[14] N.J. Messick, J.H. Kalivas, P.M. Lang, Anal. Chem. 68 (1996) 1572.
[15] K. Faber, A. Lorber, B.R. Kowalski, J. Chemom. 11 (1997) 419.
[16] M. Linder, R. Sundberg, Chemometr. Intell. Lab. Syst. 42 (1998) 159.
[17] M. Linder, R. Sundberg, J. Chemom. 16 (2002) 12.
[18] N.M. Faber, J. Chemom. 15 (2001) 743.
[19] A.J. Berger, M.S. Feld, Appl. Spectrosc. 51 (1997) 725.
[20] J.F. García, A. Izquierdo-Ridorsa, M. Toribio, G. Rauret, Anal. Chim. Acta 331 (1996) 33.
[21] N.M. Faber, J. Chemom. 15 (2001) 169.
[22] J.M.F. ten Berge, Least Squares Optimization in Multivariate Analysis, DSWO Press, Leiden, 1993.
[23] C.J. Appellof, E.R. Davidson, Anal. Chim. Acta 146 (1983) 9.
[24] J. Ferré, S.D. Brown, F.X. Rius, J. Chemom. 15 (2001) 537.
[25] B.G.M. Vandeginste, F. Leyten, M. Gerritsen, J.W. Noor, G. Kateman, J. Chemom. 1 (1988) 57.
[26] http://www.models.kvl.dk/source/nwaytoolbox/download.asp.
[27] A. Lorber, K. Faber, B.R. Kowalski, Anal. Chem. 69 (1997) 1620.
[28] N.M. Faber, Chemometr. Intell. Lab. Syst. 50 (2000) 107.
[29] J.H. Kalivas, Anal. Chem. 58 (1986) 989.
[30] J. Öhman, P. Geladi, S. Wold, J. Chemom. 4 (1990) 79.
[31] J. Öhman, P. Geladi, S. Wold, J. Chemom. 4 (1990) 135.
[32] M.J.P. Gerritsen, H. Tanis, B.G.M. Vandeginste, G. Kateman, Anal. Chem. 64 (1992) 2042.
[33] R. Bro, J. Chemom. 10 (1996) 47.
[34] N.M. Faber, R. Bro, Chemometr. Intell. Lab. Syst. 61 (2002) 133.